

1.数据集

名称	来源	数据内容	标注特点	备注
CUB200-2011	加利福尼亚理工学院	含 200 种不同类别,共 11,788 张鸟类图像数据	人工标注数据,每张图像包含 15 个局部区域位置,312 个二值属性,1 个标注框,以及语义分割图像.	最经典最常用
Stanford Dogs	斯坦福大学	含 120 种不同类别,共 20,580 张狗的图像数据	只提供标注框,这一个人工标注数据.	标注信息少
Oxford Flowers	牛津大学	两种数据库,分别含 17 种类别和 102 种类别的花,每个类别包含 40 到 258 张图像数据,总共有 8,189 张图像	只提供语义分割图像,不包含其他额外标注信息.	102 种类别的数据库比较常用,但图像数量较少
Cars	Ernesto Ramos and David Donoho	含 196 类不同品牌不同年份不同车型的车辆图像数据,一共有 16,185 张图像	只提供标注框信息.	
FGVC-Aircraft	Johns Hopkins CLSP Summer Workshop 2012	含 102 类不同的飞机照片,每一类别含有 100 张不同的照片,共有 10,200 张图片	只提供标注框信息.	

综合考虑数据集的应用频率和包含的图像数量,并且根据方法是否需要标注信息,我比较倾向 CUB200-2011 或者 Stanford Dogs 数据集

2.方法

根据《基于深度卷积特征的细粒度图像分类研究综述》罗建豪 吴建鑫综述中提到的方法,根据对比我挑选了几种算法:

基础的方法是 SIFT+BoW+SVM、POOF+SVM、Fisher+SVM 这种吗?

比较新颖的方法:

(综述中提到)

Pose Normalized CNN (Alex-Net+Fine-Tune+SVM) 需要有标注框信息和局部区域信息并且需要训练和测试, 准确率达到 85.4%

Spatial Transformer Net (Inception+Flip) 不需要标注信息的训练和测试, 准确率可以达到 84.1%

(2016 年的)

Part-Stacked CNN (ALEXNET Conv+ReLU+Pool) 训练用 BBox+Parts, 对 BBox 进行测试, 准确率达到 76.6%

Picking Deep Filter (PD+FC+SWFV-CNN) 不需要标注信息的训练和测试, VGG-VD 和 VGG-M 在 CUB200 数据及测试准确率分别达到 84.54%和 80.23%。

所以先从这几种方法中选择可行性高吗?

3.判断准则

部分检测

识别性能

本地化准确性 localization accuracy

分类精度 classification accuracy

推理效率 inference efficiency

模型解释 model interpretation.

基线改进 Improvement Over Baseline

强先验或弱先验 Strong prior or weak prior

mAP

宏平均准确率

宏平均召回率

宏平均 F1 值

不知道是否准确