

基于空间语义对象混合学习的复杂图像场景自动分类方法研究

孙显* 付琨 王宏琦

(中国科学院电子学研究所 北京 100190)

(中国科学院空间信息处理与应用系统技术重点实验室 北京 100190)

摘要: 场景分类是将多幅图像标记为不同语义类别的过程。该文针对现有方法对复杂图像场景分类性能不佳的不足,提出一种新的基于空间语义对象混合学习的复杂图像场景分类方法。该方法以多尺度分割得到的图像对象而非整幅图像为主体进行产生式语义建模,统计各类有效特征挖掘对象的类别分布信息,并通过空间金字塔匹配,构建包含层次数据和语义信息的中间向量,弥补语义鸿沟的缺陷,训练中还结合判别式学习提高分类器的可信性。在实验数据集上的结果表明该方法具备较高的学习性能和分类精度,适用于多种类型和复杂内容图像的解译,具有较强的实用价值。

关键词: 图像处理; 场景分类; 语义对象; 混合学习; 金字塔匹配

中图分类号: TN911.73

文献标识码: A

文章编号: 1009-5896(2011)02-0347-08

DOI: 10.3724/SP.J.1146.2010.00361

Spatial Semantic Objects-based Hybrid Learning Method for Automatic Complicated Scene Classification

Sun Xian Fu Kun Wang Hong-qi

(Institute of Electronic, Chinese Academy of Sciences, Beijing 100190, China)

(Key Laboratory of Spatial Information Processing and Application System Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Scene image classification refers to the task of grouping different images into semantic categories. A new spatial semantic objects-based hybrid learning method is proposed to overcome the disadvantages existing in most of the relative methods. This method uses generative model to deal with the objects obtained by multi-scale segmentation instead of whole image, and calculates kinds of visual features to mine the category information of every objects. Then, an intermediate vector is generated using spatial-pyramid matching algorithm, to describe both the layer data and semantic information and narrow down the "semantic gap". The method also combines a discriminative learning procedure to train a more confident classifier. Experimental results demonstrate that the proposed method can achieve high training efficiency and classification accuracy in interpreting manifold and complicated images.

Key words: Image processing; Scene classification; Semantic object; Hybrid learning; Pyramid matching

1 引言

场景分类是将多幅图像标记为不同语义类别的过程。场景分类技术已在计算机视觉^[1]、生命科学^[2]、遥感信息处理^[3]等方面得到广泛的应用。随着图像内容的复杂化,依靠人工目视的分类方式受到挑战。如何自动准确地完成场景分类,已成为图像处理领域的研究热点。

最初的场景分类方法大多基于图像底层特征的统计,并与监督方法相结合。Szummer 等人^[4]使用

色彩、纹理信息推断较高层次的图像特性,划分室内/外自然场景。Vailay 等人^[5]和 Greene 等人^[6]将二值分类器和低层次特征相结合,用于分层框架下的自然图像分类。这类方法利用空间信息较少,使得图像的中低层特征和高层语义之间存在较大的语义鸿沟。

近年来,不少方法通过构建中间语义层对图像场景建模,较为有效地弥合了语义鸿沟的缺陷。Oliva 等人^[7]使用粗糙度、宽阔度、伸展度等视觉感知属性描述场景的主要内容结构。Vogel 等人^[8]定义一组局部语义概念,生成局部语义概念模型计算相应语义的出现频率进行场景分类。这类方法往往需要大量的人工手动标注样本,增加了训练学习的负

2010-04-12 收到, 2010-09-27 改回

国家自然科学基金(41001285)资助课题

*通信作者: 孙显 sunxian0918@hotmail.com

担。

为此,一些方法中引入了主题模型。Li 等人^[9]将图像划分为若干稀疏区域,使用文本模型认知自然场景类别。Bosch 等人^[10]提出一种基于隐含概率语义模型的方法,选取不同图像结合判别式学习识别语义类别。然而,主题模型中并没有过多考虑不同区域间的内容关联以及空间关系。文献[11-13]则按照不同尺度划分图像,通过相邻尺度间的匹配定量分析图像的空间关系。Jiang 等人^[14]扩展了主体模型,构建多重组合提取上下文信息。这些方法在自然场景的分类中取得较好效果,但由于关注的是预先设定的网格,在统计特征信息时受先验影响较大。同时,该类方法采用的特征描述子在处理复杂场景时,利用信息稍有欠缺。

