

文章编号: 2095-2163(2019)02-0001-09

中图分类号: TP391.4

文献标志码: A

# 基于部件检测的细粒度图像分割

庞程, 姚鸿勋, 孙晓帅

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

**摘要:** 细粒度图像中物体的分割是具有挑战性的, 因为这类图像一般具有很大的表观变化和混杂的背景。大多数已有的分割方法都不能以足够的准确率将细粒度图像中物体的细小部件分割出来。然而在细粒度识别任务中, 这些细小的部件通常包含了对细粒度分类极为重要的语义信息。通过观察发现, 细粒度物体通常在类间共享相同的部件种类, 本文由此提出一种新颖的基于部件检测的细粒度图像分割方法。该方法明确地检测部件在图像中的位置, 给出部件位置假设。然后通过不断地迭代更新部件假设和分割的输出假设, 以获得更优的分割效果。实验表明本文方法能够很好地保留具有语音信息的部件, 提高细粒度分类的准确率。

**关键词:** GrabCut; 细粒度分类; 图像分割; 图像分类

## Fine-grained image segmentation based on part detection

PANG Cheng, YAO Hongxun, SUN Xiaoshuai

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

**[Abstract]** It is challenging to segment fine-grained objects due to appearance variations and clutter of backgrounds. Most of existing segmentation methods hardly separate small parts of the instance from its background with sufficient accuracy. However, such small parts usually contain important semantic information, which is crucial in fine-grained categorization. Observing that fine-grained objects almost share the same configuration of parts, the paper presents a novel part-aware segmentation method, which explicitly detects semantic parts and preserves these parts during segmentation. The paper firstly designs a hybrid part localization method, which generates accurate part proposals with moderate computation. Then the paper iteratively updates the segmentation outputs and the part proposals, which obtains better foreground segmentation results. Experiments demonstrate the superiority of the proposed method, as compared to state-of-the-art segmentation approaches for fine-grained categorization.

**[Key words]** GrabCut; fine-grained categorization; image segmentation; image classification

## 0 引言

以往的研究表明, 图像分类任务的准确率可以通过分割等预处理过程获得提升。这一结论在图像细粒度分类领域也得到了证实<sup>[1-2]</sup>。但是一般的图像分割方法并不能很好地利用细粒度研究对象的部件信息, 其分割过程往往会丢掉一些部件, 将其错误地分类成为背景。而据已有的研究分析可知, 物体的部件在细粒度视觉分析中起到至关重要的作用, 因其承载了许多具有判别力的局部特征。如果忽略这些重要的部件或者局部特征, 细粒度分类的性能反而会因为分割造成一定的损失。基于上述分析, 本文提出了一种基于部件检测的细粒度图像分割方法, 目的是利用部件信息指导分割的过程, 尽可能地保留物体的重要部件, 提高分割的准确性和后续分类的性能。

本文方法分为定位阶段和分割阶段。在定位阶段, 将对物体的部件进行定位, 获得一些部件的位置信息的假设。然后就是分割阶段, 研究利用获得的部件假设指导 GrabCut<sup>[3]</sup> 方法的分割过程, 达到优化分割的目的。具体来说, 就是在设计的优化函数中, 迭代地更新部件假设和前景分割的前景假设, 使得两者最终达到统一。即如果部件假设中认定某些像素可能属于物体的部件, 但是在分割的前景假设中这些像素属于背景, 那么就增加这些像素在分割中属于前景的概率并同时减少部件假设中其属于部件的假设。图 1 显示了本文方法和其它方法的效果。

作为该方法核心的定位阶段, 研究为其设计了 2 种定位方法, 分别是基于弱监督的部件定位方法和基于无监督的部件定位方法。但是在实际应用中, 部件的真实信息的获取通常较为困难, 只有少数

**基金项目:** 国家自然科学基金(61472103, 61772158, 61702136)。

**作者简介:** 庞程(1988-), 男, 博士研究生, 主要研究方向: 细粒度图像分类、模式识别; 姚鸿勋(1965-), 女, 博士, 教授, 博士生导师, 主要研究方向: 图像处理、模式识别、多媒体信息处理与挖掘等; 孙晓帅(1984-), 男, 博士, 讲师, 主要研究方向: 图像处理、模式识别。

收稿日期: 2018-05-22

的数据集提供了相关信息。如何在没有专家知识或者少量专家指导的情况下,自动地获取这些对于细粒度视觉分析至关重要的信息,是本文研究中拟着重考虑的问题。在基于弱监督的部件定位方法中,研究利用了数据集提供的真实部件信息训练定位模型,然后利用定位模型推理查询图像中的物体部件信息。因为研究中仅仅在训练阶段使用了真实的部件信息,而一旦定位模型训练完成后,就可以在任何图像中实现自动的部件定位。更进一步,对于连训练数据都没有真实部件信息的情况,研究继而提出了无监督的部件定位(发现)方法。该方法仅仅利用一些底层的视觉特征,挖掘图像局部区域可能存在的部件,然后利用这些部件训练定位部件模型。相关的部件发现的工作,比如文献[4]也利用了图像底层视觉特征。但是本文提出的部件发现策略首次引入了姿态一致性的约束,使得本文发现的部件能够更好地覆盖研究对象。



