

图像纹理分类方法研究进展和展望

刘丽¹ 赵凌君² 郭承玉¹ 王亮³ 汤俊¹

摘要 纹理分类是计算机视觉和模式识别领域的一个重要的基本问题,也是图像分割、物体识别、场景理解等其他视觉任务的基础. 本文从纹理分类问题的基本定义出发,首先,对纹理分类研究中存在的困难与挑战进行阐述;接下来,对纹理分类方面的典型数据库进行全面梳理和总结;然后,对近期的纹理特征提取方法的发展和现状进行归类总结,并对主流纹理特征提取方法进行了详细的阐述和评述;最后,对纹理分类发展方向进行思考和讨论.

关键词 纹理分类, 特征提取, 深度学习, 局部特征描述, 计算机视觉

引用格式 刘丽, 赵凌君, 郭承玉, 王亮, 汤俊. 图像纹理分类方法研究进展和展望. 自动化学报, 2018, 44(4): 584–607

DOI 10.16383/j.aas.2018.c160452

Texture Classification: State-of-the-art Methods and Prospects

LIU Li¹ ZHAO Ling-Jun² GUO Cheng-Yu¹ WANG Liang³ TANG Jun¹

Abstract Texture is a fundamental characteristic of many types of images. Texture classification is one of the essential tasks in the field of computer vision and pattern recognition. It is also the basis of other complex vision tasks, such as image segmentation, object recognition and scene understanding. In this paper, we first address the importance of texture classification and summarize the difficulties and challenges in the development of texture feature extraction approaches. Then we discuss the existing texture databases which are generally acknowledged as public evaluation bases for texture classification methods. Next, we review recent achievements in the study of texture feature development and provide detail discussion on prominent texture feature descriptors. Finally, we point out the future directions of texture classification.

Key words Texture classification, feature extraction, deep learning, local descriptors, computer vision

Citation Liu Li, Zhao Ling-Jun, Guo Cheng-Yu, Wang Liang, Tang Jun. Texture classification: state-of-the-art methods and prospects. *Acta Automatica Sinica*, 2018, 44(4): 584–607

作为一种重要的视觉线索,纹理广泛存在于自然界各种物体的表面,在图像中则体现为特征值强度的某种局部重复模式的宏观表现. 无论是对自然图像、遥感图像或是医学图像而言,纹理都是一种非常重要的特征. 而对于纹理特征的研究也成为图像处理、计算机视觉和模式识别领域中一个难以回避的基础性难题,一直以来广受研究者的关注. 在纹理特征分析上所取得的进展,不仅对于人类对自身视觉机理的理解和研究具有推动作用,而且对计算机视觉和模式识别领域的诸多问题起到了重要的支撑作用并得到了广泛的应用,如视觉导航、场景分类、

物体识别、人脸识别、智能视频分析、基于内容的图像和视频检索、遥感图像分析、工业检测、医学图像分析和文本分类等.

有关图像纹理特征分析的研究最早可追溯到 1962 年 Julesz 的工作^[1]. 经过 50 多年的持续研究,各类理论和算法层出不穷. 不少研究者对纹理特征提取方法进行了综述,代表性的工作有: 1993 年, Tuceryan 等^[2] 对早期的纹理特征提取方法进行总结,将其大致分为 5 类: 基于统计的特征、基于几何的特征、基于结构的特征、基于信号处理的特征和基于模型的特征. 同年, Reed 等^[3] 对自 1980 年以来的纹理特征和纹理分割方法做了简要综述. 1999 年, Randen 等^[4] 对基于信号处理的纹理特征的性能做了详细的实验比较综述. 2002 年, Zhang 等^[5] 的综述着重讨论了不变性纹理特征提取方法. 2007 年, Zhang 等^[6] 对当前几种主要的不变性纹理特征提取方法在纹理分类和物体识别中的性能进行了评估. 2008 年, Xie 等^[7] 对已有的主要纹理特征提取方法进行了简单的介绍,并没有进行归类. 2009 年, 刘丽等^[8] 对已有纹理特征提取方法进行了归类和总结,其总结的方法多为 2005 年之前.

收稿日期 2016-06-08 录用日期 2017-05-06
Manuscript received June 8, 2016; accepted May 6, 2017
湖南省自然科学基金杰出青年基金 (2017JJ1007) 资助
Supported by The Hunan Provincial Natural Science Fund for Distinguished Young Scholars (2017JJ1007)

本文责任编辑 张长水
Recommended by Associate Editor ZHANG Chang-Shui
1. 国防科学技术大学信息系统与管理学院 长沙 410000 2. 国防科学技术大学电子科学与工程学院 长沙 410000 3. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190
1. College of Information System and Management, National University of Defense Technology, Changsha 410000 2. College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410000 3. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

Pietikäinen 等^[9]、宋克臣等^[10]、刘丽等^[11] 分别对基于局部二值模式 (Local binary pattern, LBP) 纹理特征及其在纹理分析和人脸识别以及其他领域中的应用进行综述. 上述工作中, 文献 [2-3] 总结的是较早期方法, 文献 [4-6] 讨论的是某一类型的方法, 文献 [7-8] 总结的纹理特征提取方法大多为 2000 年之前的方法, 2000 年以后的方法讨论的极少, 近期综述 [9-11] 仅仅总结了基于 LBP 的纹理特征. 这些文献对当前纹理特征提取方法的论述很不全面, 没有对近十几年以来的纹理特征提取方法进行完整清晰的总结. 事实上, 近年来对纹理特征分析的研究催生了一些优秀的图像低层特征表达学习方法, 对物体识别、场景识别、人脸识别和图像分类等计算机视觉领域的其他问题有重要推动作用. 因此, 有必要对近 10 年最新的相关研究成果进行全面的综述和讨论. 本文系统综述了纹理特征提取方法的研究进展和当下亟待解决的问题, 以期为进一步深入研究纹理特征分析和理解以及拓展其应用领域奠定一定的基础.

本文从纹理分类问题的基本定义出发, 首先从实例、类别两个层次对纹理分类研究中存在的困难与挑战进行了阐述. 接下来, 对纹理分类方面的主流数据库进行了总结. 然后, 重点对近期的纹理特征提取方法进行了梳理和归类总结, 详细阐述了代表性

方法的动机、原理、优势与不足, 揭示了各种方法之间的区别与联系. 最后, 给出了我们对纹理分类的发展方向的一些思考.

1 纹理分类的难点与挑战

纹理图像分类是视觉研究中的基本问题, 也是一个非常具有挑战性的问题. 纹理图像分类是指基于待分类图像或图像区域的内容为其指定一个预先定义的纹理类别. 纹理特征描述和分类器是图像分类的两个关键环节. 对图像进行纹理特征描述是纹理图像分类的主要研究内容, 因为如果提取的纹理特征不好, 即便好的分类器也无法完成识别任务^[4,8]. 材质分类是指根据图像内容, 识别图像中物体的所属材质 (如木质、玻璃、塑料、钢铁和纤维等). 材质分类问题往往当成纹理分类问题来研究, 因此本文中我们讨论的纹理分类问题包括材质分类问题. 我们将纹理图像分类的难点与挑战分为两个层次: 实例层次和类别层次.

1) 实例层次. 就单幅纹理图像而言, 通常由于图像采集过程中光照条件、拍摄视角、拍摄距离和拍摄表面的非刚体形变等, 使得纹理表观特征产生很大的变化, 如图 1 (a)~(c) 所示, 给纹理分类算法带来难度.

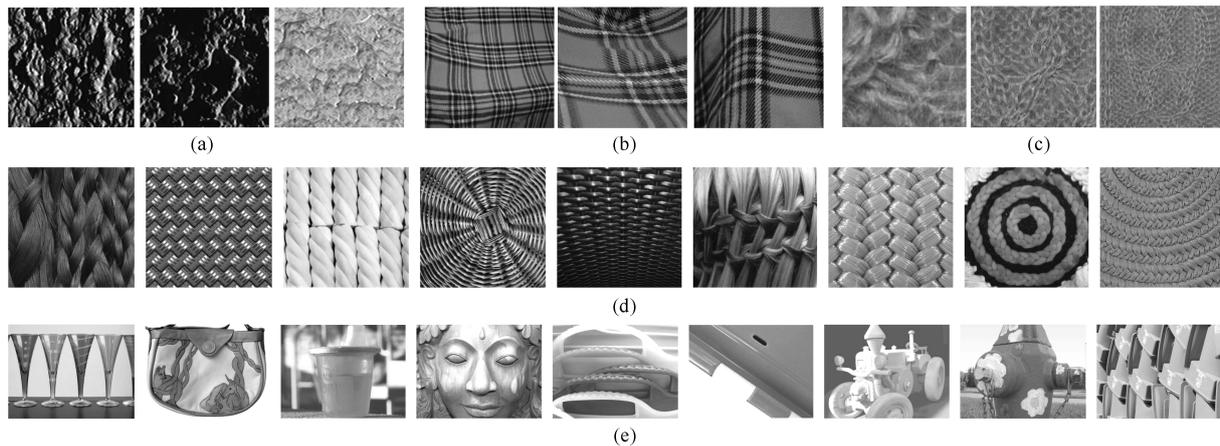


图 1 纹理识别难点示例 (实例层次: (a) 光照变化带来的影响, 图片来自 CURET 数据集第 30 类; (b) 视角变化和局部非刚性形变带来的难点, 图片来自 UIUC 数据集第 25 类; (c) 尺度变化带来的影响, 图片来自 KTHTIPS2b 数据集. 类别层次: (d) 同一类别的不同实例图像带来很大类内差异, 图片来自 DTD 数据集的 braided 类; (e) 材质识别的难点, 图片来自 FMD 数据集, 正确答案为 (从左往右): 玻璃, 皮质, 塑料, 木质, 塑料, 金属, 木质, 金属和塑料)

Fig. 1 Challenging examples of texture recognition ((a) Illumination variations, images are from the 30th category of the CURET dataset; (b) View point and local non-rigid deformation, images are from the 25th category of the UIUC dataset; (c) Scale variations, images are from the KTHTIPS2b dataset; (d) Different instance appearance variations from the same category, images are from the braided category in the DTD dataset; (e) Material classification difficulties, images are from the FMD dataset, the category for these images are as follows (from left to right): glass, cortex, plastic, wood, plastic, metal, wood, metal, and plastic, (a), (b) and (c) belong to instance-level variations; (d) and (e) belong to category-level variations.)

2) 类别层次. 困难与挑战通常来自三个方面. 首先是存在较大的类内差异, 也即属于同一类的纹理表现特征差别比较大, 其原因一方面包括前面提到的各种实例层次的变化, 更重要的是由于类内不同实例的差异, 如图 1(d) 所示; 其次是类间模糊性, 即不同类别的纹理实例具有一定的相似性, 如图 1(e) 所示, 即使让人类正确辨别图中的材质类别也是很有难度的; 再次是噪声的干扰, 在实际场景下, 可能出现模糊和随机噪声以及雨雾雪的影响等, 这使得识别问题的难度大大增加.

除了上述难点与挑战, 纹理不存在一个公认的定义也给纹理的建模和描述带来困难. 纹理图像特征描述要同时考虑以下三个相互矛盾的指标. 1) 强可区分能力. 自然界纹理图像的种类繁多, 纹理特征描述要能够表达丰富的纹理图像内容, 可以鉴别不同的纹理图像类别. 2) 高鲁棒性. 纹理特征要对上述实例层次和类别层次提到的变化因素具有高鲁棒性. 纹理特征的特征区分力和特征鲁棒性都受类内表现差异和类间表现差异影响, 大的类内表现差异和小的类间表现差异通常会导致纹理特征的可区分力减弱、鲁棒性降低. 3) 低计算复杂度和低存储. 要求特征提取能快速实现, 并且内存和硬盘占用量小, 这对于大规模分类问题以及资源有限 (CPU 性能有限、内存空间有限、电池电量有限) 的移动设备计算尤其重要.

2 纹理分类与纹理数据集

数据是视觉识别研究中一个非常重要的因素, 纹理识别也不例外, 纹理数据集在纹理识别研究中起着重要的作用. 一方面, 标准数据集便于让研究者公平地进行算法方面的比较; 另一方面, 随着大数据时代的到来, 对数据集进一步丰富、完备的要求更加迫切. 在数据足够多的情况下, 可能最简单的模型、算法都能得到很好地效果. 鉴于数据对算法的重要性, 我们将在本节对纹理分类方面的主流数据进行概述. 目前常用的基准纹理数据集主要有医学图像、自然纹理图像、材质图像和动态纹理 4 个方面, 我们重点介绍自然纹理图像数据集和材质纹理图像数据集. 在介绍不同数据库时, 将主要从数据库图像数目、类别数目、每类样本数目、图像大小、纹理图像特点和分类难度等方面进行阐述, 如表 1 所示.

Brodatz^[12] 是出现最早、使用最多、最著名的一个纹理数据库, 经常出现在纹理分类、纹理分割和纹理合成的研究工作中. 该数据库有 112 类纹理图像, 每类仅有一幅样本图像. 尽管有很多纹理分类算法都涉及到该数据库, 但是大多数研究者仅仅采用了该数据库的一个较小子集进行算法性能测试. 近期才有研究者开始用整个数据集测试分类算法性能. Brodatz 包含较多的纹理类别, 有利于评估纹理特征的鉴别能力, 但是有些纹理类别非常相似,

表 1 主流纹理分类数据库, 下载地址为: Brodatz^[13]、VisTex^[14]、CURET^[15]、Outex^[16]、KTH TIPS^[17]、UIUC^[18]、KTH TIPS2a^[17]、KTH TIPS2b^[17]、UMD^[19]、ALOT^[20]、FMD^[21]、Drexel^[22]、OS^[23]、DTD^[24]、MINC^[25]

Table 1 Widely used texture datasets and their download link: Brodatz^[13], VisTex^[14], CURET^[15], Outex^[16], KTH TIPS^[17], UIUC^[18], KTH TIPS2a^[17], KTH TIPS2b^[17], UMD^[19], ALOT^[20], FMD^[21], Drexel^[22], OS^[23], DTD^[24], MINC^[25]

数据库	图像数目	类别数目	图像尺寸	灰度/颜色	成像条件	光照变化	旋转变换	视点变化	尺度变化	图像内容	实例类别	建立年度
Brodatz	112	112	640 × 640	灰度	实验可控					物体表面	实例	1966
VisTex	167	167	786 × 512	颜色	户外		√			物体表面	实例	1995
CURET	5 612	92	200 × 200	颜色	实验可控	√	小	√		材料表面	实例	1999
Outex	8 640	320	746 × 538	颜色	实验可控	√	√	√		材质/物体	实例	2002
KTH TIPS	810	10	200 × 200	颜色	实验可控	√	小	小	√	材料表面	实例	2004
UIUC	1 000	25	640 × 480	灰度	户外可控	√	√	√	√	材料表面	实例	2005
KTH TIPS2a	4 608	11	200 × 200	颜色	实验可控	√	小	小	√	材料表面	类别	2006
KTH TIPS2b	4 752	11	200 × 200	颜色	实验可控	√	小	小	√	材料表面	类别	2006
UMD	1 000	25	1 280 × 960	灰度	户外可控	√	√		√	物体表面	实例	2009
ALOT	25 000	250	768 × 512	颜色	实验可控	√	√	√		材料表面	实例	2009
FMD	1 000	10	512 × 384	颜色	不可控	√	√	√	√	材料表面	类别	2009
Drexel	40 000	20	200 × 200	颜色	实验可控	√	√		√	材料表面	实例	2012
OS	10 422	22	不固定	颜色	不可控	√	√	√	√	材料表面	杂波	2013
DTD	5 640	47	不固定	颜色	不可控	√	√		√	纹理属性	类别	2014
MINC	2 996 674	23	不固定	颜色	不可控	√	√	√	√	材料表面	杂波	2015
MINC2500	57 500	23	362 × 362	颜色	不可控	√	√	√	√	材料表面	杂波	2015

即使人眼也难以区分开来. 该数据集展现的类内差异较小, 每一类仅有一幅样本图像, 不存在光照、旋转、视点和尺度变化的影响, 因此与实际应用图像有很大差距. VisTex 也是早期比较著名的一个数据库, 由 MIT 多媒体实验室的视觉与建模小组构建, 该数据集有 167 类纹理图像, 每一类只有一幅样本. 该数据集不是在可控的实验室环境下构建, 而是在自然光照条件下构建, 因此比较接近实际条件, 然而和 Brodatz 一样, 也是不能体现现实中的类内差异. VisTex 较多出现在纹理合成或纹理分割的论文中, 很少出现在图像级的纹理分类中. Brodatz 和 VisTex 更多地是从宏观上关注物体表面或者多个物体排列的纹理图像, 后续研究者开始关注材质表面的纹理特性以及材质在光照、旋转、视点和尺度等成像条件变化时纹理外观呈现的差异, 因此基于这两个数据集的纹理分类实验逐渐减少.

