

基于部件模型的无人机系统小样本车辆目标识别方法*

牛轶峰, 朱宇亭, 李宏男, 王 菡, 吴立珍
(国防科技大学 智能科学学院, 湖南 长沙 410073)

摘要:对地目标检测与识别是无人机系统典型任务之一,但受限于任务特殊性,往往难以获取足够的目标样本数据以实现高可靠的目标识别。为此,结合人的认知特性,提出一种基于部件模型的小样本车辆目标识别方法,可有效提高无人感知能力。采用视觉显著性检测与物体性检测相结合的检测方法,提取目标可能区域;采用基于图论的 GrabCut 方法与最大类间方差法相结合的分割方法,分割目标并提取目标内部部件;采用基于概率图模型的部件识别方法,通过将部件轮廓稀疏表示为条件随机场,并进行概率推理实现部件识别;采用基于贝叶斯的目标识别方法完成目标是否为车辆的判断。通过无人机拍摄的车辆图像验证表明,算法可在样本较少、光照变化、存在遮挡等情况下,以较高准确率检测并识别出车辆目标,同时识别算法具有一定可解释性。

关键词:无人机;目标识别;图像分割;部件模型;小样本学习

中图分类号:TP391 **文献标志码:**A **文章编号:**1001-2486(2021)01-117-10

Small sample vehicle target recognition using component model for unmanned aerial vehicle

NIU Yifeng, ZHU Yuting, LI Hongnan, WANG Chang, WU Lizhen

(College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China)

Abstract: Detecting and recognizing targets on the ground is one of the typical tasks of UAVs (unmanned aerial vehicles), but it is limited by the task particularity so that it is often difficult to obtain sufficient data about target samples to achieve highly reliable target recognition. In view of this problem, a small-sample vehicle target recognition method based on the component model was proposed, which combined the cognitive characteristics of human beings to improve the perception ability of ground targets. The possible region of the target was extracted by visual saliency detection and objectness detection, and then the GrabCut segmentation method based on the Graph theory and the maximum between-class variance was used to segment the target and to extract the components from the target. A component recognition method based on a probability map model was used to perform component recognition by sparsely representing a component outline as a conditional random field and performing probabilistic reasoning. The Bayesian-based target recognition method was used to determine whether the target was a vehicle. Verification on real images captured by the UAV showed that the algorithm can detect and identify the vehicle target with high accuracy under the condition of fewer samples, poorer illumination and certain occlusion. At the same time, the recognition method can achieve the effect of certain interpretability.

Keywords: unmanned aerial vehicle; target recognition; image segmentation; component model; small sample learning

无人机对地面车辆等运动目标的侦察与监视作为无人机系统典型任务之一得到了广泛关注^[1],目前通常由任务控制站操作员进行人工目标判读与识别,需要应用自动目标识别算法,以提高目标识别精度,降低人的负担。尽管基于图像的车辆目标识别与跟踪方法已有大量研究,但是大多研究均是对水平方向或在斜上方固定视角(如道路监控)拍摄的车辆图像进行检测和识别,这些成像条件下的目标检测识别与固定翼无人机在高空快速飞行成像条件不太相同。无人机在高

空拍摄车辆目标时会存在不同视角(如不同角度下的俯视视角),而且无人机本身处于飞行运动中,因此获得的目标图像会受到光照、抖动、运动模糊等因素影响,使得传统基于灰度等特征的方法不够稳定,给检测识别过程带来困难^[2];此外,目前无人机俯视视角下的样本数据较少,在很多军事任务中难以获取大量目标样本用于训练。因此,研究如何在小样本下实现对目标检测识别是一个有重要意义的科学问题。

近年来,深度学习方法在目标检测领域表现

* 收稿日期:2019-06-10

基金项目:国家自然科学基金资助项目(61876187)

作者简介:牛轶峰(1979—),男,河南伊川人,教授,博士,E-mail:niuyifeng@nudt.edu.cn

远超传统检测方法,如 R-CNN^[3]、Faster R-CNN^[4]、YOLO^[5]、SSD^[6]、Cascaded R-CNN^[7]、RefineDet^[8]、SNIP^[9]等算法在检测准确率和速度上都远超一些传统检测方法,因此被广泛用于无人机目标感知任务中。然而,基于深度学习的方法需要大量样本数据训练,这在目标常见且样本充足情况下可行,但对于一些样本较少的特殊目标就难用这种方法取得好的检测效果,例如,实际应用中由于各种原因往往难以获取大量无人机俯视图下的军事目标数据,不宜进行大量标定,以致难以达到满意的训练及检测效果。此外,目标感知能力不但是无人机系统自主能力的重要组成部分,也是服务于无人机操作员进行情报处理和态势感知的重要手段,而深度学习方法随着网络层数的增加,对目标的特征描述往往缺乏可解释性,这导致当机器学习的结果发生错误时,人类很难找出错误的原因。

目前,循环神经网络(Recurrent Neural Network, RNN)型生成式组成模型递归网络(Recursive Cortical Network, RCN)——使用小样本学习的方法,在基于网络文本验证创建的全自动区分计算机和人类的图灵测试上获得突破性的成果^[10]。RCN 在多个 CAPTCHA 数据库中获得

了极佳的成绩(reCAPTCHA 准确率:66.6%),整体上以 300 倍的数据有效性击败了深度学习的卷积网络模型。RCN 为解决小样本条件下实现对目标的检测识别提供了新的解决思路。

人类具有在少量的样本中学习出抽象的概念并将这些概念应用到多种用途的能力,在应对物体形变、材质变化等方面具有很好的迁移能力。同时人具有从部件层次识别物体的能力,能够根据识别到的部件信息推断出目标的类别。比如在遮挡情况下,看到一个轮子和一块车窗,能够推断出这是一辆车。考虑到人的认知机制,本文基于部件模型的思想,将目标进行部件分解,通过识别各个具有语义信息的部件并进行贝叶斯推理判断目标是否为车辆。本文以复杂背景下地面车辆目标为例,研究如何在小样本下对目标进行具有可解释性的检测识别,总体框架如图 1 所示。