图1 本文方法示意图

Fig. 1 Illustration of the proposed method

和本文提出的分割方法最相关的是文献[2]研发的共生的细粒度分类和分割方法。该方法第一次发现了细粒度图像分类和物体检测具有一定的协同作用,将部件检测和分割相结合可以提升细粒度视觉分析的性能。在设计框架中分别使用 Deformable Part Model (DPM)<sup>[5]</sup>和 GrabCut 作为其检测和分割算法。本文提出的方法与上述方法的不同主要表现在如下2个方面:

(1) DPM 仅仅给出了一些粗略的部件的检测结果,而本文方法可以利用更富含语义信息的专家定义的部件训练模型进行检测。

(2) 研究并不是在每次分割迭代之后都重新检测部件的位置,而是迭代地更新本文的部件检测假设和分割前景假设,为的是最大限度地保留尽可能多的对细粒度视觉分析有用的部件。

## 1 基于弱监督的部件检测

时下现有的部分成果已经证明了,精确的物体部件检测能够显著地提高细粒度分类的性能<sup>[1-2]</sup>。

只是这些研究一般都基于参数化模型的工作,同时也都使用了这一假设:部件的位置分布遵循的是高斯分布。但已有证明指出这个假设是不正确的<sup>[6]</sup>,部件的分布并不遵循高斯分布。而且,这些工作使用的参数化模型也不足以精确地描述一些无规律的部件分布模式。为此本文提出了一种基于弱监督的,结合参数模型和非参数模型的混合部件检测方法。该方法能够处理一些姿态变化巨大的部件分布模式,并且在较低的计算开销下能够足够精确地得到部件检测结果。

文献[6]的方法也使用了非参数的部件模型。方法中利用了底层视觉特征,比如全局的形状,来进行粗略的部件对齐。区别于该方法的是,本文方法仅仅将非参数模型的结果作为先验知识,后续还会利用参数模型进行更精确的检测。设计过程中,研究在第一阶段利用非参数部件传递,将一些训练样本的真实部件坐标赋予查询样本。而后在第二阶段,研究在传递的部件的邻域内进行精确的部件检测,使用的是参数化的部件检测子。对于传统的利用滑动窗口进行检测的方法,本文提出的混合部件检测方法的优点可表述为:

(1) 结合了参数化和非参数化模型的优点,既能处理一些姿态变化极大的情况,又能在较低计算开销的情况下给出较为准确的检测结果。

(2) 研究中第二阶段的在部件邻域内进行检测的方法,能够避免不同部件因为具有相似的表现特征而造成的检测混淆。

如图2所示,弱监督部件检测模型框架分为2个阶段,即:非参数的部件坐标传递和参数化的部件检测。对于该算法设计步骤可阐释如下。

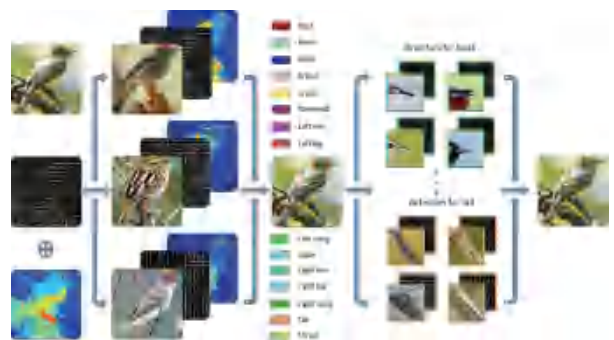


图2 弱监督部件检测框架

Fig. 2 Pipeline of the semi-supervised part detection

在非参数部件坐标传递阶段,所有样本的最小边界框中的图像被缩放到  $256 * 256$  尺寸,图像的 HOG 特征被用作底层视觉特征描述物体的姿态。对于一幅待检测部件的查询图像,研究首先在训练

样本中搜索和查询图像的姿态最相近的  $k$  个近邻样本。然后,求取这些近邻的各个部件的平均坐标,并传递给查询图片作为查询图片对应部件的可能的位置。因为通过观察分析发现,整体姿态相近的 2 个样本,其组成部件的相对位置也是非常接近的。

在参数化的部件检测阶段,假设已为每一个部件训练了一个部件检测子,研究在上一步每个可能的部件位置的邻域,使用对应该部件的检测子进行滑动窗口搜索。算法将检测子检测分数最大的位置作为检测到的部件位置。接下来将会介绍并用到训练部件检测子的方法。首先根据物体姿态对训练样本进行聚类,获得具有相近姿态的若干类别,基于此将为每一个姿态训练单独的部件检测子。然后对于第  $i$  个部件,研究收集这个姿态类别中所有样本在部件  $i$  位置的图像块作为该部件的正例训练图像,而负例训练图像则采集自背景区域。需要指出的是,即使是同一部件,因为细粒度研究对象的巨大类别数,也可能呈现出多样化的外观。如果为一个部件仅仅训练一个检测子,那么获得的检测子可能具有较低的判别能力,因为其需要同时兼顾处理多样化外观的泛化能力。于是,研究有针对性地提出为一个部件训练多个检测子。其训练图像可由上述训练图像通过聚类得到,训练使用 SVM 分类器。那么最终研究为第  $i$  个部件训练的检测子实际上就是一个包含  $SVM_1^i, SVM_2^i, \dots, SVM_m^i$  在内分类器集合。