考虑现有方法的不足,本文提出了一种新的基于空间语义对象混合学习的场景自动分类方法。该方法通过多尺度分割获取空间语义对象,能结合不同层次体现目标各类特征和上下层间的空间语义。文中通过产生式方法对语义对象建模,以概率直方图描述对象属于各类别的概率,充分考虑对象内部可能包含多类目标的情况。在此基础上,本文方法利用对象空间金字塔进行匹配计算,将高层先验知识与底层统计特征相结合,改进了中间语义层对图像内容及其分布概率的归纳能力。此外,本文方法还引入混合学习训练场景分类器,避免受样本容量的限制,提高了分类器的可信度。同时,文中对纹理、形状等多类特征的描述进行改进,通过致密对象拟合表达,并与尺度旋转不变的特征描述子融合,提高信息采集的全面性和准确性,对复杂场景图像获得了良好的效果。值得注意的是,与相关方法不同,本文方法无需详细定义各个图像的具体内容,只需标明每幅图像所属的类别,即可实现场景的自动分类,具有更高的鲁棒性,提高了训练效率。

2 研究方法

基于空间语义对象混合学习的场景自动分类方法的整个流程大致如图1所示。

2.1 空间对象的获取

本文方法中采用 Pyramid-cut 算法^[15]获取空间对象。分割算法中,金字塔顶层为图像整体,底层

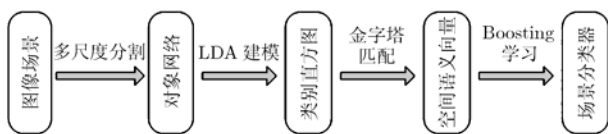


图1 本文方法的处理流程

为所有像素,中间 L 个层次将 Normalized-cut 算法与尺度因子 γ 相关联:

$$C_l = \gamma \cdot C_{l-1}, \quad l = 2, 3, \dots, L \quad (1)$$

其中 C_l 表示第 l 层的分割对象数, C_1 定义如下:

$$C_1 = \lceil \sqrt{\max(I_w, I_H)} \rceil \cdot \lambda \quad (2)$$

其中 I_w 和 I_H 分别为待处理图像的长、宽像素数, $\lceil \cdot \rceil$ 为上取整高斯函数, λ 为分割系数。

Pyramid-cut 算法从第一层起按照 C_1 完成初始分割,每个对象随后继续分割成 γ 块小对象并组成下一图层,自上而下依次递归处理。一方面能够将目标对象从复杂背景中划分出来;另一方面严格保证小尺度对象位于大尺度对象范围内,避免出现边界重叠等错误情况。

2.2 对象的语义表达

本文方法将每个对象视为主体进行 LDA (Latent Dirichlet Allocation) 建模和分析。该模型最早应用于文本分类^[16,17],在图像分类中,特征等同于文本分类中的单词。文中改进了多类特征的描述方式,同时考虑对象内部可能包含多类目标的情况,提高了信息采集的效率。

(1) 特征词汇 常见的特征描述子如 SIFT 等,需要聚类量化后才能成为特征词汇。然而,色彩、纹理等特征并不适合于聚类量化,需要通过改进,才能与尺度旋转不变的特征描述子结合处理。

(a) 色彩特征 对于样本对象 B_i ,统计色彩直方图 Hg_i ,并在训练集上度量直方图相似差异比 Dg_i :

$$Dg_i = \frac{\sum_{j=1}^{N_B^+} \text{diff}(Hg_i, Hg_j^+) / N_B^+}{\sum_{j=1}^{N_B^-} \text{diff}(Hg_i, Hg_j^-) / N_B^-} \quad (3)$$

其中 N_B^+ 和 N_B^- 代表同类和不同类对象数量,采用 χ^2 距离统计相似度 $\text{diff}(Hg_i, Hg_j)$:

$$\text{diff}(Hg_i, Hg_j) = \sum_{q=0}^{Q-1} \frac{(Hg_i(q) - hg(q))^2}{hg(q)} \quad (4)$$

$$hg(q) = \frac{Hg_i(q) + Hg_j(q)}{2}, \quad q = 0, \dots, Q-1 \quad (5)$$

其中,取 $Q = 256$ 。计算得到灰度、RGB 和 HSV 共 7 个通道的直方图相似差异比。

由于以上这些特征针对的都是图像区域,方法中采用致密特征加以改进。如文献[18]所述,在对象中心点四周取相邻的 $N \times N$ 方形,特征向量的维数为 N^2 。当方形区域越过对象边界时,以边界上的值来拟合近似。对于彩色特征,在各个通道中分别选取方形区域计算。

(b) 纹理特征 以 LM 滤波器组^[19]为基础构建纹

理特征。按照尺度、方向和相位的不同, 总计 128 个滤波器, 与图像做卷积后获得纹理特征。对于彩色图像, 滤波器分别对 HSV3 个通道进行滤波, 转换为灰度图像后再与其它滤波器进行卷积。每个点的滤波响应将被连接在一起组成一个特征, 然后采用和颜色特征相似的方法统计纹理特征的描述向量。

