

图像语义分割技术开源代码

北京电子科技学院 吴乐 金鑫

图像语义分割 (semantic segmentation), 从字面意思上理解就是让计算机根据图像的语义来进行分割, 如图 1 所示, 使得计算机能够分析输入图像 (左图), 输出像素级语义标记 (右图)。

语义在语音识别中指的是语音的含义, 在图像领域, 语义指的是图像的内容, 比如左图的语义就是三个人骑着三辆自行车; 分割的意思是从像素的角度分割出图片中的不同对象, 对原图中的每个像素都进行标记, 比如右图中粉红色代表人, 绿色代表自行车。



图 1 图像语义分割示意图

本文着重介绍几个基于深度学习的图像语义分割技术开源代码, 包括: 图像语义分割数据集和网络结构, 具体结构包括 FCN、SegNet、U-Net、DeepLab v1&v2、RefineNet、PSPNet。

1. 图像语义分割数据集

1.1. Pascal VOC 系列

PASCAL VOC 2012 系列数据集开始有 1464 张具有标注信息的训练图片, 2014 年增加到 10582 张训练图片。主要涉及了日常生活中常见

的物体, 包括汽车, 狗, 船等 20 个分类。其网址为: <http://host.robots.ox.ac.uk/pascal/VOC/voc2012/>

1.2. Microsoft COCO

该数据集共有 80 个类别。这个数据集主要用于实例级别的分割 (Instance-level Segmentation) 以及图片描述 Image Caption)。

其网址为: <http://cocodataset.org/>

1.3. Cityscapes

这是适用于汽车自动驾驶的训练数据集, 包括 19 种都市街道场景: road、side-walk、building、wall、fence、pole、traffic light、traffic sign、vegetation、terrain、sky、person、rider、car、truck、bus、train、motorcycle 和 bicycle。

该数据库中用于训练和校验的精细标注的图片数量为 3475, 同时也包含了 2 万张粗糙的标记图片。其网址为:

<https://www.cityscapes-dataset.com/>

2. 图像语义分割网络

2.1. FCN

FCN 网络发表于 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]// Computer Vision and Pattern Recognition. IEEE, 2015:3431-3440。

起源于美国加州大学伯克利分校的这项工作作为语义分割引入了端到端的全卷积网络, 在网络结构中, 重新利用 ImageNet 的预训练网络用于语义分割, 使用了反卷积层进行上采样并且引入跳跃连接来改善上采样粗糙的像素定位。其网络结构如图 2 所示,

代码的主页为: <https://github.com/MarvinTeichmann/tensorflow-fcn>。

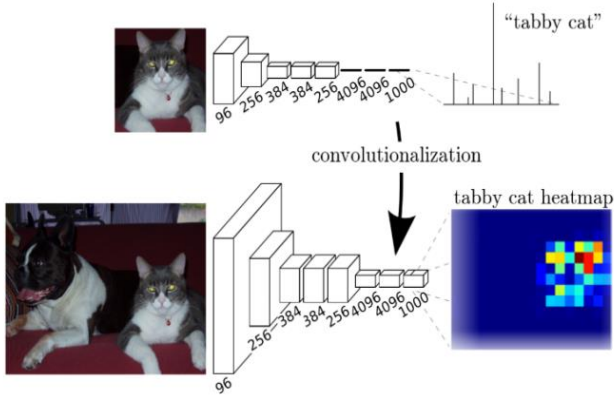


图 2 FCN 网络结构

2.2. SegNet

发表 SegNet 网络的论文为: Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):1-1。来源于美国加州大学伯克利分校的这项作为语义分割引入了端到端的全卷积网络，在构建的网络结构中，重新利用 ImageNet 的预训练网络用于语义分割，并使用了反卷积层进行上采样并且引入跳跃连接来改善上采样粗糙的像素定位。如图 3 所示为 SegNet 的网络结构，其代码的网址为：<https://github.com/zhixuhao/unet>

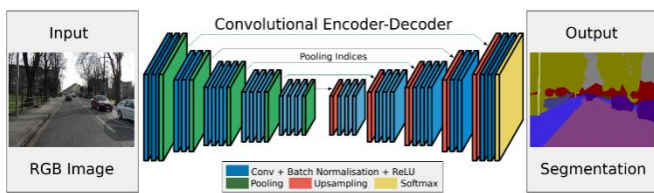


图 3 SegNet 网络结构

2.3. U-Net

U-Net 网络的论文发表于 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015:234-241。这项来源于德国

弗莱堡大学的工作为语义分割引入了更规整的网络结构，在结构上结合了不同尺度的跨层连接，通过将编码器的每层结果拼接到译码器中得到更好的结果。如图 4 所示为 U-Net 网络的结构图。其代码的网址为：<https://github.com/zhixuhao/unet>