为了评估本文提出方法的计算开销,研究假设一幅图像的大小为  $n * n$  像素,搜索的步长为 1 个像素。那么一个传统的基于滑动窗口的搜索算法搜索全图像将需要  $n * n$  个 CPU 时间。而在本文提出的算法中,研究关注的是在检测到的部件邻域内进行有限的搜索而不是全图。如果研究中设定这个邻域的大小为图像边长的  $1/m$ ,那么本文的算法仅用  $1/m^2$  的 CPU 时间就可以完成搜索,并且还避免了不同部件之间的相互干扰。从效率和准确性方面都较一般的滑动窗口方法更具优势。

## 2 基于部件检测的细粒度图像分割

这一节中,拟将论述本文提出的基于部件检测的细粒度图像分割方法。该方法以 GrabCut 分割方法为基础,利用引入的部件的位置信息以指导分割过程,达到细粒度视觉分析的要求。GrabCut 使用混合高斯模型作为分割算法的颜色模型,并不断地迭代更新颜色模型。与其不同的是,本文的分割方

法则是交替更新研究中的部件假设和颜色模型,使得部件假设和分割的前景假设尽量达成一致。本文的算法不需要任何的用户介入,仅仅需要物体的最小边界框。而物体的最小边界框可以方便地使用任意检测方法得到,这不在本文的讨论范围。

### 2.1 定义

本文方法的核心思想是:如果本应该属于物体部件的像素被分割方法错误地分类为背景,那么就增大了分割模型中这些像素属于前景的概率。这就是研究中引入部件检测的原因。同时,研究还利用部件假设产生一个物体前景掩膜用于初始化分割的颜色模型,以减少因错误的模型初始化造成的分割错误。本文分割方法的目标函数可以表示为:

$$E(s, w; z) = E^{GC}(s; z) + \alpha E^P(w) + \beta E^C(s, w; z), \quad (1)$$

其中,  $z$  表示一幅图片的所有像素的 RGB 值组成的矩阵;  $s$  表示 GrabCut 分割产生的前景掩膜;  $w = [w_1, w_2, \dots, w_n]$  存储每个像素属于前景的概率;  $E^{GC}$  表示 GrabCut 的分割能量项;  $E^P$  表示部件能量项;  $E^C$  表示一致性项,且是分割能量和部件能量交互的桥梁,负责惩罚两者不一致的情况; 超参数  $P = [p_1, p_2, \dots, p_n]$ , 用于指示每个像素在部件假设中的类别,即是属于部件还是属于背景;  $w$  是根据  $P$  进行初始化的,如果  $z_i$  代表的像素是属于部件的,那么  $w_i$  就会被初始化为 1, 否则为 0;  $\alpha$  和  $\beta$  是 2 个正的常数,用于平衡各个能量项。这里, GrabCut 能量项写作:

$$E^{GC}(s; z) = \sum_{n \in I} D_n(s_n; z_n) + \gamma \sum_{m, n \in C} V_{m, n}(z_m, z_n), \quad (2)$$

其中,  $D$  是高斯概率分布的负对数;  $V$  是平滑项;  $C$  是所有相邻像素的索引;  $\gamma$  是一个常数;  $I$  是图像中所有像素的索引。研究得到部件能量项的定义为:

$$E^P(w) = \sum_{n \in U} w_n, U = dif(P, s), \quad (3)$$

这里,如果一个像素在部件假设的记录中是属于部件的,但是又被分割结果  $s$  归类为背景,那么则称这个像素为不一致像素。  $dif(\cdot)$  返回  $I$  中所有不一致像素的索引。据此,一致性能量项可以定义为:

$$E^C(s, w; z) = \sum_{i \in U} (1 - w_i) D_i(s_i; z_i) + \sum_{j \in \bar{U}} D_j(s_j; z_j). \quad (4)$$

其中,  $U$  表示  $\bar{U}$  的补集,而这两者的并集是  $I$ 。研究中将通过交替更新分割的输出和部件假设来最

小化这个目标函数。本节分割方法的示意图可见图3。

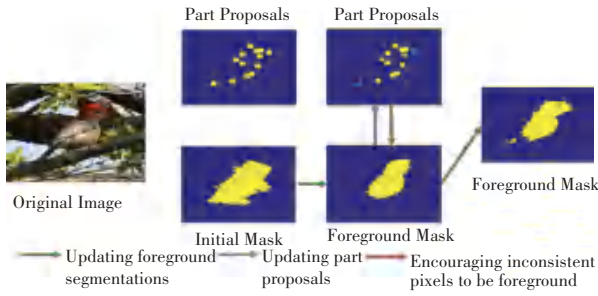


图3 基于部件检测的分割示意图

Fig. 3 Illustration of the part-aware segmentation

## 2.2 分割优化

上文的目標函数可以按照如下的方法优化,即:固定部件假设  $w$ , 更新前景分割  $s$ , 而后反过来, 如此交替, 直到目标函数的值不再有大的变化。对此可展开论述如下。

