# 语义分割的数据集

目前学术界主要有三个benchmark（数据集）用于模型训练和测试。

第一个常用的数据集是Pascal VOC系列。这个系列中目前较流行的是VOC2012，Pascal Context等类似的数据集也有用到。

第二个常用的数据集是Microsoft COCO。 COCO一共有80个类别，虽然有很详细的像素级别的标注，但是官方没有专门对语义分割的评测。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）。所以COCO数据集往往被当成是额外的训练数据集用于模型的训练。

第三个数据集是辅助驾驶（自动驾驶）环境的Cityscapes，使用比较常见的19个类别用于评测。

1、Pascal Voc 2012

标准的VOC2012数据集有21个类别(包括背景)，包含:{ 0=background，1=aeroplane, 2=bicycle, 3=bird, 4=boat, 5=bottle, 6=bus, 7=car , 8=cat, 9=chair, 10=cow, 11=diningtable, 12=dog, 13=horse, 14=motorbike, 15=person, 16=potted plant, 17=sheep, 18=sofa, 19=train, 20=tv/monitor，255= 'void' or unlabelled }这些比较常见的类别。VOC2012中用于分割的图片中，trainval包含2007-2011你那所有对应的图片，test只包含2008-2011年的图片。trainaug有10582张图片，trainval中有2913张图片，其中1464张用于训练，1449张用于验证，而测试集有1456张图片，测试集的label是不对外公布的，需要将预测的结果上传到Pascal Challenge比赛的测试服务器才可以计算MIoU的值。

1. MS COCO

COCO(Common Objects in Context)是一个新的图像识别、分割和图像语义数据集，是一个大规模的图像识别、分割、标注数据集。它可以用于多种竞赛，与本领域最相关的是检测部分，因为其一部分是致力于解决分割问题的。

该竞赛包含了超过80个物体类别，分别为：['background = 0','person=1', 'bicycle=2', 'car=3', 'motorcycle=4', 'airplane=5', 'bus=6', 'train=7', 'truck=8', 'boat=9', 'traffic light=10', 'fire hydrant=11', 'stop sign=13', 'parking meter=14', 'bench=15', 'bird=16', 'cat=17', 'dog=18', 'horse=19', 'sheep=20', 'cow=21', 'elephant=22', 'bear=23', 'zebra=24', 'giraffe=25', 'backpack=27', 'umbrella=28', 'handbag=31', 'tie=32', 'suitcase=33', 'frisbee=34', 'skis=35', 'snowboard=36', 'sports ball=37', 'kite=38', 'baseball bat=39', 'baseball glove=40', 'skateboard=41', 'surfboard=42', 'tennis racket=43', 'bottle=44', 'wine glass=46', 'cup=47', 'fork=48', 'knife=49', 'spoon=50', 'bowl=51', 'banana=52', 'apple=53', 'sandwich=54', 'orange=55', 'broccoli=56', 'carrot=57', 'hot dog=58', 'pizza=59', 'donut=60', 'cake=61', 'chair=62', 'couch=63', 'potted plant=64', 'bed=65', 'dining table=67', 'toilet=70', 'tv=72', 'laptop=73', 'mouse=74', 'remote=75', 'keyboard=76', 'cell phone=77', 'microwave=78', 'oven=79', 'toaster=80', 'sink=81', 'refrigerator=82', 'book=84', 'clock=85', 'vase=86', 'scissors=87', 'teddy bear=88', 'hair drier=89', 'toothbrush=90']。

