

最近更新：2019年9月20日

# Panoptic Segmentation

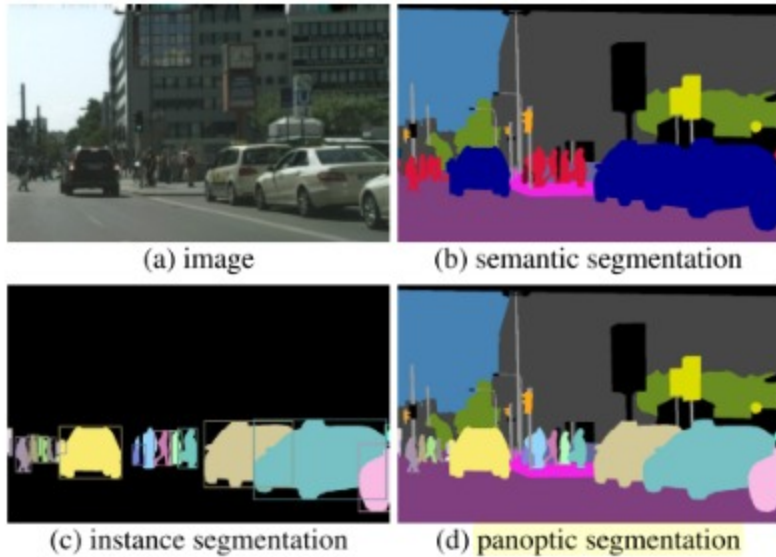
CVPR 2019 全景分割

**1. Introduction**

语义分割通过带孔全卷积网络，根据不同的stuff进行划分；实例分割则是在目标检测的基础上基于检测框进行物体的分割。然而没有一种框架可以将两者进行融合实现既能分割背景又能分割实例，而这在自动驾驶和AR技术中大有作为。由此提出全景分割任务将两者进行结合。

全景分割的思路很直观：为图像的每个像素分配语义label和类内实例id，前者用于区分语义信息，后者用于分割实例。提出全景分割时，只是启发式地将语意分割和实例分割两种任务的输出进行后处理的融合（如NMS），并以此建立PS任务的baseline。为了评价全景分割的质量，提出panoptic quality (PQ) 标准，将背景和物体的评价纳入一个完整的框架下。

全景分割区别于语义分割的特点是其能够细致划分出每个物体；区别于实例分割的是，全景分割得到的mask是互不重叠的。示意图如下：



**2. Panoptic Segmentation**

- Task format**  
全景分割的标注方法：  
像素级的标注，标出类别label和类内实例id。如果某像素的这两个信息都能匹配，则可以将该像素匹配到某个类别和实例中去；  
类外的像素可以分配空标签，即并不是所有的像素都要有语义类别。
- Stuff and thing labels**  
对于stuff和thing（背景填充和物体实例）的标签，分别采用 $L^{St}$ 和 $L^{Th}$ 表示，两者的交集是空集，并集是所有可能的label空间。这两者是互相独立不相关的（很好理解，像素属于那个类和它属于哪个物体不具有相关性）。具体而言：对于背景而言，可以拥有不同的class标签，但是属于同一个物体；而对于实例而言不仅区分class还要区分实例id。
- Relationship**  
Semantic segmentation：  
都是像素级的分类，需要为每个像素分配对应的标签。  
Instance segmentation：  
实例分割基于region的，允许重叠的segmentation，而全景分割和语义分割一样是像素级的label，不允许重叠标签的出现。
- Confidence scores**  
这一点上更像语义分割而不是实例分割，对于PS不需要置信分数评价分割质量。提到这个，作者认为语义分割和全景分割可以直接利用人工标注的label进行对比从而评价当前mask的质量；而实例分割在选择mask时评价的是分类置信度，这个并没有人工标注进行参考，因此难以把握。

