

# 图像分割

# Image Segmentation

- 图像分割 (image segmentation) : 根据某些规则将图片分成若干个特定的、具有独特性质的区域, 并抽取出感兴趣目标。
- 目前图像分割任务发展出了以下几个子领域:
  - 语义分割 (semantic segmentation)
  - 实例分割 (instance segmentation)
  - 以及刚兴起的新领域: 全景分割 (panoptic segmentation)

# 语义分割 semantic segmentation

- 从像素层次来识别图像; 为图像中的每个像素指定类别标记



dense prediction



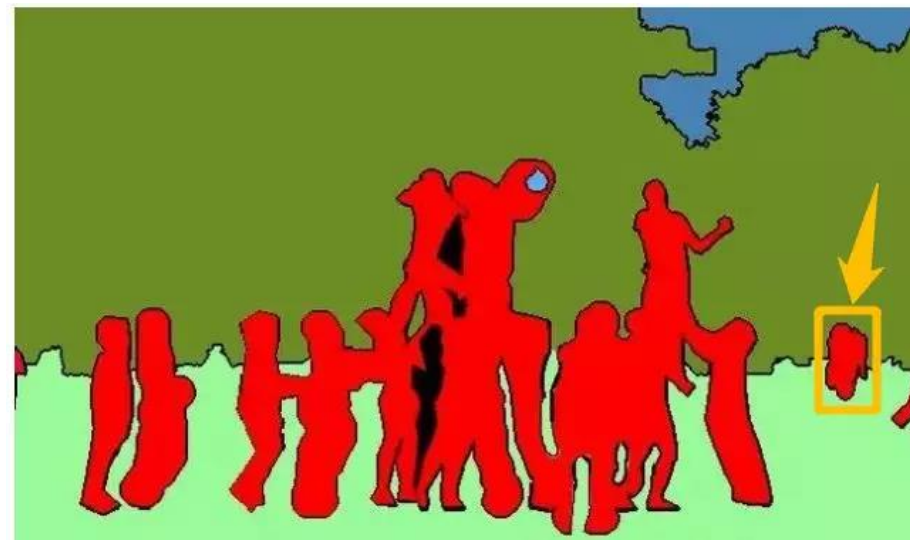
Person  
Bicycle  
Background

label *each pixel* of an image with a corresponding **class** of what is being represented  
(用相应的表示的类别来标记图像每个像素)

