计算机视觉的定义和终极目标

阿里云

定义

- 计算机视觉（Computer Vision）是一门研究如何使机器“看”的科学，也可以看作是研究如何使人工系统从图像或多维数据中“感知”的科学。

终极目标

- 计算机视觉成为机器认知世界的基础，终极目的是使得计算机能够像人一样“看懂世界”。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

计算机视觉的三个优势

阿里云

计算机视觉技术相较于人类：

- 在图像处理方面上，实现超人的准确性；
- 例：图片颜色、细节敏感度。

- 在细微变化识别方面上，性能远胜于人类；
- 例：医疗图像分析。

- 在计算能力方面上，计算速度与精确性完胜人类；
- 超级计算机。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

人类视觉的定义及其工作原理

阿里云

人类视觉

- 人类的感官之一；
- 人类获取信息最直接有效的方式。

工作原理

- 外部光线穿过角膜，再通过通过瞳孔。
- 光线穿过晶状体照射到视网膜，感光体细胞会将光线转换成电信号。
- 电信号从视网膜通过视神经传播到大脑成像。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

计算机视觉与人类视觉的相似点

阿里云

计算机视觉与人类视觉具有**相似的结构**：

计算机视觉

- 转换方式：数字化过程
- 接收器：摄像头
- 转换器：电线
- 处理器：CPU

相当于

人类视觉

- 转换方式：生理过程
- 接收器：眼睛
- 转换器：神经细胞
- 处理器：大脑

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

计算机视觉与人类视觉的不同点

阿里云

区别有以下几点：

- 一个是机器，一个是生物；
- 人类的眼睛比摄像机更加灵活；
- 人类的神经更加复杂；
- CPU只是按照人类的指示做事，人类大脑有自己的思维；
- 计算机视觉可以获取人类视觉获取不到的信息，例如：红外摄像机；
- 计算机视觉可以到人类到不了的地方，例如：太空作业。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

计算机视觉在娱乐领域的应用和作用

阿里云

应用

- 智能审核网络视频内容；
- 优化前端内容的开发和运营，创造出更多玩法。

作用

- 有效缓解视频平台的监管的巨大的压力；
- 提高软件用户体验度和活跃度；
- 为视频平台创造了新的应用场景。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

数字图像的定义及数字图像处理的内容

阿里云

数字图像

- 又称为数码图像或数位图像；
- 是用一个数字矩阵来表达客观物体的图像；
- 是由模拟图像数字化得到的；
- 是一个离散采样点的集合，每个点具有其各自的属性；
- 是以像素为基本元素的图像；
- 可以用数字计算机或数字电路存储和处理的图像。

数字图像处理包括的内容

- 图像变换；
- 图像增强；
- 图像恢复；
- 图像压缩编码；
- 图像分割；
- 图像分析与描述；
- 图像的识别分类。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像数字化的两个过程

