

# 基于卷积神经网络特征和改进超像素匹配的 图像语义分割

郭呈呈\*, 于凤芹\*\*, 陈莹

江南大学物联网工程学院, 江苏 无锡 214122

**摘要** 非参数语义分割算法易受到图像检索精度和语义类别不均衡数据集的影响而导致语义分割精度下降。针对这些问题,提出了一种基于卷积神经网络(CNN)特征和改进超像素匹配的图像语义分割算法。通过 CNN 学习得到图像特征,降维后进行图像检索,得到精度更高的检索集;利用高斯核密度估计对检索集图像的超像素加权,提升稀少类目标超像素标签的匹配精度,从而提高查询图像的语义分割精度。在 SIFTflow 和 KITTI 数据库上的实验结果显示,本文算法的每像素和平均每类语义分割精度均达到最优。

**关键词** 图像处理; 语义分割; 非参数化; 卷积神经网络特征; 高斯核

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP55.081005

## Image Semantic Segmentation Based on Convolutional Neural Network Feature and Improved Superpixel Matching

Guo Chengcheng\*, Yu Fengqin\*\*, Chen Ying

School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

**Abstract** The segmentation accuracy of nonparametric semantic segmentation algorithm, is easily affected by the image retrieval accuracy and semantic category unbalanced dataset. To solve these problems, an image semantic segmentation algorithm based on convolutional neural network (CNN) feature and improved superpixel matching is proposed. Image features are obtained through CNN learning, and images are retrieved after reducing dimensions of the features, and then the accuracy of image retrieval set can be improved. Superpixels of the images in the retrieval set are weighted by using Gaussian kernel density estimation, which increases the superpixel matching accuracy of rare classes. Therefore, semantic segmentation accuracy of query image can be improved. The experimental results on SIFTflow and KITTI datasets show that both per-pixel and per-class rates of the proposed algorithm are the best among different algorithms.

**Key words** image processing; semantic segmentation; nonparametric; convolutional neural network feature; Gaussian kernel

**OCIS codes** 100.5010; 150.1135; 110.2960

## 1 引言

图像由像素组成,图像的语义分割是将图像中的像素按照表达语义含义的不同进行分组,即为每个像素分配一个类别标签,同时完成图像分割与多目标分类识别两项任务,如行人检测、道路分割、图像标注等<sup>[1-3]</sup>。单独的机器视觉问题通过语义分割方法能够整合在统一的框架中解决,实现整体的场景

理解,因此语义分割在机器视觉领域具有重要的研究意义,已广泛地应用于机器人导航、自动驾驶、智能安防等技术领域。

随着大规模训练数据的出现以及计算机硬件的迅速发展,深度学习在图像语义分割问题中不断取得新的突破<sup>[4-8]</sup>,通过全卷积网络实现端到端的语义分割,改进的卷积-反卷积神经网络<sup>[7]</sup>解决了特征图与输入图像尺寸不匹配的问题,多尺度卷积神经网络

收稿日期: 2018-01-18; 修回日期: 2018-02-17; 录用日期: 2018-03-09

基金项目: 国家自然科学基金(61573168)、中央高校基本科研业务费专项资金(JUSRP51733B)