原图



语义分割



实例分割



全景分割

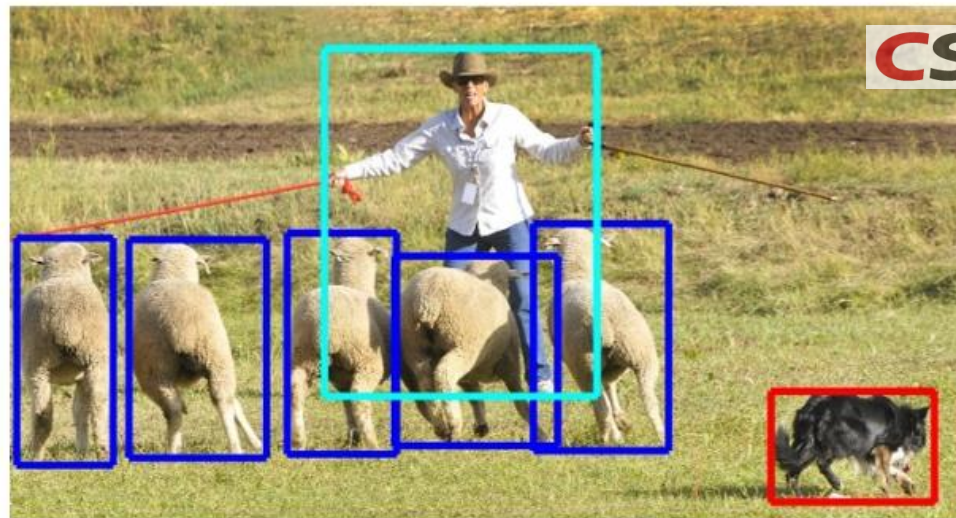


Things  
and  
stuff

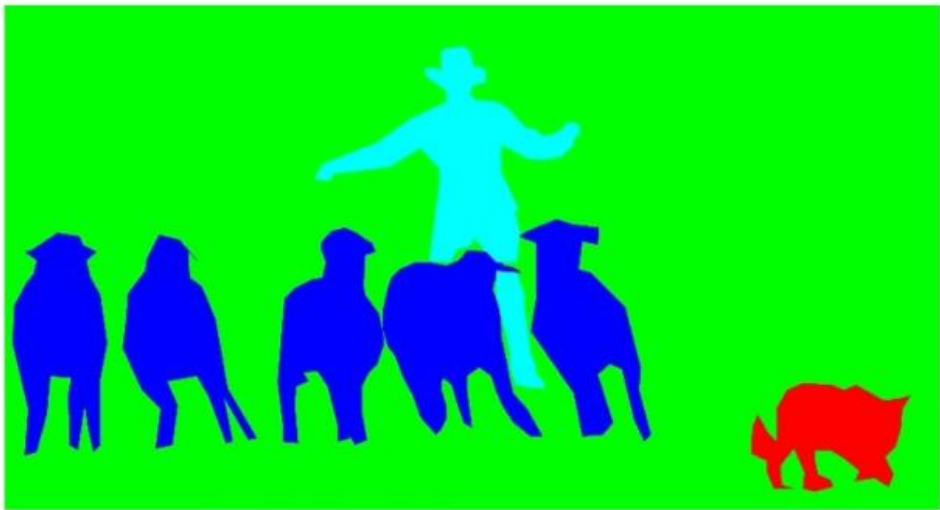




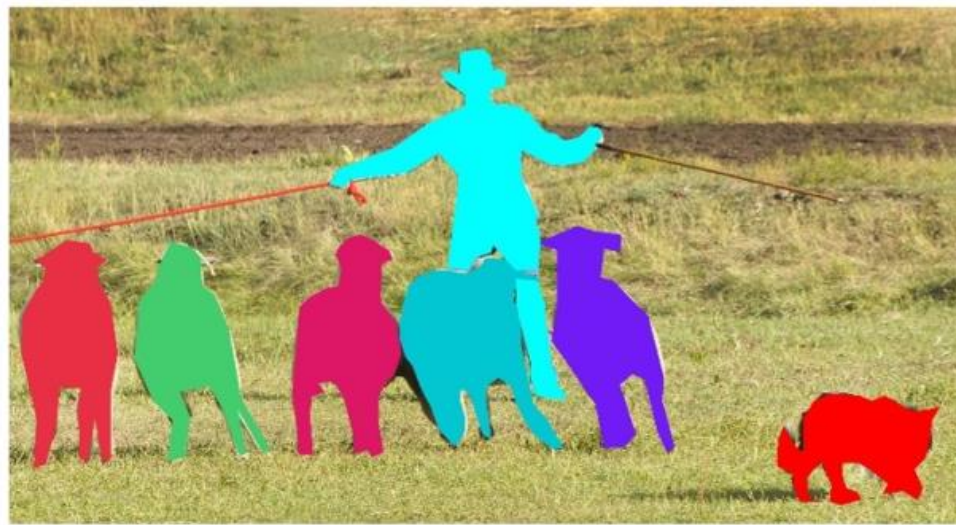
图像分类 (classification)



目标检测 (object detection)



语义分割 (semantic segmentation)



实例分割 (instance segmentation)

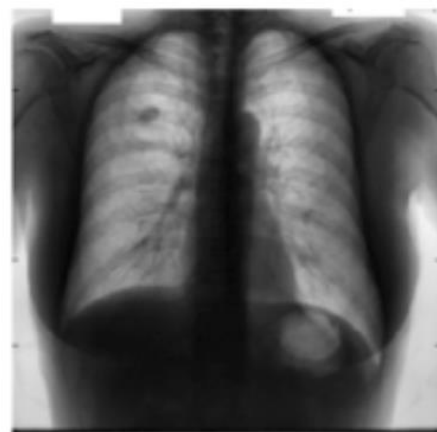
- 实例分割具有挑战性，因为它需要正确检测图像中的目标，同时还要精确地分割每个实例