91个填充类别，分别为['banner=92', 'blanket=93', 'branch=94', 'bridge=95', 'building-other=96', 'bush=97', 'cabinet=98', 'cage=99', 'cardboard=100', 'carpet=101', 'ceiling-other=102', 'ceiling-tile=103', 'cloth=104', 'clothes=105', 'clouds=106', 'counter=107', 'cupboard=108', 'curtain=109', 'desk-stuff=110', 'dirt=111', 'door-stuff=112', 'fence=113', 'floor-marble=114', 'floor-other=115', 'floor-stone=116', 'floor-tile=117', 'floor-wood=118', 'flower=119', 'fog=120', 'food-other=121', 'fruit=122', 'furniture-other=123', 'grass=124', 'gravel=125', 'ground-other=126', 'hill=127', 'house=128', 'leaves=129', 'light=130', 'mat=131', 'metal=132', 'mirror-stuff=133', 'moss=134', 'mountain=135', 'mud=136', 'napkin=137', 'net=138', 'paper=139', 'pavement=140', 'pillow=141', 'plant-other=142', 'plastic=143', 'platform=144', 'playingfield=145', 'railing=146', 'railroad=147', 'river=148', 'road=149', 'rock=150', 'roof=151', 'rug=152', 'salad=153', 'sand=154', 'sea=155', 'shelf=156', 'sky-other=157', 'skyscraper=158', 'snow=159', 'solid-other=160', 'stairs=161', 'stone=162', 'straw=163', 'structural-other=164', 'table=165', 'tent=166', 'textile-other=167', 'towel=168', 'tree=169', 'vegetable=170', 'wall-brick=171', 'wall-concrete=172', 'wall-other=173', 'wall-panel=174', 'wall-stone=175', 'wall-tile=176', 'wall-wood=177', 'water-other=178', 'waterdrops=179', 'window-blind=180', 'window-other=181', 'wood=182', 'other=183']。提供了118287张训练图片，5000张验证图片，以及超过40670张测试图片。由于其规模巨大，目前已非常常用，对领域发展很重要。实际上，该竞赛的结果每年都会在ECCV的研讨会上与ImageNet数据集的结果一起公布。它有如下特点：

* 1）Object segmentation：物体分割
* 2）Recognition in context ：上下文识别
* 3）Superpixel stuff segmentation：超分辨率的实物分割
* 4）330K images (>200K labeled)：33万张图片（超过20万有标记）
* 5）1.5 million object instances：150万个物体实例
* 6）80 object categories：80个物体类别
* 9）91 stuff categories ：91个stuff类别
* 10）5 captions per image：每张图像5个标题
* 11）250,000 people with keypoints：25万张带关节点的人物图片

COCO数据集由微软赞助，其对于图像的标注信息不仅有类别、位置信息，还有对图像的语义文本描述，COCO数据集的开源使得近两三年来图像分割语义理解取得了巨大的进展，也几乎成为了图像语义理解算法性能评价的“标准”数据集。详细介绍参考。window下安装COCO的API参考这里。注意COCO用于语义分割的API要从这里下载：https://github.com/nightrome/cocostuffapi

3、Cityscapes

Cityscapes数据集则是由奔驰主推，提供无人驾驶环境下的图像分割数据集，用于评估视觉算法在城区场景语义理解方面的性能。Cityscapes包含50个欧洲城市不同场景、不同背景、不同季节的街景的33类标注物体，包括：{'unlabeled'=0 , 'ego vehicle'=1 , 'rectification border'=2 , 'out of roi'= 3 , 'static'=4 , 'dynamic'=5 , 'ground'=6 ,'road'=7 ,'sidewalk'=8 ,parking'=9 ,'rail track'=10 ,'building'=11 ,'wall'=12 ,'fence'=13 , 'guard rail'=14 ,'bridge'=15 ,'tunnel'=16 ,'pole'=17 ,'polegroup'=18 , 'traffic light'=19 ,'traffic sign'=20 , 'vegetation'=21 , 'terrain'=22 ,'sky'=23 , 'person'=24 , 'rider'=25 , 'car'=26 ,'truck'=27 , 'bus'=28 ,'caravan'=29 ,'trailer'=30 ,'train'=31 ,'motorcycle'=32 , 'bicycle'=33 }，但是在这33个类中，评估时只用到了19个类别，因此训练时将33个类映射为19个类，评估时需要将19个类又映射回33个类上传评估服务器。这个数据需要注册账号才能下载。Cityscapes数据集共有fine和coarse两套评测标准，前者提供5000张精细标注的图像，后者提供5000张精细标注外加20000张粗糙标注的图像，用PASCAL VOC标准的 intersection-over-union （IoU）得分来对算法性能进行评价。 5000张精细标注的图片分为训练集2975张图片，验证集有500张图片，而测试集有1525张图片，测试集不对外公布，需要将预测结果上传到评估服务器才能计算mIoU值。