\* E-mail: 416564199@qq.com; \*\* E-mail: yufq@jiangnan.edu.cn

081005-1

络模型<sup>[8]</sup>则充分利用了物体间的上下文关系,弱监督语义分割方法<sup>[4-6]</sup>不依赖逐像素标注的训练图像,大大降低了获取标注图像的难度。然而,这些方法在解决语义分割的问题时需要花费大量时间训练模型的参数,很难适应不断变化的开放式数据集,而当数据集中有新的训练数据加入或者增加新的类别标签时参数模型都需要重新训练。因此,提出非参数的语义分割方法用于解决这个问题。非参数方法通过提取图像特征检索相似图像,然后进行超像素匹配将相似图像的标签转移到查询图像上,最终实现查询图像的语义分割。Liu 等<sup>[9]</sup>提出利用尺度不变特征变换(SIFT)流匹配方法将像素的标签从数据集图像转移到新的图像上,但单个像素携带的语义信息非常少且计算复杂度高。Tighe 等<sup>[10-11]</sup>提出了基于超像素的标签转移方法,并且通过加入类别分类器来进一步改善分割精度。Yang 等<sup>[12]</sup>更加关注场景中的稀少类目标,通过扩展检索集并在其中加入稀少类标签的超像素以提升稀少类目标的语义分割精度。Tung 等<sup>[13]</sup>认为超像素容易将目标碎片化,为了获得更完整的目标,利用内容自适应的窗进行标签转移。Najafi 等<sup>[14]</sup>提出了一种采样和滤波的策略,提取图像的颜色纹理等特征后计算图像间的相似性得分,根据得分采样各个类别的超像素块,利用高斯滤波操作实现超像素间的匹配。

由于在图像检索步骤采用人工选择的图像特征,现有非参数语义分割方法易造成误检,并且在类别分布不均衡的数据集中难以分割出稀少类目标。为此,本文对非参数图像语义分割算法进行了改进,利用卷积神经网络(CNN)从原始图像中学习图像特征,并通过主成分分析(PCA)对特征进行降维处理,减少特征冗余,提高检索的速度和精度;采用核密度估计改进超像素似然估计,能够更准确地为局部超像素特征分布建模,提升稀少目标类别分割精度;最后利用马尔科夫随机场(MRF)对语义分割结

果进行平滑优化,实现对查询图像的语义分割。

## 2 基本原理

非参数图像语义分割算法包括图像检索、超像素匹配和 MRF 语义推理 3 个步骤。

### 2.1 基于 CNN 特征的图像检索

图像检索用于找到与查询图像相似的图像集合,为后续的标签转移提供上下文语义信息并且提升计算效率。非参数方法通常通过人工选择图像特征(如颜色直方图、空间包络特征(GIST)、SIFT、方向梯度直方图(HOG)特征等)在数据集中检索查询图像的相似图像检索集,对于背景环境复杂的图像,利用这些特征很容易造成误检。受到卷积神经网络在图像识别任务上成功应用的启发,提出利用卷积神经网络自动学习原始图像数据的高层语义特征,并对特征降维后进行图像检索,在保证检索速度的同时提升检索精度。

对于查询图像  $I_q$ ,将图像缩放为统一的  $224 \text{ pixel} \times 224 \text{ pixel}$  作为输入,用 5 个卷积层和 2 个全连接层进行学习,在第 1、2、5 层卷积操作后使用最大池化的方法进行池化操作,最终在第 7 层输出 4096 维图像的特征向量。CNN 网络结构如表 1 所示(第一行为卷积滤波器的感受野大小,第二行为步长和填充,第三行为是否进行局部响应归一化(LRN)和最大池化操作)。

为了保证检索速度和进一步提升检索精度,对学习到的图像特征进行 PCA 降维处理,去除特征冗余,最终得到 128 维的图像特征向量用于图像检索。利用 CNN 特征计算查询图像  $I_q$  与数据集图像  $I_d$  间的余弦距离并按升序排序,取前  $K$  幅图像构成  $I_q$  的检索集  $R = \{I_d^1, I_d^2, \dots, I_d^K\}$ ,其中  $m(I_q, I_d^i) > m(I_q, I_d^{i+1})$ ,  $m(I_q, I_d)$  为查询图像与检索集图像间的相似性。检索集  $R$  包含的全部类别标签决定了查询图像的语义标签种类。

表 1 CNN 网络结构

Table 1 Network architecture of CNN

Item	Conv1	Conv2	Conv3	Conv4	Conv5	Full6	Full7
Size	$64 \times 11 \times 11$	$256 \times 5 \times 5$	$256 \times 3 \times 3$	$256 \times 3 \times 3$	$256 \times 3 \times 3$	4096	4096
Step	Stride 4, pad 0	Stride 1, pad 2	Stride 1, pad 1	Stride 1, pad 1	Stride 1, pad 1	Drop-out	Drop-out
Operation	LRN, pool	LRN, pool	—	—	Pool	—	—