(c)形状特征 通过计算图像 I 中所有像素的方向能量 OE 对形状特征建模^[20]。方向能量定义如下:

$$OE = (I \cdot f_1)^2 + (I \cdot f_2)^2 \quad (6)$$

其中 f_1 为高斯差分滤波器, f_2 为 f_1 的变形为

$$\left. \begin{aligned} f_1(x, y) &= \frac{d^2}{dy^2} \left(\frac{1}{Z} \exp\left(\frac{y^2}{\sigma^2}\right) \exp\left(\frac{x^2}{\ell^2 \sigma^2}\right) \right) \\ f_2(x, y) &= \text{Hilbert}(f_1(x, y)) \end{aligned} \right\} \quad (7)$$

其中 σ 为尺度参数, ℓ 为方向比参数, Z 为归一化常量。

(d)宽基线特征 宽基线特征^[21]能对旋转、缩放、视角等变化保持一定的稳定性, 包括最大稳定极值区域、致密灰度 SIFT 和稀疏灰度 SIFT3 类。最大稳定极值区域利用仿射协变区域表达整幅图像, 采用 Harris 检测子构建感兴趣点的椭圆形区域^[22], 并按照文献[23]中所述方法进行计算。致密灰度 SIFT 特征则在对象内以若干像素间隔取点, 每个格点中的 SIFT 描述子以圆区域计算, 半径为 4、8、12、16。稀疏灰度 SIFT 特征对图像每个灰度尺度的仿射不变区域进行计算, 利用椭圆形状变换, 以主轴为参照映射到圆形, 计算得到 128 维 SIFT 描述子。随后, 利用 Sivic 等人^[24]提出的视觉单词对, 滤除冗余单词, 按照坐标分配给各个尺度的对象。

(2)LDA 建模与求解 LDA 模型中定义单词 w_d 为 M_d 个单词中一项, 文本 $D_d = \{w_{d1}, \dots, w_{dN_d}\}$ 由 N_d 个单词组成。随机生成 $1 \times K$ 的向量 θ , 该向量满足 Dirichlet 分布 $p(\theta)$, K 为主题个数; 然后根据分布 $p(z_d | \theta)$ 随机选取 $p(w_d | z_d)$ 的第 z_{d1} 行, 根据分布 $p(w_d | z_d = z_{d1})$ 随机选取 z_{d1} 行的 w_{d1} 列。

应用于图像分类处理时, 为提高模型可处理性, 采用一个 $1 \times K$ 行向量的参数 β_1 修正分布 $p(\theta)$ 的表达式, 采用一个 $K \times |V|$ 的矩阵向量 β_2 表达 $p(w_d | z_d)$, 如图 2 所示。

$$\begin{aligned} p(\theta, z_d, w_d | \beta_1, \beta_2) \\ = p(\theta | \beta_1) \prod_{n=1}^{N_d} p(z_{dn} | \beta_1) p(w_{dn} | z_{dn}, \beta_2) \end{aligned} \quad (8)$$

采用变分估计法^[25]求解式(8), 以最大化 w_d 的对数似然得到 β_2 的迭代解, 通过 Hessian 矩阵下的 Newton-Raphson 方法^[26]得到 β_1 的估值, 解决

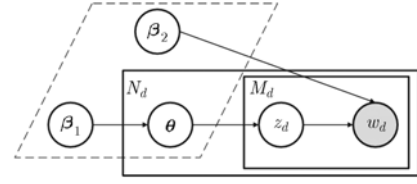


图 2 修正后的 LDA 模型

Dirichlet 分布似然的最大估计问题。

(3)对象类别语义表达 解得模型参数后, 可以得到每个对象包含各个类别主题的概率直方图分布。

2.3 对象空间金字塔匹配

Koenderink 等人^[27]采用给定位置、尺寸的高斯窗口进行金字塔合并匹配。本文中提出的对象空间金字塔匹配模型, 以分割对象为基本处理单位, 摆脱受窗口尺寸、形状等制约, 可获得更具可信度的区域特征和上下文信息。

匹配过程中, 在任一固定层次中, 当且仅当两个点落入相同网格时才认为是相互匹配的。定义 $\Gamma_k(x, y, l)$ 为第 l 层上坐标值为 (x, y) 的像素点中第 k 个主题的出现概率。将 $\Gamma_k(x, y, l)$ 记作 Γ_k^l 。由于相邻层之间的对象存在上下关联, 在 l 层上的匹配信息涵盖 $l+1$ 层上的匹配信息, 使得 l 层上的较大对象出现新的信息, 记为 $\Gamma_k^l - \Gamma_k^{l+1}$, 其中 $l = 0, 1, \dots, L-1$ 。考虑较大对象中包含较多的冗余信息, 容易对结果产生影响, 需要对关联权重进行惩罚, 定义第 l 层的权重为 $1/\gamma^{L-l}$ 。将所有层次叠加, 获取像素点在 L 个层次上累积的第 k 个主题概率和 t_k^L :

$$\begin{aligned} t_k^L &= \Gamma_k^L + \sum_{l=0}^{L-1} \frac{1}{\gamma^{L-l}} (\Gamma_k^l - \Gamma_k^{l+1}) \\ &= \frac{1}{\gamma^L} \Gamma_k^0 + \sum_{l=1}^L \frac{1}{\gamma^{L-l+1}} \Gamma_k^l \end{aligned} \quad (9)$$