### 2.2.1 更新部件假设 $w$

固定前景分割结果  $s$ , 则分割能量项  $E^{GC}$  可以被忽略, 目标函数变为:

$$\begin{aligned} \min_w \alpha E^p(w) + \beta E^c(s, w; \mathbf{z}) &= \min_w \alpha \sum_{n \in U} w_n + \\ &\beta \sum_{i \in U} (1 - w_i) D_i(s_i; \mathbf{z}_i) + \beta \sum_{j \in \bar{U}} D_j(s_j; \mathbf{z}_j) = \\ &\min_w w_n \sum_{n \in U} (\alpha - \beta D_i(s_i; \mathbf{z}_i)) + \beta \sum_{j \in \bar{U}} D_j(s_j; \mathbf{z}_j), \end{aligned} \quad (5)$$

由于式(5)的最后一项不依赖于  $w$ , 所以式(5)可以写成:

$$\min_w w_n \sum_{n \in U} (\alpha - \beta D_i(s_i; \mathbf{z}_i)). \quad (6)$$

研究选择合适的  $\alpha$  和  $\beta$ , 使得  $\alpha - \beta D_i(s_i; \mathbf{z}_i)$  总为一个正值。因此式(6)可以看成是一个  $w$  的单调递增函数。同时, 限定每次迭代中, 每个  $w_i$  的变化量或者是0、或者是一个正常数  $\lambda$ 。当  $w_i$  减小到接近于0时, 可将其设为0并更新超参数  $P_i$ , 使得  $P_i$  不再属于部件类、而是属于背景类。这种做法的研究目的是, 只要部件假设和分割的结果不符合, 那么减少被错误分到背景的部件的像素属于前景的概率。从而可以保证最终的分割结果对于一些不准确的部件检测结果造成的误差具有一定的宽容性。

### 2.2.2 更新前景分割 $s$

固定  $w$ , 则部件能量项  $E^p$  项可以忽略。目标函数变为下面的形式:

$$\min_s E(s, w; \mathbf{z}) = E^{GC}(s) + \beta E^c(s, w; \mathbf{z}) =$$

$$\begin{aligned} &\min_s \sum_{n \in I} D_n(s_n; \mathbf{z}_n) + \gamma \sum_{(m,n) \in C} V_{m,n} + \\ &\beta \sum_{i \in U} (1 - w_i) D_i(s_i; \mathbf{z}_i) + \beta \sum_{j \in \bar{U}} D_j(s_j; \mathbf{z}_j) = \\ &\min_s \sum_{i \in U} (1 + \beta - \beta w_i) D_i(s_i; \mathbf{z}_i) + \\ &(1 + \beta) \sum_{j \in \bar{U}} D_j(s_j; \mathbf{z}_j) + \gamma \sum_{(m,n) \in C} V_{m,n}. \end{aligned} \quad (7)$$

与那些被正确分割了的像素不同, 式(7)总是尽量使得不一致像素的系数变小, 相当于鼓励分割算法将不一致像素被分类为前景。需要指出, 式(7)仍然可以使用 GrabCut 的优化方法进行求解, 这里不再赘述。

总之, 研究按照如下步骤来优化式(1), 具体流程可描述为:

- (1) 根据式(6)优化部件假设  $w$ 。
- (2) 根据  $w$  更新超参数  $P$ 。
- (3) 重新估计混合高斯颜色模型。
- (4) 根据式(7)优化前景分割  $s$ 。

算法的收敛性依赖于部件假设和 GrabCut 的分割前景假设是否达到一致。从前述分析容易看出, 两者的不一致性是由不一致像素引起的。如果部件假设中属于部件的所有像素在分割假设中被归类为前景, 由此就可达到保留这些部件的目的。即不存在不一致像素, 算法已经收敛。本文的算法将设计为交替更新部件假设和分割的前景假设, 这也是一个不断减少不一致像素的过程。当然, 如果2个假设的不一致性小于一个阈值或者已经达到最大的更新次数, 此时也可认定算法已经收敛。

## 3 实验结果与分析

研究在 CUB-200-2011<sup>[7]</sup> 鸟类数据集中和 Stanford Dogs<sup>[8]</sup> 数据集测试了本文的分割方法, 这2个数据集是细粒度视觉分析领域的基准数据集。实验部分将考虑2个方面的性能: 细粒度图像分割和细粒度分类。研究内容可分述如下。

### 3.1 数据集和参数设置

CUB200-2011 数据集包括200种鸟类的11 788幅图像, 并附有每幅图像的真实分割掩膜、类别标签、部件位置坐标、图像的最小边界框, 二值化特征属性等信息。尽管研究者已经对该数据集进行了数目可观的研究, 但是其样本表现出巨大的姿态、光照、尺度变化, 仍然使得该数据集成为细粒度视觉分析领域最具挑战性的数据集之一, 而且也还有大量的研究在将其拟定用于诸如细粒度分类等一系列任

务中。在本文实验中, 研究使用数据集提供的最小边界框进行弱监督和无监督的部件检测, 然后使用基于部件检测的分割方法对图像中的物体进行分割。

Stanford Dogs 数据集具有两倍于 CUB 鸟类数据集的样本数量, 分别为 120 类狗, 收集了 20 580 个样本图像。尽管该数据集为每一类提供了更多的样本, 但是由其整合的标注信息要少于鸟类数据集, 仅仅包括有图像标签和最小边界框。更值得注意的是, 狗相比于鸟类有着与人类的更多交互, 比如前者可能会穿着衣服或者佩戴玩具装饰, 而这些显然会导致视觉分析中更多的遮挡和视觉混淆等问题。另外, 因为该数据集没有提供真实的图像分割掩膜, 就