4、Pascal-Context

Pascal-Context数据集是对于PASCAL-VOC 2010识别竞赛的扩展，包含了对所有训练图像的像素级别的标注。共有540个类，包括原有的20个类及由PASCAL VOC分割数据集得来的图片背景，分为三大类，分别是物体、材料以及混合物。虽然种类繁多，但是只有59个常见类是较有意义的。由于其类别服从一个幂律分布，其中有很多类对于整个数据集来说是非常稀疏的。就这点而言，包含这59类的子集常被选作真实类别来对该数据集进行研究，其他类别一律重标为背景。

5、KITTI

KITTI是用于移动机器人及自动驾驶研究的最受欢迎的数据集之一，包含了由多种形式的传感器得出的数小时的交通场景数据，包括高分辨率RGB、灰度立体摄像机以及三维激光扫描器。尽管很受欢迎，该数据集本身并没有包含真实语义分割标注，但是，众多的研究者手工地为该数据集的部分数据添加标注以满足其问题的需求。Alvarez等人[36,37]为道路检测竞赛中的323张图片生成了真实标注，包含三个类别：道路、垂直面和天空。Zhang等人标注了252张图片，其中140张训练、112张测试，其选自追踪竞赛中的RGB和Velodyne扫描数据，共十个类。Ros等人在视觉测距数据集中标注了170个训练图片和46个测试图片，共11个类。

6、NYUDv2

NYUDv2是2.5维数据集，它包含1449张由微软Kinect设备捕获的室内的RGB-D图像。其给出密集的像素级别的标注（类别级别和实力级别的均有），训练集795张与测试集654张均有40个室内物体的类[60]，该数据集由于其刻画室内场景而格外重要，使得它可以用于某种家庭机器人的训练任务。但是，它相对于其他数据集规模较小，限制了其在深度网络中的应用。

7、SUN-RGBD

SUNRGBD数据集由四个RGB-D传感器得来，包含10000张RGB-D图像，尺寸与PASCAL VOC一致。该数据集包含了NYU depth v2 [46], Berkeley B3DO [61], 以及SUN3D [47]数据集中的图像，整个数据集均为密集标注，包括多边形、带方向的边界框以及三维空间，适合于场景理解任务。