### 2.2 超像素匹配

在超像素匹配阶段将查询图像分割为超像素<sup>[15]</sup>  $S = \{s_i\}_{i=1}^n$ ,并提取超像素的特征  $F = \{f_i^k\}_{i=1}^n$ ,每个超像素  $s_i$  属于各个语义类别的似然

率得分由检索集图像超像素计算得到,根据朴素贝叶斯假设,在给定语义类别的情况下,每种特征  $f_i^k$  间是相互独立的,则超像素  $s_i$  属于类别  $c$  的似然概率为

$$L(s_i, c) = \frac{P(s_i | c)}{P(s_i | \bar{c})} = \prod_k \frac{P(f_i^k | c)}{P(f_i^k | \bar{c})}, \quad (1)$$

式中:  $\bar{c}$  为包括  $c$  在内的所有语义标签类别;  $f_i^k$  为  $s_i$  的第  $k$  种特征向量; 似然率  $P(f_i^k | c)/P(f_i^k | \bar{c})$  通过  $f_i^k$  在检索集中的近邻的特征密度估算得到。用  $D$  表示训练集中所有超像素的集合,  $N_i^k$  表示检索集中特征向量为  $f_i^k$  的超像素  $s_i$  的近邻超像素集合, 则

$$\frac{P(f_i^k | c)}{P(f_i^k | \bar{c})} = \frac{n(c, N_i^k)/n(c, D)}{n(\bar{c}, N_i^k)/n(\bar{c}, D)}, \quad (2)$$

式中  $n(c, X)$  为集合  $X$  中类别标签为  $c$  的超像素个数。通过最大化 (1) 式为查询图像  $I_q$  的每个超像素  $s_i$  分配一个语义标签实现图像的语义分割。

### 2.3 改进的高斯核似然估计超像素匹配

对于基本的超像素匹配方法, 当数据集数据足够多且各类别目标数量相对平衡的时候, 利用超像素特征能够很好地匹配到其局部近邻超像素, 但在实际情况中数据集通常是非常不均衡的, 如车、门等稀少类目标在数据集中的样本十分稀疏, 大大减弱了这些稀少类目标特征的表达能力, 使用规模较小的检索集进行超像素匹配会使得这个问题更加严重, 而简单地增加检索集图像数量则会引入大量干扰, 无法改善超像素的似然估计。

为此, 提出利用核密度估计为局部特征分布做更加准确的建模, 通过查询图像超像素特征与其近邻超像素特征间的距离和高斯核为每一个近邻超像素估测一个权重。假设  $f^k$  是独立同分布的, 则  $P(f^k)$  的计算公式为

$$P(f^k) = \frac{1}{Z} \sum_{j=1}^{|D|} \exp\left(-\frac{\|f^k - f_j^k\|^2}{\sigma_k^2}\right), \quad (3)$$

式中:  $Z$  为归一化因子;  $|D|$  为集合  $D$  中超像素的个数;  $\sigma_k$  为高斯核带宽; 最优带宽采用文献[16]的方法计算得到。对于查询图像超像素的特征向量  $f_i^k$ , 根据核密度估计对其近邻超像素进行加权处理, 则  $f_i^k$  的条件概率的计算公式为

$$P(f_i^k | c) = \frac{1}{n(c, D)} \sum_{j=1}^{n(c, N_i^k)} \exp\left(-\frac{\|f_i^k - f_{c,j}^k\|^2}{\sigma_k^2}\right), \quad (4)$$

式中  $f_{c,j}^k$  为标签类别为  $c$  的近邻超像素的特征向量。同理将 (4) 式中的  $c$  替换为  $\bar{c}$  即可定义  $P(f_i^k | \bar{c})$ 。若将检索集图像分割为超像素后直接作为近邻超像素集合  $N_i^k$ , 会导致估测的条件分布过于平滑, 因此计算了  $f_i^k$  和  $f_j^k$  间的欧氏距离, 取距离最近的前  $T$  个超像素组成集合  $N_i^k$ 。