在材质纹理数据库中, 最著名、最常用的一个是 CURET (Columbia-Utrecht reflectance and texture)^[26] 数据集, 它有 61 类纹理, 将每个纹理样本在可控的实验室环境下成像, 在 205 个不同的视点成像, 由于光源位置固定, 导致光照变化带来很大的影响, 使得材质出现光照反射、阴影等类内差异, 如图 1(a) 所示. Varma 等^[27] 基于该数据库构建了一个子集, 用于纹理分类的研究, 该子集后来成为用来评估纹理分类方法的基准集, 后续研究论文中出现的 CURET 大多数都指该子集. 它包含 61 类纹理图像, 每一类有 92 幅图像, 每个图像大小为 200 像素 \times 200 像素, 原始 CURET 数据库中其他视点的图像因无法获得 200 像素 \times 200 像素大小区域而被忽略. 尽管视点变化, 但是该数据集中的图像并没有呈现明显的平面内图像旋转, 且没有尺度变化. 因此, 区分力强的纹理特征 (即使没有旋转不变性) 也能在 CURET 上获得高识别率, 但是具有尺度不变性的特征性能反而会下降. 由于 CURET 没有尺度变化, 瑞典皇家理工学院 (KTH) 构建了 KTHTIPS (KTH textures under varying illumination, pose and scale, KTHTIPS)^[28], 该数据库构建一个原因是为了扩展 CURET 数据库, 从中选出 10 类材质类别, 每一类成像于 3 种不同光照、3 个姿态和 9 个尺度, 获得 81 幅图像. 此外, 虽然材质类别与 CURET 中的相同, 但是 KTHTIPS 中的实际样本与 CURET 的不同, 因此 KTHTIPS 的另一个目的是为了识别来自不同样本的成像图片. 随后, 瑞典皇家理工学院的研究团队又构建了 KTHTIPS2 数据库, 包含 11 类材质类别. 与之前的数据库不同的是, 研究者首次尝试对同类材质的不同材质实例进行成像, 每类材质选择 4 种差异很大的实例 (参见图 2).

在 3 种不同姿态、4 种不同光照和 9 种不同尺

度下成像, 得到 432 幅图像. KTHTIPS2 数据库有两个版本, KTHTIPS2a 和 KTHTIPS2b, 不同之处在于前者比后者少了 144 幅图像. 基于这个数据库的实验, 每一类中选择来自一部分材质实例的成像图像来识别剩余材质实例的图像, 这是识别没有见过的材质实例, 是此数据库的主要难点所在, 此外, 较大的尺度变化也增加了识别的难度. 尽管有姿态的变化, 但是该数据库的旋转变化很小, 因此不具有旋转不变性的纹理特征也可能获得较好的识别性能.

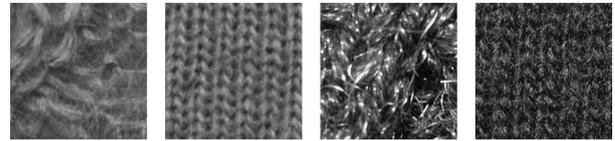


图 2 来自 KTHTIPS2 的某类图像的样本

Fig. 2 Image examples from one category in KTHTIPS2

Outex 数据库由芬兰奥卢大学机器视觉研究小组构建^[29], 该数据集包含 320 类纹理图像, 每个纹理样本成像于 3 种光照条件和 9 个旋转角度, 没有视点变化和尺度变化的影响. 在 Outex 数据库基础上, 构建了一系列子集, 作为测试纹理分类和分割算法基准集, 详见文献 [29]. 其中, Outex_TC00010 和 Outex_TC00012^[30] 两个基准集在纹理分类问题上得到广泛使用, 主要用于测试纹理特征的旋转不变性和光照不变性. 尽管原始 Outex 数据库包含 320 类纹理图像, 是类别最多的一个纹理库, 但是目前还没有研究者采用整个数据库进行纹理分类, 有极少数研究者采用接近 300 类的纹理进行分类^[31].

UIUC 纹理库^[32-33] 的构建是为了评估不变性纹理特征的性能, 包含 25 类纹理, 每一类 40 幅样本图像, 在近期被广泛使用. 该数据库的难点在于, 每类样本图像不多, 但是有很大的类内差异, 由于很大的视角变化和尺度变化, 还存在局部非刚性形变 (参见图 1(b)). 尽管 UIUC 的光照变化的影响比 CURET 要小很多, 但是其他类内差异要比 CURET 更大且更接近现实环境. UIUC 每幅图像大小为 640 像素 \times 480 像素, 分辨率比较高, 而实际中图像纹理区域多半不是高分辨率的. UMD 数据库^[34] 的大小和引起类内差异的因素同 UIUC 一样, 但是图像大小为更高分辨率 1280 像素 \times 960 像素. 该数据库的高分辨率是为了测试基于分形的纹理特征的性能. 尽管与 UIUC 类似, 但是 UMD 的分类更容易, 很多方法都可以在该数据库上达到 99% 以上的分类率 (参见表 2). ALOT (Amsterdam library of textures, ALOT) 数据库^[35] 包含 250 类纹理, 每一类 100 幅图像, 成像于 8 种不同的光照条件和 4 个不同的视角, 没有尺度变化和非刚性形变等影响, 该数据库包含的纹理类别较多. 该数据库有三

种分辨率版本,由高到低分别对应图像大小为 1536 像素 \times 1024 像素, 768 像素 \times 512 像素和 384 像素 \times 256 像素. ALOT 数据库中有些纹理类别很容易区分,也有一些很难区分. Drexel 数据库是由美国卓克索大学 Oxholm 等^[36] 创建的一个数据库,包含 20 种不同纹理,每种纹理有 2000 幅样本图像,在实验室环境中,于不同光照条件、成像距离以及多种平面内和平面外旋转变条件下成像获得. 该数据库的构建是为了研究纹理几何(即自相似性和尺度变化)基础特性,利用其进行物体表面法向估计和几何纹理分类. 目前该数据集在纹理分类中受到的关注较少.

随着纹理分类方法的进步,很多方法在上述几个数据库上所达到的性能都接近饱和(KTHTIPS2 除外),视觉识别逐渐开始处理更加真实场景的视觉问题,因而对数据库的泛化性、规模等也提出了新的要求和挑战. MIT 研究者构建了 FMD 材质数据集(Flickr material database)^[37],从网上搜集了 10 类材质类别,每一类有 100 个样本,每个样本来自于不同的材质实例,与以往的实验室可控环境下对同一个实例进行成像构建数据库的方法不同,更为接近真实场景,一些示例图片如图 1(e) 所示. FMD 与之前提到的传统材质纹理数据集有很大不同,注重于识别物体的材质如塑料、木质、纤维和玻璃等. 尽管 MIT 的研究者指出^[38],人类可以快速识别材质,但是计算机材质自动识别问题不同于物体识别、场景识别和纹理识别问题,也许仅仅依靠低层视觉特征并不能解决这个问题,然而研究者往往用纹理分类的方法来识别材质. 传统纹理特征在 FMD 上的分类性能比较差^[37],采用近期的深度学习方法可以获得较大的提升,但是性能还有很大提升空间^[39-40]. 然而, FMD 数据集仅仅包含 10 类材质,每一类的样本图像也较少,而且是人工仔细挑选的,不足以反映材质类别的类内差异,不利于特征的自动学习.

由于近期基于视觉属性物体识别获得广泛关注^[41-44],而纹理属性(Texture attributes)在描述物体时也起着重要作用. 为此,牛津大学的 Andrea Vedaldi 研究小组^[45]收集了 47 种描述纹理特征的属性,并基于此构建了一个纹理属性数据库 DTD(Describable textures dataset),每一类属性包含 120 幅图像,全库共 5640 幅图像,全部从网上搜集得到. 该数据与上文提到的数据库有很大不同,传统纹理特征在该数据库上识别率低于 50%. 最大的挑战来自于较大的类内差异,如图 1(d) 所示,其中的纹理图像均属于 braided 类,而传统意义上讲,这些纹理应该属于不同的纹理类别,因此该数据库的识别很有难度. 此外,由于纹理属性是语义描述类别,存在多义性,即同一幅纹理图像可以用多个纹理属

性进行描述,相似的纹理图片可能出现在不同的纹理属性类别中,导致类间模糊性,增加了识别难度. DTD 数据库的不足之处在于,每一类的图像数不多,不足以反映类内变化. 但总的来说,DTD 数据集的构建是用属性来描述纹理方面的一个新的尝试.

近期,康奈尔大学的研究小组 Bell 等^[46]构建了一个大规模的场景材质数据库 OS(Open Surfaces). 他们认为现实场景中的物体及表面都是由材质、纹理、形状等特征决定(比如一个厨房,包含花岗岩台面、金属材质的锅、木质地板、玻璃灯饰等). 前文提到的所有数据库每一幅图像为一种纹理类别,且纹理填满图像区域,OS 数据库的每幅图像为一个场景,包含多种材质、纹理等;其规模比以前的纹理数据库显著增大,共有 2 万多幅实际场景图像,每幅图像提供了材质、纹理等标注信息,以及多个分割开来的材质区域,总共有 10 万多个单一材质区域,可支持多种应用问题. 就材质分类问题而言,OS 存在一些局限性:OS 的材质类别不平衡,有的类别包含丰富的样本图像(如木质类有 2 万个样本图像),有的类别包含的样本图像比较少,才几十个(如水),这个问题是由于标注不平衡造成的.

为了进一步推动材质分类问题的研究, Bell 等^[47]在 OS 的基础上,构建了一个大规模材质数据库 MINC(Materials in context database),从 436 749 幅场景图像中标注出 2 996 674 幅材质图像片,属于 23 种不同材质,其中类别样本数目不平衡,最多的类别包含 564 891 个图像片,最少的类别包含 14 954 个图像片,前者约为后者的 40 倍. MINC 数据库规模显著大于 FMD,材质类别也增加到 23 类,MINC 中的图像样本如图 3 所示.



图 3 来自 MINC 数据库中的图像样本(第一行为食物类别,第二行为 foliage 类别)

Fig. 3 Image samples from the MINC database (Example images in the first row are from the food category, while those in the second row are from the foliage category)

MINC 数据库支持图片级材质分类和场景图像材质分割任务. 如图 3 所示,MINC 中的样本图像片包含上下文信息,但是图像片的中心像素的材质类别代表整个图像片的材质类别,这点与传统纹理数据库(整个图像片均为同一类别)不同. Bell 等^[47]

又从 MINC 中选择一部分图像构成了 MINC2500 数据集, 共 23 类材质, 每类包含 2500 幅图像. 牛津大学的 Cimpoi 等^[40] 也基于 OS 数据库构建数据集进行了分类和分割实验, 并进行了部分纹理属性的标注, 但是他们构建的数据集尚未公开.

我们对近期纹理分类中常用的纹理数据库进行了总结, 更多有关纹理数据库的描述可参见 Hossain 等的工作^[48]. 我们也可以发现, 物体类别越多, 导致类间差越小, 要求特征具有强可区分力, 类内差异越大, 对特征的鲁棒性提出更高要求, 分类任务越困难; 图像数目多少、图像尺寸的大小, 则直接对算法的可扩展性提出了更高的要求. 如何在有限时间内高效地处理大规模数据、进行准确的纹理分类以适用于不同的应用场景成为当前研究的关注点.

3 纹理特征提取方法研究进展

图像纹理研究已经有 50 多年的历史, 各类理论和算法层出不穷, 本节中我们对近期的纹理特征提取方法进行了简单梳理和归类总结, 并着重阐述了其中代表性的工作, 揭示不同方法之间的区别与联系.

纹理分类任务是指根据图像内容为其指定一个预先定义的纹理类别. 对图像进行纹理特征描述是纹理分类的主要研究内容. 一般说来, 纹理分类算法通过手工特征或者特征学习方法对整个图像进行全局描述, 将原始数据变换成一个合适的特征矢量, 使得分类器能够基于该特征矢量进行分类任务.

早期的纹理特征提取方法种类繁多(可参见综述 [2-4, 8] 等), 大部分基于一个共同的假定, 即纹理图像是在理想状态下获取的, 没有考虑到现实世界的纹理图像的复杂性(参见前文总结的纹理分类难点与挑战). 许多早期发表的方法在 Brodatz 数据库上进行纹理特征分类性能的评估, 而 Brodatz 数据库的局限性我们已经在前文提及. 在上个世纪 90 年代中期, Zhang 等^[5] 提倡不变性纹理分类方法的研究. 上世纪末至今, 纹理分类方法呈现一个新的局面, 研究者们把 Julesz 等的 Texton 理论^[49] 以一个新面孔重新搬上了历史舞台, 提出 Bag of textons (BoT) 的纹理分类方法. 这就是计算机视觉领域的一个重要模型——词包模型 (Bag of words, BoW). 词包模型最初产生于自然语言处理领域, 通过建模文档中单词出现的频率来对文档进行描述与表达. Csurka 等^[50] 于 2004 年首次将词包的概念引入计算机视觉领域, 由此开始大量的研究工作集中于词包模型的研究. 事实上, 词包模型方法早已经在纹理分类领域出现^[30, 32, 51], 即 BoT. 因此也可以说对纹理分类的研究催生了计算机视觉领域的词包模型. 由此开始, 大量的研究工作集中于视觉词包模型的

研究, 也多侧重于局部特征描述子设计.

传统人工设计局部特征结合视觉词包模型的框架在模式分类问题中主导地位被近几年出现的深度学习技术^[52] 所动摇. 深度学习的基本思想是通过有监督或者无监督的方式学习层次化的特征表达, 来对物体表面进行从底层到高层的描述. 卷积神经网络^[53-54] 是深度学习的一种, 已成为当前计算机视觉领域的研究热点, 它的权值共享网络结构使之更适用于图像分析, 且降低了网络模型的复杂度, 减少了权值的数量, 可以实现端到端的学习训练与分类任务, 避免了传统识别算法中复杂的人工特征提取过程. 为此, 深度卷积网络特征在纹理分类问题中也得到很好地应用, 代表性的方法还是将训练好的深度卷积神经网络特征与传统的词包模型相结合.

与此同时, 值得关注的也有基于纹理视觉属性的研究工作, 旨在缩小计算机视觉中的“语义鸿沟”问题. 底层视觉特征则只能被机器识别, 没有直接语义含义. 视觉属性是物体的一种属性, 可以直接被人类视觉感知也能被机器理解的较高层次的描述. 这项工作的开展建立在纹理特征的学习表达方法之上.

3.1 视觉词包模型

在计算机视觉领域中, BoW 最开始起源于纹理分类问题, 但后来被广泛应用于物体分类和场景分类等图像分类问题, 并逐渐形成了由局部特征描述、特征编码、特征汇聚和特征分类 4 部分组成的标准物体分类框架^[55], 如图 4 所示.