使得在该数据集上只能通过分类性能来衡量本文方法的效果。

### 3.2 实验结果与分析

#### 3.2.1 部件检测

给定 CUB200-2011 数据集中真实的部件位置标注, 研究将其与本文的部件检测方法进行对比, 以验证本文方法的有效性。文中选择了在鸟类分类中比较重要的 3 个部件(背部、后颈部和腿部)进行实验验证, 实验结果见图 4。实验得到的检测误差是通过计算本文的检测结果和真实部件坐标的欧式距离得到的。样本的最小物体边界框的尺寸采用这些误差的标准化, 标准化的样本误差将会排序用于展示。

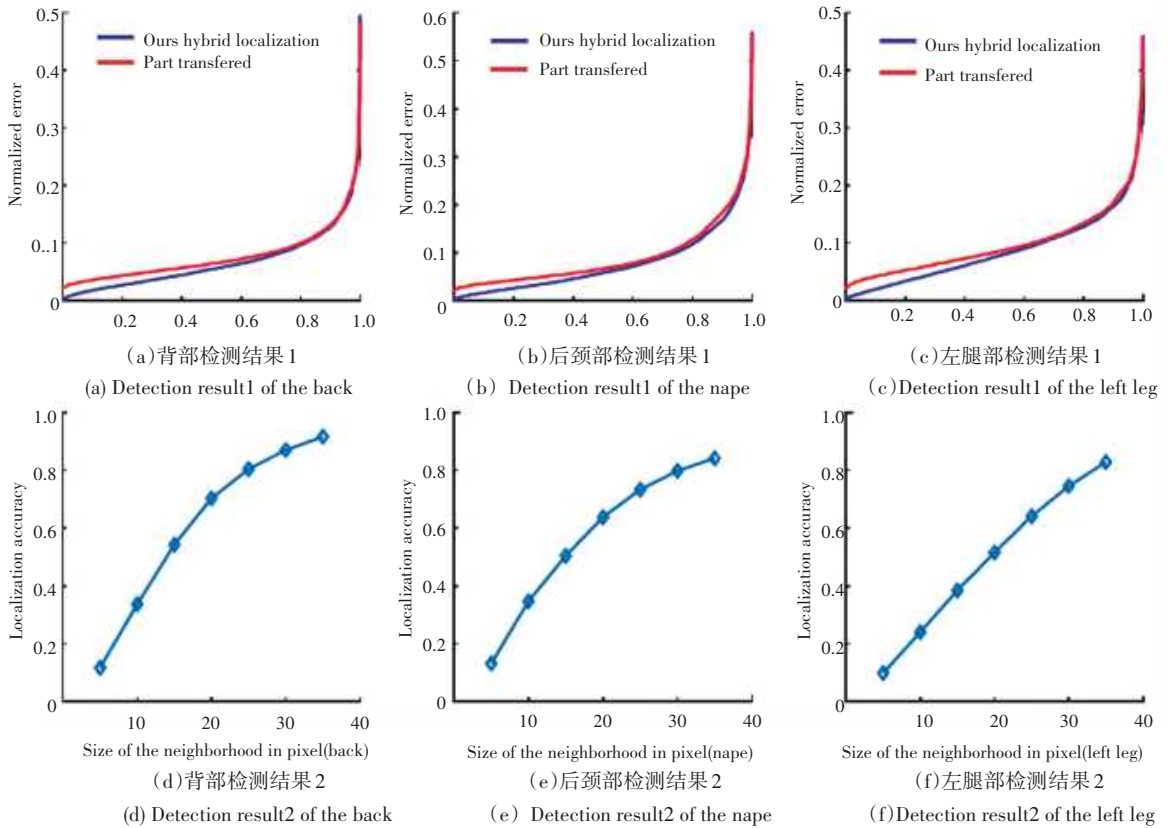


图 4 部件检测结果

Fig. 4 Results of part detection

如图 4(a) ~ (c) 图像所示, 研究中结合了非参数模型和参数模型的部件检测比仅仅使用了非参数传递的方法<sup>[4]</sup>具有更小的错误率。对比的方法还使用了真实的分割掩膜, 而本文方法却并未使用。不准确的部件检测结果会降低前景分割的性能, 但是如果研究检测到的部件处在真实部件的邻域内,

本文的算法仍有可能得到正确的分割结果。在这种情况下, 这个部件检测的结果被看成是正确的。文中的实验也展示了邻域大小对检测精度的影响, 如图 4(d) ~ (f) 所示。对此分析可知, 更大的邻域设置能够得到更高的检测准确率。在后续试验中则将选择 20 像素作为邻域的大小, 其对应的部件检测准确率为

60%,这是一个同时兼顾了准确率和计算开销的选择。

### 3.2.2 前景分割实验

研究使用文中的结合参数模型和非参数模型的部件定位方法估计样本图像中部件的位置。CUB 鸟类训练集中的真实部件信息被用于训练部件检测子,而对于狗类数据集,研究首先使用无监督的部件发现策略发现一些部件,然后再训练部件检测子。通过使用数据集推荐的训练-测试样本划分,而后对所有样本进行分割,在整个数据集上评价本文提

