基于多尺度上下文的图像标注算法

周 全[1]　王 磊[1,2]　周 亮[1]　郑宝玉[1]

摘　要　提出了一种在层次化分割框架下, 通过结合图像的底层局部特征以及高层的上下文特征, 进行图像自动语义标注的新算法. 该算法的核心思想在于对较大的图像区域的识别结果有利于对其包含的较小图像区域进行识别. 算法首先对每层分割后的图像区域进行识别, 然后利用贝叶斯定理将各层区域识别的结果通过线性加权的方式进行融合, 从而达到对整幅图像进行自动语义标注的目的. 与现有的图像标注算法相比, 仿真实验表明本文算法获得了最好的标注精度以及最快的标注速度.

关键词　图像标注与理解, 图像分割, 特征选择, 分类

引用格式　周全, 王磊, 周亮, 郑宝玉. 基于多尺度上下文的图像标注算法. 自动化学报, 2014, **40**(12): 2944−2949

DOI　10.3724/SP.J.1004.2014.02944

Multi-scale Contextual Image Labeling

ZHOU Quan[1]　WANG Lei[1, 2]　ZHOU Liang[1]　ZHENG Bao-Yu[1]

Abstract　This paper provides a novel method for image labeling by combining the local features and contextual cues in a multiple segmentation framework. Our main insight is that identifying a larger image region provides strong evidence for classifying the contained smaller ones. The proposed method weights the classification results of each image region at different levels using the Bayesian rules, which are obtained by a series of learned discriminative models based on bag of features. Multiple segmentation framework provides a robust representation, allowing a wide variety of cues to contribute to the confidence in each semantic label. Compared with previous methods, the algorithm achieves the state-of-the-art results and fastest implemental speed on the benchmark dataset.

Key words　Image labeling and understanding, segmentation, feature selection, classification

Citation　Zhou Quan, Wang Lei, Zhou Liang, Zheng Bao-Yu. Multi-scale contextual image labeling via layered segmentation. *Acta Automatica Sinica*, 2014, **40**(12): 2944−2949

众所周知, 对图像进行自动语义标注在计算机视觉和人工智能领域已经引起越来越多的关注[1−7]. 从计算机视觉角度而言, 图像标注在于为自然图像中每个像素分配一个语义标签, 以达到对图像中包含的物体进行识别以及精确分割的目的. 由于图像中的局部区域提供的信息有限, 仅仅使用这些局部特征来识别图像像素/区域往往具有歧义性, 从而导致不理想的标注结果. 为解决这些问题, 现有的图像标注算法往往运用图像中包含的上下文信息来辅助对图像像素/区域进行识别, 从而提高图像标注的性能.

目前, 条件随机场模型 (Conditional random fields, CRF)[8] 已经被广泛运用于图像标注系统的设计中. CRF 模型一般包含两个部分: 1) 基于图像局部特征的势能量项; 2) 基于两两约束的上下文势能量项. 对于第一个势能量项, 往往使用判别式分类器[1−2] 的模型进行建模; 而对于第二个势能量项, 则往往采用共生矩阵[2−3, 9]、几何上下文[10] 以及全局上下文[11] 进行建模. 虽然用 CRF 进行建模在图像标注问题上取得了一系列的进展, 但是这类模型仅仅考虑了两两约束关系的上下文信息. 而没有考虑到上下文信息具有多尺度特性, 即不同尺度的上下文可以提供不一样的图像视觉信息, 从而导致图像中蕴含的上下文信息没有得到充分地利用. 因此, 通过这种上下文特征的建模方法往往不能获得很好的标注结果.

为克服 CRF 模型在场景标注问题中的不足, 本文提出建模多尺度上下文的算法进行图像的自动语义标注. 如图 1 所示, 随着可观察到的图像区域一步步扩大, 我们能够获得的图像视觉信息也越来越多, 导致人眼对观察到的物体识别能力得到显著的提升[12]. 这充分说明, 正确识别较大图像区域有利于对其包含的较小图像区域进行识别. 因此, 在表达多尺度上下文的过程中, 需要考察不同大小区域之间的包含位置关系. 与此同时, 除了相同尺度上的区域之间存在上下文关系以外, 不同尺度上的区域之间也应该存在上下文关系. 在图 1 中, 既然直接识别图 1 (a) 中的图像区域非常困难, 那么我们可以先识别图 1 (b), 图 1 (c) 和图 1 (d), 进而再利用这些尺度的识别结果来识别图 1 (a) 中的区域. 图 2 展示了本文方法的基本流程. 首先, 本文采用超像素和层次化分割的方法来表达图像, 并根据超像素和分割区域之间的位置关系确定彼此之间的包含关系; 然后, 运用目前通用的袋特征 (Bag of feature, BoF)[13] 对多层分割后各层的图像区域进行识别; 在识别超像素的过程中, 利用贝叶斯线性加权模型, 对包含该超像素图像区域的识别结果进行加权投票, 并最终得到该超像素的语义类别标签. 其中投票的权重可以通过分类器算法自动学习得到.