# 图像分割应用

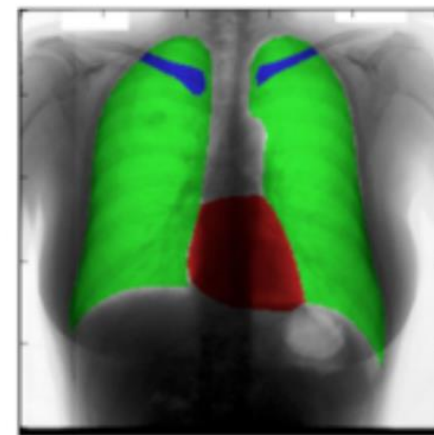
自动驾驶  
Autonomous vehicles



医学图像诊断  
Medical image diagnostics



Input Image



Segmented Image



# 语义分割任务表达



Input



- 1: Person
- 2: Purse
- 3: Plants/Grass
- 4: Sidewalk
- 5: Building/Structures

3	3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	1	1	1	1	3	3	3	5	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	5	5	5	5	5	5	5	5
5	5	3	3	3	3	1	1	3	3	5	5	5	5	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	5	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	4	4	5	5	5	5
4	4	4	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4	4
3	3	3	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	1	4	4	4	4	4	4	4	4

Semantic Labels



0: Background/Unknown

1: Person

2: Purse

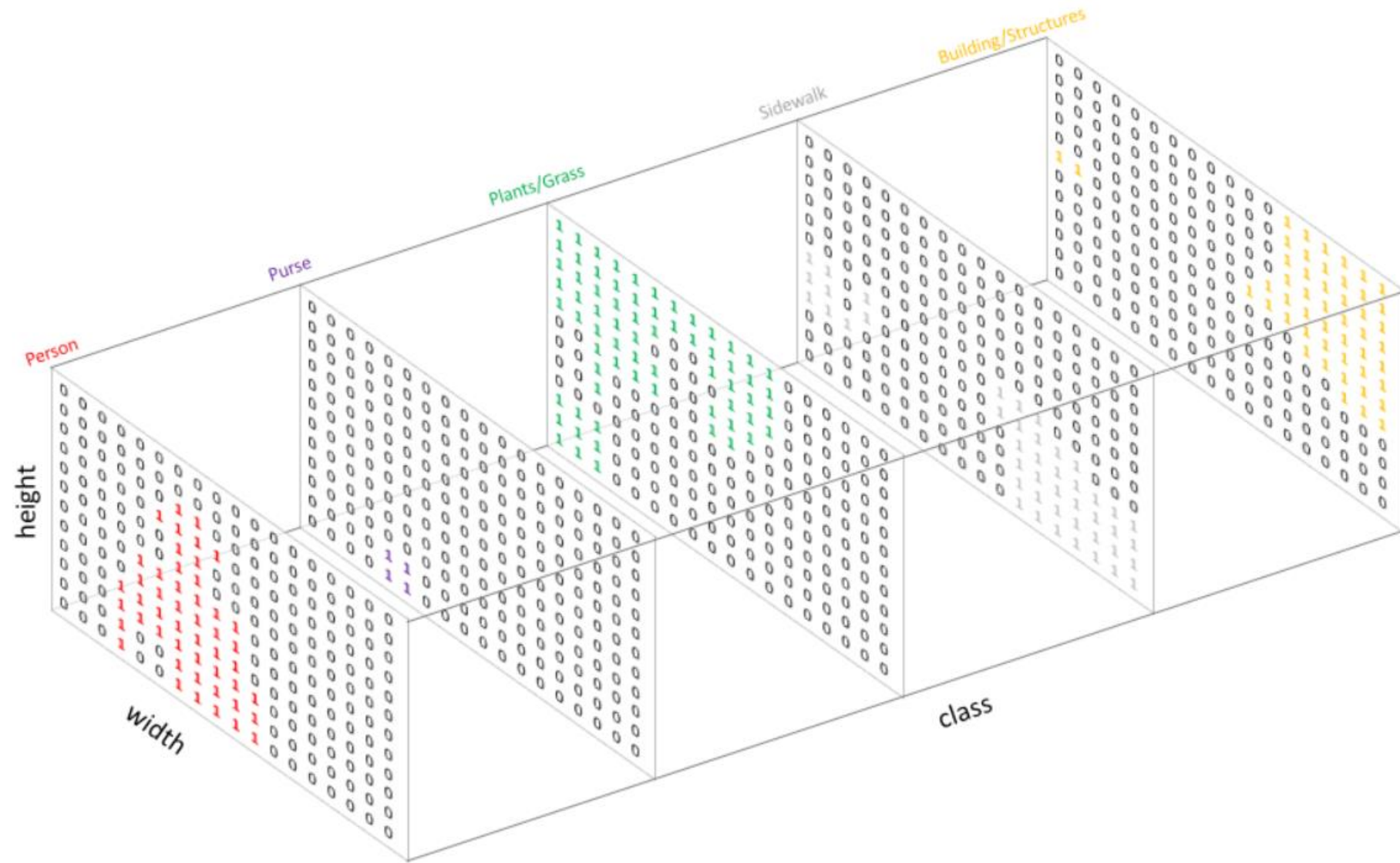
3: Plants/Grass

4: Sidewalk

5: Building/Structures

Mask(掩模): 覆盖预测单个通道, 表示图像中存在特定类别的区域





# 常用数据集

- PASCAL VOC 2012 Segmentation Competition
- COCO 2018 Stuff Segmentation Task
- BDD100K: A Large-scale Diverse Driving Video Database
- Cambridge-driving Labeled Video Database (CamVid)
