

注：此文档来源于网络，仅供同行分享学习使用，如有侵权，请联系删除！联系方式：[coolens@coolens.cn](mailto:coolens@coolens.cn)

## 深度学习中的图像分割：方法和应用

### 什么是图像分割？

图像分割是计算机视觉中的一个关键过程。它包括将视觉输入分割成片段以简化图像分析。片段表示目标或目标的一部分，并由像素集或“超像素”组成。图像分割将像素组织成更大的部分，消除了将单个像素作为观察单位的需要。图像分析有三个层次：

- **分类** - 将整幅图片分成“人”、“动物”、“户外”等类别
- **目标检测** - 检测图像中的目标并在其周围画一个矩形，例如一个人或一只羊。
- **分割** - 识别图像的部分，并理解它们属于什么对象。分割是进行目标检测和分类的基础。



## 语义分割 vs. 实例分割

在分割过程本身，有两个粒度级别：

- **语义分割** - 将图像中的所有像素划分为有意义的对象类。这些类是“语义上可解释的”，并对应于现实世界的类别。例如，你可以将与猫相关的所有像素分离出来，并将它们涂成绿色。这也被称为 **dense** 预测，因为它预测了每个像素的含义。



PREDICT  
→



● Person   ● Bicycle   ● Background

华为云社区

- **实例分割** - 标识图像中每个对象的每个实例。它与语义分割的不同之处在于它不是对每个像素进行分类。如果一幅图像中有三辆车，语义分割将所有的车分类为一个实例，而实例分割则识别每一辆车。

## 传统的图像分割方法

还有一些过去常用的图像分割技术，但效率不如深度学习技术，因为它们使用严格的算法，需要人工干预和专业知识。这些包括：

- 

**阈值** - 将图像分割为前景和背景。指定的阈值将像素分为两个级别之一，以隔离对象。阈值化将灰度图像转换为二值图像或将彩色图像的较亮和较暗像素进行区分。

- 

- 

**K-means 聚类** - 算法识别数据中的组，变量 **K** 表示组的数量。该算法根据特征相似性将每个数据点(或像素)分配到其中一组。聚类不是分析预定义的组，而是迭代地工作，从而有机地形成组。

- 

- 

**基于直方图的图像分割** - 使用直方图根据“灰度”对像素进行分组。简单的图像由一个对象和一个背景组成。背景通常是一个灰度级，是较大的实体。因此，一个较大的峰值代表了直方图中的背景灰度。一个较小的峰值代表这个物体，这是另一个灰色级别。

- 

- 

**边缘检测** - 识别亮度的急剧变化或不连续的地方。边缘检测通常包括将不连续点排列成曲线线段或边缘。例如，一块红色和一块蓝色之间的边界。

- 

## 深度学习如何助力图像分割方法

现代图像分割技术以深度学习技术为动力。下面是几种用于分割的深度学习架构：

使用 CNN 进行图像分割，是将图像的 patch 作为输入输入给卷积神经网络，卷积神经网络对像素进行标记。CNN 不能一次处理整个图像。它扫描图像，每次看一个由几个像素组成的小“滤镜”，直到它映射出整个图像。

传统的 cnn 网络具有全连接的层，不能处理不同的输入大小。FCNs 使用卷积层来处理不同大小的输入，可以工作得更快。最终的输出层具有较大的感受野，对应于图像的高度和宽度，而通道的数量对应于类的数量。卷积层对每个像素进行分类，以确定图像的上下文，包括目标的位置。

**集成学习** 将两个或两个以上相关分析模型的结果合成为单个。集成学习可以提高预测精度，减少泛化误差。这样就可以对图像进行精确的分类和分割。通过集成学习尝试生成一组弱的基础学习器，对图像的部分进行分类，并组合它们的输出，而不是试图创建一个单一的最优学习者。

**DeepLab** 使用 DeepLab 的一个主要动机是在帮助控制信号抽取的同时执行图像分割——减少样本的数量和网络必须处理的数据量。另一个动机是启用多尺度上下文特征学习——从不同尺度的图像中聚合特征。DeepLab 使用 ImageNet 预训练的 ResNet 进行特征提取。DeepLab 使用空洞卷积而不是规则的卷积。每个卷积的不同扩张率使 ResNet 块能够捕获多尺度的上下文信息。DeepLab 由三个部分组成：

- **Atrous convolutions** — 使用一个因子，可以扩展或收缩卷积滤波器的视场。
- **ResNet** — 微软的深度卷积网络(DCNN)。它提供了一个框架，可以在保持性能的同时训练数千个层。ResNet 强大的表征能力促进了计算机视觉应用的发展，如物体检测和人脸识别。
- **Atrous spatial pyramid pooling (ASPP)** — 提供多尺度信息。它使用一组具有不同扩展率的复杂函数来捕获大范围的上下文。ASPP 还使用全局平均池(GAP)来合并图像级特征并添加全局上下文信息。

**SegNet neural network** 一种基于深度编码器和解码器的架构，也称为语义像素分割。它包括对输入图像进行低维编码，然后在解码器中利用方向不变性能力恢复图像。然后在解码器端生成一个分割图像。

## 图像分割的应用

图像分割有助于确定目标之间的关系，以及目标在图像中的上下文。应用包括人脸识别、车牌识别和卫星图像分析。例如，零售和时尚等行业在基于图像的搜索中使用了图像分割。自动驾驶汽车用它来了解周围的环境。

## 目标检测和人脸检测

这些应用包括识别数字图像中特定类的目标实例。语义对象可以分类成类，如人脸、汽车、建筑物或猫。

- **人脸检测** - 一种用于许多应用的目标检测，包括数字相机的生物识别和自动对焦功能。算法检测和验证面部特征的存在。例如，眼睛在灰度图像中显示为谷地。
- **医学影像** - 从医学影像中提取临床相关信息。例如，放射学家可以使用机器学习来增强分析，通过将图像分割成不同的器官、组织类型或疾病症状。这可以减少运行诊断测试所需的时间。
- **机器视觉** - 捕捉和处理图像，为设备提供操作指导的应用。这包括工业和非工业的应用。机器视觉系统使用专用摄像机中的数字传感器，使计算机硬件和软件能够测量、处理和分析图像。例如，检测系统为汽水瓶拍照，然后根据合格 - 不合格标准分析图像，以确定瓶子是否被正确地填充。

## 视频监控 — 视频跟踪和运动目标跟踪

这涉及到在视频中定位移动物体。其用途包括安全和监视、交通控制、人机交互和视频编辑。

- **自动驾驶** 自动驾驶汽车必须能够感知和理解他们的环境，以便安全驾驶。相关类别的对象包括其他车辆、建筑物和行人。语义分割使自动驾驶汽车能够识别图像中的哪些区域可以安全驾驶。
- **虹膜识别** 一种能识别复杂虹膜图案的生物特征识别技术。它使用自动模式识别来分析人眼的视频图像。
- **人脸识别** 从视频中识别个体。这项技术将从输入图像中选择的面部特征与数据库中的人脸进行比较。

## 零售图像识别

这个应用让零售商了解货架上商品的布局。算法实时处理产品数据，检测货架上是否有商品。如果有产品缺货，他们可以找出原因，通知跟单员，并为供应链的相应部分推荐解决方案。