图 1　利用多尺度上下文进行图像标注的研究构思

Fig. 1　Motivation of using multi-scale contextual information for image labeling

1　图像表达

1.1　超像素

传统而言, 一幅图像往往通过二维 RGB 像素的图像序列来表达. 在没有给定如何聚类这些像素的先验知识的前提下, 只能通过计算图像的局部特征 (如像素的 RGB 颜色或滤波器响应) 来进行像素聚类以形成超像素图像. 如图 2 (b) 所示, 本文首先采用均值漂移算法 (Mean shift, MS)[14], 以图

收稿日期 2013-12-03　录用日期 2014-05-23
Manuscript received December 3, 2013; accepted May 23, 2014
国家自然科学基金 (61201165, 61271240, 61201164), 高等学校博士学科点专项科研基金 (20113223120002), 中国博士后科学基金 (2013M531392), 江苏高校优势学科建设工程资助项目, 东南大学移动通信国家重点实验室开放研究基金 (2011D05), 南京邮电大学科研基金 (NY210072, NY213067) 资助
Supported by National Natural Science Foundation of China (61201165, 61271240, 61201164), Specialized Research Fund for the Doctoral Program of Higher Education (20113223120002), China Postdoctoral Science Foundation (2013M531392), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, The Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (2011D05), and the Scientific Research Foundation of Nanjing University of Posts and Telecommunications (NY210072, NY213067)
本文责任编委 黄庆明
Recommended by Associate Editor HUANG Qing-Ming
1. 南京邮电大学宽带无线通信与传感网技术教育部重点实验室 南京 210003 2. 东南大学移动通信国家重点实验室 南京 210096
1. Key Laboratory of Broadband Wireless Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003　2. National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096

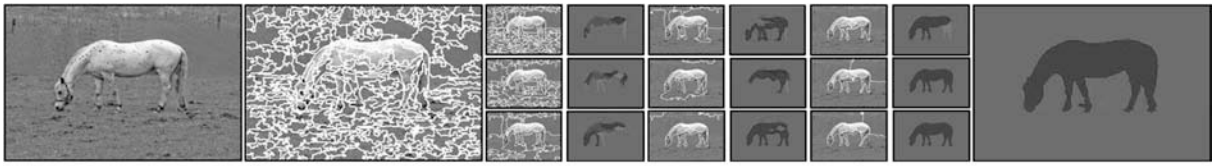


图 2 基于层次化分割的多尺度上下文图像标注算法流程

Fig. 2 Illumination of our layered segmentation framework

像像素灰度值为特征进行像素聚类以形成超像素图像. 从图 2 (b) 中可以看出, 尽管这些超像素往往具有不规则的大小和形状, 采用 MS 算法的优势在于可以将图像中同质的像素聚合到大的超像素, 而将不同质的图像区域分解到许多小的超像素. 对于一幅分辨率为 320 像素 × 210 像素图像而言, MS 算法一般可以将其过分割为大约 300 个超像素. 通过对某个特定超像素所包含的所有像素分配统一的语义标签, 就可以对整幅图像进行标注. 相对于给每个像素进行标注而言, 采用超像素的表达方法可以直接对每个超像素进行标注, 从而提高整个算法的计算效率. 与此同时, 还可以计算图像中复杂的统计特征 (如 BoF) 来计算标注的结果.

1.2 层次化分割

直观而言, 对较大区域的识别有利于辅助识别其包含的较小区域. 如图 2 (c) 所示, 如果一个较大图像区域被识别成 “马”, 那么其包含的较小的图像区域也将以很高的概率被识别成 “马”. 显然, 随着区域大小的变化, 这种上下文信息也在发生变化. 为了能够利用不同尺度的上下文信息, 本文采用层次化分割的方法来获得不同大小的图像区域. 具体而言, 本文通过更新 MS 算法的分割参数 $\{\sigma, k, \min\}$ 来产生层次化分割的结果. 这些分割参数的物理意义依次为对像素进行滤波的滤波器带宽、聚合像素之间的最远距离以及每个超像素必须包含的最少像素个数. 在每层分割结果中, 输入图像被分解成一系列的图像区块. 在实际计算过程中, 不可能遍历所有的分割参数. 因此, 本文从所有层次化分割的结果中抽取少量的过分割图像. 实验表明, 仅仅使用少量的层次化分割不仅能简化整个计算过程, 还可以达到很好的标注结果.

