

基于形式概念分析和语义关联规则的目标图像标注

顾广华^{1,2} 曹宇尧^{1,2} 崔冬^{1,2} 赵耀³

摘要 基于目标图像标注一直是图像处理和计算机视觉领域中一个重要的研究问题。图像目标的多尺度性、多形变性使得图像标注十分困难。目标分割和目标识别是目标图像标注任务中两大关键问题。本文提出一种基于形式概念分析 (Formal concept analysis, FCA) 和语义关联规则的目标图像标注方法, 针对目标建议算法生成图像块中存在的高度重叠问题, 借鉴形式概念分析中概念格的思想, 按照图像块的共性将其归成几个图像簇挖掘图像类别模式, 利用类别概率分布判决和平坦度判决分别去除目标噪声块和背景噪声块, 最终得到目标语义簇; 针对语义目标判别问题, 首先对有效图像簇进行特征融合形成共性特征描述, 通过分类器进行类别判决, 生成初始目标图像标注, 然后利用图像语义标注词挖掘语义关联规则, 进行图像标注的语义补充, 以避免挖掘类别模式时丢失较小的语义目标。实验表明, 本文提出的图像标注算法既能保证语义标注的准确性, 又能保证语义标注的完整性, 具有较好的图像标注性能。

关键词 图像标注, 形式概念分析, 语义关联规则, 共性特征, 特征融合

引用格式 顾广华, 曹宇尧, 崔冬, 赵耀. 基于形式概念分析和语义关联规则的目标图像标注. 自动化学报, 2020, 46(4): 767–781

DOI 10.16383/j.aas.c180523

Object Image Annotation Based on Formal Concept Analysis and Semantic Association Rules

GU Guang-Hua^{1,2} CAO Yu-Yao^{1,2} CUI Dong^{1,2} ZHAO Yao³

Abstract Object-based image annotation has always been an important research issue in the field of image processing and computer vision. Image annotation is very difficult because of the multi-scale and variability of the objects. Object-based image annotation has two key issues: object segmentation and object recognition. This paper proposed an object image annotation method based on formal concept analysis (FCA) and semantic association rules. Aiming at the high overlap problem of image blocks for objectness proposal generation algorithm, the idea of concept lattice in formal concept analysis was used to classify the image blocks into several image clusters according to the commonality of image blocks and mine the image category pattern. After removing the object-noise block and the background-noise block by the category probability distribution decision and the flatness decision, respectively, the final semantic object clusters are obtained. In addition, aiming at the discrimination problem of semantic objects, we firstly got common feature descriptions by fusing features of image clusters, and generated the initial object image annotation through the classifier. The semantic association rules were then mined through the semantic image annotations to perform the semantic complement of image annotations to avoid missing smaller semantic objects when mining category patterns. Experimental results show that the proposed image annotation algorithm not only ensures the precision of semantic annotation, but also ensures the integrity of semantic annotation. It has the better performance of image annotation.

Key words Image annotation, formal concept analysis (FCA), semantic association rules, common features, feature fusion

Citation Gu Guang-Hua, Cao Yu-Yao, Cui Dong, Zhao Yao. Object image annotation based on formal concept analysis and semantic association rules. *Acta Automatica Sinica*, 2020, 46(4): 767–781

收稿日期 2018-08-01 录用日期 2018-12-18
Manuscript received August 1, 2018; accepted December 18, 2018

国家自然科学基金 (61303128), 河北省自然科学基金 (F2017203169, F2018203239), 河北省高等学校科学研究重点项目 (ZD2017080), 河北省留学回国人员科技活动项目 (CL201621) 资助

Supported by Natural Science Foundation of China (61303128), Natural Science Foundation of Hebei Province (F2017203169, F2018203239), Key Foundation of Hebei Educational Committee (ZD2017080), Science and Technology Foundation for Returned Overseas People of Hebei Province (CL201621)

本文责任编辑 黄庆明

Recommended by Associate Editor HUANG Qing-Ming

1. 燕山大学信息科学与工程学院 秦皇岛 066004 2. 河北省信息传

随着互联网的发展和移动终端的普及, 人们生活中所接触到的图像数量产生了爆炸性增长。图像数据的大规模增长对图像处理技术产生了更多更高的要求。对图像语义目标进行标注将会给图像语义理解等任务带来广泛的应用。图像目标语义标

输与信号处理重点实验室 秦皇岛 066004 3. 北京交通大学信息科学研究所 北京 100044

1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004 2. Hebei Provincial Key Laboratory of Information Transmission and Signal Processing, Qinhuangdao 066004 3. Institute of Information Science, Beijing Jiaotong University, Beijing 100044

注^[1] 目的是生成目标语义关键词, 对图像进行语义描述. 语义关键词不仅能对图像内容进行简单的理解, 而且还能为其他计算机视觉任务提供有用信息, 如图像语义理解^[2]、图像目标检测^[3-4] 和多语义图像分类^[5] 等. 图像标注可以理解为图像目标内容与语义标签的映射^[6-7]. 目前标签与图像内容之间的映射方法主要有潜在语义模型^[8-9] 和语义传递模型^[10-12] 等, 这些模型在单目标图像中有着很好地标注效果, 但在多目标图像中, 将整张图像内容作为标签的映射很容易因为目标的不明确出现标注错误. 因此为了更好地标注多目标图像, 以图像目标区域内容作为标签映射的标注模型应运而生, 比如目标概率模型^[10]、多目标学习模型^[11] 等, 都在多目标图像标注上取得了很好的效果.

在多目标学习模型中最主要的问题是如何从多个图像区域中找到完整的图像目标区域. 因为在同一张图像中很有可能出现多个不同类别, 不同大小的图像目标, 因此在寻找目标区域时必须对图像进行多尺度划分, 目标建议生成算法^[12-13] 是较好的生成多尺度图像块的方法. 常见的 DPM (Deformable part model) 模型^[14] 能够很好地对目标区域进行识别, 但它只适合单目标识别, 对复杂的多目标图像并不适合. Verma 等提出基于可判别分类模型的图像标注方法^[15], 按照分类器对图像的概率分布对图像区域内容进行判别取得了比较好的结果, 但是对于背景噪声块不能进行有效的筛选. 通常的图像分块算法除了产生目标图像块以外, 还会产生大量的噪声块包括背景块 (不存在目标) 和不完整目标块 (存在不完整目标). 因此在进行目标区域识别之前必须进行噪声图像块的消除.

目标建议生成算法^[12-13] 针对图像中的语义目标生成了大量的多尺度局部区域, 在这些局部区域中存在着很多相似度极大的图像局部块, 并且本文发现, 在图像中越明确的语义目标生成的局部图像块越多. 即针对每一幅图像, 可以生成几个图像块簇, 每个簇对应一个语义目标, 目标语义越明确, 其对应的簇内包含的图像块越多且密集. 为了提高图像目标区域内容标注的准确性, 本文提出了一种基于形式概念分析 (Formal concept analysis, FCA)^[16-18] 的图像目标类别模式挖掘及图像块簇生成方法. 将所有图像块构建形式背景, 基于概念格中内涵和外延挖掘出图像块中的类别模式, 即生成多个类别图像块簇; 对簇内图像块进行特征融合, 通过分类器的类别概率分布筛选并消除不完整目标块噪声簇; 利用平坦度^[19] 判决去除背景块噪声簇, 之后筛选出完整目标块簇并进行目标识别与标注; 最后, 为了避免挖掘类模式时忽略较小的语义目标, 本文利用关联规则算法挖掘数据集中语义之间的关联

规则, 对图像标注进行语义丰富和标签补充, 完成最终的图像目标语义标注. 本文在 VOC2007 数据集^[20] 中证明了将同类目标按照特征共性进行共同判别可以有效地减少目标误判, 提高图像标注准确率.

1 基于 FCA 和特征融合的图像标注

1.1 基于 FCA 的模式挖掘

形式概念分析是一种进行数据分析和规则提取的强有力的工具. 形式概念分析从形式背景出发对组成本体的概念、属性以及从属关系等用形式化的语境表述出来, 然后根据语境构造出概念格 (Concept lattice), 即本体^[18], 从而清楚地表达出本体的结构.

定义 1. 设 U 是对象的集合, M 是属性的集合, I 是两集合 U 与 M 间的关系, 则称三元组 $K = (U, M, I)$ 为一个形式背景 (简称背景). $(u, m) \in I$ 表示对象 u 具有属性 m . 背景可以用一个矩形表表示, 如图 1(a) 所示, 每一行为一个对象, 每一行为一个属性. 若 u 行 m 列交叉为 1, 则表示对象 u 具有属性 m ; 若 u 行 m 列交叉为 0, 则表示对象 u 不具有属性 m .

定义 2. 设 $K = (U, M, I)$ 是一个形式背景, 若 $A \subseteq U, B \subseteq M$, 令 $f(A) = \{m \in M \mid \forall u \in A, (u, m) \in I\}$, 及 $g(B) = \{u \in U \mid \forall m \in B, (u, m) \in I\}$. 如果 A 和 B 满足 $f(A) = B, g(B) = A$, 则称二元组 (A, B) 为一个概念. 其中, A 称之为概念 (A, B) 的外延, B 称之为概念 (A, B) 的内涵^[16].