出的方法。试验中,每个图像样本都被缩放到最大边不大于 500 像素的尺寸。

研究将提出的基于部件检测的细粒度分割方法与基准方法 GrabCut<sup>[3]</sup>、转为细粒度图像的分割方法<sup>[2]</sup>、最先进的基于深度神经网络的一般前景分割方法<sup>[9]</sup>做了比较。基于深度网络的方法训练了一个结合图像类别标签和图像边界标签的网络,显著地提高了一般前景分割方法的性能。综上所述各种方法的分割结果即如图 5 所示。

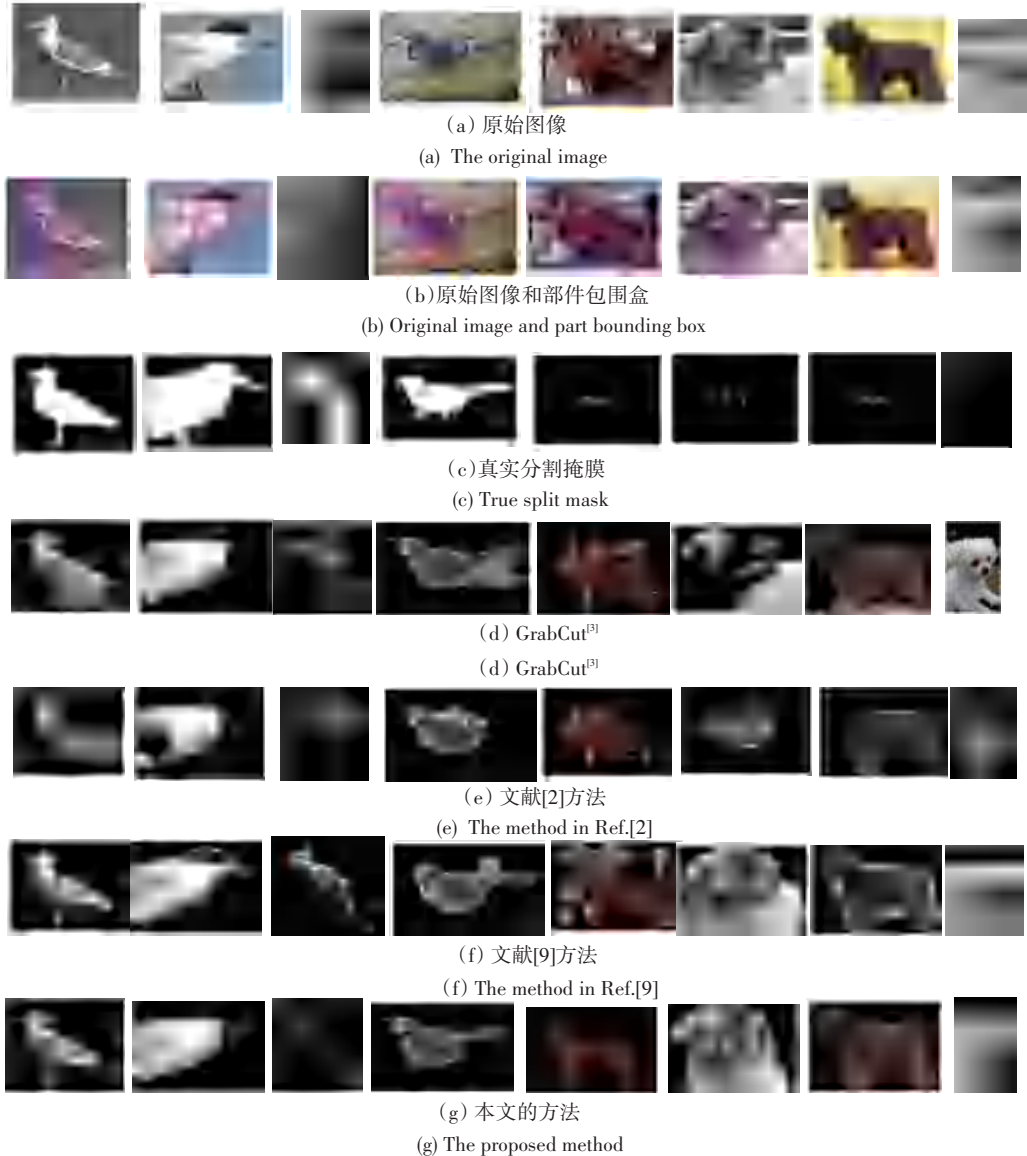


图 5 分割结果

Fig. 5 Results of segmentation

同时,图 5 还给出了文中部件检测的结果、及部分数据集附带的真实分割掩膜。由图 5 中可以看出,本文的方法能够保留一些具有语义的部件,但是

GrabCut 和 Chai 等人的方法<sup>[2]</sup>都倾向于丢弃这些重要部件。另外,本文方法通过使用部件检测结果初始化分割的颜色模型,在相当程度上减少了错误的

初始化带来的分割错误,而 Jain 等人的方法<sup>[9]</sup>的分割结果则包含了太多的背景噪声。

不仅如此,本文还研究了一些分割失败的例子,各实例运行结果如图 6 所示。其中有一些是由于错误的部件检测而引起的。可以看出,这些例子中杂乱的背景十分容易对前景的检测和分割造成干扰,而本文的对比方法在这种情况下也不能得到良好的结果。