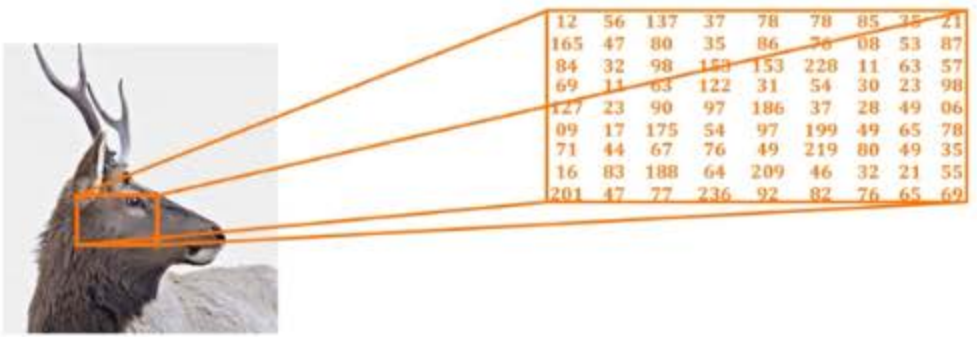
阿里云

采样

采样是将空间上连续的图像变换成离散的点，采样频率越高，还原的图像越真实。

量化

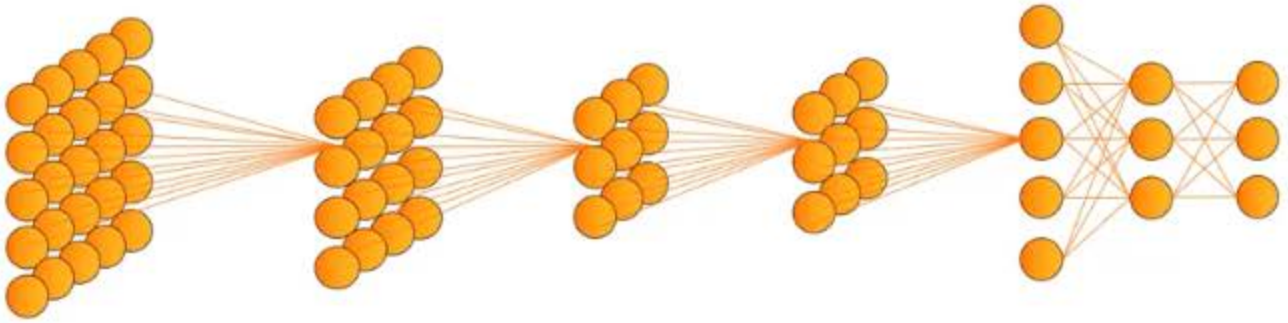
量化是将采样出来的像素点转换成离散的数量值，一幅数字图像中不同灰度值的个数称为灰度等级，级数越大，图像越清晰。



阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

计算机视觉的基础工作原理

阿里云



- 构造多层**神经网络**
- 较低层识别初级的**图像特征**
- 若干底层特征组成更**上一层特征**
- 通过多个层级的**组合**
- 最终在顶层做出**分类**

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分类、目标检测、语义分割、实例分割的介绍

阿里云

- 给定一组各自被标记为**单一类别**的图像，对一组新的测试图像的**类别进行预测**，并测量预测的**准确性结果。**

图像分类

- 给定一张图像，让计算机找出其中所有**目标的位置**，并给出每个目标的**具体类别。**

目标检测

计算机视觉技术

语义分割

- 将整个图像分成**像素组**，然后对像素组进行**标记**和**分类**；
- 语义分割是在语义上理解图像中**每个像素是什么，**还须确定每个物体的**边界。**
- 如一张“人驾驶摩托车行驶在林间小道上”的图片

实例分割

- 在语义分割的基础上进行，将**多个重叠物体**和不同背景的**复杂景象**进行**分类**；
- 同时确定对象的**边界**、**差异**和彼此之间的**关系。**

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

视频分类、人体关键点检测、场景文字识别、目标跟踪的介绍

阿里云

- 分类的对象是由**多帧图像**构成的、包含**语音数据**、**运动信息**等的视频对象
- 需要理解每帧图像**包含内容**，还需要知道上下文**关联信息**。

视频分类

- 通过人体**关键节点**的组合和**追踪**来识别人的**运动**和**行为**
- 对于**描述人体姿态**，**预测人体行为**至关重要。

人体关键点检测

计算机视觉技术

场景文字识别

- 在图像背景复杂、分辨率低下、**字体多样**、**分布随意**等情况下，将图像信息转化为**文字序列**的过程。

目标跟踪

- 在特定场景**跟踪**某一个或多个**特定感兴趣对象**的过程。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分类的定义

阿里云

定义

图像分类的核心是从给定的分类集合中给图像分配一个标签。

- 图像分类模型读取该图片；
- 生成该图片属于集合 {dog, cat, hat，mouse}中各个标签的概率。

- 82% cat
- 15% dog
- 2% hat
- 1% mouse

- 对于计算机来说，图像是一个由数字组成的巨大的**3维数组**；
- 猫的图像大小是宽240像素，高250像素，有红绿蓝3个**颜色通道**，该图像就包含了240X250X3=180000个数字，均为0-255之间的**整数**，其中0表示全黑，255表示全白。
- 图像分类的任务就是把这些上百万的数字变成一个**简单的标签**，比如“猫”。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

提问

如果要给下列图片加标签，我们可以怎么加?