在形式概念背景中, 概念表示外延中所有的对象元素都对应内涵中的所有属性. 形式概念分析通过构造概念格的方式, 从一个形式背景中计算出所有的概念, 并将其按照内涵逐层丰富的方式进行有序的整理. 通过概念格可以清晰地得到所有概念以及概念之间的所属关系, 如图 1 所示. 其中, 图 1(a) 为形式背景, 图 1(b) 为由图 1(a) 生成的概念格. 概念格中的概念 $\{1\ 4, a\ c\}$ 从属于其父概念 $\{1\ 2\ 4, a\}$ 和 $\{1\ 4, c\}$, 也即表示在概念 $\{1\ 2\ 4, a\}$ 和 $\{1\ 4, c\}$ 中的对象 $\{1\ 4\}$ 共同拥有属性 $\{a\ c\}$.

受概念格启发, 本文为目标建议生成算法^[12-13] 生成图像块设计了模式挖掘的聚类算法 (Pattern mining clustering, PMC). 算法中以图像块为对象, 以图像块的稀疏二值化特征为属性构建形式背景, 按渐近构造概念格的方式求出所有概念. 在所得概念中外延为图像块集, 内涵为这些图像块共有的特征模式. 然后利用所有概念计算出所有满足簇内具有共有性、簇间具有差异性的图像簇和类别模式.

1.1.1 构造形式背景

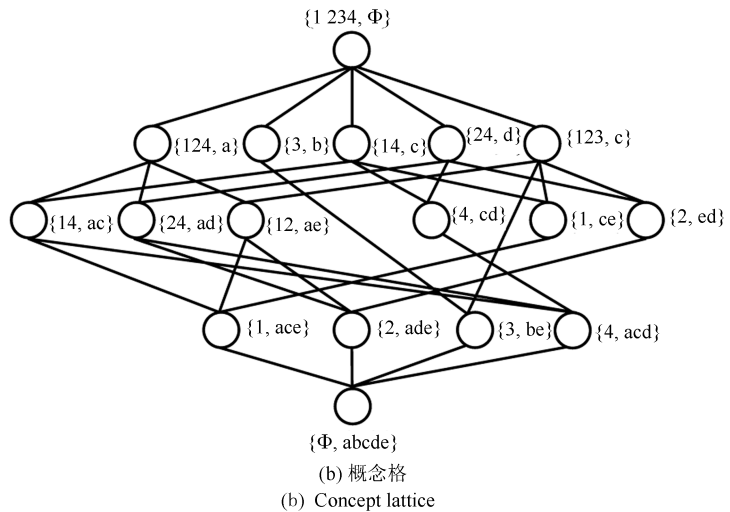
与传统的图像特征相比, 卷积神经网络 (Convolutional neural network, CNN)^[21-22] 提取的深度特征具有更好的泛化能力, CNN 众多的网络框架中 VGGNet^[22-23] 在 ILSVRC^[24] 挑战中有着非常好的表现. 最近的研究^[25-26] 已经证明从全连接层提取的 4096 维特征是一般识别任务的极佳表示. 因此本文使用图像块通过 VGG-16 全连接层的 4096 维特征构造特征形式背景.

在 CNN 特征中的前 d 个最大值几乎可以代表图像的主要特征, 并且在对其进行二值化后依然能较好地表现其特性. 为了验证该结论, 本文对 5 类图像进行分类实验. 将图像 CNN 特征中前 d 个最大值置 1, 其他值置 0, 进行稀疏化和二值化后利用 SVM (Support vector machine) 分类器进行分类对比, 对比实验结果如表 1 所示.

表 1 中 CNN- d 表示将 CNN 特征 4096 维中的前 d 个最大值置 1, 其余特征值置 0. 表 1 中最后一列为原始 CNN 特征的分类结果. 通过对比可以看出, 取 CNN 特征的前 20 个最大值进行稀疏二值化就已经能够获得较好的分类效果. 由此可见, 取 CNN 特征的前 d 个值进行稀疏二值化可以代表图像的语义特性. 因此, 本文将每个图像块的 CNN 特征进行稀疏化和二值化构造特征形式背景.

	a	b	c	d	e
1	1	0	1	0	1
2	1	0	0	1	1
3	0	1	0	0	1
4	1	0	1	1	0

(a) 形式背景
(a) Formal context



(b) 概念格
(b) Concept lattice

图 1 根据形式背景构造概念格

Fig. 1 Construction of concept lattice based on the context

表 1 CNN 特征稀疏二值化分类效果对比

Table 1 Comparison of CNN feature sparse binarization classification

分类精度	CNN-10	CNN-20	CNN-30	CNN-40	CNN-50	CNN
Accuracy	0.81	0.88	0.86	0.89	0.93	0.93

1.1.2 挖掘潜在特征模式

在特征形式背景中, 对象表示图像块, 属性表示特征点. 由形式背景构建概念格如图 1 所示, 可知每个概念的外延表示一个图像块簇, 相应的内涵表示这个图像块簇的共有特征点即特征模式. 因为本文要将具有共同特征模式的图像块进行特征融合按照其共性进行处理, 所以每个图像簇内的对象数不易太少, 如图 1 (b) 中第 4 层概念虽然具有最高的内涵维度, 但其对应的外延只有一个对象, 无法体现其共有性.

因此在计算特征模式时, 设置外延维度阈值 α , 如果一个初始概念的外延元素数小于 α , 那么其子概念的外延元素必然小于阈值. 为了快速完整地计算出所有的特征模式及其对应的图像簇, 本文按照构造概念格的算法自上而下逐层丰富内涵元素来计算所有特征模式. 当生成所有概念的外延元素数均大于阈值 α 时, 则计算出所有潜在特征模式. 经过 α 的筛选, 计算出的所有概念都具有较高维度的外延, 即挖掘出的每个潜在特征模式均对应较多的图像块, 如图 1 (b) 中概念 {1 2 4, a}, 表示挖掘出的特征模式 a 对应着第 1, 2, 4 三个图像块.

1.1.3 特征模式优化

在概念中内涵的维度越高说明对象的共有性就越强, 如图 1 (b) 中第三层中的概念 {1 4, a c} 与第二层中的概念 {1 2 4, a} 相比, 因为外延 (1, 4) 对应

的内涵维度为 2, 而外延 (1, 2, 4) 对应的内涵维度为 1, 所以对象 1 和 4 之间的共有性要大于 (1, 4) 和 2 之间的共有性. 为了确保图像簇中图像块的共有性, 设置内涵维度阈值 β . 首先, 从第 1.1.2 节中得到的所有概念中筛选出所有内涵维度大于 β 的概念集合. 经过 β 的筛选可以保证所有概念中的内涵都具有较高的维度, 相应图像簇具有较强的共有性. 然后, 为了进一步保证不同图像簇之间存在差异性, 分别计算图像簇中类别模式的相同特征元素的个数 γ . 如果 γ 大于维度较小特征模式中元素个数的 1/2, 那么说明两个图像簇具有较小的差异性, 将其图像簇和特征模式进行合并. 最终保证得到的图像簇具有类间差异性和类内相似性的特点. 通过特征模式优化后, 更新的概念格中剩余所有概念内涵的维度将保证簇内距离最小, 概念格中不同概念之间内涵的差异性将保证簇间距离最大. 按照概念格中概念对图像块进行簇划分, 如图 2 所示.

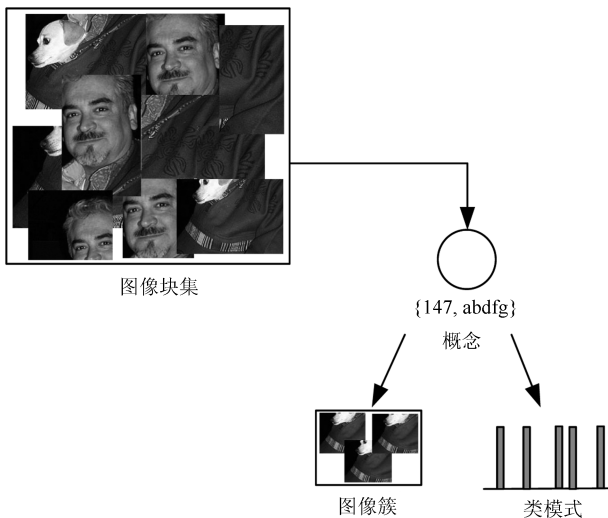


图 2 由概念格构建图像簇

Fig. 2 Image cluster construction from concept lattice

1.2 基于类别模式和特征融合的图像标注

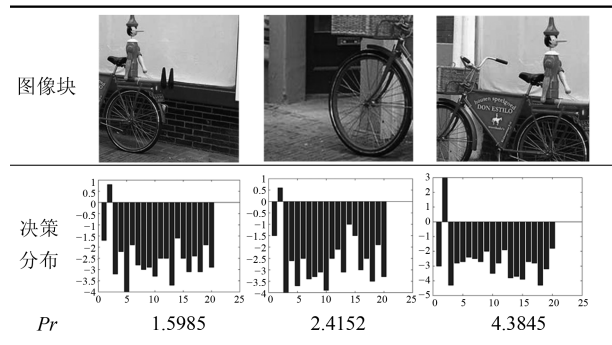
在对图像进行分块后, 存在完整目标、部分目标和背景三种图像块, 其中对图像标注有意义的完整目标块只占其中的小部分. 按照基于判别模型的图像标注算法, 让图像中所有的图像块通过用语义目标图像训练的分类器, 其中完整的目标块将生成比较集中的类别决策分布, 而其他噪声图像块生成的决策分布将会比较离散, 如表 2 所示. 本文利用该性质设置语义判别参数 Pr 来选择合适的标注图像块:

$$Pr = Max_f(x) - Max_s(x) \quad (1)$$

其中, x 为分类器对应所有类的类别决策值, $Max_f(x)$ 表示 x 中的最大值, $Max_s(x)$ 表示 x 中的次大值.