对于包含 L 个层次和 K 个主题的对象, 可将各像素点的向量叠加, 得到中间语义向量:

$$T^L = \sum_{k=1}^K t_k^L \quad (10)$$

该向量中每一维数据的概率都代表了各个主题在图像内容中存在的可能性, 如图 3 所示。匹配计算中, 所有的直方图按照主题的总权重和归一化处理, 并保证主题数目的一致性。

2.4 概率分类

得到每幅图像类别语义向量后, 加上对应的场景类别标记作为训练样本。采用 JointBoosting 算法训练场景分类器^[28], 在概率框架下实现对标记数据的联合统计和快速遍历。该算法将每个子类别视为一个弱分类器, 每一轮迭代选取一个最优弱分类

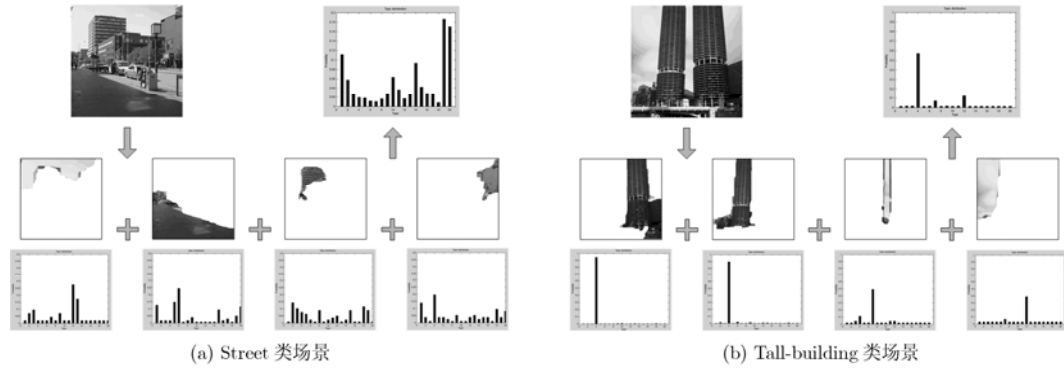


图3 部分场景中对象的主题概率分布及语义向量

器:

$$h_j(v, s) = \begin{cases} aF(v) + b, & s \in S_n \\ \psi^j, & s \notin S_n \end{cases} \quad (11)$$

其中 j 表示迭代轮次, (a, b, ϑ, ψ) 为分类器训练参数, n 为分类算法共享图中的某个结点, S_n 为共享结点 n 的类别子列表, s 为场景类别, F 为阶跃函数, 定义如下:

$$F(v) = \begin{cases} 1, & T(v) > \vartheta \\ 0, & T(v) \leq \vartheta \end{cases} \quad (12)$$

其中 $T(v)$ 表示类别语义向量 \mathbf{T} 中的第 v 维概率响应。

经过 J 轮迭代训练后, 其中 J 为需要学习的附加模型总数, 得到强分类器:

$$H(v, s) = \sum_{j=1}^J h_j(v, s) \quad (13)$$

标记过程中, 利用强分类器 H 计算待检测图像 I 属于 s 类场景的标记概率:

$$P = p(I, s) = \frac{1}{1 + \exp(-H(v, s))} \quad (14)$$

最后, 依据最大概率 $s^* = \arg \max_s p(I, s)$ 确定图像的场景类别。

3 实验分析

本文实验中主要采用两类数据集: 一类为自然场景数据集, 包括 Oliva 和 Torralba(OT)数据集^[7]、

Vogel 和 Schiele(VS)数据集^[8]、Fei-fei Li 和 Perona (FP)数据集^[9]、Lazebnik、Schmid 和 Ponce(LPS)数据集^[11]。从 FP 和 OT 数据集中分别抽取 13 类和 2 类场景组成混合数据集, 每个类别包含 200 幅图像, 其中 50 幅用于训练, 150 幅用于测试。另一类为遥感场景数据集, 该数据集中均为从 Google Earth 截取的分率为 1 m 左右的图片, 包括深海、近海、城市、农田等 4 类, 每类由 200 幅图像组成, 标记出不同的场景类别, 图像的平均尺寸为大约 600×600 像素大小, 从每个类别中取 50 幅用于训练, 其余 150 幅用于测试。