本文方法主要分为四个部分,分别为显著性区域检测、目标部件分割、目标轮廓识别和车辆识别。首先结合人的视觉注意机制,利用时域与频域相结合的方式检测出疑似车辆目标区域,之后基于部件模型和概率推断的方式对目标部件进行分割和轮廓识别,推理得出目标是否为车辆,并在不同条件下进行仿真验证。

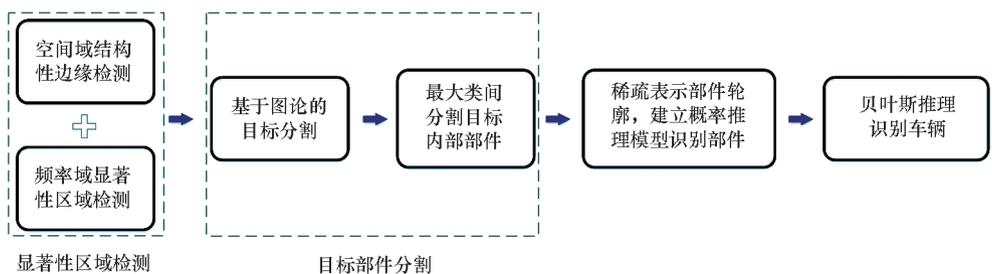


图 1 基于部件模型的车辆识别方法

Fig. 1 Vehicle identification method based on component model

1 结合人视觉注意机制的频域时域相结合的 车辆检测方法

人在一张图像中寻找目标时,不会像传统滑动窗法那样从左上角开始遍历全图进行查找,而是会一眼“纵观全局”找到可能是目标的显著性区域再进一步确认。这种机制被称为视觉注意机制,它会显著提高人对视觉信息的处理速度。在此基础上,通过结合目标自身先验信息,从显著性区域中滤除一些不具有目标基本特征的结果进而提高后续识别的准确率,在一定程度上可以弥补样本不足带来的影响。本文在识别目标前首先进行可能是目标的显著性区域检测,并利用车辆目标具有的结构性边缘特征进行约束。

显著性区域的检测方法有多种^[11],如基于空间域的 Itti 模型^[12]、AC 算法^[13]、HC 算法^[14]、EdgeBox 算法^[15]、BING 算法^[16]、Selective Search 算法^[17]等;基于频域的谱残差(Spectral Residual, SR)算法^[18]、FT 算法^[19]等。基于频域的方法通过设计高通滤波器,可以去除低频背景噪声,但往往会丢掉部分目标本身的结构性信息(例如对于矩形的车辆,频域方法检测的结果通常只会突出矩形的四个角点,而弱化或丢掉了边缘信息,见图 2(c)、(d)),而单独使用空间域的检测方法又容易受低频背景等噪声的影响(突出了物体的结构性边缘信息,但难以去除阴影、地面等信息,见图 2(b))。因此,本文将频域方法(SR 算法)与空间域方法(EdgeBox 算法)结合,从而检测出疑似车

辆目标区域。

SR 算法基本流程如算法 1 所示,基本思想是首先将图像变换到频率域,然后对低频成分进行抑制,突出目标边缘轮廓等高频信息。EdgeBox 算法首先在图像中进行结构性边缘检测,然后根据检测结果给所有候选区域打分,根据得分高低输出具有结构性的显著性区域。

算法 1 SR 算法

Alg. 1 SR algorithm

Step 1: 将图像 $f(x,y)$ 通过 FFT 变换到频域空间

Step 2: 计算频率域中的幅值谱、相位谱和对数幅值谱

$$A(u,v) = |F(f)|$$

$$P(u,v) = \text{angle}(F(f))$$

$$L(u,v) = \log(A(u,v))$$

Step 3: 计算谱残差(为 $n \times n$ 的局部均值滤波器)

$$R(u,v) = L(u,v) - h_n \cdot L(u,v)$$

Step 4: 得到显著性图

$$S(x,y) = |F^{-1}[\exp(R(u,v) + i \cdot P(u,v))]|^2$$

由于检测结果往往有较多大小各异的候选框,为进一步处理往往需要进行不确定目标个数下的目标框自动聚类。考虑到 EdgeBox 算法在同一个目标上的检测框往往相邻或者重叠,而属于

不同目标的框往往相隔较远,因此本文设计了一种迭代式的目标框自动聚类算法,如算法 2 所示,在每次迭代中用最小外接矩形取代有重叠关系的候选框,直到目标框彼此之间没有重叠,实现自动聚类。设计的 SR - EdgeBox 算法流程如图 2 所示。利用 EdgeBox 算法在空间域提取边缘,之后基于 SR 算法在频率域提取高频信息,去除低频信息,得到目标区域。

算法 2 迭代式自动聚类算法

Alg. 2 Iterative automatic clustering algorithm

Step 1: 输入 n 个候选框 box

Step 2: 将所有 box 置于队列中,选取第一个候选框 box_1 为基准,依次计算与其他 $n-1$ 个 box 是否有重叠,若有重叠则计算包围二者的最小矩形 box_{\min} ,取代 box_1 ,同时删除当前 box,然后继续依次计算 box_1 与其他 box 是否有重叠

Step 3: 当 box_1 与其他所有 box 都计算完是否重叠以及求出对应的最小包围矩形后,将 box_1 添加到最后一个 box,并删除原来的 box_1 ,使 box_2 变为 box_1 , box_3 变为 box_2 ,依此类推。然后继续执行 Step 2,该过程重复 n 次

Step 4: 输出最终 m 个可能的目标框 ($m \leq n$)