### 2.4 基于 MRF 的语义推理

为了使图像语义分割的结果更加符合上下文环境, 利用标准的 MRF 模型对分割结果进行优化。定义一个能量函数为

$$E(c) = \sum_{s_i} E_{\text{data}}(s_i, c_i) + \lambda \sum_{s_i, s_j} E_{\text{smooth}}(c_i, c_j), \quad (5)$$

式中:  $\lambda$  为平滑系数; 数据项定义为  $E_{\text{data}} = -\log[L(s_i, c_i)]$ ; 平滑项根据类别标签的共现概率定义为

$$E_{\text{smooth}}(c_i, c_j) = -\log[(P(c_i | c_j) + P(c_j | c_i))/2] \times \delta[c_i \neq c_j], \quad (6)$$

式中类别标签不同的相邻超像素的条件概率  $P(c | c')$  由训练集图像估测得到。MRF 推理采用最优图割算法计算, 最终得到一幅图像的语义分割结果。

## 3 算法流程

基于 CNN 特征和改进超像素匹配的图像语义分割算法流程图如图 1 所示。

训练阶段步骤如下:

1) 将训练图像尺寸统一缩放为  $224 \text{ pixel} \times 224 \text{ pixel}$ , 输入 CNN 网络, 经过 5 个卷积层和 2 个全连接层输出 4096 维的图像特征, 利用 PCA 降维后每幅训练图像计算出 128 维的特征向量用于测试阶段的图像检索。

2) 采用基于图的快速分割算法对训练集图像进行超像素分割, 每个超像素中大于 50% 的像素类别即为该超像素的类别。

3) 提取 4 种超像素特征, 分别为 128 维的颜色特征向量、36 维的位置特征向量、1024 维的 SIFT 特征向量以及 168 维的分层方向梯度直方图 (PHOG) 特征向量用于测试阶段的超像素匹配。

4) 用两个条件概率  $P(c_i | c_j)$  和  $P(c_j | c_i)$  代替联合概率  $P(c_i, c_j)$  计算一个超像素分配到类别  $c_i$  时其相邻超像素分配到类别  $c_j$  的类别共现概率  $E_{\text{smooth}}(c_i, c_j)$ , 并作为 MRF 能量函数的平滑项。

测试阶段步骤如下:

1) 提取测试图像的 CNN 特征并降维, 计算测试图像和训练集图像特征间的欧氏距离并按升序排序, 取距离排序结果的前  $K$  幅图像组成检索集  $R = \{I_d^1, I_d^2, \dots, I_d^K\}$ 。

2) 将测试图像分割为超像素  $s_i$  并提取超像素的特征  $f_i^k$ , 计算  $s_i$  与检索集图像超像素特征间的欧氏距离, 取前  $T$  个组成近邻超像素集合  $N_i^k$ 。

3) 通过(4)式计算超像素  $s_i$  的高斯核似然估计,取负对数后作为 MRF 的数据项。

4) 最终利用图切优化的方法进行最大后验概率推理,得到测试图像的全局语义标注,实现图像的语义分割。

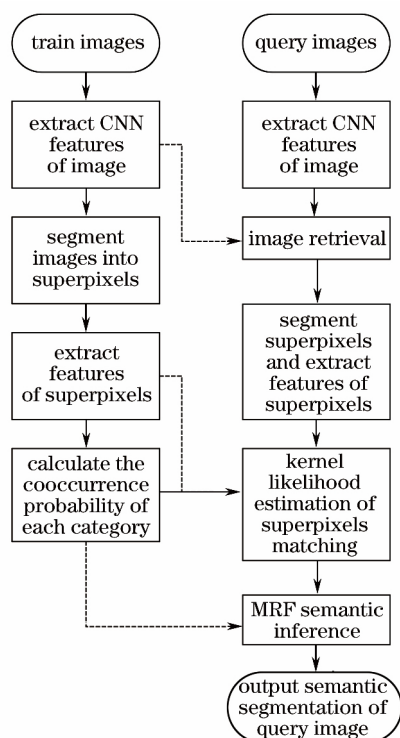


图1 算法流程图

Fig. 1 Flow chart of the algorithm

## 4 仿真实验结果与分析

为验证本文算法的有效性,仿真实验分别在 SIFTflow<sup>[9]</sup> 和 KITTI<sup>[17]</sup> 数据库上进行。SIFTflow