空间金字塔匹配中, 尺度序列和尺度因子等参数决定对象大小和数量多少。表 1 给出了使用不同参数时的分类准确率。由于训练建模中存在一些随机因素, 以 3 次训练得到分类器的平均值为准。可以发现, 当 $L = 4$, $\gamma = 4$ 时的分类准确率最佳。

LDA 建模过程中, 主题个数会影响信息表达的准确度。个数过多, 容易导致个别类别被拆分; 个数过少, 则许多类别混杂在一起, 降低模型的可靠性。图 4 给出了不同主题个数时在混合测试集上的分类准确率, 从图中可见, 针对此数据集的最佳主题个数为 20 左右。

方法中使用了多类特征描述子, 为评估各类特征在分类中所起的作用, 以某类特征被忽略后对分类准确率下降的比例作为度量依据, 下降比例越大, 说明该类特征的重要性越大。由表 2 可知, 自然场

表1 不同参数对分类准确率的影响

| 数据集 | 单尺度(%) | 空间金字塔匹配(%) | | | | | | | | |
|-----|--------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | $L = 3$ | | | $L = 4$ | | | $L = 5$ | | |
| | | $\gamma = 3$ | $\gamma = 4$ | $\gamma = 5$ | $\gamma = 3$ | $\gamma = 4$ | $\gamma = 5$ | $\gamma = 3$ | $\gamma = 4$ | $\gamma = 5$ |
| OT | 80.3 | 88.7 | 91.7 | 89.4 | 93.1 | 96.6 | 95.2 | 90.1 | 92.9 | 91.5 |
| VS | 81.1 | 83.8 | 84.7 | 84.7 | 86.5 | 89.2 | 89.2 | 85.6 | 88.3 | 87.4 |
| FP | 76.4 | 83.2 | 83.8 | 83.8 | 86.2 | 87.1 | 86.8 | 84.7 | 85.9 | 85.3 |
| LPS | 72.9 | 82.3 | 82.9 | 82.9 | 85.1 | 85.7 | 85.3 | 83.5 | 84.1 | 84.1 |

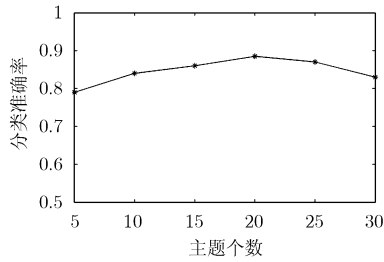


图 4 LDA 模型中主题个数对算法性能的影响

表 2 特征词汇提取中各类信息被忽略后对分类准确率下降的影响

| 特征类别 | 色彩 | 纹理 | 形状 | 致密 SIFT | 稀疏 SIFT |
|--------------|------|------|-----|---------|---------|
| 自然场景的下降比例(%) | 15.9 | 18.3 | 3.3 | 8.4 | 5.1 |
| 遥感场景的下降比例(%) | 9.6 | 12.6 | 3.0 | 11.7 | 10.5 |

景中场景构成较为简单, 色彩和纹理特征信息所起的作用较大, 而遥感场景中目标成分复杂, 受光照、旋转等变化影响较大, 因此纹理和尺度旋转不变的特征所起的作用较大。这几类特征融合在一起, 相比仅利用单一特征的处理更具全面性。

与 Boosting 算法相似的判别式学习方法还有 SVM 等。为与本文中的多分类 Boosting 算法比较, 实验中使用 SVM 进行验证, 代码来自于 LIBSVM 算法包^[29]。从表 3 可以发现, 在混合数据集上, Boosting 的分类准确率与 SVM 的分类准确率大致相当, 略高 1% 左右, 这是因为 JointBoosting 算法利用了特征共享, 具有相对稍高的多分类处理性能。类似的, 实验中也利用 BOW 模型^[30]进行比较。从表 3 发现, LDA 比 BOW 的分类准确率提高 5% 左右, 这是因为 LDA 模型能更好地统计多类主题的概率分布, 增强了信息表达的丰富性。图 5 给出了 4 种不同的组合方式, 即: LDA-Boosting、LDA-SVM, BOW-Boosting, BOW-SVM 在混合数据集上取不同数量训练图像时的分类准确率比较。本文采用的 LDA-Boosting 组合的性能最好, 受样本数量变化的