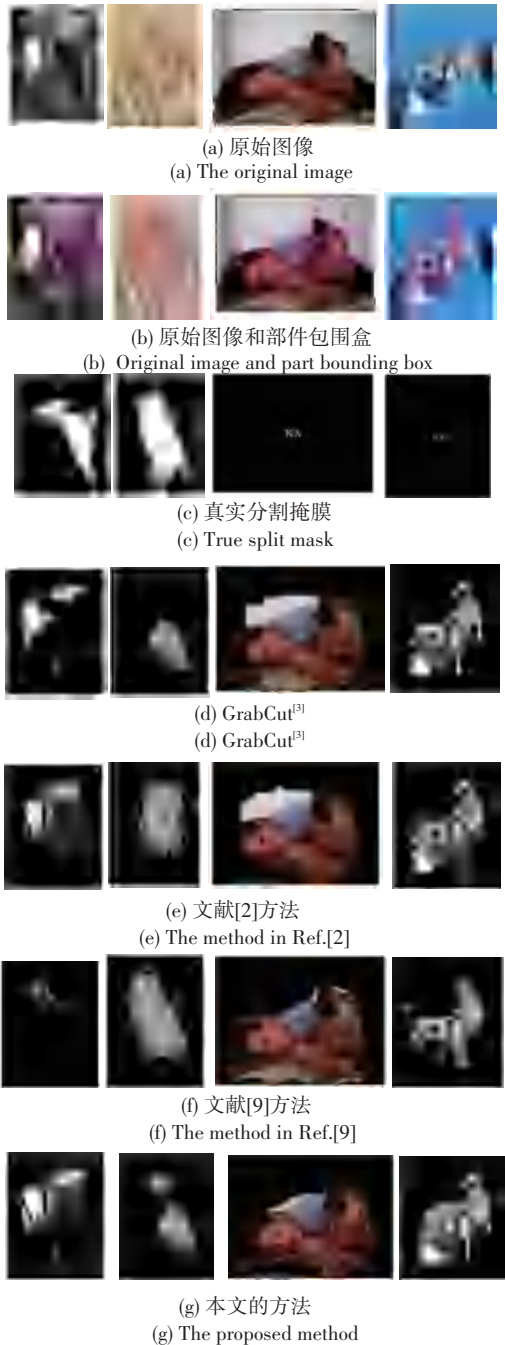


图 6 分割失败的例子

Fig. 6 Some failures of segmentation

为了显示本文分割算法的有效性,研究使用了 2 种衡量标准,分别是: PASCAL VOC<sup>[10]</sup> 重叠率 (Seg.I) 和被正确分割的前景像素和背景像素占总体像素的百分比 (Seg.II)。

本文分割方法在 CUB200-2011 数据集上的分割性能见表 1。由表 1 可知,本文方法对比 GrabCut 在 SegI 和 SegII 两项指标上分别提高了 11.1% 和 9.4%,对于 Chai 的方法分别提高了 1.9% 和 0.7%,对于 Jain 的方法分别提高了 5.2% 和 2.8%。值得注意的是,本文的方法取得了比 Jain 的基于深度神经网络的分割方法更好的前景分割结果,这也从侧面反映了一般的前景分割方法在细粒度图像分割任务中并不能获得良好效果。这里,计算 2 个分割指标的时候,是取最小物体边界框之内的图像进行计算的。以上实验基于鸟类数据集而得。对于 Stanford 狗类数据集而言,研究时未能提供真实的图像分割掩膜,因此并不能够定量地分析各个分割算法的性能。本文将在下一节中对比使用不同分割方法得到的前景图像进行图像细粒度分类的性能评估,以进一步衡量各方法分割的效果。

表 1 分割性能比较

方法	SegI	SegII
不使用分割	47.0	47.0
GrabCut <sup>[3]</sup>	64.5	74.2
文献[2]方法	73.7	82.9
文献[9]方法	70.4	80.8
<b>本文方法</b>	<b>75.6</b>	<b>83.6</b>

### 3.2.3 细粒度分类实验

从认知上来说,基于部件检测的分割方法能够更多地保留对于图像细粒度分类具有重要意义的部件,从而改进分类的效果。本文的目标是为图像的细粒度分类系统提供一个通用的分割方法,因此并没有选用那些带有特别设计的图像细粒度分类系统来验证文中算法的分割效果,这是因为要考虑到通用性方面的问题。文中专门选用了最为常用的基于视觉单词词袋的分类方法 BOW<sup>[11]</sup>。该方法对底层视觉特征进行编码得到直方图向量,对于复杂背景表现出较强的鲁棒性并且在没有任何结构先验的情况下能够取得较为理想的分类效果。而且,文中提取了图像的 SIFT、dense SIFT 和 color name 三种特征作为图像的描述子。其中,又分别为 2 种 SIFT 类的特征分别学习了 2 个由 8 000 个视觉单词组成的字典,为 color name 特征学习了由 800 个视觉单词