8、ADE20K\_MIT

ADE20K是一个场景理解的新的数据集，这个数据集是可以免费下载的。它包含151个类别（包括背景），包括各种物体（比如人、汽车等）、场景（天空、路面等），150个类别如下：['background' = 0，'wall=1', 'building, edifice=2', 'sky=3', 'floor, flooring=4', 'tree=5', 'ceiling=6', 'road, route=7', 'bed =8', 'windowpane, window =9', 'grass=10', 'cabinet=11', 'sidewalk, pavement=12', 'person, individual, someone, somebody, mortal, soul=13', 'earth, ground=14', 'door, double door=15', 'table=16', 'mountain, mount=17', 'plant, flora, plant life=18', 'curtain, drape, drapery, mantle, pall=19', 'chair=20', 'car, auto, automobile, machine, motorcar=21', 'water=22 ', 'painting, picture=23', 'sofa, couch, lounge=24 ', 'shelf=25 ', 'house=26 ', 'sea=27 ', 'mirror=28', 'rug, carpet, carpeting=29', 'field=30', 'armchair=31', 'seat=32', 'fence, fencing=33', 'desk=34', 'rock, stone=35', 'wardrobe, closet, press=36', 'lamp=37', 'bathtub, bathing tub, bath, tub=38', 'railing, rail=39', 'cushion=40', 'base, pedestal, stand=41', 'box=42', 'column, pillar=43', 'signboard, sign=44', 'chest of drawers, chest, bureau, dresser=45', 'counter=46', 'sand=47', 'sink=48', 'skyscraper=49', 'fireplace, hearth, open fireplace=50', 'refrigerator, icebox=51', 'grandstand, covered stand=52', 'path=53', 'stairs, steps=54', 'runway=55', 'case, display case, showcase, vitrine=56', 'pool table, billiard table, snooker table=57', 'pillow=58', 'screen door, screen=59', 'stairway, staircase=60', 'river=61', 'bridge, span=62', 'bookcase=63', 'blind, screen=64', 'coffee table, cocktail table=65', 'toilet, can, commode, crapper, pot, potty, stool, throne=66', 'flower=67', 'book=68', 'hill=69', 'bench=70', 'countertop=71', 'stove, kitchen stove, range, kitchen range, cooking stove=72', 'palm, palm tree=73', 'kitchen island=74', 'computer, computing machine, computing device, data processor, electronic computer, information processing system=75', 'swivel chair=76', 'boat=77', 'bar=78', 'arcade machine=79', 'hovel, hut, hutch, shack, shanty=80', 'bus, autobus, coach, charabanc, double-decker, jitney, motorbus, motorcoach, omnibus, passenger vehicle=81', 'towel=82', 'light, light source=83', 'truck, motortruck=84', 'tower=85', 'chandelier, pendant, pendent=86', 'awning, sunshade, sunblind=87', 'streetlight, street lamp=88', 'booth, cubicle, stall, kiosk=89', 'television receiver, television, television set, tv, tv set, idiot box, boob tube, telly, goggle box=90', 'airplane, aeroplane, plane=91', 'dirt track=92', 'apparel, wearing apparel, dress, clothes=93', 'pole=94', 'land, ground, soil=95', 'bannister, banister, balustrade, balusters, handrail=96', 'escalator, moving staircase, moving stairway=97', 'ottoman, pouf, pouffe, puff, hassock=98', 'bottle=99', 'buffet, counter, sideboard=100', 'poster, posting, placard, notice, bill, card=101', 'stage=102', 'van=103', 'ship=104', 'fountain=105', 'conveyer belt, conveyor belt, conveyer, conveyor, transporter=106', 'canopy=107', 'washer, automatic washer, washing machine=108', 'plaything, toy=109', 'swimming pool, swimming bath, natatorium=110', 'stool=111', 'barrel, cask=112', 'basket, handbasket=113', 'waterfall, falls=114', 'tent, collapsible shelter=115', 'bag=116', 'minibike, motorbike=117', 'cradle=118', 'oven=119', 'ball=120', 'food, solid food=121', 'step, stair=122', 'tank, storage tank=123', 'trade name, brand name, brand, marque=124', 'microwave, microwave oven=125', 'pot, flowerpot=126', 'animal, animate being, beast, brute, creature, fauna=127', 'bicycle, bike, wheel, cycle =128', 'lake=129', 'dishwasher, dish washer, dishwashing machine=130', 'screen, silver screen, projection screen=131', 'blanket, cover=132', 'sculpture=133', 'hood, exhaust hood=134', 'sconce=135', 'vase=136', 'traffic light, traffic signal, stoplight=137', 'tray=138', 'ashcan, trash can, garbage can, wastebin, ash bin, ash-bin, ashbin, dustbin, trash barrel, trash bin=139', 'fan=140', 'pier, wharf, wharfage, dock=141', 'crt screen=142', 'plate=143', 'monitor, monitoring device=144', 'bulletin board, notice board=145', 'shower=146', 'radiator=147', 'glass, drinking glass=148', 'clock=149', 'flag=150']。

它的训练集由20210张场景图片组成，验证集由2000张图片构成，测试集有3352张图片组成。color\_map文件位于：https://github.com/CSAILVision/semantic-segmentation-pytorch

1. 三个人像分割（抠图）的数据集

(1)第一个是Baidu People segmentation dataset，下载地址，这个数据集主要是用于人体整体分割。它由5387张训练图片组成，但是测试图片没有公布。因此训练时可以从5387中随机挑选500张作为验证集，然后4887张作为训练集。参考论文《Early Hierarchical Contexts Learned by CNN for image segmentation》。

(2)第二个是由论文《Automatic Portrait Segmentation for Image Stylization》的作者制作的用于大头贴分割的数据集Portrait Segmentation dataset，下载地址在这里，这个数据集主要是用于图像中大头贴的分割。这个数据集由1700张训练集组成，由300张测试集组成。

(3)此外最近supervise.ly公布了一个人像分割数据集，这个数据集规模较大，由5711张图片组成，目前还没用过。

10、人体关节分割的数据集

这个数据集是由中山大学的一个实验室创建的，可以对人体的各部分关节进行实例级别的分割，包括单人实例分割、多人实例分割以及在视频中的人体实例级别的分割。

下载地址：www.sysu-hcp.net/lip/overview.php