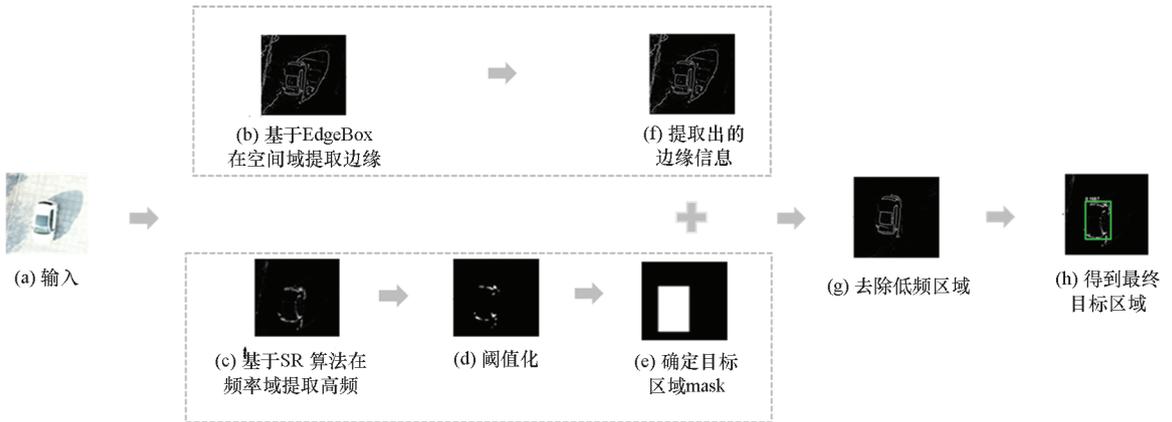


图 2 SR - EdgeBox 算法流程

Fig. 2 SR - EdgeBox algorithm flow

为了比较各类算法的效果,采用不同的算法对包含车辆的图片进行显著性检测实验,如图 3(b) ~ (d) 所示。为了验证本文算法的有效性,在存在遮挡的情况和多辆车同时存在的情况下进行了测试。由图 3 可以看出,频域 SR 显著性检测方法可以将高频信息突出,然而同时会保留树叶等非目标区域。通过在空间域用结构性边缘进行约束,可以滤除非结构性区域,得到一系列的候选框,选出最可能是结构性区域的 15 个

box,随机用不同颜色表示,如图 3(c) 所示,并且最终检测出车辆目标,如图 3(d) 所示。该方法对于遮挡和多目标问题也有较好的检测效果,如图 3(e) 和如图 3(f) 所示。

2 基于部件模型和概率推断的车辆识别方法

2.1 部件的定义

为便于对算法及结果描述,将部件定义为:在颜色维度上区别于相邻区域的封闭、单一的颜色

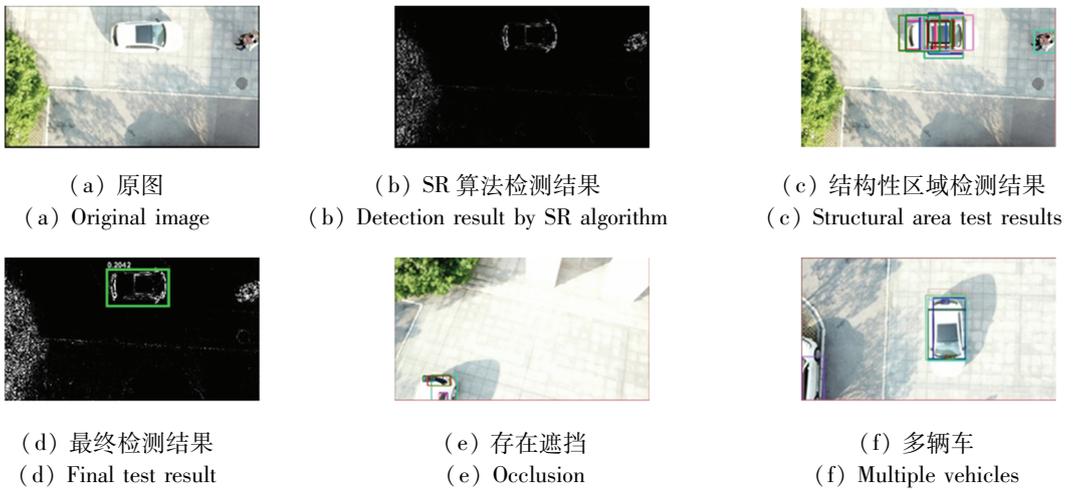


图 3 检测结果

Fig. 3 Test results

块。例如在白色车辆中,黑色的车窗、黑色的车玻璃即为部件;车轮作为独立的封闭黑色块,与地面、车身颜色均不同,也可作为部件。

2.2 目标部件分割

为进一步识别目标是否为车辆,需要提取特征进行匹配。考虑到轮廓形状相比于传统角点、颜色等特征有更好的稳定性,因此本文提取目标的部件轮廓作为特征进行识别,通过结合 GrabCut 图割算法^[20]与最大类间分割算法^[21]实现目标部件分割。本文将车辆部件分为前窗、侧窗、后窗和车轮四种。

目前,常用的图像分割方法主要有阈值分割、聚类分割、边缘检测分割、图分割以及基于人工神经网络的分割方法等。其中,阈值分割容易受到噪声和光照的影响;聚类分割容易过度分割;边缘检测分割方法的抗噪性和检测精度难以同时满足;基于人工神经网络的分割方法需要大量数据以及选择合适的网络结构;而图分割方法可以避免上述不足,故本文引入图分割算法。实验发现,交互式迭代图分割 GrabCut 算法可以仅给定一个表示目标区域的框就能较好地去除背景,分割出目标主体。GrabCut 是一种交互式图像前景、背景分割算法。该算法无须训练,给定目标框后,算法自动将框外的像素视为后景,将框内像素视为可能的前景和后景,然后对前景和后景数据分别建立高斯混合模型,利用目标各部件内像素的相似性特征,通过最小化能量函数实现对未标记的像素的推断。能量函数定义如式(1)所示。

$$E(\alpha, \theta, z) = U(\alpha, \theta, z) + V(\alpha, z) \quad (1)$$

其中, z 表示图像的像素值, α 表示图像每个像素

分割的标签, θ 表示分割模型的参数, $U(\alpha, \theta, z)$ 为表征目标图像主要内容的数据项, $V(\alpha, z)$ 为反映目标内部像素间相似性的平滑项。最小化求解的过程通过最小割方法实现。

GrabCut 算法优点在于可以较好地利用相邻像素之间的连续性进行分割,然而不足之处在于:首先,它需要用户交互实现连续像素标记;其次,因连续性约束,算法往往可以分割出目标整体,却难以进一步分割出内部部件。阈值分割的方法通常不受连续性约束,但对光照等噪声敏感。为此,本文结合两种方法进行部件分割。