1) 局部纹理特征提取

局部纹理特征描述是框架中的第一步, 可分为两类: 稀疏类方法和稠密类方法. 前者是基于兴趣点检测, 通过某种准则选择具有明确定义的、局部纹理特征比较明显的像素点、边缘、角点、重要区域等, 并且通常能够获得一定的几何不变性, 从而可以在较小的开销下得到更有意义的表达, 然后采用局部特征描述子对提取的稀疏分布的兴趣点区域进行特征提取, 以期获得更为紧凑的特征空间. 最常用的兴趣点检测算子有 Harris 角点检测子、拉普拉斯高斯斑块等^[56-57]. 用于描述兴趣点区域的常用局部纹理特征描述子有 SIFT (Scale invariant feature transform)、RIFT (Rotation invariant feature transform)、SPIN (Spin image) 等^[58]. 稀疏类方法的缺点是特征提取过程步骤较多、比较复杂; 对于某些纹理图像, 可能检测不到感兴趣区域或者检测到的感兴趣区域很少, 缺乏稳定性; 局部区域的不变性主要依赖于估计局部区域的主导梯度方向, 不仅计算复杂度高, 且估计得到的主导方向并不可靠. 稠密类方法是指从图像中逐像素或者按固定的步长进行局部特征的稠密提取,

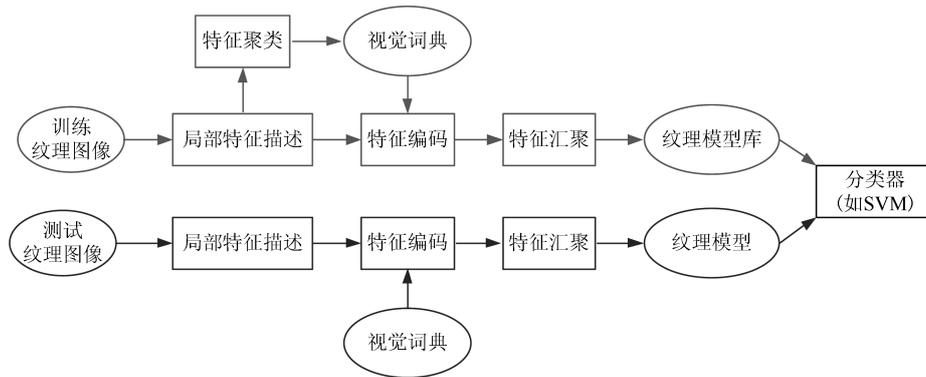


图 4 基于词包模型的纹理分类示意图

Fig. 4 Texture classification based on the BoW pipeline

大量的局部特征描述尽管具有更高的冗余度，但信息更加丰富，其中的冗余信息主要靠后面的特征编码和特征汇聚得到抽象和简并，结合使用词包模型进行有效表达后通常可以得到比稀疏类方法更好的性能，因此在纹理分类中使用较为广泛。

2) 特征编码

密集提取的底层特征中包含了大量的冗余与噪声，为提高特征表达的鲁棒性，需要使用一种特征变换算法对底层特征进行编码，从而获得更具区分性、更加鲁棒的特征表达。这一步对物体识别的性能具有至关重要的作用，因而大量的研究工作都集中在寻找更加强化的特征编码方法，重要的特征编码算法包括向量量化编码（如 k 均值）、稀疏编码、局部线性约束编码、显著性编码、Fisher vector (FV) 编码等。Huang 等^[59] 对主要特征编码方法进行了深入分析和比较，特征编码对物体检测和分类等问题起着更为重要的作用^[55]，而纹理图像具有平稳性的特征，使用最多的是简单有效的 k 均值聚类方法和 FV 方法^[60-61]，稀疏编码方法也在纹理分类中得到研究^[62-63]，其余方法使用较少。特征编码方法将不在本文中详细讨论，有兴趣的读者可参见 Huang 等^[59] 的工作。

3) 特征汇聚

特征编码后，每一个局部特征都将在视觉词典的一个或多个视觉单词上产生表达。对视觉词典上的响应进行特征汇聚和特征集整合操作得到图像的全局向量表达¹。这一步得到的图像特征表达也具有一定的不变性，同时也避免了使用特征集进行图像表达的高额代价。对于物体分类和场景分类等问题，图像通常具有较强的空间结构约束，因此 Lazebnik 等^[64] 提出的金字塔匹配 (Spatial pyramid matching, SPM) 被广泛采用。SPM 提出将图像均匀分

块，然后每个区块里面单独进行特征汇聚和整合操作，并将所有子块的特征向量串联起来作为图像最终的全局特征表达。而纹理图像更加关注的是图像中存在的一些简单的局部模式，以及这些局部模式的较有规律的重现。纹理图像的局部区域和整体具有相似性，纹理特征具有平稳性。因此，SPM 模型在纹理分类中较少采用，特征汇聚基于整幅图像进行即可。

4) 特征分类

使用最近邻分类器、支撑向量机 (Support vector machine, SVM) 等分类器进行分类。从图像提取到特征表达之后，一张图像可以使用一个固定维度的向量进行描述，接下来就是学习一个分类器对图像进行分类。常用的分类器有 SVM、最近邻、神经网络、随机森林等。SVM 是使用最为广泛的分类器之一，特别是使用了核方法的非线性 SVM，在纹理分类中基于开方核的 SVM 使用最多，常与传统 BoW 的直方图特征矢量结合使用。近期基于 FV 的图像特征表达矢量的维数高达数万甚至数十万的量级，导致无法使用非线性 SVM 进行分类，常采用线性 SVM。

综上所述，BoW 模型的核心步骤之一就是局部特征提取，在纹理分类中，局部特征提取起着最为关键的作用，因此得到广泛研究，下面对近期出现的局部纹理特征描述子归类总结。总的来说，一个优秀的局部纹理特征描述子应该具备如下性质：1) 可重复性，同一类型的纹理在不同的成像条件下，两幅图像对应的特征越多越好；2) 可区分力强，即具有很好地纹理类别区分能力；3) 局部性，易于获得对几何和成像变化的不变性；4) 高效性，局部特征提取的计算效率越高越好，以满足实时性需求；5) 数量性，提取的特征集要足够多，特征的密集度最好能在一定

¹常见的聚汇操作有取最大值和取平均值，分别是指对所有局部特征在同一个视觉单词上产生的响应取最大值或取平均值，这样所有局部特征在整个视觉词典上的响应就产生一个向量表达，该向量的维度和视觉单词的数量相同。例如在纹理分类中常用的 k 均值聚类后的视觉词典，特征编码后，图像的每个局部特征只会对一个视觉单词产生响应。那么特征汇聚就是统计图像中视觉单词出现的频率直方图，作为图像的全局表达。

程度上反映图像的内容;在上述性质中最为重要的是可重复性,其主要依赖于以下两个方面的性质:不变性和鲁棒性.前者是指局部特征不随图像的变换(如旋转、缩放、仿射变换、投影变换等成像几何)而改变;后者是指局部特征对于局部的变形(例如图像噪声、光照影响、压缩噪声、图像模糊以及非刚体形变等)不敏感;6)低维性,有利于后续环节处理,提高系统的整体效率;7)多尺度分析,可以从多尺度刻画局部纹理特征,提供更加充分的描述.

3.1.1 稀疏类纹理特征描述子

局部特征提取分两步进行,首先,从图像中检测一些感兴趣的区域;然后,选择合适的局部特征描述子从感兴趣区域提取特征.此类方法在利用多幅二维图像进行三维重建、图像匹配等任务中非常有用,曾经是计算机视觉领域的研究热点,出现了一系列感兴趣区域检测子和特征描述子,详见综述性文章^[56-58].研究者将此类方法用于纹理识别问题,最为代表性的工作是 Lazebnik 等^[32-33]和 Zhang 等^[6]的工作.

Lazebnik 等^[32-33]提出采用两种仿射不变性感兴趣区域检测方法,即 Harris 角点和 Laplacian 斑块,从纹理图像中检测一些感兴趣区域,这些区域通常是椭圆形区域;然后,通过估计感兴趣椭圆区域的主导梯度方向和尺度,将其转换成一个具有仿射不变性的圆形区域;采用两个局部特征描述子从圆形区域提取特征: RIFT 和 SPIN. SPIN 和 RIFT 描述子提取的特征分别为 32 维和 100 维,其中 RIFT 与 SIFT^[65]非常类似,具有旋转不变性,而 SPIN 计算的是软直方图,需要利用高斯加权,计算量很大.经过上述步骤,对于每一幅纹理图像,产生了两种类型的感兴趣区域和两种局部特征,得到 4 个特征通道.在每幅纹理图像的每个通道的特征空间中进行 k 均值聚类,得到局部纹元字典,基于此计算一个统计特征 $(c_1, w_1), (c_2, w_2), \dots, (c_k, w_k)$, 其中 c_i 为纹元, w_i 为图像中该纹元的权重(归类于该纹元的像素数目占整个图像像素的比例).采用测地距 (Earth mover's distance, EMD) 作为距离度量,多通道特征组合,基于最近邻分类器进行分类.

Zhang 等^[6]的工作建立在 Lazebnik 的工作基础之上,采用相同的感兴趣区域检测算子,特征描述子除了 SPIN 和 RIFT 之外,还增加了 SIFT 描述子.他们的结论指出,具有多种不变性的局部特征区域检测和描述算子不一定是最优纹理分类性能的保证,需要具体问题具体分析,因为实际中,不是所有问题都需要各种类型的不变性特征.采用具有互补性的特征描述子可以提高分类的性能.

在稀疏类方法中,由于纹理特征仅仅是从局部

感兴趣区域提取,而这两种特征点的数目要比整个图像像素数目小得多,因此特征空间更为紧致.同时,也会带来缺陷.1) 感兴趣区域的数量性不能得到保证,因为有些纹理图像中,能够检测到的感兴趣区域较少,引起不稳定性;2) 感兴趣区域并不能代表整个纹理图像,可能损失一些重要纹理模式和特征;3) 感兴趣区域的主导梯度方向和主要尺度的估计,难以准确;4) 特征提取过程复杂、繁琐,特征维数较高,且 EMD 计算量较大.

3.1.2 稠密类纹理特征描述子

1) LM 滤波器组方法

Leung 等^[51]率先研究了不同成像视角和光照条件下获得的纹理图像的识别问题.采用的滤波器组如图 5 所示,简称为 LM (Leung-Malik) 滤波器组,由简单的高斯滤波器、拉普拉斯高斯滤波器和高斯滤波器的一阶、二阶导数组成,共 48 个滤波器.一幅纹理图像经 LM 滤波器组滤波后,每一个像素点均产生一个滤波响应特征矢量,根据定义纹理图像就是局部结构的重复,因此有理由认为存在一些

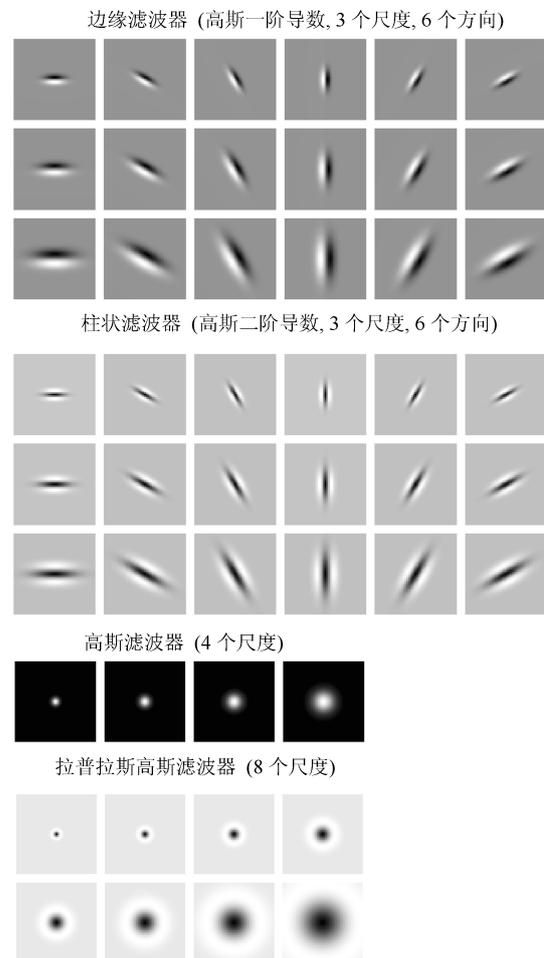


图 5 LM (Leung-Malik) 滤波器组

Fig. 5 LM (Leung-Malik) filters

有代表性的滤波特征矢量, 而每个像素上的特征只是这些代表性特征的噪声变体, 可以由它们近似表示. 基于此, 加之受 Julesz^[49] 的纹元 (Texton) 理论所启发, Leung 等^[51] 做出了一个具有影响力的研究工作: 明确地给出了二维纹元 (2D Texton) 的定义, 将其定义为滤波响应特征的聚类中心, 这个定义使得二维纹元可计算, 被后续研究者广泛采用.

这个二维纹元也就是后来计算机视觉中 BoW 模型中的视觉单词. 考虑到材质纹理在不同成像条件下呈现的三维表现特征, Leung 等将二维纹元定义扩展到三维纹元, 将三维纹元定义为一些具有代表性成像视角和光照条件下的纹理图像的滤波响应矢量串联后的聚类中心.

基于 LM 滤波器, Leung 等的纹理分类方法分为三个步骤, 其实验在原始 CURET 数据库上进行.

a) 三维纹元字典学习. 如图 6 所示. 对训练集中的每类纹理图像, 随机选取不同视角和光照条件下的 20 幅图像, 进行配准后, 分别采用 LM 滤波器进行滤波, 则每幅图像的每个像素点可提取一个 48 维滤波特征矢量, 将 20 幅图像对应像素位置特征矢量进行串联, 则可获得 960 维的 3D 特征矢量; 对所有像素点的 3D 特征矢量聚类, 聚类中心作为三维纹元; 所有训练集的纹理类别学习到的三维纹元一起组成全局三维纹元字典, 以表示所有纹理类别.

b) 纹理类别模型学习. 对每一类训练纹理图像, 与三维纹元字典学习阶段保持一致, 选择同样视点和光照条件的 20 幅纹理图像, 经配准后, 按照同样的方法获得 960 幅滤波特征图, 将每个像素点的 960 维特征按照最近邻原则进行特征编码, 将其编码到

距离最近的三维纹元; 采用简单的直方图统计方法进行特征汇聚, 获得的直方图特征矢量作为该类纹理的模型.

c) 纹理图像分类. 在分类阶段, 对于训练集中给定的纹理类别, 每一类需要重新获得不同于训练集中、但成像视角和光照条件必须与训练集相同的 20 幅新的测试样本图像配准后, 按照纹理类别模型学习中同样的方法获得待分类图像的直方图特征矢量, 与给定纹理类别的模型进行对比, 选择合适的分类器进行分类.

为了学习纹理外观如何随着成像条件的变化而改变, LM 的方法需要预先获得一系列的纹理图像, 以及其光照和成像角度等条件, 且需要进行图像配准, 这一点在实际应用中往往难以实现, 因为现实世界的纹理图像的成像条件常常是未知的. 这种方法不适合对未知成像条件的任意单幅纹理样本图像进行识别. 尽管如此, Leung 等给出了纹元的可计算定义, 在纹理分类领域产生深远影响. 值得注意的是, LM 方法其实就是最原始的计算机视觉领域的 BoW 方法. 我们这里将 LM 方法称为 BoW 方法, 后续研究者的方法大多在于局部特征描述方面不同, 采用的仍然是 BoW 架构. 在 Leung 等^[51] 提出的框架基础上, 采用相同的 LM 滤波器组, Cula 等^[66] 做了改进, 采用 2D 纹元而不是 3D 纹元, 使得能够在没有任何成像环境的先验知识的前提下, 分类单幅纹理图像.