动物 狗 柯基

水果 柠檬 黄柠檬

图像分类

1、根据大类、小类加标签
2、可以多单个或多个标签
3、不同的标签粒度和个数会形成不同的分类任务

单标签与多标签分类的区别

单标签

- 数据样本属于**一个大类**的；
- 数据进行分类后用可以**用一个值代表**；
- 单标签内有二分类（两个选项）和多分类（多个选项）；
- 例子：单标签三个样本的二分类整形（0/1）输出为：**[0,1,0]**。

多标签

- 数据样本可以划分到**几个大的不冲突主题类别**中；
- 在大主题中分别可以进行二分类和多分类问题；
- 例子：多标签(假设为两个标签)三个样本的二分类整形输出为：**[[0,1], [0,0],[1,1]]**。

跨物种语义级别的图像分类定义

阿里云

定义

- 在不同物种层次上识别不同类别的对象，如猫狗分类；
- 各个类别之间属于不同的物种或大类，往往具有**较大的类间方差**，而类内具有**较小的类内方差**；
- 多类别图像分类由传统的特征提取方法转到数据驱动的深度学习方向来，取得了较大进展。

猫

狗

马

子类细粒度图像分类的定义

阿里云

定义

- 子类细粒度分类相较于跨物种图像分类难度更大；
- 是一个大类中的子类的分类，如不同鸟的分类等；
- 在区分出基本类别的基础上，进行更精细的子类划分；
- 由于图像之间具有更加**相似的外观和特征**，受采集过程中存在干扰影响，导致数据呈现**类间差异性大，类内间差异小**，分类难度也更高。

麻雀

鹦鹉

白鸽

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

多标签图像分类的定义

阿里云

定义

- 给每个样本**一系列的目标标签**，表示的是样本各属性且不相互排斥的，预测出一个概念集合；
- 标签数量较大且复杂；
- 标签的标准很难统一，且往往类标之间相互依赖并不独立；
- 标注的标签并不能完美覆盖所有概念面；
- 标签往往较短语义少，理解困难。

猫、书包、盆栽

鸟、房屋、天空

人、盘子、手机

图像分类遇到的挑战

阿里云

虽然图像分类在大赛上的正确率已经接近极限，但在实际工程应用中，面临诸多挑战。

- 类别不均衡
- 数据集小
- 巨大的类内差异
- 实际应用环境复杂

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分类的常用数据集：CIFAR-10

阿里云

介绍

- CIFAR-10：一个用于识别普适物体的小型图像数据集；
- 包含**6万张**大小为32 x 32的彩色图像；
- 共有**10个类**，每类有6000张图；
- 共**5万**张图组成训练集合，训练集合中每一类均等且有5000张图；
- 共**1万**张图组成测试集合，测试集合中每一类均等且有1000张图；
- 10个类别：飞机（airplane）、汽车（automobile）、鸟类（bird）、猫（cat）、鹿（deer）、狗（dog）、蛙类（frog）、马（horse）、船（ship）和卡车（truck）；
- 类是完全互斥的：在一个类别中出现的图片不会出现在其它类中。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

使用的相关神经网络：LeNet-5、AlexNet

阿里云

LeNet-5

- 是**最早**的卷积神经网络之一;
- 1998年第一次将LeNet-5应用到图像分类上，在手写数字识别任务中取得了巨大成功;
- LeNet-5通过连续使用卷积和池化层的组合提取图像特征，总共5层：3层卷积和2层全连接，池化层未计入层数;
- LeNet-5是卷积神经网络的开篇大作，完成了卷积神经网络从无到有的突破。