2 图像标注模型

2.1 问题建模

假设 Λ 表示二维图像网格, 并记定义在 Λ 上分辨率为 $W \times H$ 的图像为 I_Λ, 其中 W 和 H 分别代表图像的宽度和高度. 对于输入图像 I_Λ, 假定 MS 分割算法将其分解成为 M 个超像素, 并记这些超像素为: $I_{\Lambda_1}, \cdots, I_{\Lambda_M}$, 每个超像素对应的语义标签来自于一个离散的随机变量集合: $l \in \{1, \cdots, L\}$. 记标注的结果为 $\boldsymbol{y}$, 那么, 图像 I_Λ 的标注结果 $\boldsymbol{y}$ 一共有 $\mathcal{L} = L^M$ 种可能. 本文的目标就是要在 $\mathcal{L} = L^M$ 种可能的标注结果中找到一种最优的组合 $\boldsymbol{y}^*$, 使得下式的似然概率最大化:

$$\boldsymbol{y}^* = \arg\max_{\boldsymbol{y} \in \mathcal{L}} p(\boldsymbol{y}|I_\Lambda) \tag{1}$$

由于任意两个超像素对应的图像区域彼此互不重叠, 式 (1) 中定义的标注模型 $p(\boldsymbol{y}|I_\Lambda)$ 可以假设条件独立于 $\boldsymbol{y}$, 那么, 模型 $p(\boldsymbol{y}|I_\Lambda)$ 可以进一步分解到每一个超像素上:

$$p(\boldsymbol{y}|I_\Lambda) = \prod_{i=1}^{M} p(y_i|I_{\Lambda_i}) \tag{2}$$

其中, y_i 表示给超像素 I_{Λ_i} 分配的语义标签. 在层次化分割的框架下, 收集每层包含第 i 个超像素的图像区块, 那么由这些区域组成的集合记为 $R_i = \{r_{ji}\}$. 为了获得第 i 个超像素的似然模型 $p(y_i|I_{\Lambda_i})$, 本文将 R_i 中的图像区块看成一系列隐变量, 通过贝叶斯准则, $p(y_i|I_{\Lambda_i})$ 可以定义为

$$\begin{aligned} p(y_i|I_{\Lambda_i}) = \sum_{r_{ji} \in R_i} p(y_i, r_{ji}|I_{\Lambda_i}) \propto \\ \sum_{r_{ji} \in R_i} p(r_{ji}|I_{\Lambda_i}) p(y_i|I_{\Lambda_i}, r_{ji}) \end{aligned} \tag{3}$$

这里 $p(y_i|I_{\Lambda_i}, r_{ji})$ 代表在给定超像素 I_{Λ_i}, 以及包含 I_{Λ_i} 的图像区块 r_{ji} 的条件下, 为第 i 个超像素分配语义标签的概率. 而 $p(r_{ji}|I_{\Lambda_i})$ 则代表在给定一系列超像素的条件下产生图像区块 r_{ji} 的概率. 由于没有给定任何先验知识来指导如何进行分割, 因此假设图像区块 r_{ji} 可以由超像素任意产生, 这也就意味着式 (3) 中的概率 $p(r_{ji}|I_{\Lambda_i})$ 可以看成是一个常数. 在这种假设下, 式 (3) 可以简化为

$$p(y_i|I_{\Lambda_i}) \propto \sum_{r_{ji} \in R_i} p(y_i|I_{\Lambda_i}, r_{ji}) \tag{4}$$

为了在标注模型中有效利用图像区块 r_{ji} 所提供的语义信息, 本文进一步在式 (4) 的基础上, 将 r_{ji} 的语义标签变量 y_j 作为隐变量. 再次通过贝叶斯准则对 y_j 进行积分, 式 (4) 变化为

$$\begin{aligned} p(y_i|I_{\Lambda_i}) \propto \sum_{r_{ji} \in R_i} \sum_{y_j} p(y_i, y_j|I_{\Lambda_i}, r_{ji}) \propto \\ \sum_{r_{ji} \in R_i} \sum_{y_j} p(y_j|I_{\Lambda_i}, r_{ji}) p(y_i|y_j, I_{\Lambda_i}, r_{ji}) \end{aligned} \tag{5}$$