2) MR8 方法

注意到 Leung 等^[51] 的方法的不足, Varma 等^[27] 尝试进行改进, 探索未知成像条件下无需图

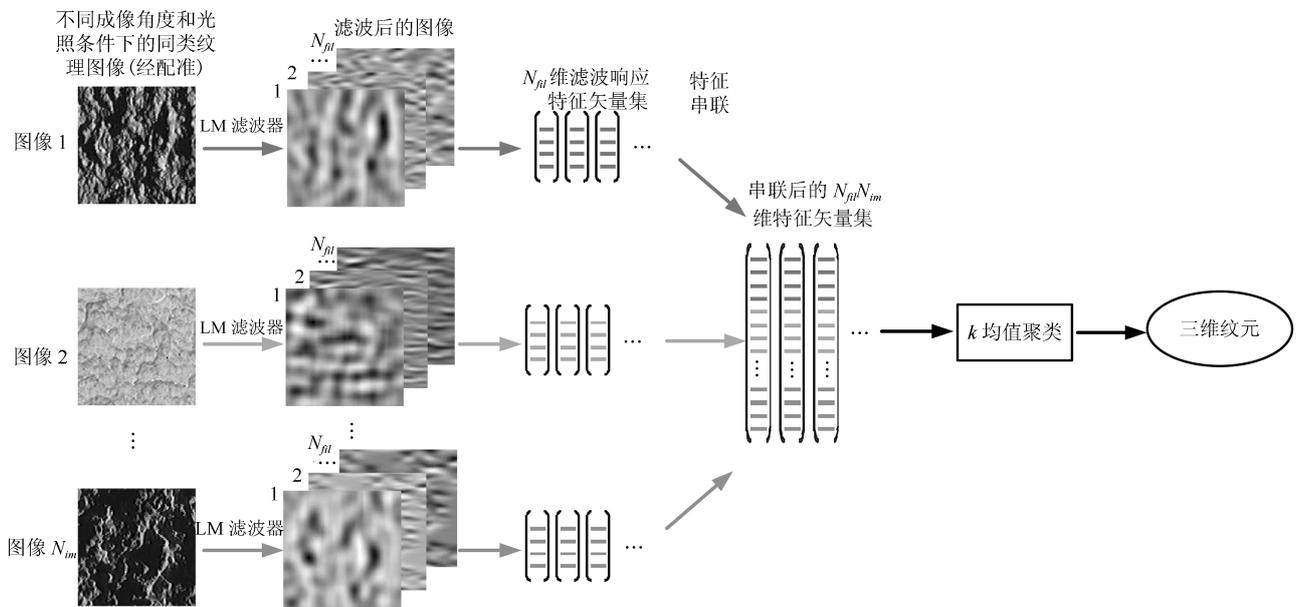


图 6 Leung 和 Malik 提出的三维纹元字典学习流程示意图

Fig. 6 Illustration for 3D texton dictionary learning with LMfilters proposed by Leung and Malik

像配准的单幅纹理样本图像的分类方法, 提出具有旋转不变性的局部纹理特征提取方法—MR8 方法, 其特征其实是基于 LM 滤波器设计的. 不同的是, Varma 等^[27]的方法不再学习三维纹元字典, 而是直接学习二维纹元字典, 而分类过程的其他步骤和 Leung 等^[51]的方法大致相同. 与图 6 类似, 在 Varma 等^[27]提出的方法中, 直接提取滤波特征, 省去特征串联的步骤, 取而代之的是将不同纹理图像的滤波器特征矢量累积, 直接进行特征聚类, 学习二维纹元. 因此, 避开了图像配准的步骤, 无需事先知道图像的成像条件, 并且可以对单幅纹理图像进行分类. MR8 特征由一组基本滤波器 (Base filter set, BFS) 导出, BFS 由来自 LM 滤波器组的 38 个滤波器组成, 保留了 LM 滤波器组中的两种各向异性滤波器: 边缘滤波器、柱状滤波器, 以及两个各项同性的滤波器: 一个高斯滤波器和拉普拉斯高斯滤波器 (参见图 5). 为了获得旋转不变性, 同时降低纹理特征维数, Varma 等从 BFS 滤波特征矢量中推导出 8 个滤波特征矢量, 并称之为 MR8 特征矢量: 各项同性滤波器保持不变, 各项异性的滤波器中, 在同尺度上的不同方向的滤波值中保留最大的滤波响应. Varma 等采用最近邻分类器, 实验表明 MR8 的纹理分类性能明显好于 LM 滤波器组和 Schmid 滤波器组, 成为后续研究者广泛用于性能比较的一个基准方法. 随后, Hayman 等^[28]的工作指出 SVM 可以进一步提升 MR8 特征的纹理分类性能.

3) Patch 特征

滤波方法一直在纹理分析中占有主导性的地位 (可参见 Randen 等^[4]的综述), 前面介绍的三种局部纹理特征描述方法也是滤波方法. Varma 等^[67]挑战滤波方法在纹理分析中的主导地位, 提出一种简单的 Patch 特征. 在局部特征提取阶段, 跳过滤波这一步骤, Varma 等^[67]直接将每个像素点周围的 Patch 表示成一个特征矢量, 输入和 Leung 等^[51]、Varma 等^[27]等采用的一样的 BoW 框架进行纹理分类. 他们的实验表明, Patch 特征的性能优于 MR8 方法, 并对 Patch 特征为何有效做了详细的理论分析. Patch 特征本身不具有旋转不变性, 这一点限制了其在实际中的应用. Varma 等^[67]采用估计局部纹理片主导梯度方向的方法来获得旋转不变性, 然而, 且不说主导梯度方向估计方法本身的鲁棒性就差, 估计每个局部纹理片的主导梯度方向的计算代价也是比较昂贵的. Patch 方法将纹理建模为局部纹理片矢量的联合分布, 认为纹理图像的所有信息应包含在纹理片矢量特征空间中. 而滤波器与图像进行卷积, 可以看成从图像中检测和匹配该滤波器的模式. 一个局部纹理片对这个滤波器响应的强度反映出当前纹理片与滤波器相似的程度, 越

相似则响应越强. 滤波方法等同于将图像纹理片特征空间投影到滤波器所表示的低维子空间中. 如果可以事先知道哪些特征具有较好的纹理类别鉴别力, 那么可以有针对性地设计一组滤波器来匹配这些特征, 这样可以得到最佳效果. 然而, 实际中这种先验信息并不可知, 滤波过程是降维过程, 固然导致局部纹理片中纹理信息的丢失, 因此这可能是 Patch 方法优于 MR8 方法的一个重要原因. 此外, 滤波方法中用到的滤波器组是固定的, 不具有任何学习能力, 因此其特征描述能力也具有局限性. 根据上述分析, 有理由认为, 最好的特征提取方法应该是可以从局部纹理片特征空间中自动学习鉴别力强的特征表达. 事实上, 近年来得到广泛关注的深度学习理论中一个重要的观点就是手工设计的底层特征描述子 (包括滤波器) 作为视觉信息处理的第一步, 往往会过早地丢失有用的信息, 直接从图像像素学习到任务相关的特征描述是比手工特征更为有效.

4) LBP 方法

早在 1996 年, Ojala 等^[68]就提出了 LBP 方法, 至 2002 年该方法发展完善^[30], 此后, LBP 在计算机视觉领域获得广泛关注, 在纹理分类和人脸识别领域得到广泛应用和研究, 催生了大量 LBP 变种方法. 目前 LBP 已经成为纹理分类和人脸识别领域主要的特征提取方法之一. 如想详细了解 LBP 方法及其各种变种方法, 可参见 Liu 等^[11]的综述.

LBP 方法其实也是 BoW 框架下的方法, 与前面 Varma 等^[27,67]采用的 BoW 框架不同之处在于纹元字典学习环节. LBP 不需要采用 k 均值等聚类方法学习纹元, 而是事先定义好了全局纹元字典, 也因此大大降低了特征提取的计算量. 在局部纹理特征提取阶段, LBP 与 Varma 等^[67]的 Patch 方法类似, 也是对中心像素及其邻域系统的联合分布进行量化以获得纹元. 具体来说, LBP 方法考虑中心像素 x_c 及其周围半径为 r 的圆形邻域上等间隔分布的 p 个邻域像素 x_0, x_1, \dots, x_{p-1} 的联合分布 $g(x_c, x_0, x_1, \dots, x_{p-1})$, 将其近似为 $g(x_0 - x_c, x_1 - x_c, \dots, x_{p-1} - x_c)$. 在 LBP 方法中, 不再采用矢量量化方法对此特征空间进行量化, 而是采用特殊的、固定的量化方式, 即将每一项 $x_i - x_c$ 与 0 进行比较, 大于等于 0 则量化为 1; 否则, 量化为 0. 如此特征空间 $(x_0 - x_c, x_1 - x_c, \dots, x_{p-1} - x_c)^T$ 被划分成 2^p 个不同的区域, 也即产生了 2^p 不同的纹元, 而每个局部特征矢量 $(x_0 - x_c, x_1 - x_c, \dots, x_{p-1} - x_c)^T$ 被映射到某一个纹元, 该方法记为 $LBP_{r,p}$. Ojala 等^[30]将其扩展到圆形邻域系统和多尺度分析, 并进一步对 2^p 种不同纹元进行聚类, 提出旋转不变 LBP 算子 $LBP_{r,p}^{ri}$ 、均匀 LBP 算子 $LBP_{r,p}^{u2}$ 以及旋转不变均匀算子 $LBP_{r,p}^{riu2}$, 特征更加鲁棒, 鉴别力增强, 直方

图维数大大降低. LBP 方法存在如下显著优势: 1) 计算复杂度低; 2) 灰度尺度不变性; 3) 易于工程实现; 4) 纹元字典无需训练学习. 此外, LBP 方法可以非常灵活地适应计算机视觉领域的其他问题. LBP 方法的出现, 给纹理分类等问题注入了新的活力, 最近两年仍然有新的 LBP 变种方法出现, 如 Liu 等^[69] 提出的 MRELBP (Median robust extended local binary pattern) 方法, Guo 等^[70] 提出的 SSLBP (Scale selective local binary patterns) 方法, Sulc 等^[71] 提出的 FFfirst (Fast features invariant to rotation and scale of texture) 方法和 Ryu 等^[72] 提出的 scLBP (Sorted consecutive local binary pattern) 方法等. 其中, MRELBP 方法具有计算复杂度低、特征维数较低、对高斯随机噪声、椒盐噪声、随机像素损毁和图像模糊等具有高鲁棒性, 同时在现有主流基准纹理数据集上获得优异分类性能. 此外, 一个值得一提的工作是, 与 LBP 对局部差分特征的量化方式不同, Sharma 等^[73] 提出对每个局部特征矢量 $(x_0 - x_c, x_1 - x_c, \dots, x_{p-1} - x_c)^T$ 特征空间采用混合高斯模型进行建模, 采用 Fisher 特征矢量进行特征汇聚, 称为局部高阶统计方法 (Local higher-order statistics, LHS). 他们的研究中仅仅考虑了一阶邻域系统 (3×3), 因此特征描述能力有限.

5) RP 和 SRP 特征

为解决传统方法在纹理特征提取的复杂度与准确性之间的矛盾, Liu 等^[74] 将随机投影 (Random projection, RP) 和压缩感知 (Compressive sensing, CS) 理论^[75-76] 引入纹理分析领域, 从理论上论证了随机投影和压缩感知在纹理图像分析领域的可行性, 提出一种基于随机投影的纹理分类方法 (称为 RP 方法), 直接挖掘反映图像具有稀疏性的本质特性的特征, 有效地降低特征提取的时间和空间复杂度. 该方法在特征提取阶段, 从原始局部纹理片中提取少数随机投影测量值作为局部纹理特征, 这些非传统的随机特征具有低维、非自适应性、信息无损和普适性等显著特点. 证明了从随机纹理特征中学习得到的纹元具有更好的表达能力. 纹元字典和纹理图像全局模型的学习和训练均在低维压缩域进行, 避开了计算耗时的高维原始纹理数据重构过程^[62-63], 大大地节省了计算时间和数据存储. 随机投影特征相比于传统降维方法如 PCA 等, 对动态数据更新具有适应性、对海量图像特征提取具有可扩展性.

在继承 RP 特征的优势前提下, 为进一步提高 RP 特征的鲁棒性和鉴别力, Liu 等^[77-78] 提出了有序随机投影 (Sorted random projection, SRP) 特征. 常用的获得局部不变性的方案可分为三类. a)

增加训练数据: 将每个局部图像区域旋转到各个角度或者缩放到各个尺度, 加入训练. 这增加了对数据的依赖, 且给数据存储、特征学习和表示以及分类算法增加困难, 尤其当数据规模过大时, 此方法显然不适合. b) 局部区域规则化: 将区域映射到一个固定半径的圆形区域从而获得尺度和仿射不变性. 为了保证旋转不变性, 估计局部图像片的主导方向, 主导方向最常用的方法是梯度方向直方图的最大值对应的方向, 主梯度方向的估计往往是不可靠的, 特别是对于缺乏明显边缘特征的平坦区域或具有多个主梯度方向的角点区域; 而尺度不变性则需要估计每个区域的尺度, 也存在类似的不足. 此外, 逐像素点计算每个局部区域的主导梯度方向也大大增加了计算量. c) 计算对方向不敏感的边缘分布直方图, 例如计算距局部区域中心像素固定距离的邻域像素的直方图, 然后对不同距离的直方图进行级联得到多层直方图, 如 SPIN 描述子, 此类方法计算复杂度高, 且鲁棒性较差. 因此, 设计简单直观、计算复杂度低的有效的具有不变性的局部特征描述子, 具有重要价值. 为了避开上述方法的不足, Liu 等^[77-78] 提出几种具有旋转不变性和灰度尺度不变性的简单、低维而又有高可区分力的 SRP 局部特征表达方法, 如图 7 所示. Liu 等^[77-78] 指出多尺度排序策略 (图 7(c)) 优于全局排序策略 (图 7(a)), 提出从图像局部区域的中心像素的邻域系统提取三种类型特征: 强度特征、径向差分和角向差分 (分别如图 7(c)~(e) 所示), 接着对强度/差分特征进行多尺度排序, 最后采用随机投影进行降维. SRP 可以有效保持局部区域的多尺度结构信息, 也能充分利用两两像素的成对信息. SRP 使得图像局部片特征空间重新分布, 在保持了不同图像类别之间的鉴别力的同时, 产生了更加紧致、对旋转变换、光照变化和尺度变化不敏感的特征聚类, 大大有利于图像视觉词典的学习以及图像全局特征表达. Liu 等^[79] 继续研究了多种 SRP 特征的融合方法.

6) BIF 特征

BIF (Basic image features) 方法^[80] 的主要思想也是将纹理图像表示成纹元字典的无序直方图分布, 与 LBP 方法一样, BIF 方法的纹元字典的构建也不需要聚类学习, 而是基于 Griffin 等^[81-82] 提出的基础图像特征 (BIF), 因此纹理特征提取过程也存在计算复杂度低的优点. Griffin 等^[81-82] 的研究指出高斯滤波器的一阶、二阶导数组成的 5 个简单滤波器 (Derivative of Gaussian, DtG) 可以有效检测图像局部对称性结构特征. 图像对 DtG 滤波器的响应值, 可以大致分为 6 类, 即称为 6 种 BIF 特征. Crosier 等^[80] 基于 BIF 进行