从表 2 中可以看出, 如果在图像块中存在较为完整的图像语义目标, 则对应的 Pr 值越高; 反之, Pr 值越低说明图像块为噪声块的可能性越高.

表 2 不同图像块的类决策分布
Table 2 Category decision distribution of different image blocks



对图像进行切割的图像目标建议算法会产生很多内容极具相似性的图像块, 如果依次对其进行判断将会产生很大的计算资源浪费, 并且直接对图像块进行判别, 存在一些个性图像块判别错误的问题. 为了进一步提高图像标注的准确率和标注效率, 本文结合第 1.1 节提出的特征模式挖掘算法, 将图像块按照特征模式集成图像簇, 并将图像簇的特征进行融合, 利用融合特征的共性进行图像标注.

在第 1.1 节中通过形式概念分析的方法对每张图像的图像块按照类别模式进行聚类, 将具有相同类别模式的图像块聚集在一起形成图像簇. 如果设待标注图像集为 $I = \{I_i\}_{i=1}^n$ 那么将得到图像簇 $C = \{C_i\}_{i=1}^n$, C_i 表示第 i 张图像生成的图像簇 $C_i = \{c_j\}_{j=1}^h$, 中 c_j 表示第 j 个图像簇. 通常情况下图像 I_i 对应的图像簇集合 C_i 中会包含能够明确表现图像语义的图像目标簇, 不能清楚表现图像语义的目标噪声簇和完全不包含图像语义的背景簇, 三种类型的图像簇如表 3 所示.

表 3 图像簇种类划分
Table 3 Dividing image cluster types

目标簇	
目标噪声簇	
背景簇	

同一图像簇中的图像块具有非常大的相似性, 如果对这些图像块依次判断类别概率分布, 这样不但浪费了巨大的计算资源, 而且也没有充分利用图像簇中图像块的共同属性. 为了在减小计算复杂度的同时利用图像簇中图像块的共性, 本文借鉴 CNN 网络中池化的思想, 将图像簇中所有图像块的 CNN 特征融合成一个具有图像块共性的图像簇特征 f_c . 对图像簇中每个图像块通过 VGG-16 模型第 14 层的全连接层得到的 4096 维特征向量进行融合如图 3 所示, 融合方法分为最大值融合和平均值融合.

最大值融合是从图像簇中所有图像块特征的每个维度中选择最大特征值组成一个新的 4096 维图像簇特征 $f_{c,max}$:

$$f_{c,max}(i) = \max(f_1(i), \dots, f_r(i)) \quad (2)$$

其中, $i = 1, 2, \dots, 4096$, f_r 表示图像簇中第 r 个图像块的 4096 维特征.

平均值融合是求图像簇中所有图像块特征每个维度的平均值组成一个新的 4096 维图像簇特征 $f_{c,mean}$:

$$f_{c,mean}(i) = \frac{1}{r} \sum_{j=1}^r f_j(i), \quad i = 1, 2, \dots, 4096 \quad (3)$$

通过图像簇特征融合后每个图像簇都将得到代

表图像块共性的簇特征 f_c . 图像簇中包含具有完整图像目标的目标簇和包含噪声的噪声簇, 为了从中选出包含完整目标的目标图像簇, 将 f_c 通过用目标图像训练好的分类器. 能够明确表现语义信息的图像簇将产生较为集中的类决策分布, 不能明确表现语义目标的噪声簇将产生较为离散类决策分布, 如表 4 所示.

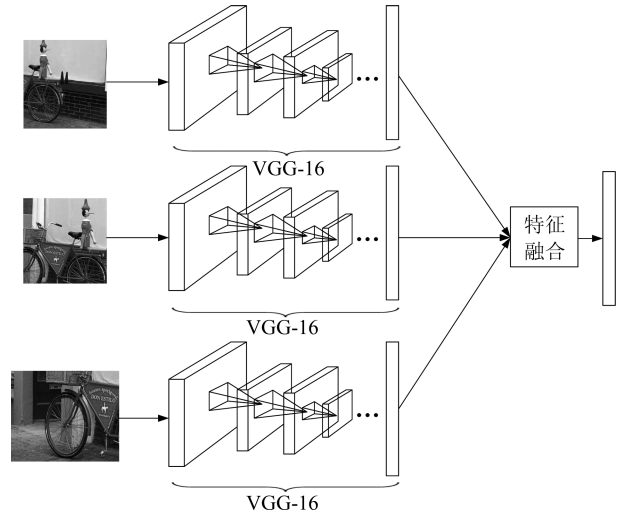


图 3 特征融合模型

Fig. 3 Feature fusion model

表 4 融合特征决策值

Table 4 Decision values of fusion feature

图像簇	非融合特征	最大值融合特征	均值融合特征

表 4 中的第 1 列分别为能够表现语义信息的目标簇、不能明确表明语义信息的目标噪声簇和没有语义信息的背景簇; 第 2 列表示每个图像簇中图像块对应的语义类决策分布 (图像簇中的每个图像块对应 1 组概率分布); 第 3 列为对图像簇图像块的特征进行最大值融合后融合特征的语义类决策分布; 第 4 列为对图像簇的特征进行均值融合后融合特征的语义类决策分布. 在所有分布柱状图中, 灰色条对应的语义类为图像簇所属的真正的类, 如果灰色条对应的决策值为所有决策值中的最大值, 则说明类别判决结果正确.

在表 4 中第 2 行第 2 列图像簇对应的图像块类决策值中可以看出, 第二个图像块类决策分布中的最大值并非正确的语义类 (即分类错误). 从第 2 行的后两列可以看出, 经过特征融合后图像簇决策值中的最大值取到了正确的语义类. 从而可以看出对图像簇进行特征融合可以将个别分类错误的图像块进行修正, 从而增加标注的鲁棒性. 利用式 (1) 分别计算两种融合方式的语义判别参数, 其中最大值融合特征的语义判别参数 $Pr_{\max} = 3.7508$ 和均值融合特征的语义判别参数 $Pr_{\text{mean}} = 2.4528$. 对比 Pr_{\max} 和 Pr_{mean} 可以看出, 最大值融合方式的语义判别正确概率要优于均值融合方式, 更有利于语义判别.

第 3 行的图像簇为目标噪声簇, 从第 2 列图像块分类结果可以看出, 这些图像块均无法取得正确的分类效果. 针对这种目标噪声图像簇, 我们希望其融合特征具有更加均匀的语义判别概率, 这样更有利于对其去噪. 本文分别计算了两种融合方式的融合特征进行语义判别概率分布的方差, 均值融合特征和最大值融合特征语义判别概率分布的方差分别为: $V_{\text{mean}} = 0.6150$ 和 $V_{\max} = 2.0857$, 由此可以看出, 均值融合特征语义判别得到的概率分布明显比最大值融合方式得到的概率分布均匀. 本文算法利用式 (1) 来计算图像簇的语义判别参数 Pr , 分别计算得到最大值融合特征的语义判别参数为 $Pr_{\max} = 1.2$, 均值融合特征的语义判别参数为 $Pr_{\text{mean}} = 0.48$. 由此进一步说明, 均值融合特征算法更有利于对图像簇进行模糊去噪.

第 4 行的图像簇为颜色和纹理较为单一的背景图像簇, 由第 2 列可知, 通过类决策值可以看出这些图像块具有明确的语义判别结果, 但是均为错误的语义分类结果. 通过实验发现, 针对背景图像簇, 利用均值融合方法得到的图像簇特征无法利用模糊分类对其进行去噪.

为解决上述问题, 本文利用背景区域的颜色和纹理分布相对单一的特点, 根据平坦度将其与目标区域进行区分. 图像平坦度可以用来评估图像的复

杂程度. 平坦度可以由其灰度值分布表示, 像素灰度值分布越均匀平坦度越大, 灰度值分布波动越大平坦度越小. 为了计算图像平坦度, 将图像划分成尺寸为 $m \times n$ 的多个局部框. 定义局部熵 H 如式 (4) 所示, 利用局部熵来表示每个局部框平坦度的大小. 对 H 设置阈值 th , 如果局部框的 $H > th$, 判定此局部框区域是平坦区域并设置其平坦度 $F = 1$, 否则判定为不平坦区域, 设为 $F = 0$.

$$H = \frac{-\sum_{i=1}^m \sum_{j=1}^n \log_2 q_{ij}}{\delta_w \log_2 mn} \quad (4)$$

$$q_{ij} = \frac{p_{ij}}{\sum_{i=1}^m \sum_{j=1}^n p_{ij}} \quad (5)$$

其中, p_{ij} 是局部框内第 (i, j) 个像素的灰度值. δ_w 是局部框内像素灰度值的方差.

设图像簇中存在 l 个图像块, 每个图像块被划分了 q 个局部框, 则图像簇的平坦度被定义为:

$$F_r = \frac{\sum_{j=1}^k \sum_{i=1}^l F_{ij}}{l \times q} \quad (6)$$

其中, F_{ij} 表示图像簇中第 j 个图像块中第 i 个局部框的平坦度, 可以看出 $F_c \in [0, 1]$.