首先对应用视觉显著性算法分割出的疑似目标区域(见图 3(d))的目标整体利用最大类间方差法求取阈值进行二值分割(本文假定车辆车身与部件可按颜色分为两类),在二值化图像中的相同区域内进行给定长度下的划线操作,用自动化的方式取代用户交互过程,并将得到的线连同其标签(前景或后景)输入 GrabCut 算法继续迭代,直至分割出目标部件。其中,最大类间方差法求取分割阈值的方法如算法 3 所示。

算法 3 最大类间方差法求阈值

Alg. 3 Method of maximum interclass variance for threshold

Step 1: 分别计算目标与背景像素所占比例 ω_1 以及平均灰度 μ

$$\omega_1 = N_1 / (M \times N), \omega_0 = N_0 / (M \times N)$$

$$\mu = \omega_1 \times \mu_1 + \omega_0 \times \mu_0$$

Step 2: 记两类的方差为 g

$$g = \omega_1 \times (\mu_1 - \mu) + \omega_0 \times (\mu_0 - \mu)$$

Step 3: 寻找分割阈值 T

$$T_{thr} = \operatorname{argmax}_T g$$

为了验证本文提出的分割算法效果,与其他

几种分割算法 (GrabCut 算法、Ostu 算法以及 Selective Search(图中简称为 SS)中的分割方法)进行了对比。考虑篇幅原因,测试图选取了具有代表性的三张图,图 4、图 5、图 6 分别表示目标车辆存在少量遮挡、大量遮挡和不存在遮挡的分割

结果。通过实验结果对比可以看出,本文通过随机划线算法将 GrabCut 和 Ostu 算法相结合,在分割目标时可以较理想地将物体主要部件分割出来,而其他算法往往难以将部件很好地分割出来,或者存在过度分割的情况。

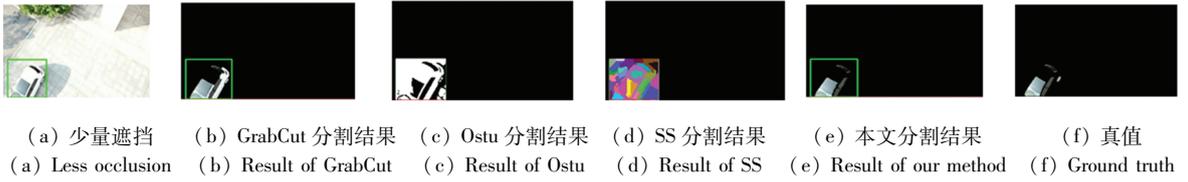


图 4 少量遮挡车辆分割结果对比

Fig. 4 Comparison of vehicle segmentation results under less occlusion

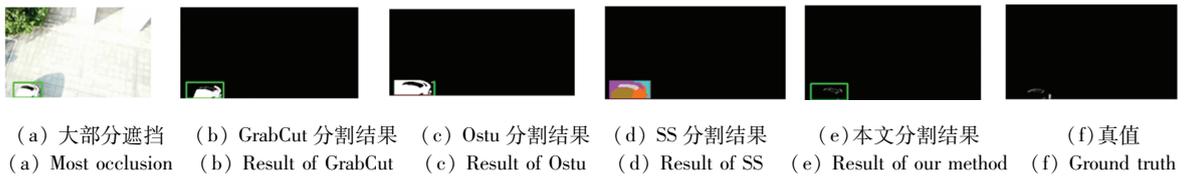


图 5 大部分遮挡车辆分割结果对比

Fig. 5 Comparison of vehicle segmentation results under most occlusion

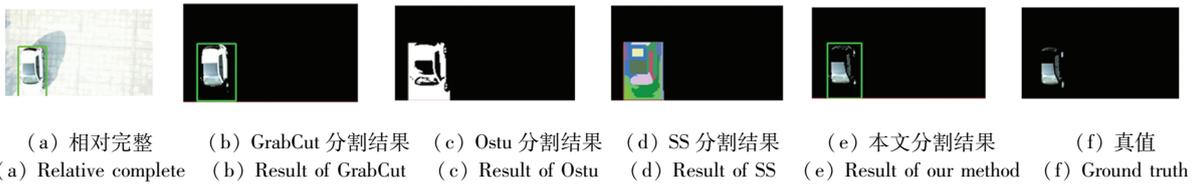


图 6 相对完整车辆分割结果对比

Fig. 6 Comparison of vehicle segmentation results without occlusion

2.3 目标轮廓识别

分割出部件后需要根据轮廓形状进行部件识别。为解决形变影响轮廓识别率的问题,本文对训练部件轮廓建立图模型,通过对轮廓进行特征稀疏表示并利用图模型的方法推断识别目标部件轮廓。由于概率图模型在推理时对局部的像素偏移有较好的容纳性,即局部的像素偏移不会对整体轮廓相似度计算造成很大影响,并且每一个样本都可以训练得到一个图模型,因此,本文方法可以在小样本下更好地适应部件轮廓存在局部形变的识别问题。

本方法建立两层网络表征目标轮廓,如图 7 所示,其中包含特征层(记为 F)和池化层(记为 H),分别表示轮廓上各点的方向和轮廓可能产生的不同形变及位移。通过提取轮廓上各点的方向作为特征,进行稀疏表示,实现用表示方向的特征节点对目标轮廓稀疏离散表示。同时,利用轮廓的连续性建立节点之间的约束关系,从而建立表征目标轮廓的图模型。在图模型上利用消息传递

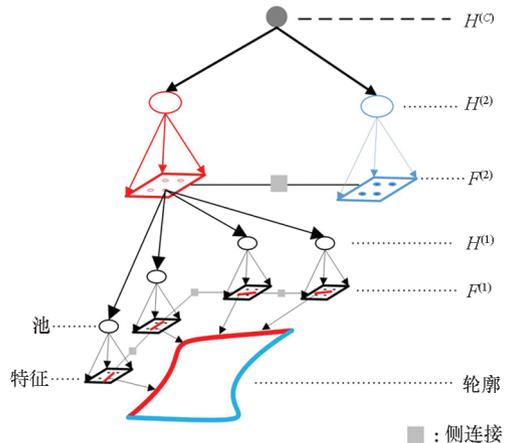


图 7 表征目标轮廓的网络

Fig. 7 Network characterizing the target contour

的概率求取方法计算待识别图像与已知训练样本图模型的匹配程度,作为目标轮廓的识别相似度。在一个共有 C 层的模型中,每一层的变量都只依赖上一层,故整个结构可用条件随机场描述^[22],用联合概率分布的对数形式可以表示为如式(2)所示。

$$\begin{aligned} & \log p(F^{(1)}, H^{(1)}, \dots, F^{(C)}, H^{(C)}) \\ &= \log p(F^{(1)} | H^{(1)}) + \log p(H^{(1)} | F^{(2)}) + \dots + \\ & \quad \log p(F^{(C)} | H^{(C)}) + \log p(H^{(C)}) \end{aligned} \quad (2)$$