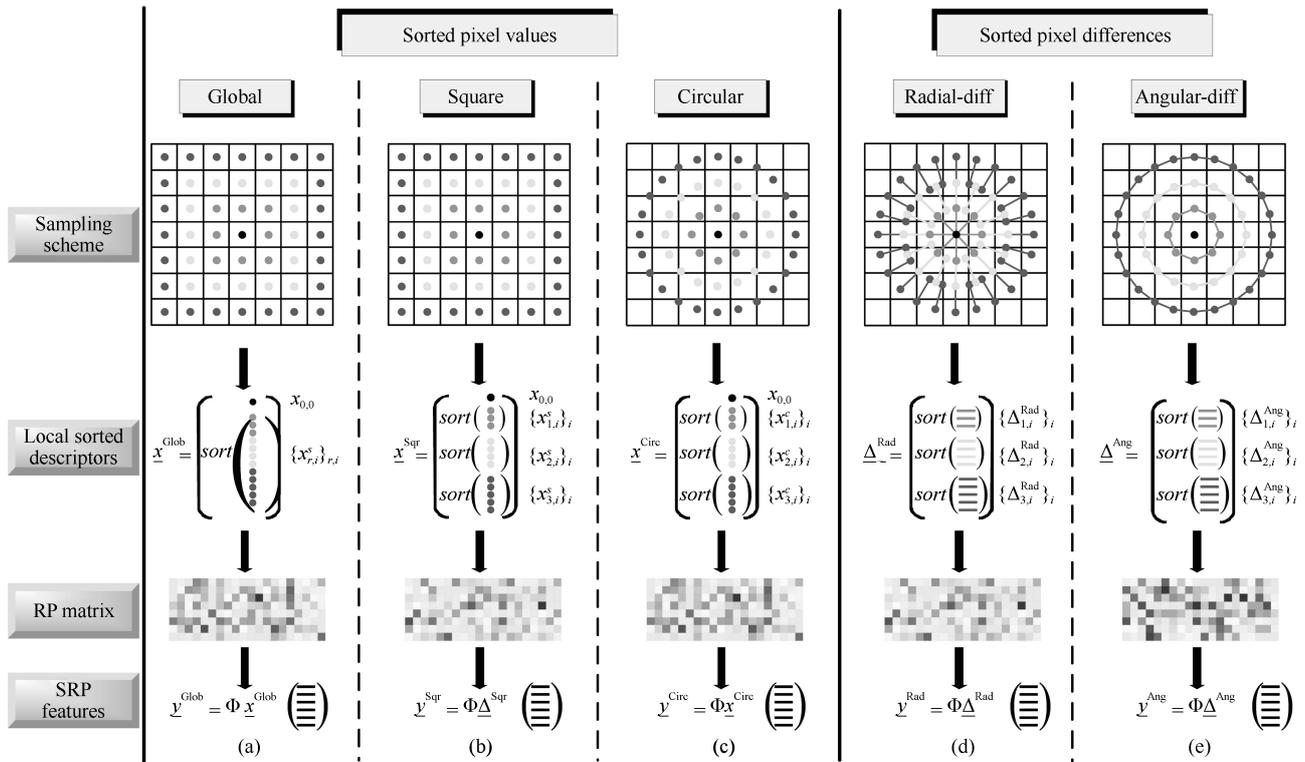


图 7 SRP 描述子示意图

Fig. 7 Illustration of SRP descriptors

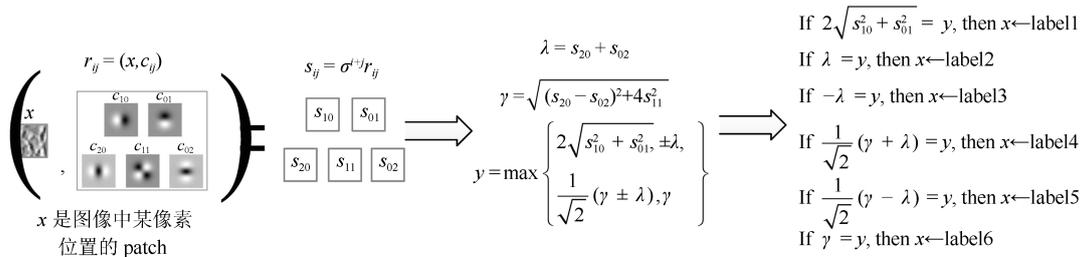


图 8 BIF 局部特征提取过程示意图

Fig. 8 Illustration of BIF feature extraction

有效纹理分类. BIF 特征提取过程可总结为如图 8 所示, 图像中的每个像素经过 DtG 滤波器滤波后, 获得响应值 s_{ij} ; 接着执行图 8 中的第三步, 每个像素处可以计算得到 6 个特征值: $2\sqrt{s_{10}^2 + s_{01}^2}$, λ , $-\lambda$, $(\gamma - \lambda)/\sqrt{2}$, $(\gamma + \lambda)/\sqrt{2}$, γ ; 将该像素点分类为这 6 个值中最大值的那一类². 如此, BIF 方法用简单的方式对滤波响应特征空间进行了量化.

图像中的每个像素点都被分为 6 个标签中的一个, 对标签进行频率直方图统计则可以得到一个全局特征矢量, 不过该特征矢量维数仅仅为 6. Crosier 等^[80] 注意到这个 6 维的特征矢量, 表达能力有限, 纹理分类效果很不理想, 尽管特征自身具有不变性,

但仅仅是单一尺度上的特征提取, 而多尺度分析被认为在纹理分类中起着重要的作用. 为此, Crosier 等^[80] 提取多尺度 BIF 特征, 建议提取 4 个不同尺度的 BIF 特征, 并且计算特征图的联合分布直方图, 可以获得 1 296 (6^4) 维的全局特征表示. Crosier 等^[80] 实验表明, 多尺度 BIF 特征具有较强的特征表达能力, 结合简单的最近邻分类器, 可以达到很好地纹理分类准确率. Timofte 等^[83] 也是基于 BIF 特征, 采用了类似多层金字塔特征汇聚方法和基于稀疏、联合表示的分类方法, 进一步提升了分类性能.

7) WLD 特征

WLD (Weber law descriptor) 局部特征描述

²比如说这 6 个值中最大值为 γ , 则该像素点用 γ 的类别标签 label6 表示.

方法由 Chen 等^[84] 提出, 主要受韦伯定律 (Weber's law) 所启发: 人类对光照、声音等刺激信号的感知辨别不仅与刺激信号强度的绝对变化相关, 而且与刺激信号原来的强度本身有关系. 换言之, 人类对刺激信号强度变化的辨别能力取决于信号变化的相对值, 而不是绝对值. WLD 包括两个组成部分: 刺激变化 (Differential excitation) 部分和方向部分. 前者是基于每个像素与其邻域系统像素的强度变化之和与像素本身强度的比值, 后者是该像素的梯度方向. 基于这两个成分构建两个直方图特征矢量, 作为全局特征表达. WLD 与 LBP 和 SIFT 均有相似之处.

具体地, 如图 9 所示, 刺激变化部分度量的是该像素处邻域系统像素强度变化的相对值: $\xi(x_c) = \arctan[\sum_{i=0}^{p-1} (\frac{x_i - x_c}{x_c})]$.

而方向成分是该像素处的梯度方向: $\theta(x_c) = \arctan \frac{x_7 - x_3}{x_5 - x_1}$.

x_0	x_1	x_2
x_7	x_c	x_3
x_6	x_5	x_4

图 9 WLD 描述子—阶邻域系统示意图

Fig. 9 First order neighborhood in WLD

对 ξ 进行聚类, 将所有的 θ 量化成 T (如 $T = 8$) 个方向. 这样, 可得到两个特征图. 最终统计这两种特征的一种二维直方图作为图像的全局特征表达. WLD 特征支持多分辨分析. WLD 直方图特征矢量在纹理分类和人脸检测中进行了性能测试, 取得不错的效果, 但是其性能与主流纹理特征描述子有些差距, 其优点是思想直观简单, 计算复杂度低.

3.1.3 基于分形的纹理特征描述子

理论上, 分形模型是一种很好地处理多尺度问题的数学模型. 然而, 在早期的纹理特征提取方法中, 分形维数并不是一种很好地纹理特征描述子^[8]. 主要原因是: 1) 自然纹理图像并不是真正意义上的分形, 也即它们不是在任意尺度上都具有相同的结构特征. 2) 分形维数是一个积分度量, 它需要根据图像多个尺度上的信息进行估计, 且分形维数的估计依赖于尺度和估计方法, 具有不稳定性. 3) 传统的基于分形维数的纹理分类方法没有利用纹理的全局统计特性, 缺乏有效性. 4) 分形维数特征对重要的纹理基本特征如边缘、角点和均匀斑块区域等缺乏鉴别力.

近期, 受 BoW 方法的启发, 研究者对分形方法重新进行思考, 提出多分形谱 (Multi-fractal spectrum, MFS) 方法^[34, 85-86], 对纹理图像的视点变化、非刚性变化以及局部光照变化等具有不变性.

最基本的 MFS 方法由 Xu 等^[34] 提出. MFS 首先定义在简单的图像特征上, 如强度特征、梯度特征和 LoG 特征, 不妨假设提取了 n 个特征. 基本思想是, 首先, 将单个图像特征 (图像强度、梯度或 LoG 滤波特征) 利用 k 均值进行聚类, 聚类中心 (即纹元) 数目假设为 k ; 然后对该特征图进行特征编码, 获得纹元标签图; 接着, 将纹元标签图继续分解成 k 个二值特征图: 对每个纹元类别, 把当前纹元类别标记为 1, 其余纹元类别标记为 0, 按照此方式做 k 次可得 k 个的二值图; 对每个二值图, 进行多分辨分析, 采用计盒法估计分形维数, 一共可得 k 个分形维数值, 构成一个 k 维特征矢量, 作为基于该特征导出的分形维数谱特征矢量. 由于, 基于原始图像一共提取了 n 种不同的特征, 那么可以获得 n 个 k 维分形维数谱特征矢量, 将其进行串联, 得到的联合特征矢量则称为 MFS 特征, 维数为 nk . Xu 等^[34] 提出的 MFS 特征仅仅基于最简单的强度和梯度特征, 因此纹理描述能力有限. 随后, 基于此框架, Xu 等^[85-86] 做了进一步改进, 采用更优秀的局部特征描述子如 SIFT、小波变换等代替简单的图像强度或梯度特征, 继而计算 MFS 特征矢量, 提出了 OTF^[85] 和 WMFS 方法^[86], 纹理特征表达能力增强. 近期, Quan 等^[87] 将 MFS 的思想与 LBP 方法相结合. MFS 方法较传统的分形方法纹理分类性能上有了明显提高, 然而还是具有分形方法固有的缺点. 此外, 整个纹理特征提取过程复杂繁琐, 而且依赖于纹理图像类别, 需要高分辨率图像才能获得稳定的 MFS 特征, 对于低分辨率、小尺寸纹理图像分类效果不理想.

3.2 基于深度卷积网络的方法

在过去的几十年, 设计一个模式识别系统往往需要丰富的领域知识来设计特征提取方法, 将原始数据变换成一个合适的特征矢量, 使得分类器能够基于该特征矢量进行分类等任务^[88]. 深度学习是近年来出现的一种具有里程碑意义的数据表示学习技术^[52]. 与传统人工设计特征的重要区别在于, 深度学习的本质是通过多层非线性变换从大数据中以有监督或者无监督的方式学习层次化的特征表达, 对图像进行从底层到高层的描述. 2012 年, 多伦多大学欣顿 (Hinton) 小组采用深度卷积神经网络 (Deep convolutional neural network, DCNN) 赢得了 ImageNet 图像分类的比赛^[53], 且分类准确率大幅领先. Krizhevsky 等^[53] 采用的 DCNN 模型与上世纪 90 年代初期成功用于数字手写识别的 CNN 模型^[54] 结构上基本上是一致的, 之所以赢得比赛, 这与大规模数据集 ImageNet^[89] 的构建以及并行计算技术的发展是分不开的. 自此以后, 深度学习技术开始引起学术界和工业界的广泛关注, 得到广泛应用和研究.

作为深度学习的一个重要类别,短短几年间,DCNN 已经在包括物体识别、图像分类、语音识别等诸多领域都取得了飞跃性的进展.有关深度学习技术的综述性论文很多^[55,90-91],本文从纹理分类问题出发,结合传统纹理滤波特征提取方法来理解 DCNN 方法,并总结近期基于深度卷积网络的纹理分类方法.

传统的基于滤波的纹理特征提取^[4]过程示意图如图 10 所示,通常包括三个步骤:卷积滤波(Filtering,典型如 Gabor 滤波器组)、非线性(Nonlinearity,典型如取模、Sigmoid 函数等)和汇聚(Pooling,如取局部平均值或者高斯加权平均值),得到的特征可以直接用于像素级纹理分类.DCNN 实际上是重复应用卷积滤波、非线性和汇聚三个操作算子.在 DCNN 中,卷积层可以看作滤波器组,其结构随着网络深度的增加而变得更加复杂;常用的非线性操作算子有 Sigmoid、tanh、ReLU 等,可以增强整个网络的表达能力;而其汇聚操作一般是降采样操作,通过取局部区域的最大值或平均值来达到降采样的目的,这样一方面可以忽略图像细微的细节并获得一定的不变性,另一方面可获得较宏观的特征图以便于进一步提取更加整体的特征;在网络的最后通常会增加几个全连通层和一个分类器(如 Softmax 分类器等).DCNN 网络中卷积层的滤波器是各个位置共享的,可以降低参数的规模,这是符合基于滤波器提取图像特征的传统思想的.

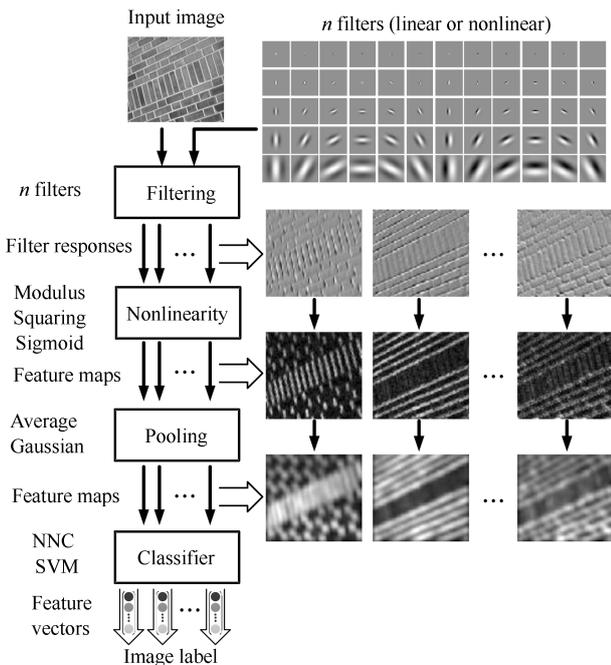


图 10 传统基于滤波方法的纹理分类流程示意图

Fig. 10 Illustration of traditional texture classification based on filtering methods

传统的基于滤波器的纹理特征提取方法实际上

只包含了一个卷积层、一个非线性层和一个汇聚层,且特征提取过程不具有自动学习能力;在采用 BoW 模型时,局部纹理特征提取过程则相当于一个卷积层,对底层特征进行特征编码则相当于非线性层,而特征汇聚操作则等同于汇聚层(或者也可将特征编码和特征汇聚统称为汇聚操作).而 DCNN 中包含多层的由简单逐渐复杂的滤波卷积,可以进行更为复杂的特征变换,并且具有学习能力,其学习过程是有监督进行的,滤波器权重可以根据数据与任务不断进行调整,从而学习到与具体任务更合适的特征表达.因此,不难理解,DCNN 具有更为强大的特征表达能力,能够在多种图像识别任务中展现出出色性能.

目前,深度卷积网络在纹理分类领域的应用研究最有代表性的工作是牛津大学视觉几何研究小组 Cimpoi 等^[39,45,49]提出的 DCNN 特征与传统特征汇聚方法 Fisher vector (FV) 相结合的方法;法国著名数学家 Stéphane Mallat 研究小组提出不变性散射卷积网络(Scattering convolutional network, ScatNet)^[92-94]; Chan 等^[95]提出的一个简单的基于主成分分析的 PCANet、基于随机投影的 RandNet 和基于独立成分分析的 LDANet; Gatys 等^[96]提出的基于 DCNN 特征的 Gram 矩阵、Lin 等^[97-98]提出的双线性 CNN 模型(Bilinear CNN).

1) 基于 DCNN 特征的 FV 特征汇聚

近期不少研究表明,在大规模图像数据库如 ImageNet^[89] 和 MITPlaces^[99] 上训练好的 DCNN 模型可以作为一种通用的图像特征描述方法,并且在计算机视觉的多种应用中取得很好的效果^[100-102]. Cimpoi 等^[39-40,45]将这个思想应用于纹理识别与分割问题,做了一系列的研究工作,最后指出 DCNN 特征与 FV 相结合的方法的性能最好.具体地, Cimpoi 等^[39-40,45]将基于 ImageNet 数据库训练好的 DCNN 模型(考虑了 AlexNet^[53]、VGG-M^[103] 和 VGG-VD^[104] 三种常用模型)直接在纹理图像上进行特征提取,没有基于纹理数据库进行网络微调训练,没有重新训练 DCNN 网络.他们对 DCNN 网络中每一个卷积层提取的局部特征以及最后一个全连接层提取的特征都做了详尽实验分析,结合 BoW 模型中的特征编码和汇聚方法包括传统 BoW (k 均值聚类)、LLC、VLAD 和 FV 等进行全局纹理图像特征表示,他们的结论表明 FV 优于其他编码方法,在很多纹理和材质数据集上取得优秀的纹理分类结果;DCNN 网络层次越深,纹理分类准确率越高,即最后一个卷积层的分类准确率最高,优于全连接层的特征; AlexNet^[53]、VGG-M^[103] 和 VGG-VD^[104] 三种常用模型中, VGG-VD 性能最好. DCNN 特征不足之处在于提取的特征缺乏旋转不变性和光照不

变性, 特征提取过程计算和存储复杂度均较高, 而且全局纹理表示特征矢量维数极高 (高达 65 536 维), 仅适合采用线性 SVM 进行分类. VGG-VD+FV 方法在现有很多基准纹理数据集上达到了最高分类准确率, 在材质数据集如 KTHTIPS2b、FMD、DTD 等上优势更为显著, 并且在物体检测、场景分类和图像微分类等问题上也取得不错的效果, 作者还将该方法用于场景材质分割问题.