数据库共有 2688 幅室外图像,其中包含了 2488 幅训练图像和 200 幅测试图像,所有图像均为 256 pixel × 256 pixel,包含 33 种语义类别标签。KITTI 数据库为复杂的交通道路场景数据库,由车载摄像机拍摄得到,选取其中 533 幅作为训练图像,100 幅作为测试图像,将数据集的语义标签合并为 7 类。数据库中的图像均提供了逐像素标注的语义标签作为基准。实验环境为 Intel i5-6200U 处理器、8G 内存,采用 Matlab R2016a 编程。

### 4.1 SIFTflow 数据库实验结果

实验采用逐像素分割标注精度和每类的分割标注精度作为算法的主要评价指标。对于每幅测试图像,利用 CNN 特征进行图像检索能够使用更小的检索集达到更优的语义分割精度,文献[10]和文献[13]的检索集  $K$  大小分别为 200 和 400,而本文算法仅需检索 40 幅图像组成检索集,计算效率更高。MRF 能量函数的平滑项系数  $\lambda=5$ ,近邻超像素集合  $T=20$ ,根据语义分割结果分别计算逐像素和各个类别的分割标注精度,并与文献[9~13]提出的非参数场景语义分割算法进行比较,对比结果如表 2 和表 3 所示。

表 2 不同算法在 SIFTflow 数据库上的语义分割精度对比  
Table 2 Comparison of semantic segmentation accuracy of different algorithms on SIFTflow dataset

Algorithm	$K$	Per-class accuracy / %	Per-pixel accuracy / %
Ref. [9]	85	—	75.27
Ref. [10]	200	34.16	77.03
Ref. [12]	40	32.40	78.61
Ref. [13]	400	42.21	77.98
Proposed	40	43.08	78.65

表 3 不同算法在 SIFTflow 数据库上部分类别的语义分割精度对比

Table 3 Accuracy comparison for several classes of semantic segmentation of different algorithms on SIFTflow dataset %

Algorithm	Building	Mountain	Sky	Field	Car	Grass	Door	Person	Crosswalk
Ref. [10]	84.21	68.02	92.06	46.41	50.28	35.76	10.03	11.96	22.89
Ref. [12]	85.44	<b>71.01</b>	92.05	50.06	50.13	42.98	11.74	13.85	23.09
Ref. [13]	85.63	70.07	92.81	59.06	<b>51.02</b>	43.11	13.68	15.96	<b>23.14</b>
Proposed	<b>89.92</b>	70.80	<b>93.06</b>	<b>67.47</b>	48.36	<b>47.37</b>	<b>17.49</b>	<b>16.52</b>	19.80

在 SIFTflow 数据库中本文算法的每像素和每类的语义分割精度分别为 78.65% 和 43.08%,比文献[13]的算法提高了 0.9% 和 2.1%,在几种语义分割算法中均达到了最优,能够更好地分割不同语义类别的目标。相比于每像素分割精度,平均每类的分割精度较低是因为数据库图像包含的各个语义类别标签的数量十分不均衡,如图 2 所示,建筑、树、山、天空和海洋 5 种语义类别占了整体的 80%,其

余 28 类仅占整体的 20%,因此分割标注这些稀少类别的难度很大。本文算法能够有效提升稀少类别的分割精度,由表 3 对比结果可知,草地、门、人等稀少类别的精度均有提高,同时,建筑、天空等普通类别的分割精度也有所提升。