经过对图像簇计算平坦度可以得到, 包含目标的图像簇具有较小的平坦度值, 而颜色和纹理较为单一的背景图像簇的平坦度值较大, 如图 4 所示.

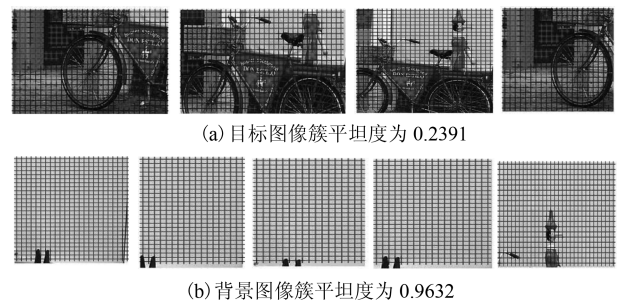


图 4 目标图像簇和背景图像簇的平坦度对比

Fig. 4 Flatness comparison of object image clusters and background image clusters

在图 4 中深色网格区域为不平坦区域, 浅色网格区域为平坦区域, 其中图 4(a) 为目标图像簇, 图像簇中的图像块均包含较为完整的语义目标, 其平坦度为 0.2391. 图 4(b) 为背景图像簇, 图像簇中的图像块均为颜色和纹理较为单一的背景块, 其平坦度为 0.9632. 由此可以看出, 目标图像簇的平坦度要远远小于背景图像簇, 通过对图像簇设置合适的阈值 T_c 可以从图像簇中去掉背景噪声图像簇.

从表 4 可以看出均值特征融合更适合对图像簇进行去噪处理, 最大值特征融合更适合对图像簇进行语义判别处理. 因此本文将利用均值特征融合和平坦度判决先对所有图像簇进行去噪处理, 分别去掉目标噪声簇和背景噪声簇, 保留目标簇, 即能具有明确语义信息的图像簇.

经过对图像簇的模糊去噪和平坦度去噪后, 得到具有明确目标语义的图像簇, 最后将图像簇内的图像进行最大值特征融合得到图像簇特征矩阵 f_m . f_m 中每个特征代表一种语义目标类别, 将 f_m 通过事先训练好的 SVM 目标分类器, 可以得到每个融合特征的语义标签集 w . 然后按照图像簇中图像块的索引判断图像块所属的待标注图像, 将语义标签 w 传递给待标注图像. 整个图像标注过程如图 5 所示.

1.3 基于关联规则的语义补充

在第 1.1 节中通过形式概念分析的方法挖掘类模式时, 有可能存在某些概念, 其外延和内涵中的元素均小于阈值 α 和 β , 即说明这些概念对应的图像块将不属于任何图像簇, 本文将这些不属于任何图像簇的图像块的集合称之为缺失图像簇. 因为目标建议算法生成的图像局部框以图像语义目标区域为主, 生成的和图像语义目标区域相关的局部框比较密集, 和图像背景区域相关的局部框将比较稀疏, 如图 6 所示. 所以根据第 1.1 节中挖掘出的图像簇类别模式生成的图像簇主要为图像语义目标区域图像块的集合, 换言之, 缺失图像簇中的图像局部块主要为图像的背景块.

但是, 在图像中可能存在尺寸很小的语义目标

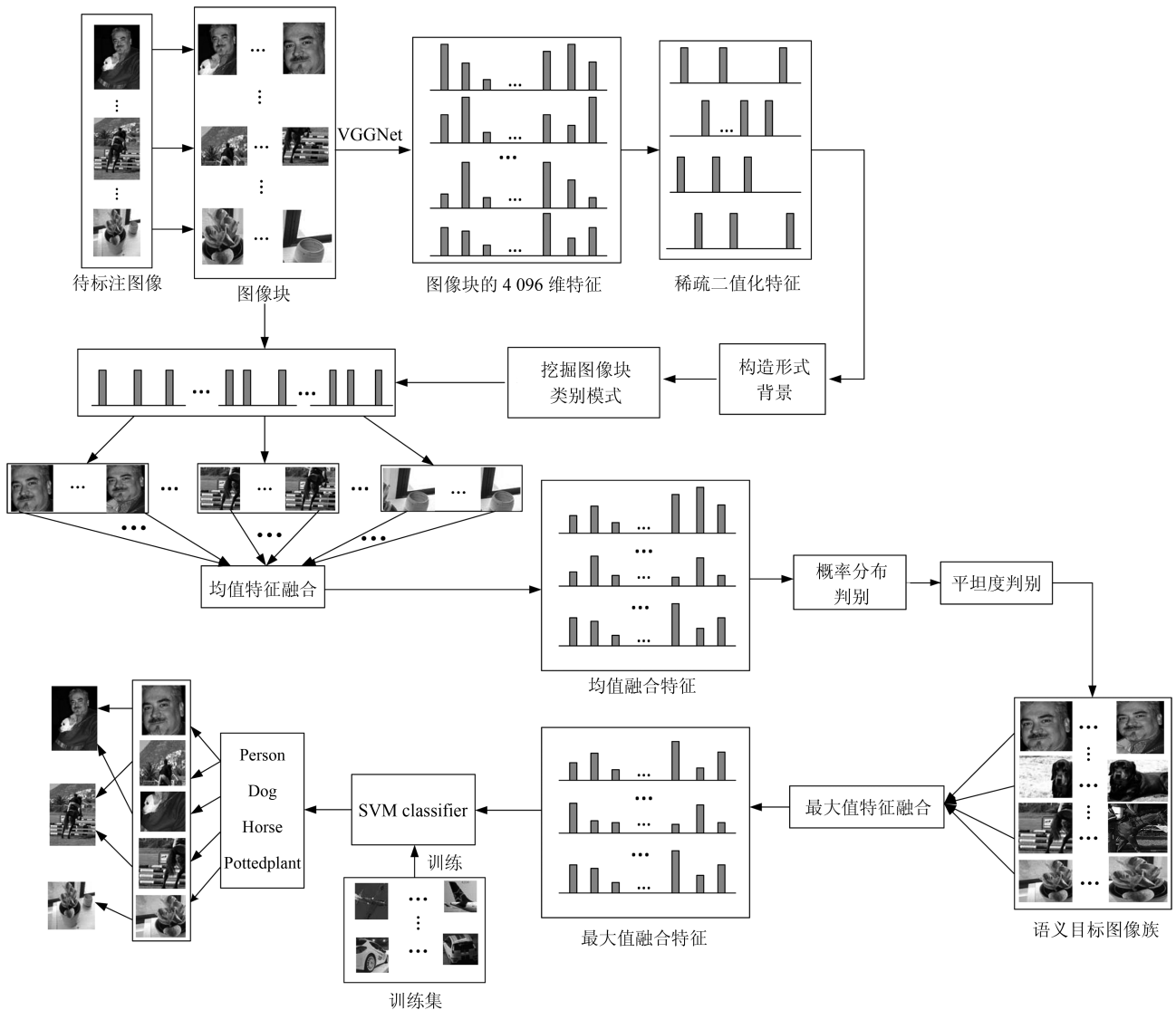


图 5 基于类别模式和特征融合的图像标注

Fig. 5 Image annotation based on category pattern and feature fusion



图 6 目标建议算法生成的图像局部框

Fig. 6 Local blocks generated by the objectness proposal generation algorithm

(如 ‘bottle’、‘cup’ 等), 可能因为目标过小在分割中存在较少的图像块, 这样在挖掘类别模式时就有可能造成小尺寸语义目标的缺失. 本文发现这种尺寸较小的语义目标一般在图像中都不会独立出现, 它们都会和图像中的主要语义目标 (大尺寸语义目标) 一起出现, 也就是说这种易缺失的语义目标与其他语义目标存在一定的关联规则. 因此本文利用 Apriori 算法^[27] 从大量图像标签中挖掘出标签之间的关联规则.

1) 频繁项集: 定义 $A = \{a_1, a_2, a_3, \dots, a_z\}$ 表示数据集 A 中包含所有事务中共有的 z 个元素. 如果一个事务 T 是 A 的子集, $T \subseteq A$. 定义一个事务数据集 $D = \{T_1, T_2, T_3, \dots, T_N\}$, 其中事务数据集 D 包含 N 个事务 (N 通常很大). 给定一个项目 X , $X \subseteq A$, 如果我们对事务数据集中包含 X 的事务 T 感兴趣, $T \in D$. 定义 X 的支持度为:

$$\text{supp}(X) = \frac{|\{T|T \in D, X \subseteq T\}|}{N} \in [0, 1] \quad (7)$$

如果 $\text{supp}(X) > \text{supp}_{\min}$, X 被称为频繁项集.

2) 关联规则: 定义一个关联规则 $X \rightarrow Y$, 表示为项目 X 对于项目 Y 的相关性. 本文比较感兴趣的是在事务数据集 D 中发生项目 X 时同时发生项目 Y 的可能性有多大. 为了对 X 与 Y 的相关性进行量化, 定义关联规则 $X \rightarrow Y$ 的信任度为:

$$\text{conf}(X \rightarrow Y) = \frac{\text{supp}(X \cup \{Y\})}{\text{supp}(X)} = \frac{|\{T|T \in D, (X \cup \{Y\}) \subseteq T\}|}{|\{T|T \in D, X \subseteq T\}|} \in [0, 1] \quad (8)$$

在应用中挖掘到的关联规则应该是相对可靠的, 所以关联规则中的信任度值一般比较大.