2) ScatNet

Bruna 等^[92-94] 提出的 ScatNet 被很多科学家 (如 Yann LeCun 等) 认为是深度学习背后的理论分析和数学解释. 而 ScatNet 最成功的一个应用即为纹理分类问题. 与 DCNN 类似的是, ScatNet 的基本思想也是重复应用如图 10 所示的卷积滤波、非线性和汇聚三个算子. 而 ScatNet 中的滤波器为 Gabor 小波或 Haar 小波, 非线性操作为滤波响应取模, 特征汇聚为取平均值. 选择小波的主要原因是平移不变性和其对图像局部非刚性形变的稳定性.

ScatNet 计算示意图如图 11 所示, 基本原理是: 在第 0 层, 仅对原始图像进行简单的高斯加权平均操作, 信息量很有限; 在第一层, 利用多尺度多方向的 Gabor 小波进行卷积并取模, 模值特征图一方面经过高斯特征加权平均以后作为输出特征图, 另一方面作为下一层输入, 经过 Gabor 小波再次卷积取模. 如此进行. 深层的小波系数包含更为丰富的信

息. Bruna 等^[92] 指出第三层小波系数的能量几乎可以被忽略, 因此 ScatNet 深度往往两层即可; $\psi_{\lambda_{k+1}}$ 的频率要低于 ψ_{λ_k} 频率, 即沿着频率下降的方向进行小波分解, 否则小波系数的能量是几乎可忽略不计的. Bruna 等^[92] 把输出的每幅特征图, 进行全图求和, 多少个特征图就形成一个多少维的特征矢量, 作为纹理图像的最终特征表达, 结合 PCA 分类器, 进行分类. Sifre 等^[93-94] 对 ScatNet 进行了拓展研究, 增加了旋转不变性和尺度不变性特征的提取方法, 以及 Rigid-motion 散射方面的研究, 在多个传统基准纹理数据集 (CURET、UIUC、UMD 和 KTHTIPS 等) 上性能优异.

与 DCNN 不同的是, ScatNet 的网络深度只有两层, 滤波器随着网络深度的增加保持不变; 并且 ScatNet 是预先设计好的小波, 整个网络结构没有学习能力, 这点和传统滤波特征提取方法一样; 但是 ScatNet 更加灵活, 可以设计提取具有各种不变性 (平移不变性、旋转不变性、非局部刚性形变不变性和尺度不变性等) 特征. 然而, ScatNet 的特征提取过程非常耗时, 是目前最耗时的一种纹理特征提取方法, 甚至比基于 CPU 的 VGG-VD 特征提取还要慢. 不过, ScatNet 最后的特征维数比较低, 特征分类过程计算复杂度很低. ScatNet 比较适合于具有平稳性特征的传统纹理分类, 不太适合于平稳性差的材质纹理识

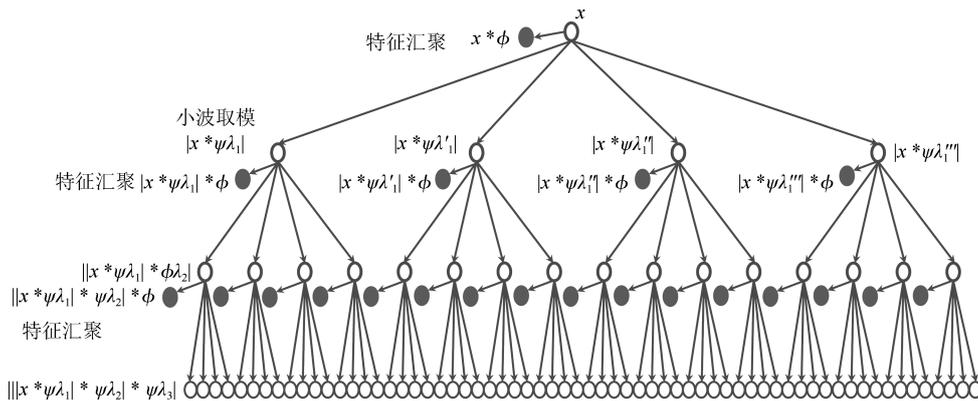


图 11 ScatNet 计算示意图 (图示了三层散射结构. x 为原始图像, ψ 为多尺度多方向的 Gabor 小波 (例如常用的 4 个尺度 8 个方向), 图中可以看成仅画出了 4 个尺度的卷积, 方向滤波器的卷积没有画出; ψ 为随着层深度变化可改变的高斯低通滤波器, 等同于高斯加权平均特征汇聚的作用, 可以获得局部特征不变性; 白色圆点为小波卷积后取模, 用于下一层再次进行小波卷积并取模操作; 黑色圆点是在白色圆点基础上进行局部特征汇聚操作, 为最终输出的特征图)

Fig. 11 Illustration of the 3-level scattering structure of ScatNet (x is the original image, and ψ is the multi-scale and multi-directional Gabor wavelet (e.g., the commonly used five scales and eight orientations). In this figure we only show the convolution in four scales and do not show the convolution in different orientations. ψ is a low-pass Gaussian filter, which changes with the depth of layers, and is equivalent to the feature pooling of the Gaussian weighted average to locally obtain invariance. The white dot is to take modulus after convolution by wavelet, which is then used for the next layer and also take the modulus. The black dot represents feature pooling for the output from the white dot, and then is used as the final feature mapping.)

别, 如 FMD、MINC、DTD 等最新的纹理数据集.

3) PCANet、RandNet 和 LDANet

受 ScatNet^[92-94] 方法的思想所启发, Chan 等^[95] 提出一种类似的方法, 命名为 PCANet. PCANet 与 ScatNet 相比, 如图 11 所示, ScatNet 采用的是事先定义好的小波滤波器如 Gabor 小波, PCANet 的 PCA 滤波器是从图像局部 Patch 特征空间学习得到, 第 $k+1$ 层的 PCA 滤波器基于第 k 层的局部 Patch 空间学习获得; ScatNet 采用的是小波取模操作, PCANet 采用的是 PCA 滤波器; ScatNet 每一层特征图采用高斯加权平均进行局部特征汇聚 (黑色圆点), 然后在此基础上, 对整幅特征图求和, 作为最终特征表达中的一维, 而 PCANet 每一层的特征图没有使用局部特征汇聚操作, 对拥有相同父节点的子节点特征图采用 LBP 方法进行特征汇聚³, 输出的是 LBP 直方图特征矢量, 对同一层的所有 LBP 直方图特征矢量进行串联作为最终特征表达, 输入分类器. 在 PCANet 框架下, 作者还提出了 RandNet 和 LDANet, 分别采用的是随机滤波器和从局部 Patch 空间学习得到的 DCA 滤波器. 在人脸识别数据集、图像分类数据集、纹理数据集和 MNIST 数据集上均进行了测试, 结论表明可以作为一种简单的 Baseline 方法. 与 ScatNet 相比, PCANet 方法特征提取速度要快很多, 尽管 PCA 滤波器的学习过程和滤波过程需要一些时间, 但是后续的 LBP 特征汇聚速度很快. 根据我们自己的实验, PCANet 方法不具有特征不变性, 在很多纹理数据集上的分类性能均比较差.

4) Bilinear CNN

Lin 等^[97-98] 提出 BCNN 的模型框架如图 12 所示, 利用两个 DCNN 网络⁴ 从图像中进行特征提取 (不妨假设两个 DCNN 模型均是 VGG-VD 模型), 在每个模型的最后一个卷积层, 将会输出 $k=512$ 个大小为 $\sqrt{n} \times \sqrt{n}$ 大小的特征图, 即每个像素点均产生一个 512 维的特征矢量, 将每个像素位置对应的两个模型输出的特征矢量做外积, 可获得一个 65 536 (512^2) 维的特征矢量, 外积捕获了不同通道特征之间的两两互信息, 提供了更加充分的表达; 对所有 n 个像素处的 65 536 维特征矢量求和, 进行无序累积以获得平移不变性, 作为最后的特征表达, 然后对其进行符号平方根归一化处理 and l_2 归一化处理, 最后与分类器相连接.

BCNN 模型中使用的两种 DCNN 模型可相同, 也可不同, 可以直接进行端到端的训练, 而 DCNN+FV 方法^[39-40, 45] 则无法执行端到端训练;

BCNN 模型仅使用了卷积层和特征汇聚层, 没有使用全连接层, 因此对输入图像大小没有限制; Lin 等^[97] 还揭示了 BCNN 方法与 BoW、FV 和 VLAD 等方法之间的联系, 但这种联系不是很紧密. BCNN 模型在细微图像分类、纹理分类和人脸识别问题中取得比较成功的应用, 获得一定关注. 其在纹理分类中的性能与 DCNN+FV 方法^[39-40, 45] 接近, 但是也仅仅是利用已有的在 ImageNet 上训练好的 DCNN 模型进行特征提取, 并没有基于纹理数据库进行端对端的训练; 双线性特征矢量维数很高, 带来计算和存储挑战.

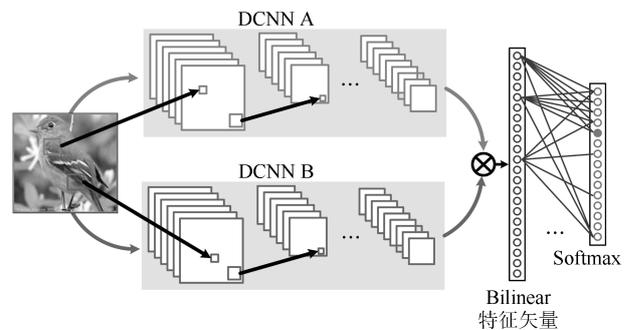


图 12 Bilinear CNN 模型结构示意图

Fig. 12 Illustration of the Bilinear CNN architecture

5) Gram 矩阵

Gatys 等^[96] 提出基于 DCNN 的纹理合成方法, 用到的模型与 Portilla 等^[105] 提出的一个重要纹理合成模型很类似. Gatys 等^[96] 采用在 ImageNet 上训练好的 DCNN 模型作为局部纹理特征分析方法, 从源纹理图像逐层提取所有卷积层的滤波特征, 如图 13 所示, 在网络的每一层, 计算不同滤波响应通道之间的两两相关特征, 即每个像素点的滤波特征矢量与自身做外积, 得到的特征矩阵称为 Gram 矩阵. 值得注意的是, 这里的 Gram 矩阵与 Lin 等^[97-98] 提出的 BCNN 模型中的双线性矢量是一样的. 和 Lin 等^[97-98] 一样, 将同层所有像素的 Gram 矩阵求和, 获得无序统计量. 这样, 所有层的 Gram 矩阵特征组成了源图像的一个平稳的、充分的特征表达. 纹理合成的目的是在图像空间中搜索一幅新的图像, 使得其与源图像在每一层上具有相同的 Gram 矩阵特征表达. 如图 13 所示, G^l 与 \hat{G}^l 分别是源图像 \underline{x} 和待合成图像 $\hat{\underline{x}}$ 的第 l 层的 Gram 矩阵, E_l 是同层的 Gram 矩阵之间的均方距离度量, 那么纹理合成代价函数则是所有层的 E_l 的加权和: $\varepsilon(\underline{x}, \hat{\underline{x}}) = \sum_{l=0}^L \omega_l E_l$. 该方法计算代价比较昂贵, 然而这是基于 DCNN 进行纹理图像合成的首次成功

³LBP 模式的计算是基于 PCA 滤波通道进行的, 而不是同一幅特征图的邻域进行的, 也即同一幅图经过 n 个 PCA 滤波器分解后, 得到 n 幅特征图, 每个像素对应一个 n 维滤波特征矢量, 对每一个 n 维的特征与 0 进行比较, 可获得一个 LBP 模式, 对所有像素的 LBP 模式进行统计可以获得一个维数为 2^n 的直方图特征矢量.

⁴仅由卷积层、非线性层和特征汇聚层组成, 不包含全连接层.

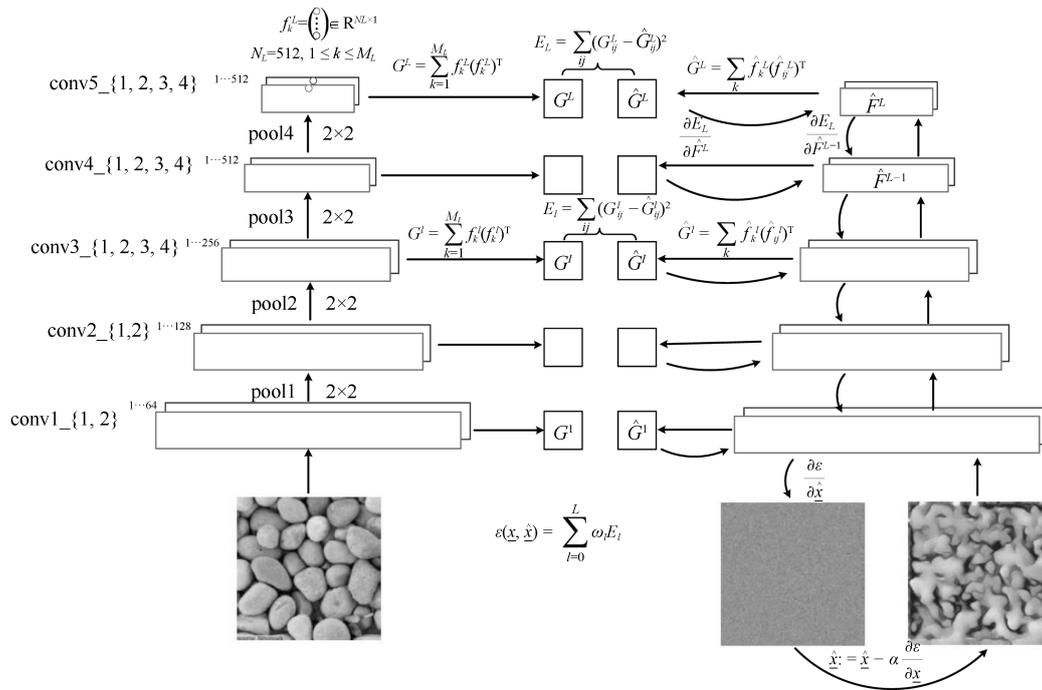


图 13 基于 VGG-VD 模型进行纹理合成示意

Fig. 13 Texture synthesis based on VGG-VD model

尝试, 具有重要的理论意义, 有助于理解深度学习的层次化特征表达, 增强其对纹理特征的可解释性, 开启了一个有意义的研究方向; 此外, 提出的 Gram 矩阵作为纹理特征表达, 作者建议其可以用于纹理分类和物体识别等任务, 而 Lin 等^[97-98]的工作证实了这一点。

3.3 基于纹理视觉属性的方法

近期, 视觉属性的研究成为计算机视觉和机器学习领域的新热点^[41-44]。视觉属性是底层视觉特征的抽象和总结, 是连接底层视觉特征与高层语义类别之间的物体一般化高层描述。底层视觉特征则只能被机器识别, 没有直接语义含义。视觉属性是物体的一种属性, 可以直接被人类视觉感知, 也能被机器理解的较高层次的描述, 例如帽子的形状或者球的颜色。基于视觉属性的目标描述比仅仅采用一个类别标签来表示物体详细得多, 因此在一些高级应用 (如理解语义检索中的复杂查询条件、场景语义理解、自动图像标题生成、迁移学习和领样本学习等) 中起着关键的作用。近期, 李菲菲团队开启了视觉基因计划^[106], 试图通过对图像进行详尽的视觉属性等标注, 推动图像高级语义理解、图像认知和图像问答等问题的研究。尽管目前视觉属性尚属于初步发展阶段, 却是一个非常有前景的研究方向。

纹理视觉属性在物体和场景描述中起着关键的作用, 如图 14 所示, 豹子是“斑点的”动物、斑马和老虎可以描述为“斑纹的”动物等。然而, 关于纹

理视觉属性的研究很少, 尽管可以追溯到 1981 年 Julesz 关于纹理感知的研究工作^[49], 但是后继研究很少。有研究者提出少数几种粗略的纹理属性如粗细度和线性等^[107]来描述纹理。

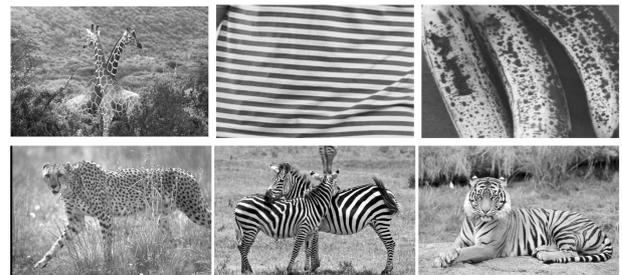


图 14 生活中常见的丰富纹理特征的物体 (可以用纹理视觉属性进行描述: 网状的、斑点的、条纹的、点状的、斑纹的)

Fig. 14 Objects with rich textures in our daily life (We can use texture attributes to describe them: mesh, spotted, striated, spotted, striped.)