由于图像区块 r_{ji} 总是包含超像素 I_{Λ_i}, 因此式 (5) 中第一项可以认为独立于 r_{ji}. 而对于第二项而言, 识别超像素 I_{Λ_i} 不需要利用区块 r_{ji} 的信息, 所以可以认为其独立于 r_{ji}. 从图 1 (c) 中可以看出, 超像素本身所携带的图像信息有限, 不足以对其进行识别. 一块白色的超像素区域有可能是动物 “马” 身上的一块毛皮, 也可以是一栋 “建筑物” 的墙面. 相比之下, 区块 r_{ji} 的语义信息 y_i 不随超像素 I_{Λ_i} 图像信息的变化而变化, 因而是一种识别 I_{Λ_i} 的鲁棒的上下文特征. 则式 (5) 进一步简化为

$$p(y_i|I_{\Lambda_i}) \propto \sum_{r_{ji} \in R_i} \sum_{y_j} p(y_j|r_{ji}) p(y_i|y_j) \tag{6}$$

通过提取图像区块 r_{ji} 的图像特征, 就可以计算为其分配标签 y_j 的归一化概率. 这说明在图像标注过程中, 对较大区域的识别有利于辅助识别其包含的较小区域. 比如, 如果对图像中一个较大区域识别成 "草地", 那么这块区域所包含的较小区域也应该具有较高的概率倾向于被识别成 "草地". 这正是式 (6) 中第二个概率密度函数 $p(y_i|y_j)$ 所代表的物理意义. 注意式 (6) 中的求和运算仅仅考虑层次化分割中那些包含超像素 I_{Λ_i} 的图像区块 r_{ji}, 那些不包含 I_{Λ_i} 的图像区块则不必代入式 (6) 进行计算, 这大大减少了模型训练过程中的计算负担. 一旦学习得到模型 $p(y_j|r_{ji})$ 和 $p(y_i|y_j)$, 那么最优的标注结果 $\boldsymbol{y}^*$ 可以通过为每个超像素分配概率最高的语义标签得到.

2.2 图像特征

为学习模型 $p(y_j|r_{ji})$, 本文采用非常流行的 BoF 特征[9]. 对于如图 3 (a) 所示的输入图像, BoF 特征首先使用一系列带通滤波器对图像像素进行滤波, 将这些图像像素投影到高维的特征空间. 然后使用 K 均值聚类算法[15] 对特征空间中的特征点进行聚类, 并将其聚类到 T 个不同的聚类中心. 本文称这些聚类中心为 "纹理基元", 那么那些具有相似颜色和纹理特性的像素就可以聚类到同一纹理基元, 而不同颜色特性的像素就可以聚类到不同纹理基元. 将这些纹理基元进行可视化, 就可以获得如图 3 (b) 所示的 "纹理基元图像". 从图中可以看出, 纹理基元图像已经可以对原始图像进行一个预分割. 在本文的层次化分割框架下, 通过统计所有纹理基元的直方图分布情况, 就可以对每个图像区块进行编码, 进而提取这些图像区域的纹理特性. 由于在聚类过程中本文采用 T 个不同的聚类中心, 因此图像区块 r_{ji} 可以表示成一个 T 维向量.

图 3 输入图像及其对应的纹理基元图像

Fig. 3 Illumination of original image and corresponding texton image

另一方面, 为学习模型 $p(y_i|y_j)$, 就需要知道 r_{ji} 的标签类别信息. 为获得这些信息, 本文直接将训练好的分类器模型 $p(y_j|r_{ji})$ 的输出, 即 r_{ji} 在每个类别上的识别置信度, 作为高层语义特征. 由于语义标签随机变量的取值只有 L 种可能, 因此, 第 i 个超像素对应的特征向量是一个 L 维向量.

2.3 分类器学习算法

本文使用逻辑回归的自适应增强模型 (Logistic regression version of adaboost, LRA)[16] 来学习模型 $p(y_j|r_{ji})$ 以及 $p(y_i|y_j)$. 为简化符号, 本文统一使用 $p(y|\boldsymbol{x})$ 来表示 $p(y_j|r_{ji})$ 以及 $p(y_i|y_j)$, 其中 $\boldsymbol{x}$ 表示第 2.2 节中提出的特征向量. 那么 $p(y|\boldsymbol{x})$ 可以表示为