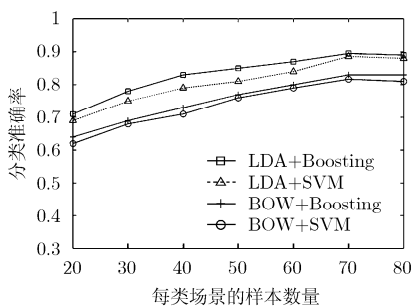


图 5 不同训练样本数量对算法性能的影响

表 3 不同产生式和判别式方法的分类准确率比较

| | LDA(%) | | BOW(%) | |
|------|----------|------|----------|------|
| | Boosting | SVM | Boosting | SVM |
| 自然场景 | 88.5 | 87.6 | 84.0 | 82.7 |
| 遥感场景 | 83.1 | 82.5 | 78.9 | 77.3 |

影响较小, 这也在一定程度上证明了这种组合方式的优越性。

判别式学习的效率对于整个算法性能有较大的影响。取混合数据集用于实验, 图 6 显示了 Boosting 和 SVM 算法的性能曲线。从图中发现, Boosting 结合 LDA 的方法准确率稍高; 随着类别数目的增加, 两种方法的准确率均开始下降, 而 SVM 结合 LDA 的下降速度更快。因而, 本文提出的 Boosting 结合 LDA 的混合学习方法能更有效的处理多类场景的分类问题。

图 7 给出了本文方法对 OT 数据集中部分图像场景的分类结果, 最后一列为错分的图像。部分图像由于目标相似, 场景差异不大而出现错分, 例如城区场景包含的建筑物占据了图像主体而被错分为高楼等。如果适当增加训练样本或者改进特征描述的种类, 可以进一步提高分类精度。本文方法还对遥感场景进行分类实验, 部分结果如图 8 所示。可以发现该方法对于这 4 类场景的分类准确率均较好, 特别是深海和农田两类图像。近海区域由于背景噪声干扰大, 准确率有所降低, 部分城市场景中包含大量的植被, 也容易被错分为农田类。在实际应用中, 可以适当结合周围场景的语义信息和一定的先验知识, 进一步提高分类精度。

此外, 还将本文方法与其它相关方法进行性能比较。从表 4 中发现, 本文方法对于较难场景的区分度优于其它方法。文献[8]中的方法将图像划分为矩形区域, 因此训练时需要将近万个矩形区域进行人工标记, 而文献[9]中的方法只采用单层次处理, 没有考虑图像中分层次的空间信息, 准确率较低。相比之下, 文献[7,11,12]的方法采用中间语义层表达

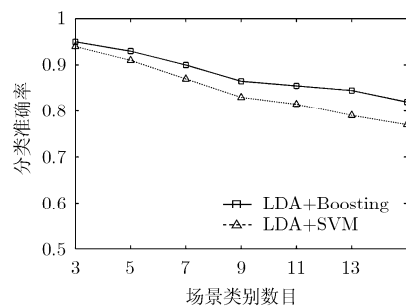


图 6 不同场景类别数量对算法性能的影响



图7 部分自然场景图像的分类结果

表4 本文方法与其它相关方法在多个场景数据集上的分类准确率比较

| 数据集 | 类别 | 训练图像 | 测试图像 | 本文方法(%) | 相关方法(%) | |
|-----|--------|------|------|---------|----------------------|----------------------|
| OT | 8 | 800 | 1888 | 89.2 | 87.8 ^[12] | 83.7 ^[7] |
| OT | 4类自然场景 | 1000 | 472 | 93.3 | 93.9 ^[12] | 89.0 ^[7] |
| OT | 4类人造场景 | 1000 | 216 | 95.7 | 94.8 ^[12] | 89.0 ^[7] |
| VS | 6 | 600 | 100 | 88.6 | 88.3 ^[12] | 74.1 ^[8] |
| FP | 13 | 1300 | 2459 | 86.3 | 85.9 ^[12] | 65.2 ^[9] |
| LSP | 15 | 1500 | 2986 | 85.1 | 83.7 ^[12] | 81.4 ^[11] |

图像内容及目标分布,较为符合人类的认知观念,但在标记过程中,需要明确图像中各个类别的属性标记,而本文方法中以隐含主题来代替,无需指明类别的具体标记,具有更高的效率。

4 结束语

本文提出了一种基于空间语义对象混合学习的复杂图像场景分类方法。该方法采用多尺度结构来表达图像数据,综合考虑了不同层次对象间的上下文关系,并采用对象空间金字塔匹配,改进了中间

语义层对图像中特征和空间信息的归纳能力。本文方法还利用判别式学习训练场景分类器,避免受样本容量限制的缺陷,提高了场景分类器的可信性。

该方法在自然和遥感场景数据集上取得了较好的分类效果,适用于多种类型和复杂内容图像的解译,具有较强的实用价值。以此为基础的图像语义分类、遥感目标区域搜索等结果可为其它的高级应用提供重要的初级结果或判断依据,例如图像精细分类、目标准确提取等。此外,本文提出的方法也可以推广到视频监控和基于视觉的人机交互等应用。