本文利用 Apriori 算法从大量图像的语义标注词中挖掘出语义之间的关联规则. 首先将所有图像语义标签的类别进行统计得到所有事务共有的元素集合 A . 在挖掘关联规则时所有事务都必须以布尔型数据表示^[24], 所以本文将每张图像对应的语义标签作为一个事务对应 A 中元素构建布尔型事务数据集 D . 然后根据 Apriori 算法由简单到复杂对语义之间的关联规则进行挖掘, 得到语义关联规则集. 将体系第 1.2 节得到的图像初始标注作为发生项目 X ,

根据 X 从语义关联规则集中提取与 X 相关联的潜在语义 Y . 由于缺失图像簇中的大部分图像块为背景块, 无法使用分类器识别出准确的补充语义, 而本文利用关联规则挖掘出的潜在语义 Y 利用信任度 conf 保证了语义的可靠性, 所以对缺失图像簇中的缺失语义与潜在语义 Y 求交集可以得到最终可信的补充语义. 语义补充的详细过程如图 7 所示.

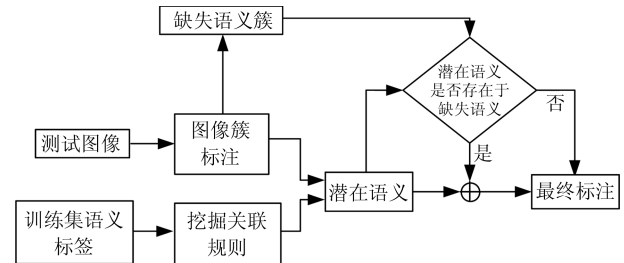


图 7 基于关联规则的语义补充

Fig. 7 Semantic complement based on association rules

2 实验结果及分析

2.1 实验数据集及参数设定

在本文使用 VOC 2007 数据集对实验方法进行验证, VOC 2007 数据集中包括 5 011 张训练样本图像, 4 952 张测试样本图像. 数据集中共有 aeroplane、bicycle、bird、boat、bottle、bus、car、cat、chair、cow、diningtable、dog、horse、motorbike、person、pottedplant、sheep、sofa、train、tvmonitor 20 类语义目标, 并提供了语义目标在图像中的区域位置. 实验中将训练集图像中的语义目标切割作为 SVM 分类器的训练样本. 考虑到每个语义类中的图像数不同, 每次实验从各目标类图像训练样本中随机选取 250 张图像作为训练样本, 从测试样本中随机选取 500 张图像作为测试样本. 重复 5 次实验求取平均值作为实验结果.

实验环境: 64 位 Windows 7 系统, CPU 主频 6.4 GHz, 16 GB 内存, MATLAB 2015a.

本文利用选择性搜索算^[12] 法对图像进行切割, 选择性搜索算法通过考虑颜色、纹理、尺寸和空间交叠这 4 个参数尽可能地使切割的图像块包含语义目标. 考虑到语义目标在图像中所占的大小和尺寸, 本文在实验中对选择性搜索生成图像块时在大小和纵横比上添加了限制, 使生成目标块的纵横比在 $1/4 \sim 4$ 之间, 目标块大小在原图的 $1/8 \sim 2/3$ 之间.

在挖掘类模式模块中, 本文设置了三个参数特征稀疏化阈值 d 、外延维度阈值 α 和内涵维度阈值 β . 针对参数值的设置, 本文从数据集中随机抽取 20 幅图像, 通过 10 次实验取平均值对三个参数进行交叉验证.

在第 1.1 节中, 本文从特征表征角度已经验证了当 $d \geq 20$ 时, CNN- k 特征将可以很好地代表图像语义特性. 本节将通过类模式挖掘算法的时间复杂度 $O(d)$ 来对 d 值进一步分析, d 值与算法时间复杂度 $O(d)$ 的关系如图 8 所示. 从图 8 中可以看出, 随着 d 值的增加, 算法时间复杂度 $O(d)$ 成指数倍增长. 为了使算法在较小时间复杂度下取得较好的实验结果, 本文设置参数 $d = 20$.

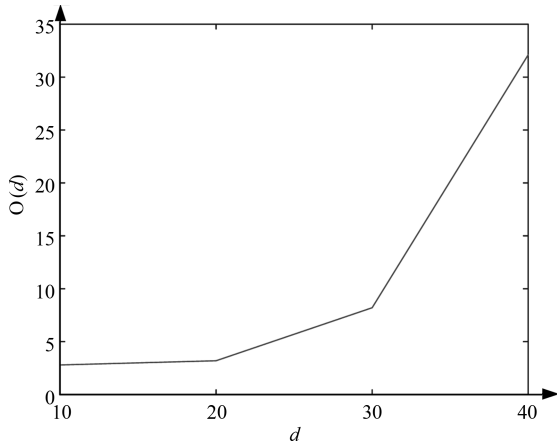


图 8 d 与算法时间复杂度 $O(d)$ 的关系

Fig. 8 The relationship between d and time complexity $O(d)$

类模式挖掘算法的最终目的是将图像中具有相似特性的图像块进行聚类, 算法中外延维度阈值参数 α 用于保证图像簇中的最少元素个数, 可以根据实验需求直接确定 $\alpha = 3$. 内涵维度阈值参数 β 为用于保证同一图像簇中元素的相似度, β 的改变将直接影响到算法的聚类效果. 本文使用 Silhouette 指标和平均图像簇数量 $Mc(\beta)$ 对算法中 β 的取值进行验证, 实验结果如表 5 所示.

$$sil(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (9)$$

其中, $a(i)$ 为样本 i 与类内所有其他样本的平均距离, $b(i)$ 为样本 i 到其他类样本平均距离的最小值. Silhouette 指标为所有样本 $sil(i)$ 的平均值, Silhouette 取值在 $(-1, 1)$ 之间, 值越大说明聚类效果越好.

从表 5 中可以看出, 随着内涵维度阈值 β 的增加, Silhouette 指标随之增加, 说明图像簇的聚类效

果越来越好, 但同时平均图像簇数量 $Mc(\beta)$ 却随之下降, 说明随着 β 的增加, 构建图像簇过程中缺失的图像块越来越多. 在保证良好聚类效果的同时, 尽量保证图像语义的完整, 本文在实验中设置参数 $\beta = 5$.

在平坦度去噪模块中, 计算图像簇平坦度时借鉴文献^[16] 设置的参数 $m = n = 16$. 为了得到准确的平坦度阈值 th 进行去噪, 本文从实验数据中选择部分图像块作为平坦度验证集如图 9(a) 所示, 其中具有明显背景的图像块作为正例, 具有明显前景的图像块作为负例, 分类实验中每次从正例和负例中各随机抽取 20 张图像块进行交叉验证. 在 $[0, 1]$ 范围内改变 th 值, 得到每个 th 值对应的分类准确率 p 如图 9(b) 所示, 从而选择最优阈值 th .

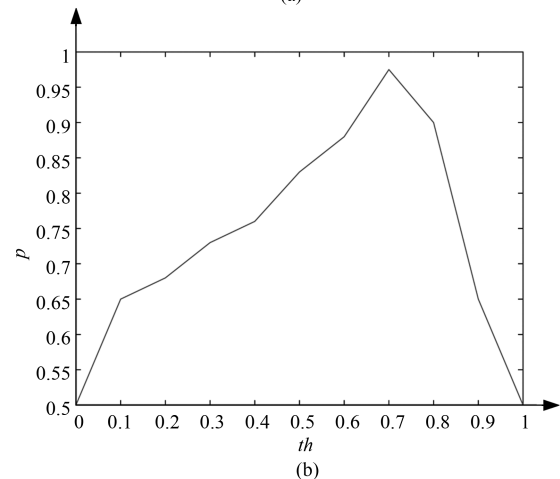
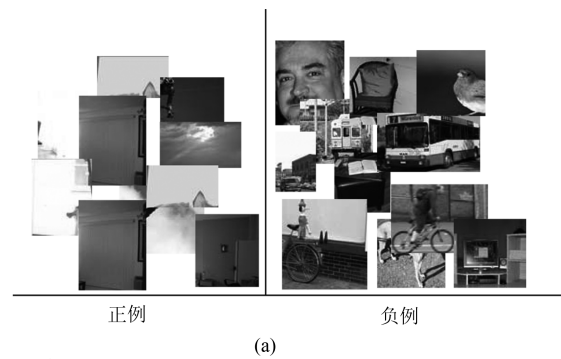


图 9 平坦度阈值 th 对实验性能的影响

Fig. 9 Performance on the flatness threshold th

从图 9(b) 中可以看出, 当 th 值取 0 时, 模型将所有测试图像块判断为平坦, 所以分类准确率为

表 5 参数 β 对实验性能的影响

Table 5 Performance on parameter β

β	2	3	4	5	6	7	8
Silhouette	0.16	0.232	0.33	0.63	0.662	0.72	0.761
$Mc(\beta)$	6	5.4	5.1	4.9	4.2	3.4	2.8

0.5. 随着 th 值的增加, 分类准确率也随之增加, 当 $th = 0.7$ 时分类准确率达到最高值为 0.975. 之后准确率开始下降, 当 $th = 1$ 时模型将所有图像块判断为不平坦, 所以分类准确率为 0.5.