表 2 中的  $K$  为检索集大小,采用的 CNN 特征图像检索方法检索精度更高,适当减少检索集图像数量不影响检索精度,因此能够用规模更小的检索

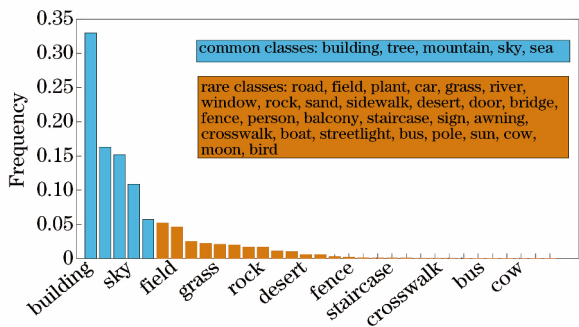


图 2 SIFTflow 数据库各类别频率分布  
Fig. 2 Frequency distributions of classes on SIFTflow dataset

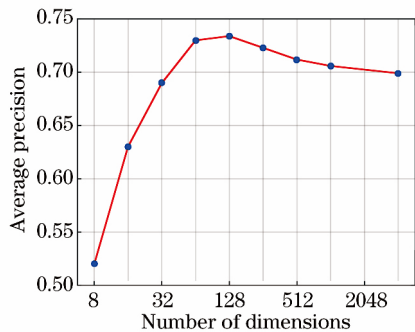


图 3 特征维度对图像平均检索精度的影响  
Fig. 3 Influence of feature dimensions on average retrieval precision of image

集达到更高的语义分割精度。在提取 CNN 特征用于图像检索前,对 CNN 特征进行了降维操作,这一

步骤不仅能够提高计算速度,对于检索精度也有一定程度的提升。图 3 所示为特征的维度与平均检索精度的关系。当将 4096 维的图像特征压缩至 128 维后进行检索,平均检索精度提升了 5.7%,这是因为直接提取的 CNN 特征存在冗余,导致计算图像间距离时出现误差,降低了检索精度。

4.2 KITTI 数据库实验结果

KITTI 数据库图像尺寸不同,需要先将图像转换为 1226 pixel×370 pixel。与 SIFTflow 数据库类似,KITTI 图像的语义类别分布也是不均衡的,如图 4 所示,出现频率占总体 80% 的类别为道路、树木、天空等普通类,而建筑、草地、指示牌、车、行人等稀少类仅占总体的 20%。表 4 为几种算法的语义分割精度对比。

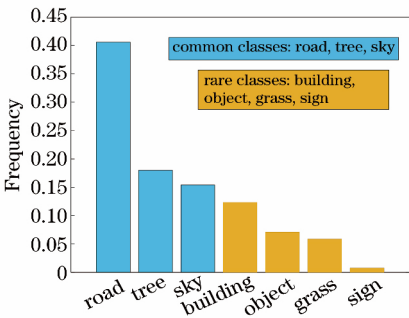


图 4 KITTI 数据库各类别频率分布  
Fig. 4 Frequency distributions of classes on KITTI dataset

表 4 不同算法在 KITTI 数据库上的语义分割精度对比

Table 4 Comparison of semantic segmentation accuracy of different algorithms on KITTI dataset									%
Algorithm	Road	Sky	Tree	Building	Grass	Sign	Object	Per-class	Per-pixel
Ref. [10]	<b>78.60</b>	79.62	98.75	49.68	4.27	13.90	22.54	49.62	73.27
Ref. [12]	76.82	79.26	98.75	57.77	4.9	14.02	31.37	51.84	74.99
Proposed (CNN)	74.79	<b>90.45</b>	<b>98.80</b>	58.45	<b>10.75</b>	13.13	39.69	55.15	77.56
Proposed (kernel)	78.01	88.25	98.11	60.56	4.54	13.46	40.23	54.74	77.68
Proposed (full)	77.27	90.02	98.77	<b>61.80</b>	9.08	<b>14.20</b>	<b>41.37</b>	<b>56.07</b>	<b>79.17</b>

(Note: full=CNN+kernel)

表 4 中对比了文献[10]、[12]算法与本文算法,其中具体展现了本文的两点改进对最终语义分割结果的影响,即利用 CNN 特征进行图像检索主要提升了普通类别的语义分割精度,如天空和树木类别分割精度达到了 90.45%和 98.80%,而采用高斯核进行超像素匹配则提升了稀少类别的语义分割精度,如建筑和物体类别的分割精度为 60.56%和 40.23%。综合这两点改进,最终本文算法的每像素和每类别分割精度在几种算法中为最优,分别比文