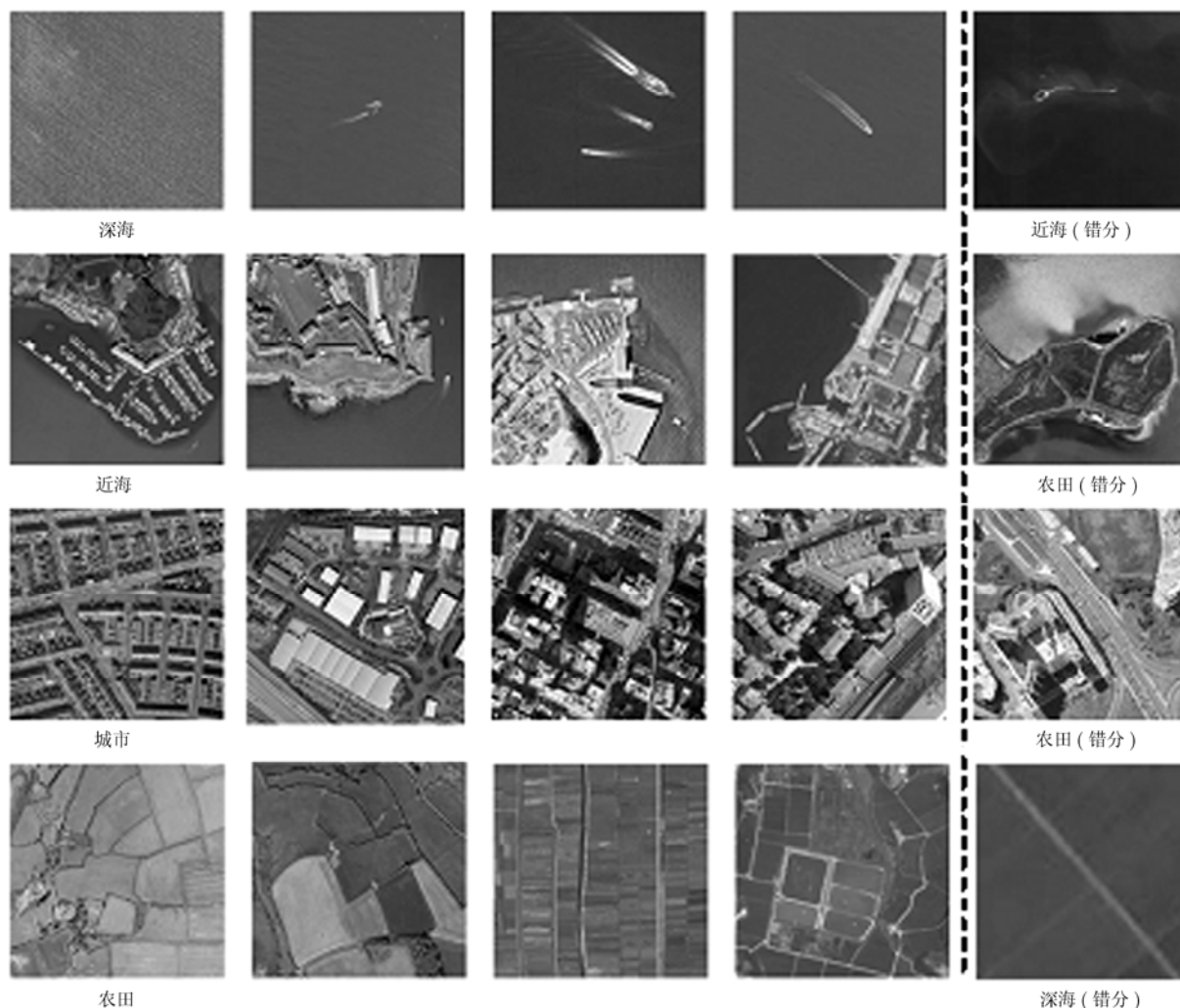


图8 部分遥感场景图像的分类结果

参 考 文 献

- [1] 程环环, 王润生. 面向自然场景分类的贝叶斯网络局部语义建模方法. 信号处理, 2010, 26(2): 234-240.
Cheng Huan-huan and Wang Run-sheng. Bayesian network based local semantic modeling for categorization of natural scenes, *Signal Processing*, 2010, 26(2): 234-240.
- [2] 顾志伟, 吴秀清, 荆浩等. 一种基于特征选择的医学图像检索方法. 中国生物医学工程学报, 2007, 26(1): 30-34.
Gu Zhi-wei, Wu Xiu-qing, and Jing Hao. A feature selection based approach in medical image retrieval. *Chinese Journal of Biomedical Engineering*, 2007, 26(1): 30-34.
- [3] 杨俊, 王润生. 遥感的道路的场景感知与分类检测/ 计算机辅助设计与图形学学报, 2007, 19(3): 334-339.
Yang Jun and Wang Run-sheng. Scene perception and classified detection for roads in remote sensing images. *Journal of Computer-aided Design & Computer Graphics*, 2007, 19(3): 334-339.
- [4] Szummer M and Picard R. Indoor-outdoor image classification. *IEEE Int. Workshop on Content-based Access of Image and Image and Video Databases*, Bombay, 1998: 42-51.
- [5] Vailaya A, Figueiredo M, Jain A, and Zhang H. Image classification for content-based indexing. *IEEE Transaction on Image Processing*, 2001, 10(1): 117-130.
- [6] Greene M R and Oliva A. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive Psychology*, 2009, 58(2): 137-176.
- [7] Oliva A and Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 2001, 42(3): 145-175.
- [8] Vogel J and Schiele B. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 2007, 72(2): 133-157.
- [9] Fei-fei Li and Perona P. A bayesian hierarchical model for learning natural scene categories. *Proc. of IEEE Computer Vision and Pattern Recognition*, San Diego, USA, 2005, 2: 524-531.

- [10] Bosch A, Munoz X, and Marti R. A review: which is the best way to organize/classify image by content? *Image and Vision Computing*, 2007, 25(6): 778–791.
- [11] Lazebnik S, Schmid C, and Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, USA, 2006, 2: 2169–2178.
- [12] Duygulu P, Barnard K, and Freitas J F D, *et al.* Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. *European Conference on Computer Vision*, Copenhagen, Denmark, 2002: 349–354.
- [13] Qin J and Yung N H C. Scene categorization via contextual visual words. *Pattern Recognition*, 2010, 43(5): 1874–1888.
- [14] Jiang Yue, Chen Jin, and Wang Run-sheng. Fusing local and global information for scene classification. *Optical Engineering*, 2010, 49(4): 1–10.