在语义补充模块中, Apriori 关联规则挖掘算法中存在最小支持度 $supp_{min}$ 和最小信任度 $conf_{min}$ 两个阈值参数. $supp_{min}$ 设置的太小将会导致过大的时间复杂度, 设置的过大将导致计算出的频繁项集过少而不利于挖掘关联规则. 同时, $conf_{min}$ 设置的过小将导致挖掘出的关联规则不准确, 设置的过大将导致挖掘出的关联规则太少.

针对 $supp_{min}$ 参数, 本文使用 VOC2007 训练集中的语义词进行不同阈值的频繁项集计算, 使用时间复杂度 $O(supp_{min})$ 和频繁项个数 $N(supp_{min})$ 进行评判, 如表 6 所示.

从表 6 中可以看出, 随着 $supp_{min}$ 的增加, 时间复杂度 $O(supp_{min})$ 在持续减少, 同时挖掘出的频繁项个数 $N(supp_{min})$ 也在减少. 当 $supp_{min}$ 从 2×10^{-4} 增加到 1×10^{-3} 时, 时间复杂度出现骤减, 后随着 $supp_{min}$ 的增加, 时间复杂度的减少较为缓慢. 考虑到当 $supp_{min} = 1 \times 10^{-3}$ 时, 频繁项个数也较为丰富, 比较适合挖掘关联规则, 所以本文设置 $supp_{min} = 1 \times 10^{-3}$.

针对 $conf_{min}$ 参数, 本文利用 VOC2007 训练集中的语义词进行语义挖掘, 然后使用测试集中的语义词进行语义补充, 通过计算挖掘关联规则数 $N(conf_{min})$ 和补充准确率 $P(conf_{min})$ 对其进行评判, 如表 7 所示. 从表 7 中可以看出, 随着阈值参数 $conf_{min}$ 的增加, 挖掘出的关联规则数 $N(conf_{min})$ 随之减少. 虽然当 $conf_{min}$ 取值较大时可以得到很准确的语义补充, 但此时关联规则数 $N(conf_{min})$ 太少, 将起不到语义补充的作用. 因此, 为了在满足语义补充要求的同时保证语义补充的准确率, 本文取 $conf_{min} = 0.6$.

综上所述, 本文实验参数设置如下: 在挖掘图像

簇的类别模式时设置稀疏二值化阈值 $d = 20$, 外延维度阈值 $\alpha = 3$, 内涵维度阈值 $\beta = 5$; 在计算平坦度时 $m = n = 16$, 阈值 $th = 0.7$; 在语义补充时挖掘关联规则算法中最小支持度 $supp_{min} = 1 \times 10^{-3}$, 最小信任值 $conf_{min} = 0.6$.

2.2 基于模式挖掘的图像聚类

本文在第 1.1 节提出了基于模式挖掘的聚类算法 (PMC). 为了验证该算法的优越性, 本文从 VOC2007 数据集中随机选择 100 张测试图进行切割处理, 然后分别利用 PMC、 k -means 和 AP (Affinity propagation) 算法进行聚类, 通过对三种聚类算法的 Silhouette 值和时间复杂度 $O(ct)$ 进行比较来验证本文聚类算法 PMC 的优越性. 在实验中 k -means 算法需要指定聚类中心的 k 值, 为了使实验效果更具对比性, 本文将设置 k 值等于每张图像 PMC 生成图像簇的个数, 实验结果如表 8 所示.

从表 8 可以看出, 模式挖掘具有较大运算量, 所以 PMC 算法在时间复杂度上高于 k -means 和 AP 算法. 但是 PMC 聚类算法的 Silhouette 值要明显高于 k -means 和 AP 算法, 这说明 PMC 在多目标图像块的聚类效果上明显优于其他聚类算法. 在图像标注算法中图像簇的准确性尤其重要, 所以本文 PMC 聚类算法更适合.

2.3 图像标注评判标准

本文使用准确率 P 、召回率 R 和 F 值作为图像标注实验的性能评判标准, 分别定义为:

$$P = \frac{1}{M} \sum_{j=1}^M \frac{Correct(w_j)}{Predicted(w_j)} \times 100\% \quad (10)$$

$$R = \frac{1}{M} \sum_{j=1}^M \frac{Correct(w_j)}{Truth(w_j)} \times 100\% \quad (11)$$

$$F = \frac{2P \cdot R}{P + R} \times 100\% \quad (12)$$

表 6 $supp_{min}$ 对实验性能的影响

Table 6 Performance on parameter $supp_{min}$

$supp_{min}$	2×10^{-4}	1×10^{-3}	2×10^{-3}	3×10^{-3}	4×10^{-3}
$O(supp_{min})$	2.04	0.57	0.31	0.19	0.16
$N(supp_{min})$	281	135	96	75	73

表 7 $conf_{min}$ 对实验性能的影响

Table 7 Performance on parameter $conf_{min}$

$conf_{min}$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$N(conf_{min})$	122	87	71	55	37	16	10	5
$P(conf_{min})$	0.46	0.54	0.62	0.72	0.84	0.86	0.94	0.98

表 8 三种聚类算法的比较

Table 8 Comparison of three clustering algorithms

	PMC	<i>k</i> -means	AP
<i>o(ct)</i>	3.20	1.58	1.31
Silhouette	0.68	0.33	0.39

其中, M 为待标注图像的总数目, w_j 为第 j 张待标注图像生成的语义标签, $Correct(w_j)$ 为生成语义标签 w_j 中正确标签的数目, $Predicted(w_j)$ 为生成语义标签 w_j 的总标签数, $Truth(w_j)$ 为第 j 张待标注图像真值标签的数目.

同时, 本文还利用 mAP 来作为图像标注的另一性能评判标准, mAP 定义为:

$$mAP = \frac{1}{N} \sum_{i=1}^N \frac{Correct(m_i)}{Predicted(m_i)} \times 100\% \quad (13)$$

其中, N 表示数据集中共有 N 类语义标签, 本文中选用 VOC2007 数据集进行实验, 即 $N = 20$; m_i 为包含第 i 类语义标签的图像. $Correct(m_i)$ 表示 m_i 中语义标签类别正确的图像数, $Predicted(m_i)$ 表示所有预测为 m_i 的图像数.

2.4 图像标注实验

本文从图像语义目标的局部特征标注出发, 利用平坦度消除背景噪声簇, 利用形式概念分析和特征融合将图像块判决提升到图像簇判决: 为了保证标注的语义正确性的同时进一步提升语义标注的完整性, 本文基于挖掘的语义关联规则进行语义补充. 为了验证所提方法的有效性, 本文设计了 4 个图像标注实验: 基于局部图像块的图像标注实验 (Image block annotation, IBA)、基于图像块和平坦度的图像标注实验 (Image block and flatness annotation, IBFA)、基于图像簇和平坦度的图像标注实验 (Image cluster and flatness annotation, ICFA) 和基于 ICFA 及语义补充的图像标注实验 (Image cluster and flatness annotation + semantic complement, ICFA + SC).

其中, 基于局部图像块的图像标注实验 (IBA) 是通过目标建议生成算法生成的局部图像块预处理后, 计算图像局部图像块的类别分布概率来选择具有明确语义信息的图像块, 通过分类器对图像进行语义标注; 基于图像块和平坦度的图像标注实验

(IBFA) 是在计算图像局部图像块的类别分类概率后, 进一步计算局部图像块的平坦度, 通过平坦度去掉背景噪声, 然后通过分类器对图像进行语义标注; 基于图像簇和平坦度的图像标注实验 (ICFA) 是利用第 1.1 节的模式挖掘算法将图像局部块划分为少量图像簇, 然后通过特征融合算法得到图像簇特征, 之后利用图像簇特征对图像簇进行概率分布和平坦度去噪, 最后使用分类器对图像簇进行语义判别; 基于 ICFA 及语义补充的图像标注实验 (ICFA + SC), 在 ICFA 的基础上结合 Apriori 关联规则挖掘算法从 VOC 2007 训练集中挖掘出语义之间的关联规则, 如表 9 所示, 基于此设计了语义补充算法以保证标注语义的完整性.

表 9 第 1 行表示已经存在的语义, 第 2 行为由存在语义推测出来的关联语义, 第 3 行为存在语义与关联语义之间的相关度, 即存在语义推测出关联语义的准确度. 从表 9 中可以看出大部分挖掘出的关联规则符合人类理解, 比如 $\{\text{diningtable}\} \rightarrow \{\text{chair}\}$, $\{\text{bicycle, bus}\} \rightarrow \{\text{person}\}$. 但也存在个别关联规则并不符合人类理解, 比如 $\{\text{pottedplant, bottle}\} \rightarrow \{\text{person}\}$, 这是因为 VOC 2007 数据集中图像语义信息较为简单, 语义之间存在的关联规则较弱, 并且图像数量和语义类个数较少, 所以在挖掘关联规则时容易针对此数据集产生过拟合现象.

在第 1.2 节本文提出了最大值融合和均值融合两种特征融合的方法, 从表 4 可以看出这两种特征融合方法得到的融合特征均可以保留图像簇中图像块的共性, 从而对图像簇中个别错误标注进行纠错, 增加图像标注的鲁棒性. 通过对两种融合方法的进一步比较, 本文发现均值融合得到的融合特征更适合对图像簇进行去噪处理, 而最大值融合特征在对图像簇进行语义判别时更具优势. 所以为了得到更好的标注结果, 本文将两种特征融合方法进行组合, 利用均值融合方法进行图像簇模糊去噪, 最后的图像簇语义判别则使用最大值融合算法.