$$p(y|\boldsymbol{x}) \propto \exp\{H^l(\boldsymbol{x})\} \tag{7}$$

其中, $H^l(\boldsymbol{x})$ 是 N 个弱分类器 $h_n^l(\boldsymbol{x})$ 的分类置信输出的累加和:

$$H^l(\boldsymbol{x}) = \sum_{n=1}^{N} h_n^l(\boldsymbol{x}) \tag{8}$$

弱分类器 $h_n^l(\boldsymbol{x})$ 采用 2 节点的决策树模型, 并且具有如下形式:

$$h_n^l(\boldsymbol{x}) = f_{n,\text{left}}^l \mathbf{1}_{[x_n^j \le \tau_n]} + f_{n,\text{right}}^l \mathbf{1}_{[x_n^j > \tau_n]} \tag{9}$$

其中, x_n^j 代表第 n 个弱分类器选择 $\boldsymbol{x}$ 中的第 j 个成分变量, τ_n 代表自适应门限, 而 $f_{n,\text{left}}^c$ 和 $f_{n,\text{right}}^c$ 代表决策树模型中两个节点的加权对数比. 这些参数可以通过文献 [12] 中的算法自动学习得到. 整个算法流程如算法 1 所示.

算法 1. 训练决策树模型算法

Require: $D_1, \cdots, D_m$: 训练样本集合; $\omega_1^1, \cdots, \omega_m^1$: 训练样本对应的初始化权重; $y_1, \cdots, y_m \in \{1, -1\}$: 训练样本对应的语义标签集合; n: 决策树模型节点个数; N: 训练算法迭代次数.

Ensure: $T_1, \cdots, T_N$: 决策树模型; $f_1^1, \cdots, f_N^n$: 每个树模型的节点的加权对数比.

1. **For** $t = 1 \sim N$ **do**

2. 根据初始的样本权重 $\boldsymbol{w}_t$ 学习 n-节点的决策树模型 T_t.

3. 为 T_t^k 分配对数似然比: $f_t^k = \frac{1}{2}\log\frac{\sum_{i:y_i=1, D_i\in T_t^k}\omega_i^t}{\sum_{i:y_i=-1, D_i\in T_t^k}\omega_i^t}$.

4. 更新样本权重: $\omega_i^{t+1} = \frac{1}{1+\exp(y_i\sum_{t'=1}^{t} f_{t'}^{k_{t'}})}$, 对于 $k_{t'}$, 有 $D_i \in T_{t'}^{k_{t'}}$.

5. 归一化权重 $\sum_i \omega_i^{t+1} = 1$.

6. **End for**

采用决策树模型的优势在于能够自动进行特征选择, 与此同时可以在特种空间中构建一个分类超平面, 对输入的特征向量进行分类. 另一方面, 与传统的自适应增强模型相比, LRA 模型虽然具有不同的权重更新准则, 但是却可以使得每个弱分类器的输出结果自动归一化. 这有利于提高整个学习算法的计算速度. 在实际操作中, 本文对每类物体单独训练一个分类器. 比如, 在训练 "树木" 这类物体的模型时, 所有属于 "树木" 这类物体的样本用于产生正样本训练集合, 而其他类别的所有样本用于产生负样本训练集合. 这样一共可以训练得到 L 个分类器模型. 这些分类器的输出通过 $\frac{\exp\{H^l(X)\}}{\sum_l \exp\{H^l(\boldsymbol{x})\}}$ 进行归一化处理.

3 仿真实验以及结果

本文采用 MSRC 21 类物体数据集 (Microsoft Research Center, MSRC 21-class dataset)[1] 和莲花山计算机视觉研究院 15 类物体数据集 (Lotus Hill Institute, LHI 15-class dataset)[17] 作为测试平台. 其中, MSRC 数据集包含 370 幅自然图像; 而 LHI 数据集包含 150 幅自然图像. 实验过程中, 两个数据集中 40 % 的图像用来做训练数据, 10 % 的图像用来做交叉验证数据, 50 % 的图像用来做测试数据. 其他参数设置如下: 超像素的分割参数为 $\{\sigma = 1, k = 1, \min = 300\}$, 层次化分割更新的步长为 $\{2, 1, 200\}$; 纹理基元的个数 (即图像区域的特征维数) 为 $T = 500$; 层次化分割的层数为 $S = 9$ 层.

3.1 定性以及定量的结果

本节首先将本文算法与现有图像标注算法在标注精度以

及标注效率上做定性和定量的对比. 图 4 展示了本文算法在 LHI 15-class 数据集[17] 和 MSRC 21-class 数据集[1] 上的混沌矩阵对比实验结果. 表 1 还对比了本文算法与现有算法在

表 1　在 MSRC 21-类数据集[1] 和 LHI-15 类数据集[17] 上的平均标注精度与运行时间对比.

