**一、基础作业内容**

1. 什么是图像分割？简单说明其用途

就是把图像分成若干个特定的、具有独特性质的区域:根据灰度、颜色、纹理,形状以及语义等特征把图像划分成若干交集为空集的区域,并集为全集，而且是有限覆盖,这些特征在同一集合内呈现出相似性，而在不同区域间呈现出明显的差异性。其实就是 对图像上的每一个像素点分类

1. 传统图像分割常见算法有哪几种？

 阈值分割法

 区域生长方法

 分裂合并方法

 能量方法：蛇模型、水平集方法

 聚类分割方法

 图割

1. 阅读Deep Snake文献，说明算法原理和思想。

因为maskrcnn为代表的逐像素的分割有很多缺点,可以用contour来表示物体的shape的方法解决,所以作者不由回想起snake这样的contour经典工作.

Deep Snake，其实就是用深度学习的方法实现了传统snake。传统snake做图像分割的时候要求先给定一个initial contour。这个contour大概围绕着目标物体。

传统snake的一个很大问题是他的目标函数和optimization都是handcrafted的，对数据的噪声比较敏感，容易收敛到局部最优点。为了解决这个问题，Deep Snake用deep learning来做这个优化过程。

图示

描述已自动生成

主要idea:给定初始轮廓，在每个顶点(a)处提取图像特征。由于轮廓是一个循环图，因此在轮廓(b)上使用循环卷积进行特征学习。蓝色、黄色和绿色节点分别表示输入特征、循环卷积核和输出特征。最后，在每个顶点上回归偏移量，使轮廓变形以匹配对象边界(c)。

手机屏幕截图

低可信度描述已自动生成

(a) Deep Snake由三部分组成:backbone、a fusion block, and a prediction head。它以一个轮廓作为输入，并输出面向顶点的偏移量来对轮廓进行变形。(b)在Deep Snake的基础上，我们提出了两阶段的管道分割实例:初始轮廓提取和轮廓变形。检测器所提出的方框给出了一个菱形轮廓，其四个顶点通过Deep Snake移动到物体的极值点。一个八边形是基于极值点构造的。Deep Snake以八边形为初始轮廓，对其进行迭代变形，以匹配目标边界。

Reference: https://zhuanlan.zhihu.com/p/134111177

**二、选做内容**

1. 查阅基于深度学习的分割算法，查阅两个篇文献，分别写出算法的主要思想。

The surprising impact of mask-head architecture on novel class segmentation

ICCV 2021

Vighnesh Birodkar, Zhichao Lu, Siyang Li, Vivek Rathod, Jonathan Huang

今天的实例分割模型在大型注释数据集上训练时非常准确，但大规模地收集掩码注释代价高昂。我们解决了部分监督的实例分割问题，在这个问题中，我们可以训练所有类别的(非常便宜的)包围盒，但只对类别的子集使用掩码。在这项工作中，我们关注的是一个流行的模型家族，它将可区分的裁剪应用到一个feature map，并根据产生的裁剪预测一个mask。在这个家族下，我们研究了Mask R-CNN，并发现它的默认策略是结合建议和groundtruth box来训练mask-head，而仅仅使用groundtruth box来训练mask-head，大大提高了它在新类上的性能。这种训练策略还允许我们利用可选的mask-head架构，我们利用明显更深的现成架构(如ResNet、Hourglass模型)来替代2-4层的典型mask-head。虽然在完全监督模式下训练时，许多架构的表现类似，但我们的主要发现是，它们可以以截然不同的方式推广到新类。我们将mask-head泛化到不可见类的能力称为 strong mask generalization effect，并表明在没有任何特定模块或损失的情况下，我们可以在部分监督COCO实例分割基准测试中取得最先进的结果。最后，我们证明了我们的效果是普遍的，适用于底层检测方法(包括基于anchor、无anchor或根本没有检测器)和不同的backbone网。代码和预先训练的模型可以在https://git.io/deepmac上找到。

BEiT: BERT Pre-Training of Image Transformers

我们引入了一个自监督的视觉表示模型BEiT，它代表了来自vision transformer的双向编码器表示。在自然语言处理领域开发的BERT之后，我们提出了一个masked的图像建模任务来预训练vision transformer。具体来说，在我们的预训练中，每幅图像都有两个视图，即图像斑块（如16x16像素），和视觉标记（即离散的标记）。我们首先将原始图像 "标记化 "为视觉标记。然后，我们随机屏蔽一些图像斑块，并将其送入骨干转化器。预训练的目的是根据被破坏的图像斑块来恢复原始视觉标记。在对BEiT进行预训练后，我们通过在预训练的编码器上添加任务层，直接对下游任务的模型参数进行微调。图像分类和语义分割的实验结果表明，我们的模型取得了与以前的预训练方法有竞争力的结果。例如，基础尺寸的BEiT在ImageNet-1K上取得了83.2%的最高准确率，在同样的设置下，明显超过了从头开始的DeiT训练（81.8%）。此外，大尺寸的BEiT仅在ImageNet-1K上就获得了86.3%的准确率，甚至超过了在ImageNet-22K上进行了监督预训练的ViT-L（85.2%）。代码和预训练的模型可在https://aka.ms/beit。