献[12]的算法提升了 8.2%和 5.6%。  
图 5 列出了几组典型的实验结果图。相比于文献[12]算法,本文算法对于稀少类目标的分割精度更高,如第 4 幅图中的建筑类和第 6 幅图中的草地和标志类,并且对于普通类也有较好的分割效果,如第 2 幅图中的天空类和树木类。但也能发现算法受到强烈光照和阴影的影响,易产生错误的分割标注结果,如第 5 幅图没能分割出右侧阴影下的车。



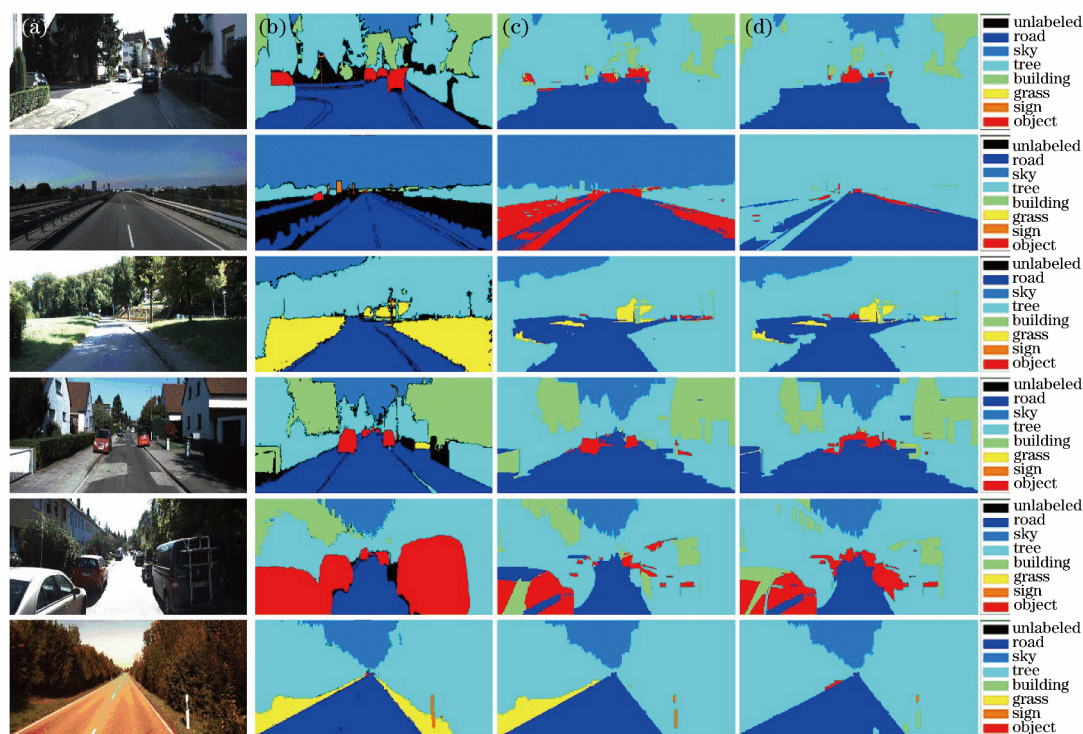


图 5 KITTI 数据库中部分实验结果图。(a)原图像；(b)正确分割标注图像；(c)本文算法；(d)文献[12]算法

Fig. 5 Part of experimental results on KITTI dataset. (a) Original image; (b) ground truth; (c) proposed algorithm; (d) algorithm of Ref. [12]

## 5 结 论

现有的非参数语义分割方法存在两点局限:1) 采用人工选择的图像全局特征进行图像检索导致检索集中干扰图像过多影响语义分割结果;2) 对于语义类别不平衡的数据库难以分割标注稀少类别的超像素。针对这些问题,本文提出两点改进:1) 利用预训练好的卷积神经网络学习图像特征,以此代替人工选择的特征用于图像检索提升普通类语义分割精度;2) 使用高斯核密度估计为局部特征分布做更加准确的建模,以此提升稀少类语义分割精度,最终提升图像整体的逐像素语义分割精度和各个类别的平均分割精度。结果表明,在 SIFTflow 数据库和 KITTI 数据库上所提算法的分割精度均达到了最优。但本文算法对于光照强烈、阴影遮蔽多的图像的分割标注结果仍不理想,有待进一步研究。