AlexNet

- AlexNet将LeNet的思想发扬光大，把CNN的基本原理应用到了很深很宽的网络中。
- 成功使用ReLU作为CNN的激活函数，并验证其效果优异;
- 训练时使用数据增强和Dropout随机忽略一部分神经元，以避免模型过拟合，提升泛化能力;
- 在CNN中使用重叠的最大池化，提升了特征的丰富性;
- 提出了LRN层，增强了模型的泛化能力。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分类在图片搜索引擎中的应用

阿里云

- 应用图像分类技术可以开发各种图片搜索引擎；
- 图片搜索引擎能通过用户上传图片，应用图像分类技术，识别出图片的内容并进行分类；
- 搜索互联网上与这张图片相同或相似信息的其他图片资源进行校对和匹配，识别图片的内容并提供相关信息。

- 用户上传图片
- 应用图像分类技术识别和分类

- 搜索具有相同或相似信息的图片资源
- 对图片资源进行校对和匹配

识别分类结果：玫瑰花

- 获得识别结果
- 得到图片内容和相关信息

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分类在垃圾分类中的应用-智能环卫

阿里云

- 为了破解传统分类投放模式可能存在的乱扔垃圾等问题，可在传统垃圾分类投放站点部署摄像头进行智能化改造；
- 阿里云提出“智能环卫”产品，提供垃圾分类投放点AI智能检测分析功能；
- 有效针对垃圾桶内的未破袋垃圾包、残余垃圾袋等进行检测和识别，检测效率高，真实环境下检测准确率超过95%。

目标检测的定义

阿里云

定义

目标检测就是识别图中有哪些物体，确定他们的类别并标出各自在图中的位置。

- 目标检测模型读取该图片；
- 寻找识别出图中的物体目标，对其进行定位，框起和标注。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分类与目标检测的区别

阿里云

图像分类：整幅图像经过识别后被分类为**单一的标签**。

目标检测：除了识别出图像中的一个或多个目标，还需要找出目标在图像中的**具体位置**。

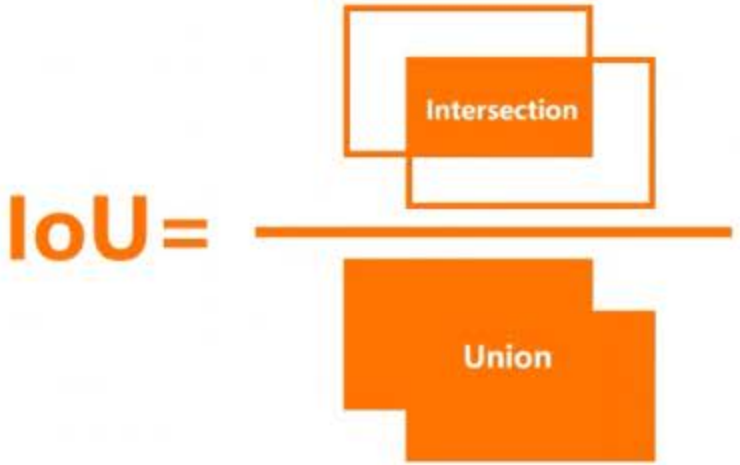
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

交并比：IoU

阿里云

定义

- 真实边界框：训练集中，人工标注的物体边界框；
- 预测边界框：模型预测到的物体边界框；
- 交并比：在分子项中，是真实边界框和预测边界框**重叠的区域**（Intersection）。分母是一个并集（Union），或者更简单地说，是由预测边界框和真实边界框所**包括的区域**。两者相除就得到了最终的得分。



阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

精确度（Precision）和召回率（Recall）

阿里云

定义

- 精确度指目标检测模型判断该图片为正类，该图片确实是正类的概率；
- 召回率是指的是一个分类器能把所有的正类都找出来的能力；
- 在这里需要明白什么是正类：

预测 / 真实	正（P）	负（N）
正（P）	TP(真正例)	FN(假负例)
负（N）	FP(假正例)	TN(真负例)

- 精确度衡量的公式为：$$Precision = \frac{TP}{TP + FP}$$
- 召回率衡量的公式为：$$Recall = \frac{TP}{TP + FN}$$

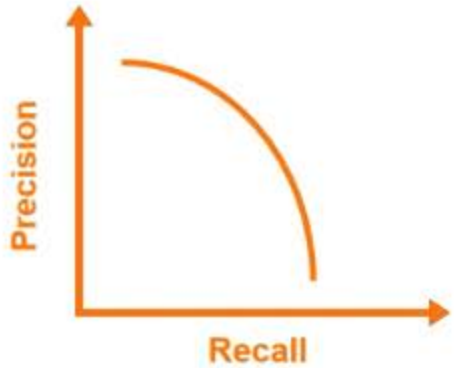
阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

平均精度值：mAP

阿里云

定义

- mAP，mean Average Precision, 即各类别平均精度均值；
- mAP是把每个类别的AP都单独拿出来，然后计算所有类别AP的平均值，代表着对检测到的目标平均精度的一个综合评价。



- 每一个类别都可以根据Recall和Precision绘制一条曲线，那么AP就是该曲线下的面积，而mAP则是多个类别AP的平均值，这个值介于0到1之间。mAP是目标检测算法里最重要的一个评估指标。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

目标检测遇到的挑战

目标数量问题

- 在图片输入模型前不清楚图片中有多少个目标，无法知道正确的输出数量。

目标大小问题

- 目标的大小不一致,甚至一些目标仅有十几个像素大小，占原始图像中非常小的比例。

如何建模

- 需要同时处理目标定位以及目标物体识别分类这两个问题。

目标检测的常用数据集：PASCAL VOC

阿里云

介绍

- PASCAL VOC：一个常用于目标检测的小型图像数据集；
- 包含**11530张**彩色图像，标定了**27450个**目标识别区域；
- 从初始4个类发展成最终的**20个类**；
- 在整个数据集中，平均每张图片有2.4个目标；
- 20个类别：
  - 动物：人、鸟、猫、狗、牛、马、羊；
  - 运载工具：飞机、自行车、船、巴士、汽车、摩托车、火车；
  - 物品：瓶子、椅子、餐桌、盆栽、沙发、电视机。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

使用的相关神经网络：CenterNet

CenterNet

- CenterNet结构优雅简单，直接检测目标的中心点和大小;
- CenterNet把目标检测任务看作三个部分：
  - 寻找物体的中心点;
  - 计算物体中心点的偏移量;
  - 分析物体的大小;
- CenterNet检测速度和精度相比于先前的框架都有明显且可观的提高，尤其是与著名的目标检测网络YOLOv3作比较，在相同速度的条件下，CenterNet的精度比YOLOv3提高了大约4个点。

目标检测在智慧交通中的应用-高速云控

阿里云

- 智慧交通是目标检测的一个重要应用领域，主要包括如下场景：

交通异常事件检测

- 检测各种交通异常事件，如车辆占用应急车道、车辆驾驶员的驾驶行为等；
- 第一时间将异常事件上报给交管部门，提高处理效率。

高速云控

- 在智慧交通场景下，阿里云提出高速云控解决方案；
- 依托阿里云计算平台，通过智能高速引擎和交通视觉计算，有效地在高速交通态势、事件处置闭环等交通应用场景实现智慧云控。

目标检测在智慧交通中的应用-智慧眼

阿里云

交通流量监控与红绿灯配时控制

- 通过目标检测算法，对道路视频图像进行分析；
- 根据分析车流量，调整红绿灯配时策略，提升交通通行能力。

四川高速 SECDG

AthenaEyes 智慧眼

四川高速 x 智慧眼

- 基于达摩院AI算法与高德深度融合的“智慧眼”高速管控平台实现路况事件感知与处置自动联动闭环；
- 2019年春运成都绕城高速拥堵平均下降10%,事件智能发现达31.3%,公众出行感知提升26%。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分割的定义