Table 1　Performance comparison on MSRC 21-class[1] and LHI 15-class[17] datasets

方法	平均标注精度 (%)		执行时间 (分钟)	
	MSRC	LHI	MSRC	LHI
本文方法	**79.4**	**82.08**	**31.2**	**19.50**
文献 [2] 算法	77.7	76.25	39.31	24.76
文献 [3] 算法	76.5	74.91	37.51	23.19
文献 [4] 算法	74.7	71.80	228.32	142.70
文献 [13] 算法	70.4	68.37	93.79	58.62
文献 [1] 算法	72.2	62.07	60.48	37.08

MSRC 21-class 数据集和 LHI 15-class 数据集上的标注性能. 从图 4 和表 1 可以看出, 与所有现有标注算法相比, 本文算法获得了最好的标注精度. 这充分说明, 在层次化分割的框架下, 本文算法能够充分利用图像中蕴含的多尺度上下文信息.

与此同时, 表 1 还对比了本文算法与其他算法的标注效率. 由于本文采用了线性加权模型, 因而比其他的图模型具有更快的计算速度. 假设层次化分割的层数为 S, 需要识别的物体类别数为 L. 那么, 本文推理的复杂度为 $\mathrm{O}(SL)$. 相比之下, 传统图模型的推理计算需要反复迭代, 非常耗时. 而本文算法在计算图像的区域特征后, 直接用训练好的贝叶斯线性加权模型就可以获得最终的标注结果.

为进一步证明本文算法的优势, 图 4 展示了两个数据集中一些自然图像标注的定性实验结果. 从图 5 中可以看出, 即使在前景物体呈现出较大视角变化, 背景物体具有较大的类间外观差异, 本文算法依然能够取得很好的标注结果. 另外, 图 5 的最右边还展示了一些标注不太好的实验结果. 比如 “树木” 被错误地识别成 “建筑物”, 而一些 “水” 对应的像素被错误地标注成 “树木”. 其原因可能在于这些类别的物体之间具有很大的类间相似性.

3.2　多尺度上下文对标注的影响

图 2 和图 6 展示了在 LHI 和 MSRC 两个数据集中多尺度上下文对标注性能影响的两个例子. 图 6 中, 如果直接利用超像素进行标注, 场景中的 “飞机” 几乎识别错误. 但是, 随着较大尺度上下文的引入, 整幅图像的标注精度在不断得到提升. 这是因为与超像素相比, 较大的图像区域往往能够获得更多的图像信息从而得到更加准确的语义信息. 当将这些不同尺度上的语义信息作为上下文辅助识别超像素时就能提高识别的性能. 相比之下, 较小的图像区域只能获得有限的图像信息, 在识别过程中具有较高的歧义性.

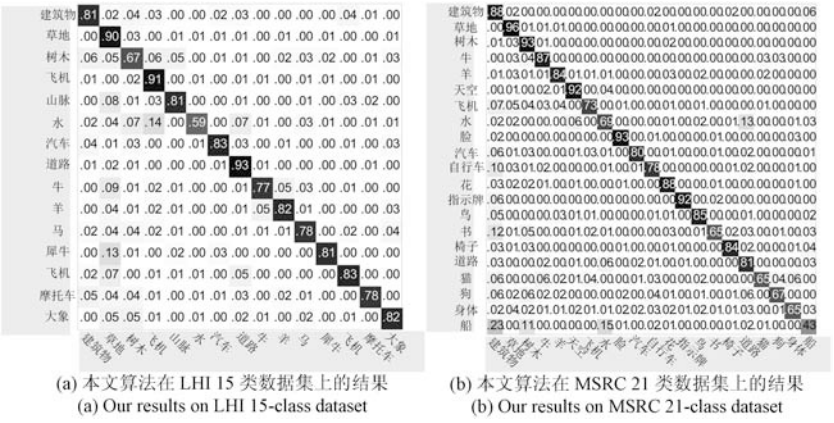
3.3　参数分析

其次, 本节还分析纹理基元个数以及层次化分割层数这些参数对标注性能的影响. 图 7 (a) 中展示了标注性能随着纹理基元个数变化而变化的结果, 而图 7 (b) 中则展示了标注性能随着层次化分割层数变化而变化的结果. 可以看出, 当纹理基元达到 500 的时候, 本文算法的标注精度达到峰值, 随着纹理基元个数进一步增大, 标注性能逐步降低. 其主要原因可能是本文模型已经过拟合. 另一方面, 当层次化分割的层数超过 9 层时, 本文算法的标注性能并没有显著性的增加. 这充分说明, 只需要少量的层次化分割就可以达到较好的标注结果.