- Cityscapes Dataset
- Mapillary Vistas Dataset
- ApolloScape Scene Parsing

# PASCAL VOC 2012

PASCAL: pattern analysis, statistical modelling and computational learning

VOC: visual object classes

VOC2012数据集分为20类，包括背景为21类，分别如下：

- 人：人
- 动物：鸟、猫、牛、狗、马、羊
- 车辆：飞机、自行车、船、巴士、汽车、摩托车、火车
- 室内：瓶、椅子、餐桌、盆栽植物、沙发、电视/监视器

挑战任务：

1. Classification/Detection Competitions
2. Segmentation Competition
3. Action Classification Competition
4. ImageNet Large Scale Visual Recognition Competition
5. Person Layout Taster Competition

官网链接：<http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html>



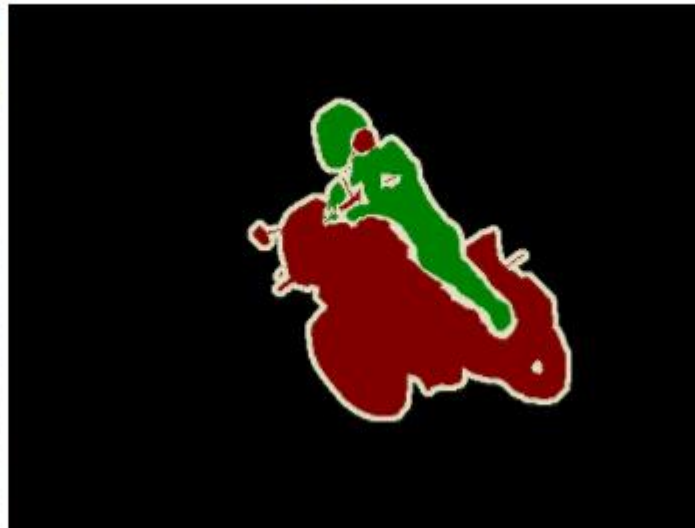
## Segmentation Competition

- **Segmentation:** Generating pixel-wise segmentations giving the class of the object visible at each pixel, or "background" otherwise.

Image



Objects



Class



# COCO

- MS COCO的全称是Microsoft Common Objects in Context, 起源于是微软于2014年出资标注的Microsoft COCO数据集, 与ImageNet 竞赛一样, 被视为是计算机视觉领域最受关注和最权威的比赛之一。
- 而在ImageNet竞赛停办后, COCO竞赛就成为是当前物体识别、检测等领域的一个最权威、最重要的标杆, 也是目前该领域在国际上唯一能汇集Google、微软、Facebook以及国内外众多顶尖院校和优秀创新企业共同参与的大赛。