阿里云

定义

- 图像分割就是把图像分成若干个特定的、具有独特性质的区域并提出感兴趣目标的技术和过程；
- 图像分割包括：**语义分割**、**实例分割**和**全景分割**。

- 图像作为分割算法的输入，输出一组区域；
- 区域可以表示为一种掩码（灰度或颜色），其中每个部分被分配一个唯一的颜色或灰度值来代表它，如下例子所示：

• 输入图像

• 分割算法运算

• 输出分割结果

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

语义分割的定义

阿里云

定义

- 语义分割是在**像素级别上的分类**，属于同一类的像素都要被归为一类；
- 语义分割是从像素级别来理解图像的。
- 如下的照片，属于猫的像素都要分成一类，属于狗的像素也要分成一类，除此之外还有背景像素也被分为一类。

- 语义分割

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

实例分割的定义

阿里云

定义

- 实例分割比语义分割更进一步；
- 对于语义分割来说，只要将所有同类别（猫、狗）的像素都归为一类；
- 实例分割还要在具体类别（猫、狗）像素的基础上区分开**不同的实例**（短毛猫、虎斑猫、贵宾犬、柯基犬）。

- 实例分割

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

全景分割的定义

阿里云

定义

- 全景分割是语义和实例分割的相结合；
- 每个像素都被分配一个类（比如：狗），如果一个类有多个实例，则可知道该像素属于该类的哪个实例（贵宾犬/柯基犬）。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分割遇到的挑战

阿里云

分割边缘不准

- 因为相邻临的像素对应感受野内的图像信息太过相似导致。

样本质量不一

- 样本中的目标物体具有多姿态、多视角问题，会出现物体之间的遮挡和重叠；
- 受场景光照影响，样本质量参差不齐。

标注成本高

- 对于数据样本的标注成本非常高，而且标注质量难以保证不含有噪声。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

图像分割的常用数据集：COCO

阿里云

介绍

- COCO：一个常用于图像分割的大型图像数据集；
- 包含**33万张**彩色图像，标定了**50万个**目标实例；
- 具有80个目标类、91个物品类以及25万个人物关键点标注；
- 每张图片包含5个描述；
- 每一类的图像多，利于提升识别更多类别位于特定场景的能力；
- 类别包括： person(人) 、 bicycle(自行车) 、 car(汽车) 、 motorbike(摩托车) 、 aeroplane(飞机) 、 bus(公共汽车) 、 train(火车)、 truck(卡车) 、 boat(船) 、 traffic light(信号灯) 、 fire hydrant(消防栓) 、 stop sign(停车标志) 、 parking meter(停车计费器) 、 bench(长凳) 、 bird(鸟) 、 cat(猫) 、 dog(狗) 、 horse(马) 、 sheep(羊) 、 cow(牛) 等等。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER

使用的相关神经网络：FCN

阿里云

FCN

- FCN全卷积神经网络是图像分割的基础网络;
- 全卷积神经网络，顾名思义网络里的**所有层都是卷积层**;
- 卷积神经网络卷到最后特征图尺寸和分辨率越来越小，不适合做图像分割，为解决此问题FCN引入上采样的方法，卷积完之后再上采样到大尺寸图;
- 为避免层数不断叠加后原图的信息丢失得比较多，FCN引入一个跳层结构，把前面的层特征引过来进行叠加;
- FCN实现了端到端的网络
  - 端到端学习是一种解决问题的思路，与之对应的是多步骤解决问题，也就是将一个问题拆分为多个步骤分步解决，而端到端是由输入端的数据直接得到输出端的结果。

阿里云培训中心
ALIBABA CLOUD TRAINING CENTER