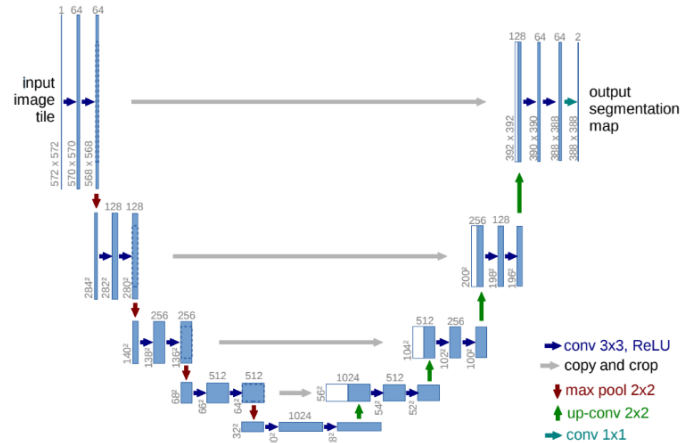


图 4 U-Net 网络结构

2.4. DeepLab v1&v2

DeepLab v1&v2 的 2 篇经典参考论文包括:

(1) Chen L C, Papandreou G, Kokkinos I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. Computer Science, 2014(4):357-361。

(2) Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 40(4):834-848。

来源于 Google 的 DeepLab 的这项工作为在语义分割上使用了膨胀卷积，提出了在暗黑空间金字塔池化 (ASPP) 的基础上来融合不同尺度的信息，并使用全连接的条件随机场来构建网络架构。如图 5 所示为 DeepLab 的原理分析说明。其代码网址为：<https://github.com/tensorflow/models/tree/master/research/deeplab>

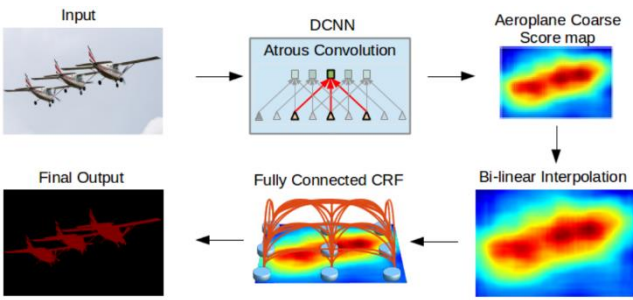


图 5 DeepLab 原理分析

2.5. RefineNet

RefineNet 网络的推荐参考论文为: Lin G, Milan A, Shen C, et al. RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:5168-5177。

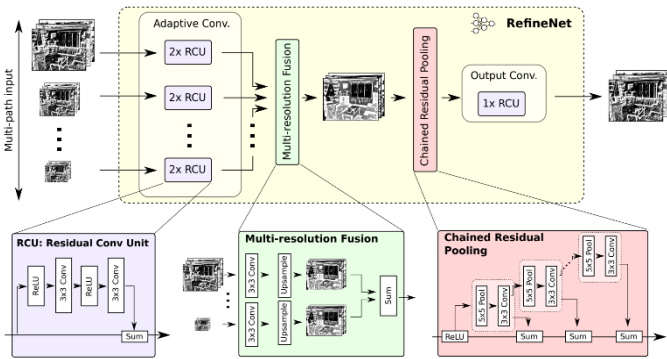


图 6 RefineNet 网络结构

来源于澳大利亚阿德莱德大学的这项工作精心设计了译码模块, 其所有模块遵循残余连接设计, 使用残差链接显式将各个下采样层和后面

的网络层结合在一起。这样, 网络高层的语义特征可以直接从底层的卷积层中获得进一步的细化 (refine)。如图 6 所示为 RefineNet 的网络结构。其代码的网址为: <https://github.com/guosheng/refinenet>

2.6. PSPNet

PSPNet 网络的推荐参考论文为: Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:6230-6239。来源于香港中文大学的这项工作为语义分割提出了金字塔池化模块来聚合图片信息。针对 FCN 全卷积网络分割面临的诸多问题, 该工作从多尺度入手, 提出了金字塔模型来提取多尺度的信息, 达到了 State-of-the-art 的结果, 此外网络使用附加的损失函数优化最后的效果。如图 7 所示为 PSPNet 网络的结构图。其代码的网址为: <https://github.com/hszhao/PSPNet>

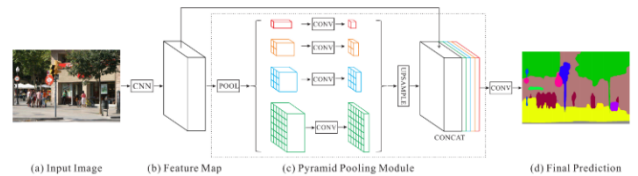
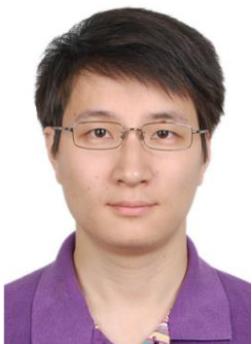


图 7 PSPNet 网络结构

(责任编辑: 贾同)



金鑫:

现于北京电子科技学院从事教学与科研工作, 担任北京电子科技学院可视计算与安全实验室负责人。研究方向为计算机视觉、虚拟/增强现实、人工智能安全。个人主页: www.jinxin.me



吴乐:

现于北京电子科技学院攻读研究学位, 主要的学习和研究方向包括计算美学、计算机图形图像处理。Email: gaoxing0031@163.com

计算机视觉中的交通标志数据集

厦门大学 曲延云 陈怡姿

交通标志的识别有着显著的颜色和形状特征，主要有指示、提示和警示车辆及行人的功能。交通标志识别一直以来是无人驾驶技术的一个重要组成部分，对驾驶行为的决策起着重要的作用。随着深度学习的发展，以及数据库建设的完善，深度学习成为交通标志检测和识别的主流方法。

本文针对交通标志识别或检测任务中常用的交通标志数据集进行总结，并给出链接。

1. 德国交通标志识别标准库

数据集地址：<http://benchmark.ini.rub.de/?section=gtsrb&subsection=dataset>

德国交通标志识别标准库（German Traffic Sign Recognition Benchmark, GTSRB）在 2011 年交通标志识别比赛（IJCNN）中被首次公开，该数据库可用于交通标志分类与识别。GTSRB 主要包括限速标志、禁令标志、解除禁令标志、指示标志、危险标志、特殊标志等 6 大类别，这 6 大类别标志包含 43 个细分类别。数据库中图像大小从 15*15 到 250*250 不等。图 1 给出了 GTSRB 的交通标志图像示例。



图 1. 德国交通标志标准库数据示例

在 GTSRB 的官方数据库中，训练集共有 39209 张图像，测试集共有 12630 张图像。同一

类的图像存放在同一文件夹下，并以 PPM 格式存储，每张图像中有且仅有一个交通标志，每个文件夹下有 CSV 格式的标注信息，标注内容为图像的大小与交通标志的位置以及类别。GTSRB 官方网站还提供预先计算的 HOG 特征、Haar-like 特征和颜色直方图的 TXT 文件。

2. 瑞典交通标志数据库

数据集地址：<http://www.cvl.isy.liu.se/research/datasets/traffic-signs-dataset/>

该数据库由 LINKÖPING UNIVERSITY 的计算机视觉实验室于 2011 年发布。实验人员利用一个车载彩色摄像头在瑞典高速公路和城市道路对真实场景进行拍摄。该摄像头镜头焦距 6.5 毫米、130 万像素。取景路程超过 350KM，共记录 20000 多帧图像。对于采集的视频，每隔 4 帧进行人工标记，总共涵盖 3488 个交通标志。该数据库可被用于交通标志的检测与识别。瑞典交通标志数据库数据示例如图 2 所示。



图 2. 瑞典交通标志数据库数据示例

3. 清华 - 腾讯交通标志数据集 (Tsinghua-Tencent100K)

清华-腾讯交通标志数据集的网络地址为：<http://cg.cs.tsinghua.edu.cn/traffic-sign/>

清华-腾讯交通标志数据库（Tsinghua-Tencent 100K）的图像来自腾讯街景全景。整个数据集包含 100000 幅图像，共包含 30000 个交通标志实例。原始全景图由 6 台单反相机拍摄然后拼接而成，并使用了曝光调整等技术处理图像。全景图的上 25%和下 25%被裁剪掉（交通标志基本不出现），其余部分垂直分割成 4 个子图像。

该交通标志数据库选取了腾讯街景全景中 5 个不同城市的 10 个区域（包括了每个城市的市区和郊区）。数据集图像中的交通标志均由人工标出。该数据库可被用于交通标志的检测与识别。与瑞典交通标志库相比，清华-腾讯 100K 中出现的交通标志实例都各不相同，而瑞典交通标志库由于是从视频序列中提取交通标志图像，会出现许多相似图像。清华-腾讯 100K 如图 3 所示。



图 3. 清华-腾讯 100K 数据示例

4. 比利时交通标志数据集

数据集地址：<https://btsd.ethz.ch/shareddata/>

比利时交通标志数据集（KUL Belgium TS Dataset）是一个大型数据集，有 10000 多个交通标志注释，数千个不同的交通标志。用 8 个高分辨率车载摄像机在比利时的法兰德斯地区的城市环境中拍摄，记录了 4 个视频序列，总计超过 3 小时，同时还有交通标志注释，图像以 JP2 格式存储。

同时该网站还提供了交通标志分类使用的数据集 Belgium TSC（Belgium TS for Classification），是 Belgium TS 数据集的一个子集，包含 62 类交通标志。其中训练集有 4591 个图像，测试集有 2534 个图像。与德国交通标志数据库相同，同一类别的交通标志归属于同一文件夹下，以 PPM 格式存储，每个文件夹下有 CSV 格式的标注信息，标注内容为图像的大小与交通标志的位置以及类别。Belgium TS 数据示例如图 4 所示。



图 4. 比利时交通标志数据集示例

（责任编辑：李策）

曲延云



教授，厦门大学信息科学与技术学院，主要研究方

向为计算机视觉、模式识别、机器学习等。

陈怡姿

现为厦门大学信息科学与技术学院攻读硕士学位。目前主要的学习和研究方向为计算机视觉、图像恢复。