其中,特征层中的每一个变量值表示该特征存在与否,与当前层其他特征变量相互独立(将轮廓稀疏表示后暂不考虑相邻特征的依赖关系),因此对于特征层,有

$$\log p(F^{(l)} | H^{(l)}) = \sum_{f'r'c'} \log p(F_{f'r'c'}^{(l)} | H^{(l)}) \quad (3)$$

其中, f', r', c' 分别表示 F 层特征、行坐标和列坐标。池化层中的每一个变量表示上一层对应特征可能的取值情况,变量的取值依赖于上一层父特征节点变量的取值,在不考虑同一层变量之间的相互约束(图 7 示意的侧连接)时,有

$$\begin{aligned} & \log p(H^{(l)} | F^{(l+1)}) = \sum_{frc} \log p(H_{frc}^{(l)} | F^{(l+1)}) \\ &= \sum_{frc} \log p(H_{frc}^{(l)} | F_{f'r'c'}^{(l+1)}) \end{aligned} \quad (4)$$

其中, f, r, c 分别表示 H 层特征、行坐标和列坐标。

然而,同一个特征所具有的不同形变由于依赖于上一层对应的父特征节点变量的取值,因此不同位移、形变之间实际上存在相互约束,考虑到这一点,池化层可以建模为:

$$\log p(H^{(l)} | F^{(l+1)}) = \sum_{f'r'c'} \log p(H_{frc}^{(l)} | F_{f'r'c'}^{(l+1)}) \quad (5)$$

式(5)表示每一个池化层的节点受它对应特征节点取值的影响。

考虑到部件轮廓结构简单,故建立包含两个特征层和两个池化层的网络进行描述,如图 7 所示($C=2$)。其中第一个特征层表示轮廓上不同方向的边缘段,第一个池化层表示每个边缘可能存在的不同位移变化和方向变化;而第二个特征层表示由边缘构成的轮廓,第二个池化层表示轮廓可能存在的不同状态(形变)。

在进行部件识别时,需要先对样本进行学习以获得图结构中的特征节点、池化节点以及节点之间的约束关系,在此基础上再对待识别的目标部件进行推理识别^[2]。

为了将目标的轮廓表示为特征节点,需要提取出轮廓上各像素点的方向。本文采用 16 个方向不同的正负高斯滤波器对目标图像进行卷积,从而得到 16 个方向上的响应(如图 8 所示,图 8(c)和图 8(d)为其中两个方向上的响应),之后通过非极大值抑制将响应最大的方向作为该像素点的方向。

为了将边缘离散化以便于计算推理,本文使用稀疏表示。稀疏表示采用贪婪稀疏算法最小化代价函数:

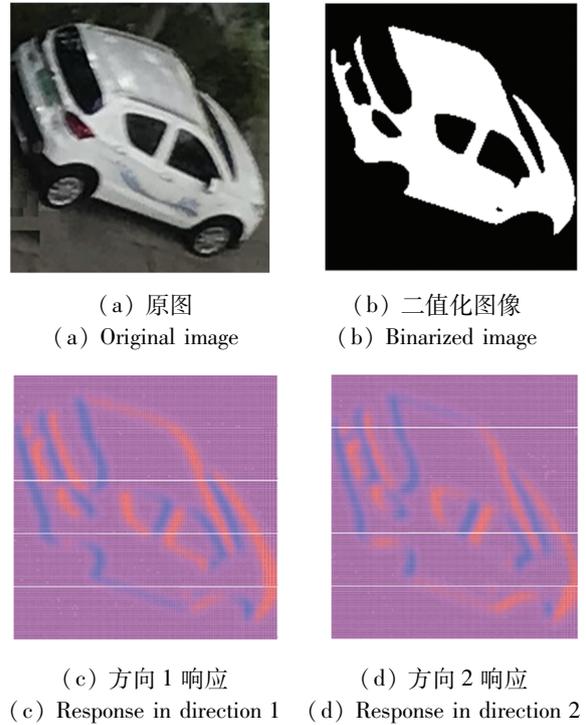


图 8 提取边缘方向特征

Fig. 8 Extract edge orientation features

$$J_{\text{sparse}}(S; X) = k|S| + \log p(X|S) \quad (6)$$

其中, X 为目标图像, $S = \{(f_1, r_1, c_1), \dots, (f_n, r_n, c_n)\}$ 为各层不同位置上的特征集合, k 为参数。贪婪稀疏算法的描述如算法 4 所示。其中收敛条件为整个轮廓都完成用特征节点表示或者未表示部分小于既定阈值。详细的图模型稀疏表示收敛性判定可参考文献[19]。

算法 4 贪婪稀疏算法

Alg. 4 Greedy sparse algorithm

```

Step 1: procedure SPARSIFY ( $m_{\text{bu}}(F^{(1)}), \text{Model}$ )
Step 2: Sparsefication  $\leftarrow \emptyset$ 
Step 3: Unexplained  $\leftarrow m_{\text{bu}}(F^{(1)})$ 
Step 4: while True do
Step 5: Messages =
          TRUNCATEDFORWARDPASS(Unexplained)
Step 6: MostActiveFeature = arg max  $m_{\text{bu}}(F^{(C-1)})$ 
Step 7: if  $m_{\text{bu}}(F^{(C-1)})(\text{MostActiveFeature}) <$ 
          Threshold then
Step 8: return Sparsefication
Step 9: end if
Step 10: Sparsefication  $\leftarrow$  Sparsefication  $\cup$ 
         {MostActiveFeature}
Step 11: ExplainedLocs  $\leftarrow$  BACKWARDPASS
         (Messages, MostActiveFeature)
Step 12: Unexplained  $\leftarrow$  SUPPRESSAROUND
         (Unexplained, ExplainedLocs)
Step 13: end while
Step 14: end procedure