为了验证特征组合融合算法的优越性, 本文分别设计了基于最大值融合的 ICFA 实验、基于均值融合的 ICFA 实验和基于组合融合的 ICFA 实验. 其中, 基于最大值融合的 ICFA 实验是在对图像簇进行模糊去噪和最后语义判别时, 均使用最大值融合算法得到图像簇的融合特征; 基于均值融合的

表 9 VOC 2007 数据集中部分语义关联规则

Table 9 Partial semantic association rules in the VOC 2007 data set

存在语义	bicycle	diningtable	bicycle, bus	bottle, chair	pottedplant, bottle
关联语义	persion	chair	person	diningtable	person
相关度	0.68	0.70	0.92	0.67	0.70

ICFA 实验是在对图像簇进行模糊去噪和语义判别时均使用均值融合算法得到图像簇的融合特征: 基于组合融合的 ICFA 实验利用均值融合算法得到融合特征对图像簇进行模糊去噪, 之后利用最大值融合算法得到融合特征进行语义判别. 表 10 对三种特征融合的标注结果进行对比.

从表 10 中可以看出, 因为均值融合在标注实验中能够更好地去除更多的噪声图像簇, 所以和最大值融合相比有着更好的表现, 最大值融合方法是选择特征中的特殊值组成融合特征, 这种融合特征更能体现出特征之间的差异性, 在语义分类中有着较好的表现, 但正是这种特性使其不利于对图像簇进行去噪, 所以相对有着最低的准确率. 最后本文结合均值融合和最大值融合的优点利用均值融合对图像簇进行去噪, 利用最大值融合对图像簇进行语义判别, 在表 10 中得到了最优的实验结果.

表 10 三种特征融合方式对比

Table 10 Comparison of three feature fusion methods

	<i>P</i>	<i>R</i>	<i>F</i>
最大值融合	0.59	0.61	0.60
均值融合	0.64	0.58	0.61
组合融合	0.72	0.56	0.63

表 11 中提供了 4 种方法 (IBA、IBFA、ICFA 和 ICFA + SC) 的标注实验结果, 并从 *P*、*R*、*F* 和时间复杂度 $O(t)$ 来验证平坦度和特征融合在标注实验中的有效性.

表 11 标注实验结果对比

Table 11 Comparison of annotation results

	<i>P</i>	<i>R</i>	<i>F</i>	$O(t)$
IBA	0.44	0.72	0.55	10.94
IBFA	0.46	0.75	0.57	11.46
ICFA	0.72	0.56	0.63	10.51
ICFA + SC	0.72	0.62	0.67	11.74

从表 11 的前 4 列对 IBA、IBFA、ICFA 和 ICFA + SC 这 4 种方法进行性能分析. 从第 2 行和第 3 行来看, 对图像块添加了平坦度判决后准确率和召回率有了略微的提升. 那是因为大部分的噪声块都能够通过类别概率分布区分, 只有少量的颜色和纹理单一的背景块需要通过平坦度来进行区分, 由于这种背景噪声块的数量较少, 所以实验结果只有很少的提高. 从第 3 行和第 4 行可以看出, 利用形式概念分析的方式将图像块生成图像簇后, 使用图像簇的融合特征进行标注实验可以很大程度上提升标注准确率. 因为考虑到计算量的问题, 所以对类别模式挖掘时将 *d* 设置的较小, *Pr* 设置的较大, 造

成语义类别有一定的缺失, 所以在 ICFA 标注中召回率 *R* 相对较小.






但综合实验效果来看, 将图像提升到图像簇利用相同语义块之间的共性来进行识别标注对语义标注效果有着很大的提升. 为了尽量减少类别模式对标注语义缺失的影响, 本文将 ICFA 算法结合语义关联规则进行语义补充 (ICFA + SC). 从表 11 中的第 4 行和第 5 行可以看出, 经过语义补充后语义标注在准确率没有降低的前提下, 召回率有了较为明显的提升.

表 11 中的 $O(t)$ 列对 IBA、IBFA、ICFA 和 ICFA + SC 这 4 种方法进行了时间复杂度分析. IBFA 相比 IBA 多了一个平坦度去噪模块, 所以前者时间复杂度高于后者; 对比 IBFA 和 ICFA 可知, ICFA 相较于 IBFA 时间复杂度略有下降, 这是因为 IBFA 算法在模糊去噪模块和语义判别模块都需要对图像的所有图像块进行遍历, 而 ICFA 是对图像簇进行处理, 虽然在图像簇生成需要耗费一定的时间, 但是后续过程中采用特征融合方法生成簇特征, 使得模糊去噪模块和语义判决模块耗时较少, 整体上 ICFA 耗时少于 IBFA; ICFA + SC 方法在 ICFA 的基础上增加了语义补充模块, 所以 ICFA + SC 时间复杂度稍高于 ICFA.

为了更清楚的对比几种实验方法的标注效果, 在表 12 中提供了 VOC 2007 数据集中部分图像的标注示例.

表 12 图像标注示例

Table 12 Annotation examples

待标注图像	IBA	IBFA	ICFA	ICFA + SC
	bicycle chair sofa	bicycle chair	bicycle	bicycle
	motorbike person train	motorbike person train	motorbike person	motorbike person
	diningtable person pottedplant	diningtable person pottedplant	diningtable person	diningtable person chair
	diningtable pottedplant sofa tvmonitor	diningtable pottedplant tvmonitor	diningtable pottedplant chair	diningtable pottedplant chair
	diningtable person sofa pottedplant	diningtable person sofa pottedplant	diningtable person sofa	diningtable person sofa chair

从表 12 的第 2、3 列可知, 对图像块进行平坦度处理可以去除一些图像中个别错误的词汇; 将第 4 列与前两列对比可以看出, 按照类别模式将具有相似性的同类图像块聚集成图像簇, 利用图像簇中的融合特征进行去噪和判决在很大程度上提高了图像标注的准确性; 比较表 12 的后两列可以发现, 从训

练集中挖掘出来的语义关联规则可以对标注语义提供有效的语义丰富。

为了进一步评测本文提出特征组合融合算法和平坦度去噪的有效性, 本文对数据集中每类的标注结果进行了准确率统计, 通过对 IBA、IBFA、ICFA 算法标注的每类准确率和 mAP 对比来验证特征组合融合算法和平坦度去噪的有效性 (需要说明的是, 由于语义补充算法对标注词语义的完整性有较大改善, 对标注词的准确性并无明显改善, 如表 11 所示, 所以此处没有对比 ICFA + SC 算法)。此外, 本文算法与多语义分类算法 AMM^[28] 和 CNN-SVM^[29] 进行比较, 来说明 ICFA 算法的优越性, 实验结果如图 10 所示。

由图 10 可以看出, 对于所有的 20 类图像, ICFA 算法的准确率均高于 IBA 和 IBFA 算法。虽然在个别少数类图像的标注结果中, ICFA 算法的准确率略低于 AMM 算法或 CNN-SVM 算法, 比如 cat 和 chair 类, 但是在大部分图像类中 ICFA 的标注结果具有优势, 并且从 mAP 上来看, ICFA 均高于其他方法。此外, 通过 IBFA 和 IBA 的准确率比较可以看出, 在图像块去噪过程中添加了平坦度去噪模块后, 在背景较为单一的 plane 和 bird 类中 IBFA 的准确率相比较 IBA 有着较为明显的提升, 这也说明

了平坦度去噪在标注实验中具有一定的贡献度。

3 结论

本文从图像语义目标的局部特征出发, 提出了基于形式概念分析类别模式挖掘算法。以图像块为对象, 以图像块的稀疏二值化特征为属性构建形式背景, 构建概念格, 依据概念的外延和内涵, 对图像块进行聚类并归化为不同的图像簇, 根据图像簇内图像块之间的共性实现图像语义目标的标注。特征的同性由特征融合来实现。本文通过实验发现最大值特征融合和均值特征融合两种融合方法中前者得到的融合特征更有利于对图像簇进行语义判别, 后者得到的融合特征更有利于对图像簇进行模糊去噪, 从而设计了组合特征融合算法得到了最优的特征融合框架。在挖掘类别模式时由于内涵阈值的限制使得图像簇中可能存在个别语义缺失, 本文通过挖掘语义之间的关联规则进行语义补充, 使语义标注在不影响准确率的情况下提升语义标注的召回率。本文方法也存在不足之处, 比如本文方法依赖提前训练好的分类器, 并且由于训练数据集中语义类别较少和数据集数量较少, 导致挖掘出的语义关联规则不够完备等。