- [15] Sun Xian, Long Hui, and Wang Hong-qi. An automatic interpretation approach for high resolution urban remote sensing image based on objects-based boosting model. *The 5th IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (URBAN 2009)*, Shanghai, May, 2009: 1–5.
- [16] Steyvers M. Combining feature norms and text data with topic models. *Acta Psychologica*, 2009, 133(3): 234–243.
- [17] Larlus D and Jurie F. Latent mixture vocabularies for object categorization and segmentation. *Image and Vision Computing*, 2009, 27(5): 523–534.
- [18] Bosch A and Zisserman A. Scene classification using a hybrid generative/discriminative approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(4): 712–727.
- [19] Leung T and Malik J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 2001, 43(1): 29–44.
- [20] Malik J and Perona P. Pre-attentive texture discrimination with early vision mechanisms. *Optical Society of America*, 1990, 7(2): 923–932.
- [21] Yue S, Wang Q, and Zhao R. Robust wide baseline feature point matching based on scale invariant feature descriptor. *Proceedings of the international conference on Intelligent Computing*, Shanghai, China, 2008: 329–336.
- [22] Martin R, Marfil R, Nunez P, Bandera A, and Sandoval F. A novel approach for salient image regions detection and description. *Pattern Recognition Letters*, 2009, 30(16): 1464–1476.
- [23] Li C and Ma L. A new framework for feature descriptor based on SIFT. *Pattern Recognition Letters*, 2009, 30(5): 544–557.
- [24] Sivic J, Russell B C, Efros A A, Zisserman A, and Freeman W T. Discovering objects and their location in images. *Proc. of International Conference on Computer Vision*, San Diego, USA, 2005: 370–377.
- [25] Jordan M, Ghahramani Z, Jaakkola T, and Saul L. Introduction to variational methods for graphical models. *Machine Learning*, 1999, 37(2): 183–233.
- [26] Minka T. Estimating a Dirichlet distribution. Technical report, M.I.T., 2000.
- [27] Koenderink J and Doorn A V. The structure of locally orderless images. *International Journal of Computer Vision*, 1999, 31(2/3): 159–168.
- [28] Liu H, Liu L, and Zhang H. Boosting feature selection using information metric for classification. *Neurocomputing*, 2009, 73(3): 295–303.
- [29] Chang C and Lin C. LIBSVM: a library for support vector machines, 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [30] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 2001, 41(2): 177–196.
- 孙 显: 男, 1981 年生, 博士, 研究方向为遥感图像处理、人工智能.
- 付 琨: 男, 1974 年生, 研究员, 研究方向为信号与信息处理、遥感图像处理.
- 王宏琦: 男, 1964 年生, 研究员, 研究方向为信号与信息处理、图像处理.