组成的字典。故而,研究最终得到的特征编码是一个 16 800 维的向量。

该数值就在 CUB200-2011 鸟类数据集上使用不同的分割方法获得图像进行图像细粒度分类的结果将如图 7 所示,该数值就是前  $r$  个排序结果中出现正确分类的准确率。首先,从所有  $r$  的取值对应的结果来看,前景分割明显有利于细粒度分类准确率的提高。究其原因即在于前景分割减少了背景噪声对特征提取和识别的干扰,故而使得所有的分割方法在准确率方面都超过了不使用分割而直接分类的方法。进一步分析可知,当  $r$  的值增大时,使用 GrabCut 的方法在准确率上已经渐渐地不再逊色于文献[2]的方法。这是因为尽管文献[2]的方法在减少背景噪声方面表现出色,但是也同时丢失了许多具有判别力的部件。如此推理也意味着随着  $r$  的增大,保留越多的部件,就可得到更高的分类准确率。文献[9]的方法表现比较稳定,但是在某些情况下却表现欠佳,因为该方法并没有如其它方法那样对细粒度分类对象的结构,比如部件进行探究。该方法常常在前景中包含了较多的背景噪声,一定程度上损害了分类的性能。

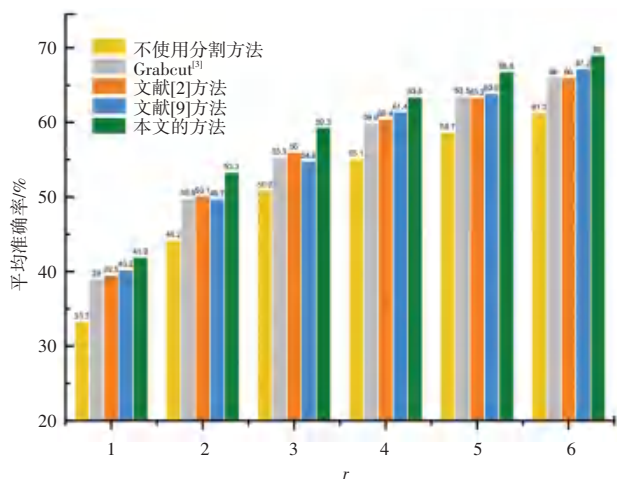


图 7 CUB200-2011 分类结果

Fig. 7 Classification results on CUB200-2011

此外,在 Stanford Dogs 狗类数据集上使用不同的分割方法获得图像进行图像细粒度分类的结果即如图 8 所示,相关设置和鸟类数据集一样。但是该数据集的样本却表现出更多的遮挡和混淆,因为狗和人类有更多的交互行为,这也在一定程度上对分类性能造成影响。由图 8 中可以看出,本文的方法在各种情况下的准确率都高于其它方法,因此就证明了本文研究引入部件检测以保留重要部件和优化

初始分割颜色模型的有效性。

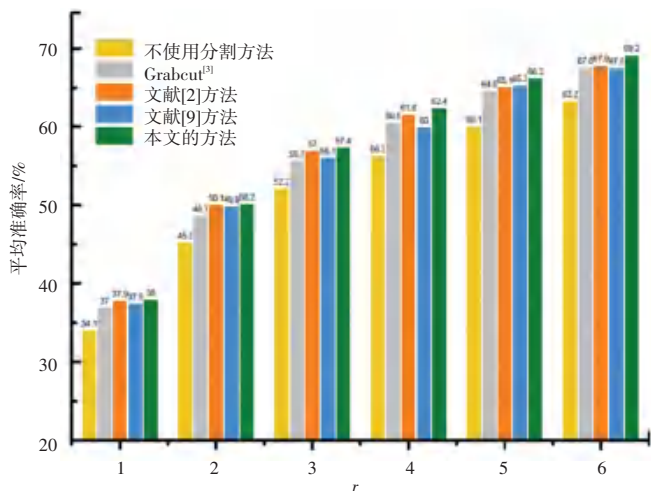


图 8 Stanford Dogs 分类结果

Fig. 8 Classification results on Stanford Dogs

## 4 结束语

本文提出了一种基于弱监督的部件检测的细粒度图像分割方法。该方法利用部件检测得到的部件假设来指导前景分割,从而保留更多的对细粒度分类有用的部件,提升细粒度分类的性能。研究设计了一种基于弱监督的部件检测方法,该方法使用数据集训练样本的部件标注训练了一个结合参数模型和非参数模型的混合型部件定位模型,而后便可以在任意的查询图像中自动检测定位预先定义的具有语义信息的部件。基于上述部件检测的分割算法则是最大程度地利用了部件假设来指导细粒度图像的前景分割,以求保留检测到的对细粒度分类具有重要作用的部件。本文的分割算法在 GrabCut 自动分割的基础上增加了部件假设能量项和一致性能量项,部件假设项及其相关的像素类别超参数能够不断修正部件假设,达到迭代优化的效果。实验证明了本文方法的有效性,在分割和分类性能上超过了基准方法 GrabCut、转为细粒度分类设计的分割方法以及先进的基于深度学习的一般图像分割方法。

## 参考文献

- [1] ANGELOVA A, ZHU S. Efficient object detection and segmentation for fine-grained recognition [C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR, USA:IEEE, 2013:811-818.
- [2] CHAI Y N, LEMPITSKY V, ZISSERMAN A. Symbiotic segmentation and part localization for fine-grained categorization [C]//2013 IEEE International Conference on Computer Vision (ICCV). Australia:IEEE, 2013: 321-328.