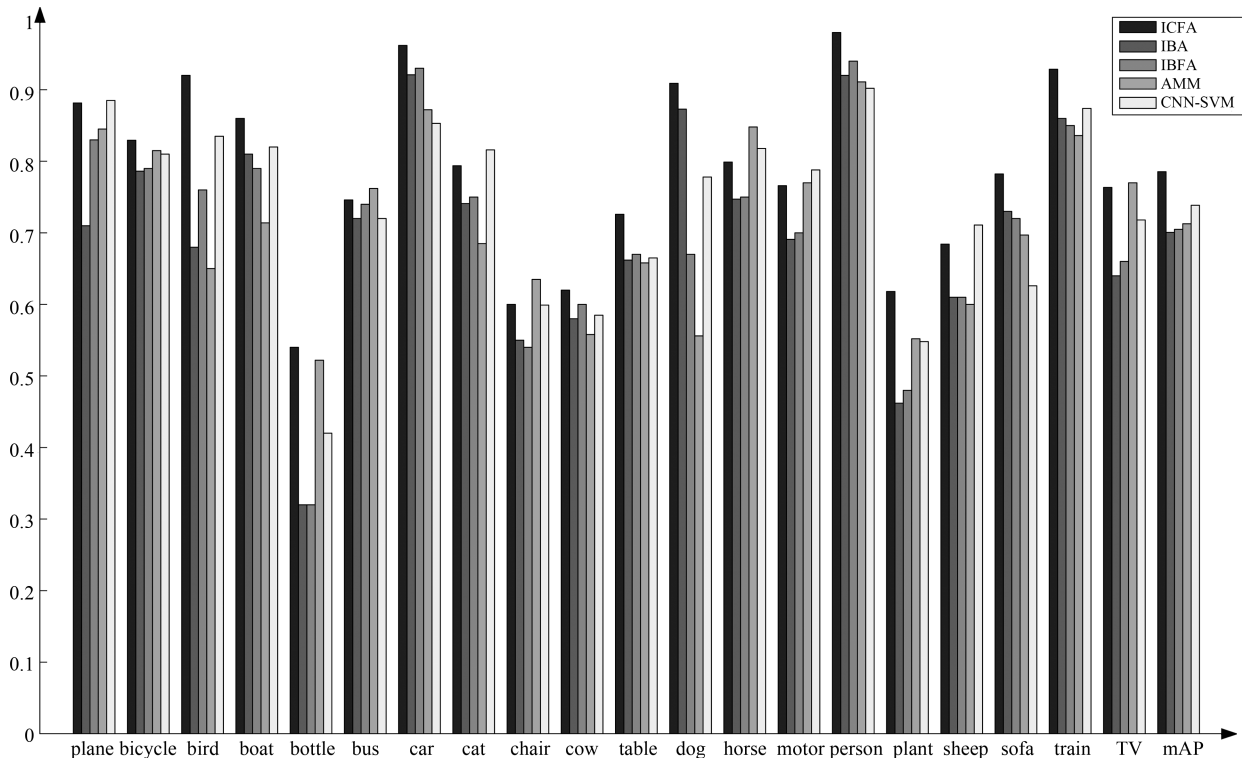


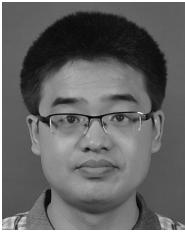
图 10 图像标注方法准确率比较

Fig. 10 Precision comparison of several annotation methods

References

- 1 Duygulu P, Barnard K, Freitas J F G D, Forsyth D A. *Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary*. Berlin: Springer, 2002. 97–112
- 2 Qu S, Xi Y, Ding S. Visual attention based on long-short term memory model for image caption generation. In: Proceedings of the 2017 the Chinese Control and Decision Conference. Chongqing, China: IEEE, 2017. 4789–4794
- 3 Lin T Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii, USA: IEEE Computer Society, 2017. 2117–2125
- 4 Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, **39**(6): 1137–1149
- 5 Wang J, Yang Y, Mao J, Huang Z H. CNN-RNN: a unified framework for multi-label image classification. In: Proceedings of Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA: IEEE, 2016. arXiv: 1604.04573
- 6 Tang J, Li H, Qi G J, Chua T S. Image annotation by graph-based inference with integrated multiple single instance representations. *IEEE Transactions on Multimedia*, 2010, **12**(2): 131–141
- 7 Wu B, Jia F, Liu W, Ghanem B, Lyu S. Multi-label learning with missing labels using mixed dependency graphs. *International Journal of Computer Vision*, 2018, **126**(8): 875–896
- 8 Kong X, Wu Z, Li L J, Zhang R, Yu P S, Wu H, et al. Large-scale multi-label learning with incomplete label assignments. *ArXiv preprint*, 2014
- 9 Jin C, Jin S W. Image distance metric learning based on neighborhood sets for automatic image annotation. *Journal of Visual Communication & Image Representation*, 2016, **34**(C): 167–175
- 10 Chen Y, Zhu L, Yuille A, Zhang H. Unsupervised learning of probabilistic object models (POMs) for object classification, segmentation, and recognition using knowledge propagation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2009, **31**(10): 1747–1761
- 11 Yang C, Dong M. Region-based image annotation using asymmetrical support vector machine-based multi-instance learning. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 17–22
- 12 Uijlings J R R, van de Sande K E A, Gevers T, Smeulders A W M. Selective search for object recognition. *International Journal of Computer Vision*, 2013, **104**(2): 154–171
- 13 Cheng M M, Zhang Z, Lin W Y, Torr P. Bing: binarized normed gradients for objectness estimation at 300 fps. In: Proceedings of the Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014. 3286–3293
- 14 Felzenszwalb P F, Mcallester D A, Ramanan D. A discriminatively trained, multiscale, deformable part model. In: Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, Alaska, USA: IEEE, 2008. 24–26
- 15 Moran S, Lavrenko V. A sparse kernel relevance model for automatic image annotation. *International Journal of Multimedia Information Retrieval*, 2014, **3**(4): 209–229
- 16 Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts. *Order Sets D Reidel*, 1982, **83**: 314–339
- 17 Thomas, J. Cook, K. A visual analytics agenda. *IEEE Transactions on Computer Graphics and Applications*, 2006, **26**(1): 12–19
- 18 Tsoumakas G, Katakis I, Vlahavas I. *Mining Multi-Label Data*. US: Springer, 2010: 667–685
- 19 Yang J, Yang F, Wang G, Li M. Multi-channel and multi-scale mid-level image representation for scene classification. *Journal of Electronic Imaging*, 2017, **26**(2): 023018
- 20 Girish K, Premraj V, Ordonez V, Dhar S, Li S, Choi Y. Baby talk: understanding and generating simple image descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(12): 2891–2903
- 21 Jia X, Shen L, Zhou X, Yu S. Deep convolutional neural network based HEp-2 cell classification. In: Proceedings of International Conference on Pattern Recognition. Cancun, Mexico: IEEE, 2017
- 22 Rajkomar A, Lingam S, Taylor A G, Blum M, Mongan J. High-throughput classification of radiographs using deep convolutional neural networks. *Journal of Digital Imaging*, 2017, **30**(1): 95–101
- 23 Bai C, Huang L, Pan X, Zheng J, Chen S. Optimization of deep convolutional neural network for large scale image retrieval. *Neurocomputing*, 2018: S0925231218304648
- 24 Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 2014, **115**(3): 211–252
- 25 Smirnov E A, Timoshenko D M, Andrianov S N. Comparison of regularization methods for imageNet classification with deep convolutional neural networks. *AASRI Procedia*, 2014, 6: 89–94
- 26 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, 2014: 512–519
- 27 Pang J, Huang J, Qin L, Zhang W, Qing L, Huang Q. Rotative maximal pattern: a local coloring descriptor for object classification and recognition. *Information Sciences*, 2017, 405
- 28 Chen Q, Song Z, Dong J, Huang Z, Hua Y, Yan S. Contextualizing Object Detection and Classification. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado USA: IEEE Computer Society, 2011. 1585–1592

- 29 Lu X, Chen Y, Li X. Hierarchical recurrent neural hashing for image retrieval with hierarchical convolutional features. *IEEE Transactions on Image Processing*, 2018, 1(27): 106–120



顾广华 燕山大学信息科学与工程学院教授. 2013 年获得北京交通大学信号与信息处理专业博士学位. 主要研究方向为图像理解, 图像检索. 本文通信作者.
E-mail: guguanghua@ysu.edu.cn

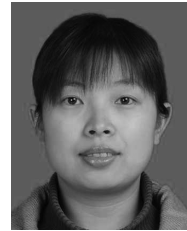
(GU Guang-Hua Professor at the School of Information Science and Engineering at Yanshan University. He received the Ph. D. degree in Signal and Information Processing from Beijing Jiaotong University in 2013. His research interest covers image understanding and image retrieval. Corresponding author of this paper.)



曹宇尧 燕山大学信息科学与工程学院硕士研究生. 2016 年获得燕山大学电子信息工程专业学士学位. 主要研究方向为多语义图像识别, 形式概念分析.
E-mail: cyg19921129@163.com

(CAO Yu-Yao Master student at the School of Information Science and Engineering at Yanshan University. He received his bachelor degree from Yanshan University in

2016. His research interest covers multi-semantic image recognition and formal concept analysis.)



崔冬 燕山大学信息科学与工程学院副教授. 2011 年获得燕山大学检测技术与自动化装置专业博士学位. 主要研究方向为医学信号处理.

E-mail: cuidong@ysu.edu.cn

(CUI Dong Associate professor at the School of Information Science and Engineering, Yanshan University. She received her Ph. D. degree in detection technology and automation equipment from Yanshan University in 2011. Her main research interest is medical signal processing.)



赵耀 北京交通大学信息科学研究所教授. 1996 年获得北京交通大学信号与信息处理专业博士学位. 主要研究方向为多媒体技术.

E-mail: yzhao@bjtu.edu.cn

(ZHAO Yao Professor at the Institute of Information Science at Beijing Jiaotong University. He received his Ph. D. degree in signal and information processing from Beijing Jiaotong University in 1996. His main research interest is multimedia technology.)

2016. His research interest covers multi-semantic image recognition and formal concept analysis.)