另外一个影响标注性能的因素是层次化分割的参数. 为此, 图 7 (c) 和图 7 (d) 分别画出了在 LHI 和 MSRC 两个数据集上识别精度随着初始分割参数 $\{\sigma, k, \min\}$ 以及更新步长的变化曲线. 可以看出, 在 LHI 数据集上当 $\{\sigma = 1, k = 1, \min = 300\}$ 以及更新步长为 $\{2, 1, 200\}$ 时, 获得最优标注结果, 本文算法在 MSRC 上具有类似的结论.

4　结论

本文提出了一种基于层次化分割框架的图像标注算法. 通过线性组合的方式来结合各层的识别标注结果, 能够有效地结合图像中蕴含的局部特征和上下文特征. 实验结果表明, 在 LHI-15 类数据集上, 本文算法在取得最优标注精度的前提下, 大幅降低了标注所需要的计算成本. 虽然如此, 本文算法依然存在较大的改善空间. 在现有基础上, 下一步将扩展本文的线性模型以融合物体检测的结果, 进一步提高标注的精度.



(a) 本文算法在 LHI 15 类数据集上的结果
(a) Our results on LHI 15-class dataset

(b) 本文算法在 MSRC 21 类数据集上的结果
(b) Our results on MSRC 21-class dataset

图 4　本文算法在 LHI 15 类数据集和 MSRC 21 类数据集上混沌矩阵的对比实验 (其平均识别精度分别为 79.4 % 和 82.08 %)

Fig. 4　The confusion matrices over LHI 15-class dataset and MSRC 21-class dataset (The average recognition accuracies are 79.4 % and 82.08 % over two datasets, respectively.)

图 5 通过本文算法对自然图像的标注结果 (其中上半部分是 LHI 15 类数据集的标注结果, 下半部分是 MSRC 21 类数据集标注的结果. 为便于检验标注的结果, 不同类别的物体通过不同深浅加以区分. 与此同时, 各自的语义类别叠加在对应的图像区域上.)

Fig. 5 Illumination of some labeling results on LHI 15-class dataset (up panel) and MSRC 21-class dataset (bottom panel) (For clarity, textual labels have also been superimposed on the resulting segmentations and different gray level denotes different category.)

图 6 多尺度上下文对图像标注的影响 (第一列给出的是原始图像及其对应的人工标注的结果. 第二列给出的是仅使用超像素携带的图像信息进行标注的结果. 剩下的是每次过分割的结果以及依次加入不同尺度上下文对图像标注结果带来的改进. 数据来源于 MSRC 数据集.)

Fig. 6 Effect of inducing multi-scale contextual information (The first column gives the original image and corresponding ground truth. The second column is the labeling result only using image features of superpixels. The rest images show the object recognizable maps by gradually increasing the number of segmentation layers. Example comes from MSRC dataset.)

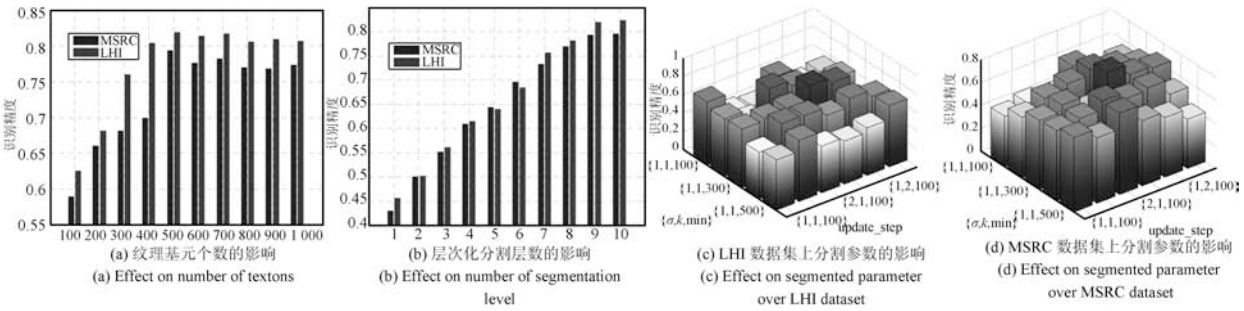


图 7 不同参数对标注精度的影响

Fig. 7 Effects on the performance with different parameters

References

1 Shotton J, Winn J W, Rother C, Criminisi A. Textonboost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 2009, **81**(1): 2−23