```

通过将一定长度的边缘用一个特定方向的特征节点表示,目标轮廓可以用较少的特征节点表示,如图9所示,其中第二列表示每个节点边缘方向特征,第三列和第四列分别表示该特征节点在图像中的横、纵坐标。

0	12	50	112
1	9	52	108
2	13	53	116
3	8	56	109
4	14	57	118
5	12	59	97
6	8	60	111
7	14	61	120
8	12	62	93
9	15	62	101
10	8	64	111

图9 轮廓稀疏化表示

Fig. 9 Outline sparse representation

边缘连续性约束学习需要计算相邻特征节点之间的欧式距离,并将其作为节点之间的约束权重添加到模型中。最终,每个训练样本会建立起表征目标轮廓的具有权重的图模型,如图10所示。

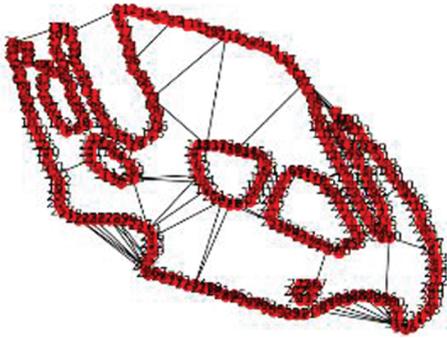


图10 轮廓图模型

Fig. 10 Contour model

2.4 目标推理

对样本学习后便可以在得到的图模型上推理出待识别目标与训练样本的匹配程度。推理过程采用最大积信念传播方法,利用消息传递机制在图模型上计算观测到待识别图像条件下的后验概率,进而得到目标轮廓与训练样本的相似度。在本文网络结构中消息定义为

$$m_{bu}(F_{f'c'}^{(l+1)} = 1) = \max_{\{s_j \in \text{const}(F_{f'c'}^{(l+1)})\}} \sum_j m_{bu}(H_j^{(l)} = s_j) \quad (7)$$

其中, m_{bu} 表示自下而上传递的消息, s_j 为每个特征节点可取的状态(不同位移及形变)。

最大积信念传播方法是一种可以在树结构图模型和带环的图模型中都取得较好效果的推理方法。方法将学习到的图模型表示为概率因子图的

形式,为

$$p(x) = p(x_1, x_2, \dots, x_n) = \prod_c \phi_c(x_c) \quad (8)$$

其中: c 表示图模型中的因子,包含了 x_1, x_2 等相互连接的变量; $\phi_c(x_c)$ 是对应的势能函数,恒正。

信念传播的目的是找到最大后验概率下的所有变量取值,即找到 $x^* = \arg \max_x p(x_1, x_2, \dots, x_n)$ 。方法是通过定义消息的概念然后用消息传递机制进行求解。定义 $\hat{m}_{c \rightarrow i}(x_i)$ 表示因子 c 中除去变量 x_i 外其他变量传递给 x_i 的消息。消息更新方式为

$$\hat{m}_{c \rightarrow i}(x_i)^{\text{new}} = \max_{x_{c \setminus i}} \{ \log \phi_c(x_i, x_{c \setminus i}) + \sum_{j \in c \setminus i} [\mu(x_j) - \hat{m}_{c \rightarrow j}(x_j)] \} \quad (9)$$

其中: $x_{c \setminus i}$ 表示因子 c 中去掉了变量 x_i 的其他变量; $\mu(x_i)$ 表示变量 x_i 的最大边缘估计,定义为

$$\mu(x_i) = \sum_{c: i \in c} \hat{m}_{c \rightarrow i}(x_i) \quad (10)$$

消息更新到变量取值收敛时或迭代到一定次数后,通过计算特征节点网络的联合概率分布可得到待识别目标轮廓与训练样本的相似度,作为识别结果。

2.5 车辆目标识别

为了能进一步识别目标并能适应遮挡问题,本文在部件识别的基础上利用贝叶斯推理进行车辆识别。考虑到无人机观测目标可能存在遮挡等情形,同时为简化问题,假设各个部件观测相互独立,采用朴素贝叶斯推理。

定义 $y \in \{0, 1\}$ 表示目标框内是否为汽车, $y = 1$ 表示是, $y = 0$ 表示不是。将待识别的目标用属性 x 表示为 $x = \{wheel, window_{side}, window_{top}, window_{back}\}$, 其中属性分别表示车轮、侧车窗、天窗、后窗。由于观测视角原因,表示车轮的属性值取值为 $\{0, 1, 2\}$, 其余属性取值为 $\{0, 1\}$, 表示是否检测到。根据朴素贝叶斯分类,是否为车的概率为

$$\begin{aligned} p(y|x) &= \frac{p(x|y)p(y)}{p(x)} \propto p(x|y)p(y) \\ &= p(wheel|y)p(window_{side}|y) \cdot \\ &\quad p(window_{top}|y)p(window_{back}|y)p(y) \end{aligned} \quad (11)$$

其中, $p(y)$ 表示目标框内是否为车的先验概率,这由具体场景和检测算法的准确率决定。

通过比较 $p(y = 1|x)$ 与 $p(y = 0|x)$ 的大小推理目标是否为车辆,若前者大,则认为是车辆,否则认为不是车辆。其中,检测结果是车辆的概率为

$$\begin{aligned}
 p(y=1|x) &= \frac{p(x|y=1)p(y=1)}{p(x)} \propto p(x|y=1)p(y=1) \\
 &= p(\text{wheel}|y=1)p(\text{window}_{\text{side}}|y=1) \cdot \\
 &\quad p(\text{window}_{\text{top}}|y=1)p(\text{window}_{\text{back}}|y=1) \cdot \\
 &\quad p(y=1) \tag{12}
 \end{aligned}$$