## 参 考 文 献

- [1] Ye G L, Sun S Y, Gao K J, *et al.* Nighttime pedestrian detection based on faster region convolution neural network [J]. *Laser & Optoelectronics Progress*, 2017, 54(8): 081003.  
叶国林, 孙韶媛, 高凯珺, 等. 基于加速区域卷积神

经网络的夜间行人检测研究[J]. *激光与光电子学进展*, 2017, 54(8): 081003.

- [2] Wang M, Xiao L, Yang F. Blind road segmentation based on saliency detection and improved projective dictionary pair [J]. *Laser & Optoelectronics Progress*, 2017, 54(4): 041001.  
王民, 肖磊, 杨放. 基于显著性检测和改进投影字典对的盲道分割[J]. *激光与光电子学进展*, 2017, 54(4): 041001.
- [3] Shi C J, Ruan Q Q. Feature selection with enhanced sparsity for web image annotation [J]. *Journal of Software*, 2015, 26(7): 1800-1811.  
史彩娟, 阮秋琦. 基于增强稀疏性特征选择的网络图像标注[J]. *软件学报*, 2015, 26(7): 1800-1811.
- [4] Bearman A, Russakovsky O, Ferrari V, *et al.* What's the point: semantic segmentation with point supervision[C]. *European Conference on Computer Vision*, 2016: 549-565.
- [5] Wei Y C, Liang X D, Chen Y P, *et al.* STC: a simple to complex framework for weakly-supervised semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(11): 2314-2320.
- [6] Shi Z Y, Yang Y X, Hospedales T M, *et al.* Weakly-supervised image annotation and

- segmentation with objects and attributes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2525-2538.
- [7] Gao K J, Sun S Y, Yao G S, *et al.* Semantic segmentation of night vision images for unmanned vehicles based on deep learning [J]. Journal of Applied Optics, 2017, 38(3): 421-428.  
高凯珺, 孙韶媛, 姚广顺, 等. 基于深度学习的无人车夜视图像语义分割[J]. 应用光学, 2017, 38(3): 421-428.
- [8] Liu D, Liu X J, Wang M Z. Semantic segmentation with multi-scale convolutional neural network[J]. Remote Sensing Information, 2017, 32(1): 57-64.  
刘丹, 刘学军, 王美珍. 一种多尺度 CNN 的图像语义分割算法[J]. 遥感信息, 2017, 32(1): 57-64.
- [9] Liu C, Yuen J, Torralba A. Nonparametric scene parsing via label transfer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2368-2382.
- [10] Tighe J, Lazebnik S. Superparsing: scalable nonparametric image parsing with superpixels[C]. European Conference on Computer Vision, 2010: 352-365.
- [11] Tighe J, Lazebnik S. Finding things: image parsing with regions and per-exemplar detectors [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3001-3008.
- [12] Yang J M, Price B, Cohen S, *et al.* Context driven scene parsing with attention to rare classes[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2014: 3294-3301.
- [13] Tung F, Little J J. Scene parsing by nonparametric label transfer of content-adaptive windows [J]. Computer Vision and Image Understanding, 2016, 143: 191-200.
- [14] Najafi M, Namin S T, Salzmann M, *et al.* Sample and filter: nonparametric scene parsing via efficient filtering[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016: 607-615.
- [15] Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation [J]. International Journal of Computer Vision, 2004, 59(2): 167-181.
- [16] Shimazaki H, Shinomoto S. Kernel bandwidth optimization in spike rate estimation[J]. Journal of Computational Neuroscience, 2010, 29: 171-182.
- [17] Menze M, Geiger A. Object scene flow for autonomous vehicles [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3061-3070.