究竟人们用哪些语义单词来描述纹理特征图像? 也就是说哪些语义单词可以表达高层纹理视觉属性? 这是关键问题, 然而相关研究寥寥无几。1997 年, Bhushan 等^[108]对此问题开展了研究, 分析了常用的英文单词与纹理感知属性之间的关系, 以及纹理属性与纹理图像之间的对应关系, 给出了一个包含 98 个单词的纹理属性集, 称其可以用来描述大量的纹理模式。Bhushan 等^[108]从心理学角度研究这 98 种纹理属性, 从其字面意义的相似性等进行

聚类, 最后得到 11 种纹理属性. Bhushan 等^[108] 的工作近期得到跟踪研究, Matthews 等^[31] 提出采用 Bhushan 等^[108] 给出的 11 种常用纹理属性来描述纹理. 如果仅对纹理图像进行纹理属性的二值标注, 即判断是否是这 11 种属性中的某种属性, 显然描述能力不足. Matthews 等^[31] 建议采用一个量化值来表达纹理图像对某种纹理属性的归属度, 例如可以用一个数值来表示某纹理图像中纹理模式规则的程度. 为此, 需要估计纹理图像对某种属性归属度的具体方法. 一方面, Matthews 等^[31] 提出直接从底层纹理特征估计纹理属性归属度, 考虑的底层特征包括灰度共生矩阵、Gabor 小波统计量和均匀 LBP 模式; 另一方面, 基于观察者的视觉感知估计纹理属性归属度, 用以衡量基于底层特征估计的方法的准确性. 然而, 要让观测者主观地进行纹理属性归属度的绝对度量是一件困难的事情. 为此, Matthews 等^[31] 采用了相对属性 (Relative attributes)^[43-44] 来标注, 通过比较任意两幅图像对同一种纹理属性的归属度大小进行标注, 然后从相对属性中学习纹理属性的绝对归属度度量. Matthews 等^[31] 研究中采用的底层特征并不能代表纹理特征发展的现状, 因此其结论并没有得到底层纹理特征与较高层的纹理语义属性之间存在紧密相关性, 然而这是这个方向的一个重要尝试, 值得进一步深入研究. 此外, Matthews 等^[31] 的实验表明, 纹理属性归属度特征有助于提高纹理图像检索的性能.

在 Bhushan 等^[108] 的工作基础之上, Cimpoi 等^[40, 45] 也开展了纹理属性方面的初探性研究. 与 Matthews 等^[31] 的研究不同, Cimpoi 等^[40, 45] 的目的是想定义一个通用的、可以描述大多数纹理特征的纹理属性集, 从纹理图像可以鲁棒地估计这些纹理属性, 并探索其应用. 为此, 基于 Bhushan 等^[108] 确定的 98 种纹理属性, 从中选择 47 种更适合纹理特征的属性, 并建立了一个纹理数据库 DTD (参见前面章节). 基于 DTD, Cimpoi 等^[40, 45] 对现有多数纹理特征的分类性能进行了实验比较, 并提出了新的纹理特征 (DCNN+FV), 前文做了介绍, 这里不再赘述.

最后, 表 2 总结了近期主流纹理分类方法在常用的基准纹理数据集上报道的分类正确率. 表 2 中总结的方法在每个数据集上采用了相同的训练/测试数据划分方式. 此外, 近期有研究者对多种 LBP 类型的纹理特征和近期基于深度卷积网络的特征做了比较详尽的纹理分类性能评估, 感兴趣读者可参见^[109].

4 对纹理分类的思考

近期, 纹理分类研究在理论和算法上已经取得

了一系列研究进展, 例如计算机视觉领域最常用的词包模型、Fisher vector 方法、LBP 方法、ScatNet 方法 (被认为是深度学习方面的数学解释) 等, 纹理特征描述子和一般的图像特征描述子的界限逐渐缩小, 很多纹理特征描述子在人脸识别、物体分类和场景分类等问题中也得到较多应用, 促进了相关领域的发展. 然而, 对于实际开放环境中多种多样的自然图像纹理分析和理解, 仍然需要在理论和算法方面开展创新研究, 以更好地学习和表达纹理特征, 并且在完善理论分析的同时, 推进纹理特征提取方法在实际工程中的广泛应用. 其中有待进一步研究解决的重要课题包括:

1) 大规模纹理数据库构建问题. 在大数据时代, 开放环境下的海量图像视频数据存在的纷繁复杂的易变性将给传统人工设计纹理特征带来巨大挑战, 而深度学习模型具有强大的数据表达能力, 无疑将会对大数据背景下的整个视觉的研究产生极大的影响, 也必然会将纹理分类的研究推向新的高度. 深度学习模型依赖于大规模数据库. 遗憾的是, 据我们所知, 目前纹理分类领域尚不存在一个类似于 ImageNet 这样的大规模纹理数据库, 这成为制约纹理分析与理解的一个瓶颈. 如表 2 所示, 很多纹理分类方法在现有的主流纹理数据集上的性能已经趋于饱和, 除了最近的几个数据库如 FMD 和 DTD. 近期的图像分类、物体检测与识别、人脸识别和场景识别等问题取得的重要进展均与相应的大规模数据集如 ImageNet、LFW 和 MITPlaces 的构建密不可分, 因此大规模纹理数据库的构建将使得整个计算机视觉领域受益. 尽管基于已有的大规模图像数据库训练得到的 DCNN 网络在纹理分类中达到了 State of the art 性能, 然而, 大量研究表明, 如果数据条件允许, 重新训练 DCNN 网络或者微调 DCNN 网络均可以进一步提升性能, 有利于学习到能够反映问题本质属性的特征. 因此, 如何构建一个好的大规模纹理数据库, 将是纹理研究进一步发展的一个重要问题.

2) 有关纹理的定义问题. 纹理不存在一个公认的定义, 这已经不是一个新的问题. 然而, 要构建一个大规模的纹理数据库, 这个问题不可回避. 传统意义上来说, 纹理是图像中特征值强度的某种局部模式的重复以及其宏观规律, 局部模式重复和平稳性是其主要特点. 然而, 近期材质识别问题也当作纹理分类问题来研究, 典型的如 FMD 和 MINC 数据集的构建. 而 MIT 研究者 Sharan 等^[38] 指出材质识别是不同于物体识别、场景识别和纹理识别的一个独特问题, 仅仅依靠纹理特征并不能解决材质分类问题. 因此, 有关纹理的定义, 仍然是一个值得大家思考的问题, 对于建立一个好的大规模纹理数据库

表 2 近期主流分类方法报道的纹理分类性能总结 (数据都是原文报道的结果, 带 * 标记的数据是引自近期综述性论文^[6])
 Table 2 Performance summary of recent dominant classification methods on texture classification (All results are quoted directly from original papers, except for those marked with *, which are from a recent review paper^[6].)

Method	Dataset	Outex_TC10	Outex_TC12	Brodatz	CUReT	KTHTIPS	UIUC	UMD	KTHTIPS2	ALOT	FMD	DTD	
LBP ^[30]	TPAMI2002	96.1	97.2										
MRS ^[27]	IJCV2005					97.4							
Lazebnik et al. ^[33]	TPAMI2005			88.2	72.5*	91.3*	96.0						
Zhang et al. ^[6]	IJCV2007			95.4	95.3	95.5	98.7						
Mellor et al. ^[110]	TPAMI2008			89.7									
MFS ^[34]	IJCV2009							92.7	93.9				
OTF ^[85]	CVPR2009							97.4	98.5				
WMFS ^[86]	CVPR2010							98.6	98.7				
Patch ^[67]	TPAMI2009			92.9*	98.0	92.4*	97.8						
WLD ^[84]	TPAMI2009									64.7			
BIF ^[80]	IJCV2010				98.6	98.5	98.8						
RP ^[74]	TPAMI2012				98.5								
SRP ^[77]	PR2012			96.3	98.5	97.7	96.3	99.1					
Timofte et al. ^[83]	BMVC2012			97.3	99.4	99.4	99.0	99.5				55.8	
Ce Liu ^[37]	IJCV2013											55.6	
ScatNet ^[93]	CVPR2013		98.8		99.8	99.4	99.4	99.7					
SRP-RCA ^[79]	TCSVT2015			96.8	99.4	99.1	98.6	99.3				53.2	
PCANet ^[95]	TIP2015				99.6								
MRELBP ^[69]	TIP2016	99.8	99.6		99.0			99.4	77.9	99.1			
AlexNet+FV ^[40]	IJCV2016				98.5			99.2	99.7	77.9	99.1	67.2	62.9
VGG-M+FV ^[40]	IJCV2016				98.7			99.6	99.9	79.9	99.4	73.5	66.8
VGG-VD+FV ^[40]	IJCV2016				99.0			99.9	99.9	88.2	99.5	79.8	72.3
BCNN ^[98]	CVPR2016								77.9			81.6	72.9

也起着关键作用。

3) 适合纹理分析与理解的深度卷积神经网络研究. DCNN 已经在图像分类、物体检测和识别等领域取得巨大成功, 直观上来看, DCNN 的网络结构很好地符合这样一个客观事实: 物体是层级组成的, 高级的特征是通过对低级特征的组合来实现的. 具体而言, 局部边缘的组合形成基本图案, 这些图案形成物体的局部, 继而再形成整个物体. DCNN 学习到的特征也很好地反映了这样一个事实, 其结构其实较适合于物体分类等问题. 但是, 这种图像整体形状的分析 and 越来越复杂的特征未必适合于纹理分析, 这仍然要回归到纹理图像的本质特性问题, 纹理分类更加关注的是图像中存在的一些简单的局部模式, 以及这些局部模式的较有规律的重现. 注意到 DCNN 提取的卷积层的特征其实与纹理分析领域常用的滤波方法非常类似, 比如第一层提取的边缘特征, 类似于纹理分析中常用的 Gabor 滤波器. 中间层也是类似于滤波方法, 只是提取特征复杂度增加. 因此, 我们有理由推测一个典型的 DCNNs 网络, 即使不改变其网络结构, 将其直接在一个纹理大数据

库上进行训练, 可以自动地学习到一些能够反映纹理图像本质属性的特征, 也有助于增强 DCNN 网络的可解释性, 通过纹理特征分析、纹理合成和可视化分析等方法深入分析纹理特征在 DCNN 网络中究竟是如何表达的. 我们相信应该能够设计出适合纹理分析的 DCNN 网络结构, 这也是一个值得探索的研究方向.

4) 纹理图像的高效分类方法研究. 目前常用的纹理分类方法是采用基于 ImageNet 训练好的 DCNN 模型作为特征提取, 然后利用 Fisher 向量将图像表示成一个高维特征矢量, 通常为几万维甚至更高. 随着大数据时代的到来、硬件技术的发展, 也使得在更大规模的数据库上进行研究和评测成为必然. 高维度导致高存储需求和高计算复杂度, 特别是资源受限的领域 (如移动设备上的识别) 中, 目前主流的纹理图像分类方法无法胜任, 因此还需要继续努力研究面向大规模纹理图像分类的高效特征提取方法. 在处理大规模的图像数据库类似于 ImageNet, 需要很大的内存, 则需要并行计算、云计算等大规模计算平台, 大规模计算平台成本昂贵. 因

此,设计具有较低维度、但是表示能力和现有方法相似(至少不显著降低)的纹理特征表示方法,具有重要意义。

虽然深度学习方法让很多问题取得显著进展,然而目前深度学习模型还存在着模型复杂度高,计算强度高问题,基本都需要 GPU 硬件的额外提速。因此,设计简单有效的 DCNN 结构,也是一个值得探索的研究方向。可以继续对传统纹理特征提取方法如 LBP 和 ScatNet 等开展深入研究,有可能启发我们设计出性能更好的深度模型,例如 Courbariaux 等^[111]小组近期关于二值化 CNN 方面的工作。

5) 开放环境下鲁棒纹理分类问题。实际应用中,光照变化、图像旋转、尺度变化、非刚性形变、局部遮挡、图像随机噪声(高斯随机噪声、椒盐噪声)、图像模糊等因素,都给纹理分类方法带来很大挑战,因此增强纹理特征的鲁棒性和易用性也是面向实际应用必须考虑的一个问题。现有的纹理分类方法对图像噪声和图像模糊等因素的影响,考虑还比较少。针对上面提到的各种环境干扰因素,也缺乏一个整体的架构把已有的各种方法对这些因素的鲁棒性做一个全面性的比较和实验评估。

6) 纹理图像语义理解问题。ImageNet 项目的创建者李菲菲教授近期已经开启视觉基因项目^[106],目的是为了对图像进行深度语义分析、理解和认知,其中图像视觉属性方面的研究就起着一个关键的作用。而纹理作为图像的一个基本特征,纹理视觉属性的研究也具有重要的意义。然而纹理视觉属性方面的研究目前还处于起步阶段,也有许多问题值得思考和解决。首先,在纹理视觉属性定义方面,哪些是人类常用来描述纹理特征的纹理视觉属性?是否存在一个通用的、公认的纹理视觉属性库?其次,在纹理视觉属性描述方面,属性描述与底层特征之间的对应关系如何?如何结合底层特征来提高识别准确度?如何将属性集用于描述具体对象?在小样本学习方面,如何利用现有的属性特征库来“扩充样本”,构建一个更为精确的识别模型?