同理,可求得不是车辆的概率。

3 实验结果

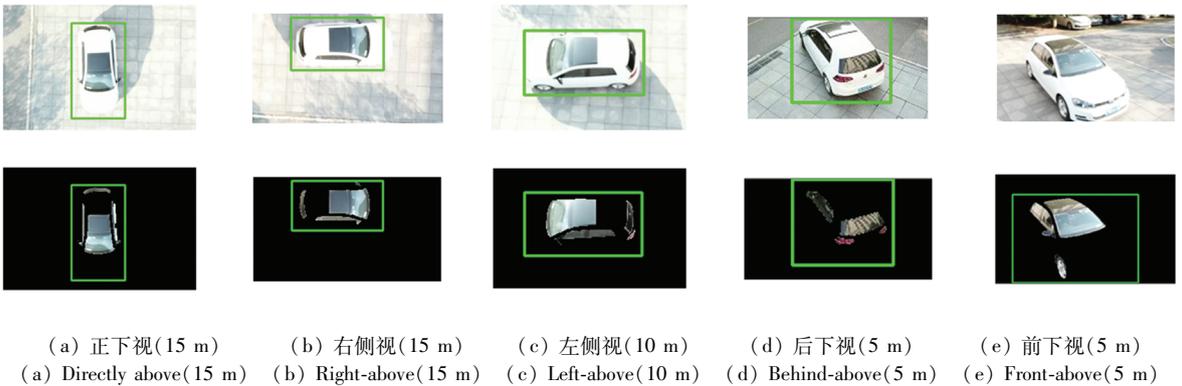
为了验证本文提出的小样本条件下车辆识别算法,采用不同条件下的车辆照片进行实验,分别验证该识别算法在视角变化、存在遮挡以及目标车辆发生变化时的效果,并且与当前较为流行的支持向量机(Support Vector Machine, SVM)算法的识别效果进行对比分析。

3.1 不同视角下车辆目标识别

为了验证识别算法在不同视角不同高度下的

识别效果,本文选用无人机在不同视角下拍摄的含有车辆的图像进行测试,其中概率推理所需的先验概率均为该场景的统计值。

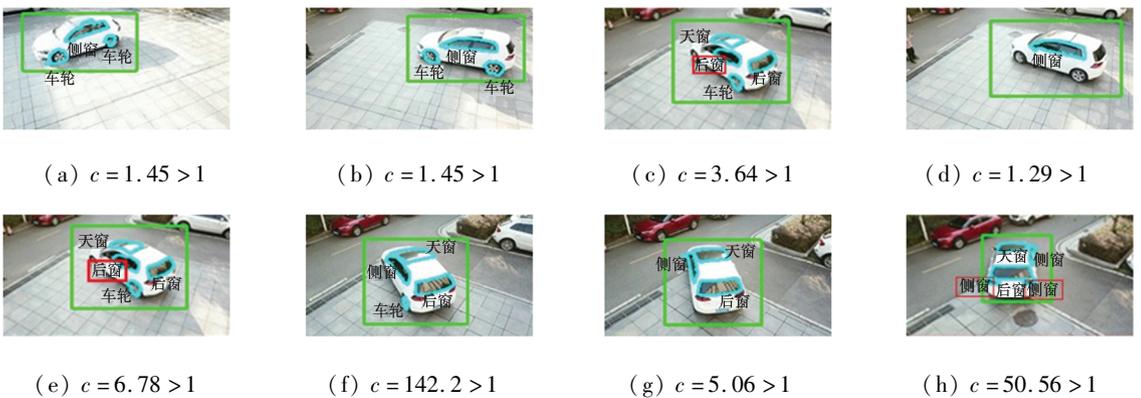
检测结果如图 11、图 12、表 1 所示。其中图 11 表示不同视角下的检测和分割结果,图 12 表示车辆部件检测的结果以及利用贝叶斯推理得出的车辆识别结果,当 $c > 1$ 时算法检测目标是否为车辆的结果为“是”。表 1 表示相机在不同视角及高度下的识别正确率。由结果可以看出,算法在不同视角、不同高度下都有较高的识别准确率。其中,侧视视角下的识别更高一些,这是因为该视角可以较多地观测并识别出车辆部件;相比之下,前下视和后下视视角下识别率较低,这是由观测到部件较少且形变较大导致。



(a) 正下视(15 m) (b) 右侧视(15 m) (c) 左侧视(10 m) (d) 后下视(5 m) (e) 前下视(5 m)
(a) Directly above(15 m) (b) Right-above(15 m) (c) Left-above(10 m) (d) Behind-above(5 m) (e) Front-above(5 m)

图 11 不同视角及检测、分割结果

Fig. 11 Detection and segmentation results from different perspectives



(a) $c = 1.45 > 1$ (b) $c = 1.45 > 1$ (c) $c = 3.64 > 1$ (d) $c = 1.29 > 1$
(e) $c = 6.78 > 1$ (f) $c = 142.2 > 1$ (g) $c = 5.06 > 1$ (h) $c = 50.56 > 1$

图 12 车辆检测与识别结果

Fig. 12 Vehicle detection and recognition results

表 1 不同视角及不同高度下的识别率

Tab.1 Recognition rate at different viewing angles and heights

拍摄高度/m	不同拍摄角度下的识别率			
	正下视	左(右)侧视	前下视	后下视
5	0.81	0.93	0.78	0.68
10	0.85	0.94	0.77	0.72
15	0.83	0.88	0.74	0.70

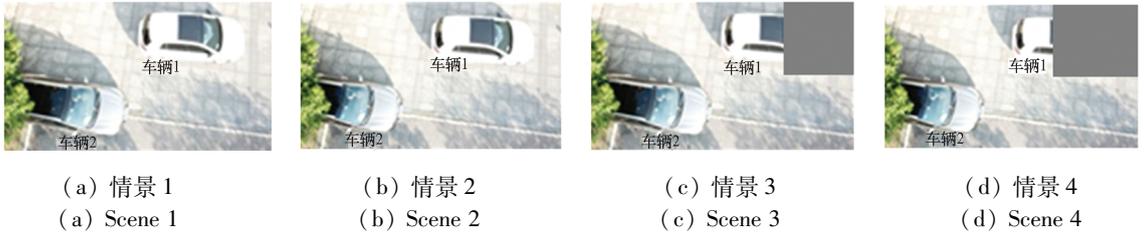


图 13 不同车辆以及不同遮挡率

Fig.13 Different vehicles and different occlusion rates

表 2 不同遮挡率下的识别率比较

Tab.2 Comparison of recognition rates under different occlusion rates

测试数据集	SVM 识别率	本文算法识别率
无遮挡	0.45	0.62
车辆 1	遮挡率 50%	0.56
	遮挡率 90%	0.23
无遮挡	0.37	0.64
车辆 2	遮挡率 50%	0.46
	遮挡率 90%	0.19