2 Tu Z W, Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(10): 1744−1757

3 Gould S, Rodgers J, Cohen D, Elidan E, Koller D. Multi-class segmentation with relative location prior. *International Journal of Computer Vision*, 2008, **80**(3): 300−316

4 Gould S, Fulton R, Koller D. Decomposing a scene into geometric and semantically consistent regions. In: Proceedings of the 12th IEEE Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 1−8

5 Jiang Li-Xing, Hou Jin. Image annotation using the ensemble learning. *Acta Automatica Sinica*, 2012, **38**(8): 1257−1262
(蒋黎星, 侯进. 基于集成分类算法的自动图像标注. 自动化学报, 2012, **38**(8): 1257−1262)

6 Zhang Su-Lan, Guo Ping, Zhang Ji-Fu, Hu Li-Hua. Automatic semantic image annotation with granular analysis method. *Acta Automatica Sinica*, 2012, **38**(5): 688−697
(张素兰, 郭平, 张继福, 胡立华. 图像语义自动标注及其粒度分析方法. 自动化学报, 2012, **38**(5): 688−697)

7 Yang Dong, Zhou Xiu-Ling, Guo Ping. Image annotation with Bayesian universal background model. *Acta Automatica Sinica*, 2013, **39**(10): 1674−1680
(杨栋, 周秀玲, 郭平. 基于贝叶斯通用背景模型的图像标注. 自动化学报, 2013, **39**(10): 1674−1680)

8 Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 2008 IEEE Conference on Machine Learning. Helsinki, Finland: IEEE, 2008. 282−289

9 Galleguillos C, Rabinovich A, Belongie S. Object categorization using co-occurrence, location and appearance. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1−8

10 Hoiem D, Efros A A, Hebert M. Geometric context from a single image. In: Proceedings of the 2005 IEEE Conference on Computer Vision. Beijing, China: IEEE, 2005. 654−661

11 He X M, Zemel R S, Ray D. Learning and incorporating top-down cues in image segmentation. In: Proceedings of the 2006 Europe Conference on Computer Vision. Berlin Heidelberg: Springer, 2006. 338−351

12 Medin D L, Schaffer M M. Context theory of classification learning. *Psychological Review*, 1978, **85**(3): 207−238

13 Yang L, Meer P, Foran D J. Multiple class segmentation using a unified framework over mean-shift patches. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007. 1−8

14 Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5): 603−617

15 Elkan C. Using the triangle inequality to accelerate k-means. In: Proceedings of the 2003 IEEE Conference on Machine Learning. Washington D. C., USA: IEEE, 2003. 147−153

16 Collins M, Schapire R, Singer Y. Logistic regression, adaboost and Bregman distances. *Machine Learning*, 2002, **48**(1−3): 253−285

17 Yao B, Yang X, Zhu S C. Introduction to a large scale general purpose groundtruth dataset: methodology, annotation tool, and benchmark. In: Proceedings of the 2009 Energy Minimization Methods in Computer Vision and Pattern Recognition. Berlin, Heidelberg: Springer-Verlag, 2007. 169−183

周 全 南京邮电大学通信与信息工程学院讲师, 博士. 主要研究方向为图像/视频处理, 计算机视觉, 机器学习与模式识别. 本文通信作者. E-mail: quan.zhou@njupt.edu.cn
(**ZHOU Quan** Ph. D., Lecturer at the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. His research interest covers image/vedio processing, computer vision, machine learning, and pattern recognition. Corresponding author of this paper.)

王 磊 南京邮电大学通信与信息工程学院副教授, 博士. 主要研究方向为无线通信. E-mail: wanglei@njupt.edu.cn
(**WANG Lei** Ph. D., associated professor at the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. His research interest covers wireless communications.)

周 亮 南京邮电大学通信与信息工程学院教授, 博士. 主要研究方向为多媒体通信, 多媒体信号处理. E-mail: liang.zhou@njupt.edu.cn
(**ZHOU Liang** Ph. D., professor at the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. His research interest covers multimedia communication and multimedia signal processing.)

郑宝玉 南京邮电大学通信与信息工程学院教授, 博士. 主要研究方向为多媒体通信, 多媒体信号处理. E-mail: zby@njupt.edu.cn
(**ZHENG Bao-Yu** Ph. D., professor at the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. His research interest covers multimedia communication and multimedia signal processing.)