**3. Panoptic Segmentation Metric**

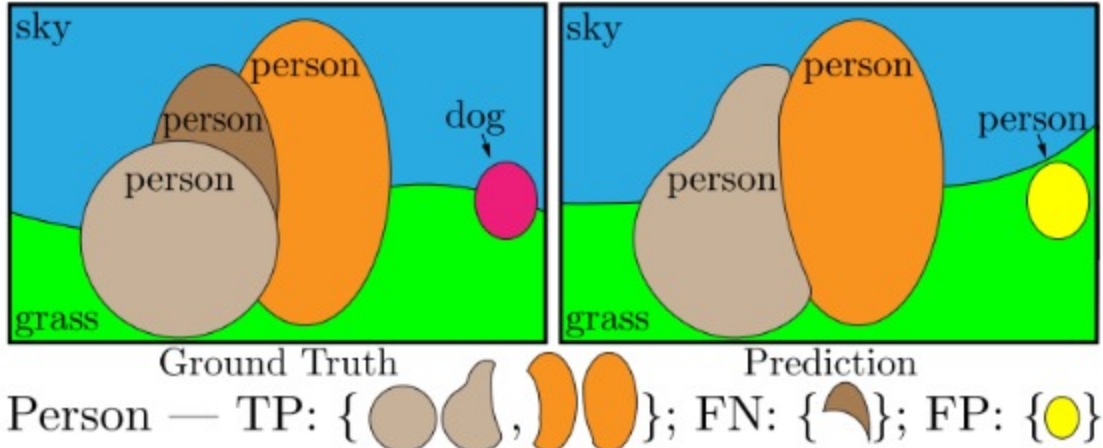
用于衡量全景分割效果的指标应具有：完备性；可解释性；简洁性。由是提出了PQ指标。

**3.1 Segment Matching**

定义match：预测的segmentation和gt的iou大于0.5，说明两者can match。再结合全景分割的不可重叠性，不难得到：最多只有一个预测的segmentation可以match gt。

**3.2 PQ Computation**

PQ的计算类似mAP，也是类内求取，然后求类间的平均值，以便不敏感类别不平衡。对于每一类，可以根据gt与预测的segmentation分为三类（下图描述）：



通过上述三类可以计算得到PQ值公式：

$$PQ = \frac{\sum_{(p,g) \in TP} IoU(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

式中出去FP与FN后，剩下的式子描述的是match的segmentation的平均IoU，加上FP与FN是为了惩罚match失败的分割实例。

有意思的是，对上述式子进行简单的恒等变化：

$$PQ = \underbrace{\frac{\sum_{(p,g) \in TP} IoU(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}$$

第一项评价的是match分割的质量，第二项类似于F1得分。

- Void labels**  
gt中可能出现两种像素标注为空的情况：超出类别的像素和模糊不清的像素（难以分类）。在评估结果时，这些空的标签不予以评估。具体而言：  
(1) 在matching部分，预测出为void的像素会被移出prediction并不参与IoU计算；  
(2) matching后，unmatched prediction按照一般情况会计算FP FN，但是对于空标签情况，如果该prediction含有的void像素块超过一定匹配阈值就会被移除，并不算作FP计算得分。
- Group labels**  
有时区分相同语义类别的实例个体标注比较困难，因此有提出组标签的标注方法。但对于PQ计算而言：  
(1) matching部分不使用组标签，而是严格区分实例  
(2) matching后，对于包含一部分相同类别像素点的unmatched predicted segments，这一部分将被去除并不视作false positives

**3.3 Comparison to Existing Metrics**

- Semantic segmentation metrics：  
衡量语义分割的标准有像素级精度，平均精度，IoU。但是其只专注于像素级的划分，不能反映物体实例级别的分割性能。
- Instance segmentation metrics：  
度量为AP，主要是引入了置信度分数confidence score对检测目标进行打分。（个人理解：其实两者不是完全的隔绝，实例分割也有用IoU监督的，而confidence score是否能够反映mask的真实质量也遭受过质疑，这个标准也不是固定的，需要多思考，存在，不一定合理。）
- Panoptic quality  
PQ的度量可以分解成SQ和RQ，SQ反映了语义分割的像素级IoU性能，RQ专注于检测识别的效果，因此将两者统一到一个框架下。

分割效果：