<http://cocodataset.org/>

目前为止有语义分割的最大数据集，提供的类别有 80 类，有超过 33 万张图片，其中 20 万张有标注，整个数据集中个体的数目超过 150 万个。





## BDD100K: A Large-scale Diverse Driving Video Database

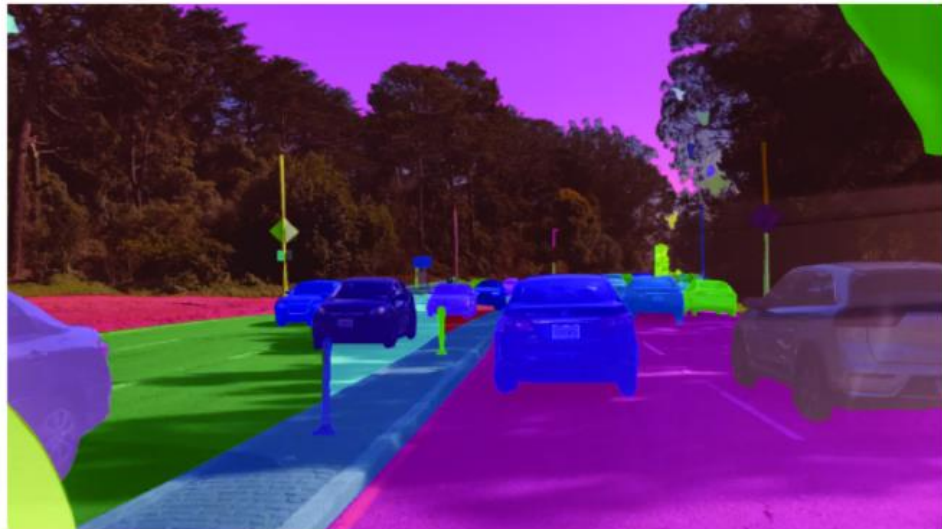
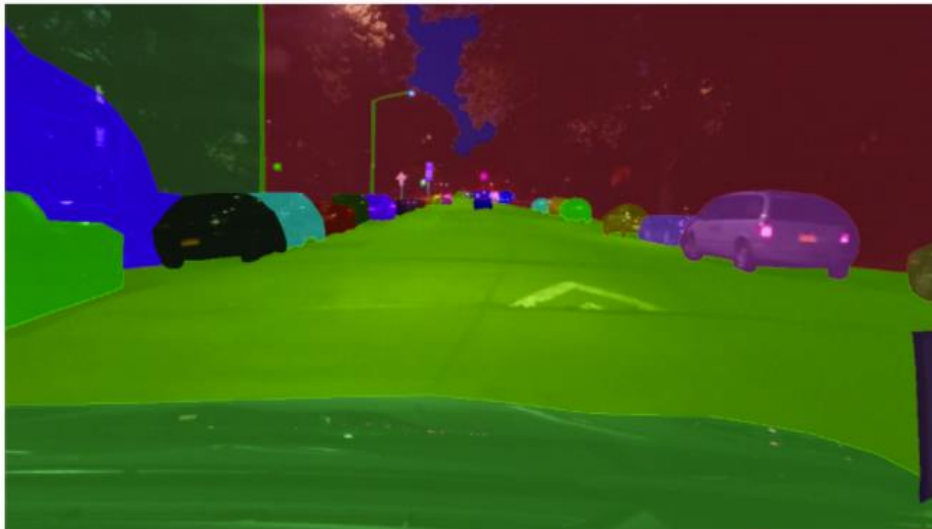
<https://bair.berkeley.edu/blog/2018/05/30/bdd/>

2018年5月伯克利大学AI实验室（BAIR）发布了目前最大规模、内容最具多样性的公开驾驶数据集 BDD100K，同时设计了一个图片标注系统。

BDD100K 数据集包含10万段高清视频，每个视频约40秒，720p，30 fps。每个视频的第10秒对关键帧进行采样，得到10万张图片（图片尺寸：1280×720），并进行标注。

	KITTI	Cityscapes	ApolloScape	Mapillary	BDD100K
# Sequences	22	~50	4	N/A	100,000
# Images	14,999	5000 (+2000)	143,906	25,000	120,000,000
Multiple Cities	No	Yes	No	Yes	Yes
Multiple Weathers	No	No	No	Yes	Yes
Multiple Times of Day	No	No	No	Yes	Yes
Multiple Scene types	Yes	No	No	Yes	Yes