5 结论

纹理分类是计算机视觉与模式识别领域的一个基础问题,在计算机视觉研究中具有重要的理论意义和实际应用价值,同时目前也存在诸多困难与挑战。本文对纹理分类领域主流数据库进行了总结和评述,对近期纹理特征提取方法进行了详细的梳理和评述,对主流方法进行了详尽的阐述,并揭示了其间内在联系。以此为基础,对纹理分类未来的发展方向进行了分析与展望。我们有理由相信,纹理分类领域的发展必然会促进计算机视觉领域相关问题的发

展,乃至整个计算机视觉领域的发展。

References

- 1 Julesz B. Visual pattern discrimination. *IRE Transactions on Information Theory*, 1962, **8**(2): 84–92
- 2 Tuceryan M, Jain A K. Texture analysis. *Handbook of Pattern Recognition and Computer Vision*. Singapore: World Scientific, 1993. 235–276
- 3 Reed T R, Dubuf J M H. A review of recent texture segmentation and feature extraction techniques. *CVGIP: Image Understanding*, 1993, **57**(3): 359–372
- 4 Randen T, Husoy J H. Filtering for texture classification: a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, **21**(4): 291–310
- 5 Zhang J G, Tan T N. Brief review of invariant texture analysis methods. *Pattern Recognition*, 2002, **35**(3): 735–747
- 6 Zhang J G, Marszalek M, Lazebnik S, Schmid C. Local features and kernels for classification of texture and object categories: a comprehensive study. *International Journal of Computer Vision*, 2007, **73**(2): 213–238
- 7 Xie X H, Mirmehdi M. A galaxy of texture features. *Handbook of Texture Analysis*. London: Imperial College Press, 2008. 375–406
- 8 Liu Li, Kuang Gang-Yao. Overview of image textural feature extraction methods. *Journal of Image and Graphics*, 2009, **14**(4): 622–635
(刘丽, 匡纲要. 图像纹理特征提取方法综述. *中国图象图形学报*, 2009, **14**(4): 622–635)
- 9 Pietikäinen M, Zhao G Y. Two decades of local binary patterns: a survey. *Advances in Independent Component Analysis and Learning Machines*. Amsterdam, The Netherlands: Elsevier, 2015.
- 10 Song Ke-Chen, Yan Yun-Hui, Chen Wen-Hui, Zhang Xu. Research and perspective on local binary pattern. *Acta Automatica Sinica*, 2013, **39**(6): 730–744
(宋克臣, 颜云辉, 陈文辉, 张旭. 局部二值模式方法研究与展望. *自动化学报*, 2013, **39**(6): 730–744)
- 11 Liu L, Fieguth P, Guo Y L, Wang X G, Pietikäinen M. Local binary features for texture classification: taxonomy and experimental study. *Pattern Recognition*, 2017, **62**: 135–160
- 12 Brodatz P. *Textures: A Photographic Album for Artists and Designers*. New York: Dover, 1966.
- 13 Brodatz textures [Online], available: <http://www.ux.uis.no/~tranden/brodatz.html>, April 18, 2018
- 14 Vision texture [Online], available: <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>, April 18, 2018
- 15 CURET: columbia-Utrecht reflectance and texture database [Online], available: <http://www.cs.columbia.edu/CAVE/software/curet/html/about.php>, April 18, 2018
- 16 Outex texture database [Online], available: http://www.outex oulu.fi/index.php?page=outex_home, April 18, 2018

- 17 The KTH-TIPS and KTH-TIPS2: image databases [Online], available: <http://www.nada.kth.se/cvap/databases/kth-tips/download.html>, April 18, 2018
- 18 UIUC Database [Online], available: http://www-cvr.ai.uiuc.edu/ponce_grp/data/, April 18, 2018
- 19 Viewpoint invariant texture description [Online], available: <http://www.cfar.umd.edu/~fer/website-texture/texture.htm>, April 18, 2018
- 20 ALOT Database [Online], available: http://aloi.science.uva.nl/public_alot/, April 18, 2018
- 21 Flickr material database (FMD) [Online], available: <http://people.csail.mit.edu/ce-liu/CVPR2010/FMD/>, April 18, 2018
- 22 DRExel Database [Online], available: <https://www.cs.drexel.edu/~kon/texture/>, April 18, 2018
- 23 Bell S, Upchurch P, Snavely N, Bala K. OpenSurfaces [Online], available: <http://opensurfaces.cs.cornell.edu/>, April 18, 2018
- 24 Describable textures dataset (DTD)[Online], available: <http://www.robots.ox.ac.uk/~vgg/data/dtd/>, April 18, 2018
- 25 Bell S, Upchurch P, Snavely N, Bala K. Material recognition in the wild with the materials in context database [Online], available: <http://opensurfaces.cs.cornell.edu/publications/minc/>, April 18, 2018
- 26 Dana K J, Van Ginneken B, Nayar S K, Koenderink J J. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics*, 1999, **18**(1): 1–34
- 27 Varma M, Zisserman A. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 2005, **62**(1–2): 61–81
- 28 Hayman E, Caputo B, Fritz M, Eklundh J O. On the significance of real-world conditions for material classification. In: Proceedings of the 8th European Conference on Computer Vision. Prague, Czech: Springer, 2004. 253–266
- 29 Ojala T, Mäenpää T, Pietikäinen M, Viertola J, Kyllönen J, Huovinen S. Outex-new framework for empirical evaluation of texture analysis algorithms. In: Proceedings of the 16th International Conference on Pattern Recognition. Quebec City, Canada: IEEE, 2002. 701–706
- 30 Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(7): 971–987
- 31 Matthews T, Nixon M S, Niranjan M. Enriching texture analysis with semantic data. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, USA: IEEE, 2013. 1248–1255
- 32 Lazebnik S, Schmid C, Ponce J. A sparse texture representation using affine-invariant regions. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Madison, Wisconsin, USA: IEEE, 2003. II-319–II-324
- 33 Lazebnik S, Schmid C, Ponce J. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(8): 1265–1278
- 34 Xu Y, Ji H, Fermüller C. Viewpoint invariant texture description using fractal analysis. *International Journal of Computer Vision*, 2009, **83**(1): 85–100
- 35 Burghouts G J, Geusebroek J M. Material-specific adaptation of color invariant features. *Pattern Recognition Letters*, 2009, **30**(3): 306–313
- 36 Oxholm G, Bariya P, Nishino K. The scale of geometric texture. In: Proceedings of the 12th European Conference on Computer Vision (ECCV). Florence, Italy: Springer, 2012. 58–71
- 37 Sharan L, Liu C, Rosenholtz R, Adelson E H. Recognizing materials using perceptually inspired features. *International Journal of Computer Vision*, 2013, **103**(3): 348–371
- 38 Sharan L, Rosenholtz R, Adelson E H. Accuracy and speed of material categorization in real-world images. *Journal of Vision*, 2014, **14**(9): Article No. 12
- 39 Cimpoi M, Maji S, Vedaldi A. Deep filter banks for texture recognition and segmentation. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, Massachusetts, USA: IEEE, 2015. 3828–3836
- 40 Cimpoi M, Maji S, Kokkinos I, Vedaldi A. Deep filter banks for texture recognition, description, and segmentation. *International Journal of Computer Vision*, 2016, **118**(1): 65–94
- 41 Farhadi A, Endres I, Hoiem D, Forsyth D. Describing objects by their attributes. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, Florida, USA: IEEE, 2009. 1778–1785
- 42 Patterson G, Xu C, Su H, Hays J. The SUN attribute database: beyond categories for deeper scene understanding. *International Journal of Computer Vision*, 2014, **108**(1–2): 59–81
- 43 Parikh D, Grauman K. Relative attributes. In: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011. 503–510
- 44 Kumar N, Berg A, Belhumeur P N, Nayar S. Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(10): 1962–1977
- 45 Cimpoi M, Maji S, Kokkinos I, Mohamed S, Vedaldi A. Describing textures in the wild. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 3606–3613
- 46 Bell S, Upchurch P, Snavely N, Bala K. Material recognition in the wild with the materials in context database. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, Massachusetts, USA: IEEE, 2015. 3479–3487

- 47 Bell S, Upchurch P, Snavely N, Bala K. OpenSurfaces: a richly annotated catalog of surface appearance. *ACM Transactions on Graphics*, 2013, **32**(4): Article No. 111
- 48 Hossain S, Serikawa S. Texture databases — a comprehensive survey. *Pattern Recognition Letters*, 2013, **34**(15): 2007–2022
- 49 Julesz B. Textons, the elements of texture perception, and their interactions. *Nature*, 1981, **290**(5802): 91–97
- 50 Csurka G, Dance C, Fan L, Willamowski J, Bray C. Visual categorization with bags of keypoints. In: Proceedings of the 2004 Workshop on Statistical Learning in Computer Vision. Prague, Czech: ECCV, 2004. 1–22
- 51 Leung T, Malik J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 2001, **43**(1): 29–44
- 52 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, **313**(5786): 504–507
- 53 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS). Lake Tahoe, Nevada, USA: ACM, 2012. 1097–1105
- 54 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324
- 55 Huang Kai-Qi, Ren Wei-Qiang, Tan Tie-Niu. A review on image object classification and detection. *Chinese Journal of Computers*, 2014, **36**(6): 1225–1240
(黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述. 计算机学报, 2014, **36**(6): 1225–1240)
- 56 Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, et al. A comparison of affine region detectors. *International Journal of Computer Vision*, 2005, **65**(1–2): 43–72
- 57 Tuytelaars T, Mikolajczyk K. Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 2008, **3**(3): 177–280
- 58 Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1615–1630
- 59 Huang Y Z, Wu Z F, Wang L, Tan T N. Feature coding in image classification: a comprehensive study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(3): 493–506
- 60 Perronnin F, Larlus D. Fisher vectors meet neural networks: a hybrid classification architecture. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 3743–3752
- 61 Sánchez J, Perronnin F, Mensink T, Verbeek J. Image classification with the fisher vector: theory and practice. *International Journal of Computer Vision*, 2013, **105**(3): 222–245
- 62 Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 2006, **54**(11): 4311–4322
- 63 Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A. Discriminative learned dictionaries for local image analysis. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Anchorage, Alaska, USA: IEEE, 2008. 1–8
- 64 Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2006. 2169–2178
- 65 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
- 66 Cula O G, Dana K J. 3D texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 2004, **59**(1): 33–60
- 67 Varma M, Zisserman A. A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(11): 2032–2047
- 68 Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 1996, **29**(1): 51–59
- 69 Liu L, Lao S Y, Fieguth P W, Guo Y L, Wang X G, Pietikäinen M. Median robust extended local binary pattern for texture classification. *IEEE Transactions on Image Processing*, 2016, **25**(3): 1368–1381
- 70 Guo Z H, Wang X Z, Zhou J, You J N. Robust texture image representation by scale selective local binary patterns. *IEEE Transactions on Image Processing*, 2016, **25**(2): 687–699
- 71 Sulc M, Matas J. Fast features invariant to rotation and scale of texture. In: Proceedings of the 2014 European Conference on Computer Vision (ECCV). Zurich, Switzerland: Springer, 2014. 47–62
- 72 Ryu J, Hong S, Yang H S. Sorted consecutive local binary pattern for texture classification. *IEEE Transactions on Image Processing*, 2015, **24**(7): 2254–2265
- 73 Sharma G, Juriea F. Local higher-order statistics (LHS) describing images with statistics of local non-binarized pixel patterns. *Computer Vision and Image Understanding*, 2016, **142**: 13–22
- 74 Liu L, Fieguth P. Texture classification from random features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(3): 574–586
- 75 Candes E J, Tao T. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Transactions on Information Theory*, 2006, **52**(12): 5406–5425
- 76 Donoho D L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, **52**(4): 1289–1306

- 77 Liu L, Fieguth P, Clausi D, Kuang G Y. Sorted random projections for robust rotation-invariant texture classification. *Pattern Recognition*, 2012, **45**(6): 2405–2418
- 78 Liu L, Fieguth P, Kuang G Y, Zha H B. Sorted random projections for robust texture classification. In: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011. 391–398
- 79 Liu L, Fieguth P W, Hu D W, Wei Y M, Kuang G Y. Fusing sorted random projections for robust texture and material classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, **25**(3): 482–496
- 80 Crosier M, Griffin L D. Using basic image features for texture classification. *International Journal of Computer Vision*, 2010, **88**(3): 447–460
- 81 Griffin L D, Lillholm M. Symmetry sensitivities of derivative-of-Gaussian filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(6): 1072–1083
- 82 Griffin L D, Lillholm M, Crosier M, Van Sande J. Basic image features (BIFs) arising from approximate symmetry type. In: Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision (SSVM). Voss, Norway: Springer, 2009. 343–355
- 83 Timofte R, Van Gool L. A training-free classification framework for textures, writers, and materials. In: Proceedings of the 23rd British Machine Vision Conference (BMVC). Surrey, Guildford, UK: BMVA, 2012.
- 84 Chen J, Shan S G, He C, Zhao G Y, Pietikainen M, Chen X L, et al. WLD: a robust local image descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1705–1720
- 85 Xu Y, Huang S B, Ji H, Fermüller C. Combining powerful local and global statistics for texture description. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, FL, USA: IEEE, 2009. 573–580
- 86 Xu Y, Yang X, Ling H B, Ji H. A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, California, USA: IEEE, 2010. 161–168
- 87 Quan Y H, Xu Y, Sun Y P, Luo Y. Lacunarity analysis on image patterns for texture classification. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, Ohio, USA: IEEE, 2014. 160–167
- 88 Jain A K, Duin R P W, Mao J C. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(1): 4–37
- 89 Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S A, et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015, **115**(3): 211–252
- 90 Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(8): 1798–1828
- 91 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, **521**(7553): 436–444
- 92 Bruna J, Mallat S. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(8): 1872–1886
- 93 Sifre L, Mallat S. Rotation, scaling and deformation invariant scattering for texture discrimination. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, Oregon, USA: IEEE, 2013. 1233–1240
- 94 Sifre L, Mallat S. Rigid-motion scattering for texture classification. *International Journal of Computer Vision*, 2014.
- 95 Chan T H, Jia K, Gao S H, Lu J W, Zeng Z N, Ma Y. PCANet: a simple deep learning baseline for image classification? *IEEE Transactions on Image Processing*, 2015, **24**(12): 5017–5032
- 96 Gatys L A, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS). Montreal, Canada: MIT Press, 2015. 262–270
- 97 Lin T Y, RoyChowdhury A, Maji S. Bilinear CNN models for fine-grained visual recognition. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 1449–1457
- 98 Lin T Y, Maji S. Visualizing and understanding deep texture representations. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 2791–2799
- 99 Zhou B L, Lapedriza A, Xiao J X, Torralba A, Oliva A. Learning deep features for scene recognition using places database. In: Proceedings of the 2014 Advances in Neural Information Processing Systems (NIPS). Montreal, Canada: Neural Information Processing Systems, 2014.
- 100 Razavian A S, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR). Columbus, Ohio, USA: IEEE, 2014. 512–519
- 101 Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, Ohio, USA: IEEE, 2014. 580–587
- 102 Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. OverFeat: integrated recognition, localization and detection using convolutional networks. In: Proceedings of the 2014 International Conference on Learning Representation (ICLR). Banff, Canada: ICLR, 2014.
- 103 Chatfield K, Simonyan K, Vedaldi A, Zisserman A. Return of the devil in the details: delving deep into convolutional nets. In: Proceedings of the 2014 British Machine Vision Conference (BMVC). Nottingham, UK: BMVA, 2014.

- 104 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 2015 International Conference on Learning Representations (ICLR). San Diego, CA, USA: ICLR, 2015.
- 105 Portilla J, Simoncelli E P. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 2000, **40**(1): 49–70
- 106 Krishna R, Zhu Y K, Groth O, Johnson J, Hata K, Kravitz J, et al. Visual genome: connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 2017, **123**(1): 32–73
- 107 Tamura H, Mori S, Yamawaki T. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, 1978, **8**(6): 460–473
- 108 Bhushan N, Rao A R, Lohse G L. The texture lexicon: understanding the categorization of visual texture terms and their relationship to texture images. *Cognitive Science*, 1997, **21**(2): 219–246
- 109 Liu L, Fieguth P, Wang X G, Pietikäinen M, Hu D W. Evaluation of LBP and deep texture descriptors with a new robustness benchmark. In: Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer, 2016. 69–86
- 110 Mellor M, Hong B W, Brady M. Locally rotation, contrast, and scale invariant descriptors for texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(1): 52–61
- 111 Hubara I, Courbariaux M, Soudry D, El-Yaniv R, Bengio Y. Binarized neural networks, Advances in neural information processing systems, 2016. 4107–4115



刘 丽 国防科学技术大学信息系统与管理学院副教授. 主要研究方向为图像理解, 计算机视觉, 模式识别. 本文通信作者.

E-mail: liuli_nudt@nudt.edu.cn

(**LIU Li** Associate professor at the College of Information System and Management, National University of Defense Technology. Her research interest covers image understanding, computer vision, and pattern recognition. Corresponding author of this paper.)



赵凌君 国防科学技术大学电子科学与工程学院副教授. 主要研究方向为遥感信息处理, 合成孔径雷达目标自动识别. E-mail: nudtzlj@163.com

(**ZHAO Ling-Jun** Associate professor at the School of Electronic Science and Engineering, National University of Defense Technology. Her research interest covers remote sensing information processing, and SAR automatic target recognition.)



郭承玉 国防科学技术大学信息系统与管理学院博士研究生. 主要研究方向为图像理解, 计算机视觉, 模式识别. E-mail: sdlwgcy@126.com

(**GUO Cheng-Yu** Ph.D. candidate at the College of Information System and Management, National University of Defense Technology. Her research interest covers image understanding, computer vision, and pattern recognition.)



王 亮 中国科学院自动化研究所模式识别国家重点实验室研究员. 主要研究方向计算机视觉与模式识别. E-mail: wangliang@nlpr.ia.ac.cn

(**WANG Liang** Professor at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research interest covers computer vision and pattern recognition.)



汤 俊 国防科学技术大学信息系统与管理学院信息工程国防科技重点实验室讲师. 主要研究方向为智能交通系统, 航迹规划, 深度学习. E-mail: jun.tang@e-campus.uab.cat

(**TANG Jun** Lecturer at the Science and Technology on Information Systems Engineering Laboratory, College of Information System and Management, National University of Defense Technology. His research interest covers intelligent transportation systems, trajectory planning, and machine learning.)