从实验结果可以看出,SVM 分类方法在本实验中的识别率较低,这是因为 SVM 属于统计学习的算法,在对特征描述较多的目标分类时需要大量的训练数据,而实验中仅用了 10 个样本进行训练,因此效果不好。而在相同样本数量的前提下,本文算法的识别效率更好,特别是遮挡较少时识别率较高,这是因为本文方法不依赖于大量数据训练(使用到的部分先验概率在实际中可根据经验、统计知识迁移得到),同时使用更稳定的轮廓边缘作为目标的表示,因此识别算法更具有适应性。随着遮挡率的增加,识别率逐渐下降,这是因为检测到的部件在逐渐减少,在遮挡率为 50% 左右时,还可以检测出一个或两个部件,但遮挡率再增加时,往往难以再识别出部件。由于 SVM 算法和本文识别算法都属于基于统计学习的方法,在

3.2 不同遮挡条件下存在相似车辆的目标识别

为验证本文方法在遮挡条件下对车辆的识别效果以及在样本数量较少时的识别效果,本文与 SVM 识别方法进行对比。实验选取的测试图像包含两辆相似车辆,且具有不同程度的遮挡,包括自然条件下的树木遮挡以及人为在目标上添加的遮挡,如图 13 所示。实验中,训练样本个数为 10,测试样本数为 100,统计算法正确将车辆目标识别出的比例。实验结果如表 2 所示。

训练和识别过程中只利用小样本数据集,没有参考其他大样本数据集,所以识别率普遍不高,后续可以再进行改进。

4 结论

本文提出了一种基于部件模型的小样本车辆识别方法,该方法对车辆目标具有很好的识别效果,并且具有一定的可解释性。首先使用图像分割的方法将检测区域中的目标进行部件分割。其次对分割的部件进行识别。考虑到部件形变情况,提出利用 RCN 概率图模型方法识别部件,通过对部件轮廓用特征节点稀疏表示,并考虑轮廓特征节点的连续性约束,实现了对形变轮廓的较高准确率的识别。最后结合先验概率知识,应用贝叶斯理论根据识别出的部件轮廓推理识别目标车辆。本文通过采用低空拍摄的图像进行算法验证,通过对训练目标在不同视角下的图像以及其他相似目标车辆的图像进行识别,并与 SVM 方法进行比较,证明本文提出的方法在小数据样本下针对颜色和结构相对规则的车辆具有较高的识别准确率,同时具有一定的泛化能力,在目标车辆发生变化(光照变化、存在遮挡等)时,也能够具有较好的识别效果。对于高空目标识别,基于无人机机载高清传感器拍摄的机载图像,同样可进行小样本识别算法的训练和实现,但需要进一步增加网络层数,并优化推理模型。

参考文献 (References)

- [1] 沈林成, 牛轶峰, 朱华勇. 多无人机自主协同控制理论与方法[M]. 2 版. 北京: 国防工业出版社, 2018.
SHEN Lincheng, NIU Yifeng, ZHU Huayong. Theories and methods of autonomous cooperative control for multiple UAVs[M]. 2nd ed. Beijing: National Defense Industry Press, 2018. (in Chinese)
- [2] 朱宇亭. 面向混合主动感知的无人机对车辆目标识别与跟踪方法[D]. 长沙: 国防科技大学, 2018.
ZHU Yuting. Mixed-initiative recognition and tracking of ground vehicles using UAV images[D]. Changsha: National University of Defense Technology, 2018. (in Chinese)
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580–587.
- [4] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779–788.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]// Proceedings of the European Conference on Computer Vision, 2016: 21–37.
- [7] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 6154–6162.
- [8] ZHANG S F, WEN L Y, BIAN X, et al. Single-shot refinement neural network for object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4203–4212.
- [9] SINGH B, DAVIS L S. An analysis of scale invariance in object detection—SNIP[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 3578–3587.
- [10] GEORGE D, LEHRACH W, KANSKY K, et al. A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs [J]. Science, 2017, 358(6368): 10.1126/science.aag2612.
- [11] 孙晓帅, 姚鸿勋. 视觉注意与显著性计算综述[J]. 智能计算机与应用, 2014(5): 14–18.
SUN Xiaoshuai, YAO Hongxun. A survey of visual attention and saliency computation [J]. Intelligent Computer and Applications, 2014(5): 14–18. (in Chinese)
- [12] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254–1259.
- [13] ACHANTA R, ESTRADA F, WILS P, et al. Salient region detection and segmentation [C]// Proceedings of the 6th International Conference on Computer Vision Systems, 2008: 66–75.
- [14] CHENG M M, ZHANG G X, MITRA N J, et al. Global contrast based salient region detection [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011: 409–416.
- [15] ZITNICK C L, DOLLAÁR P. Edge boxes: locating object proposals from edges [C]// Proceedings of the European Conference on Computer Vision, 2014: 391–405.
- [16] CHENG M M, ZHANG Z M, LIN W Y, et al. BING: binarized normed gradients for objectness estimation at 300 fps[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 3286–3293.
- [17] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104(2): 154–171.
- [18] HOU X D, ZHANG L Q. Saliency detection: a spectral residual approach[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1–8.
- [19] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1597–1604.
- [20] ROTHER C, KOLMOGOROV V, BLAKE A. “GrabCut”: interactive foreground extraction using iterated graph cuts[C]// Proceeding of the ACM Transactions on Graphics, 2004: 309–314.
- [21] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62–66.
- [22] KUMAR S, HEBERT M. Discriminative random fields: a discriminative framework for contextual interaction in classification[C]// Proceedings of the 9th IEEE International Conference on Computer Vision, 2003: 1150–1157.