# Full-frame Instance Segmentation

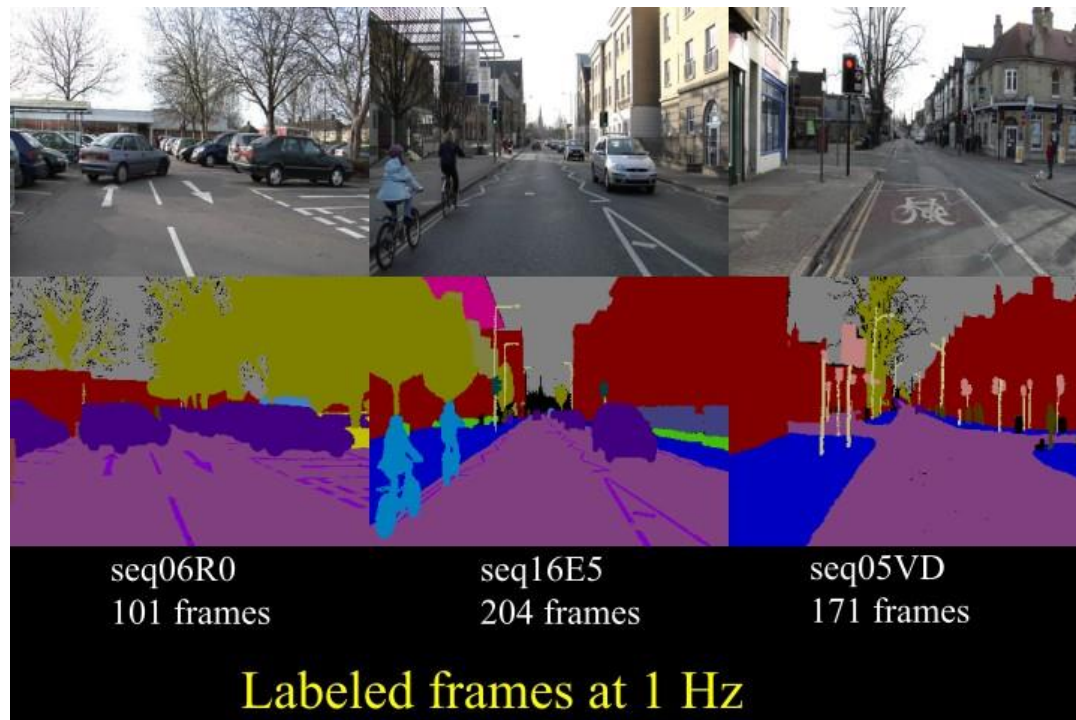


# Cambridge-driving Labeled Video Database (CamVid)

<http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/>

CamVid是第一个具有目标类别语义标签的视频集合。数据库提供32个ground truth语义标签，将每个像素与语义类别之一相关联。

该数据库解决了对实验数据的需求，以定量评估新兴算法。数据是从驾驶汽车的角度拍摄的。







包含戴姆勒在内的三家德国单位联合提供，包含50多个城市的立体视觉数据；像素级标注；提供算法评估接口。



# Mapillary Vistas

Mapillary Vistas是世界上最大最多样化的像素精确和特定实例标注的街道级图像公开数据集。



# ApolloScape Scene Parsing

- 百度公司提供的ApolloScape数据集将包括具有高分辨率图像和每像素标注的RGB视频，具有语义分割的测量级密集3D点，立体视频和全景图像。
- Scene Parsing数据集是ApolloScape的一部分，它为高级自动驾驶研究提供了一套工具和数据集。场景解析旨在为图像中的每个像素或点云中的每个点分配类别（语义）标签。



	KITTY	Citycapes	Mapillary	BDD100K	ApolloScape			
原始的数据量	14,999	250,00	25,000	120,000,000	143,906			
标注的数据量 (矩形框级别)	14,999	无	无	100,000	无			
标注的数据量 (像素级别)	400	5,000精确 20,000粗略	25,000	10,000	143,906			
场景复杂度 (每张图平均)	矩形框级别	像素级别	像素级别	矩形框级别	像素级别			
	人： 0.8 车： 4.1	人： 7.0 车： 11.8	未知	人： ~1.3 车： ~11.0	难度	易	中	难
					人	1.1	6.2	16.9
					车	12.7	24.0	38.1
多样性	好/中等天气条件 白天	50个城市 好/中等天气条件 白天	6大洲 多样的天气条件 白天和夜晚	美国的4个区域 多样的天气条件 白天和夜晚	中国2个城市的4个区域 多样的天气条件 白天 独特的Tricycle类别			
三维信息	有（三维矩形框级别）	无	无	无	有(三维点级别)			
视频标注	有（矩形框级别）	无	无	无	有（像素级别）			
车道线信息	无	无	有（仅二维） 2类	有（仅二维） 8类	有（二维和三维） 28类			
地理定位	厘米级精度	未知	城市级别精度	米级精度	厘米级精度			