1. 什么是数字图像处理？

一幅图像可以定义为一个二维函数f(x, y)，其中x和y是空间平面坐标，任意一对空间坐标(x, y)处的幅值f称为图像在该点的强度或灰度。当x, y和灰度值f都是**有限的离散量**时该图像为数字图像。

数字图像处理是指借助于**数字计算机**来处理**数字图像**。数字图像处理界定为其**输入和输出都是图像的处理**，它包含从**图像中提取特征**的处理，直至包含**各个目标的识别**。例如获取含有文本的区域的一幅图像，对这幅图像进行**预处理**，**提取各个字符**，以适合计算机进行处理的形式**描述这些字符**，以及**识别这些字符**。

1. 图像处理的基本步骤？

图像获取

图像增强（对图像进行某种操作，使结果在特定应用中比原图像更为合适）

图像复原（以图像退化的数学或概率模型为基础）

彩色图像处理

小波变换（以不同分辨率来表示图像的基础）

压缩（或少图像存储量或降低传输图像的带宽）

形态学处理（提取图像中用于表示和描述形状的成分）

分割（将一幅图像划分为各个组成部分或目标）

特征提取（包括特征检测和特征描述。特征检测是指寻找一幅图像中的特征、区域或边界。特征描述是指对检测到的特征规定量化属性，比如检测一个区域的角点，并用它们的方向和位置描述这些角点）

图像模式分类

1. 图像处理系统的组成

获取数字图像->专用图像处理硬件->计算机->图像处理软件->大容量存储、图像显示器

获取数字图像需要两个子系统。第一个子系统是**物理传感器**，其作用是对**成像目标辐射的能量产生响应**。第二个子系统是**数字化仪**，其作用是把**物理感知设备的输出转换为数字形式**。例如，在数字视频摄像机中，传感器（CCD芯片）**产生一个与光强成比例**的输出。数字化仪将这些输出转换为数字数据。

**专用图像处理硬件**通常由刚刚提及的**数字化仪**和**执行其他原始操作的硬件**如**算术逻辑单元(ALU)**组成，算术逻辑单元对整个图像并行执行**算术与逻辑运算**。如何使用ALU的一个例子是与数字化一样快的图像平均操作，这一操作的目的是降低噪声。这种类型的硬件有时称为前端子系统，其显著特点是速度快。换句话说，该单元执行要求快速数据吞吐的功能(如以 30帧/秒的速率数字化和平均视频图像)，而典型的主机无法胜任这一工作。在执行密集矩阵运算的图像处理系统中，一个或多个GPU很常见。

计算机：通用计算机

图像处理软件

图像处理应用：提供大容量存储，1024x1024的图像，每个像素灰度为8bit时需要1MB，存储主要分三类——短期存储、在线存储和档案存储，短期存储可以使用计算机内存或者专用存储板；在线存储通常采用磁盘或光介质存储器。在线存储的关键因素是对存储的数据的频繁存取。

第二章 数字图像基础

1. 图像感知与获取

我们感兴趣的大多数图像，都是由“照射”源和形成图像的“场景”元素对光能的反射或吸收产生的。将照射能量转换为数字图像的三种主要传感器的原理是：组合输入电能和传感器对正被检测能量类型的响应，将入射能量转换为电压。输出电压波形是传感器的响应，将传感器响应数字化，得到一个数字量。

简单来说，就是输入电压->获得传感器响应（电压波形）->数字化

1. 图像取样和量化

获取图像的本质是由**感测的数据**生成**数字图像**。

多数传感器的输出是连续的电压波形，波形的幅度、空间、特性都与正被感测的物理现象相关，需要把连续感测的数据**转换为数字形式**以产生一幅数字图像。

转换包括两种处理：取样和量化

2.1 取样和量化的基本概念

对于一幅连续图像f，xy坐标是连续的，幅度也是连续的

**取样：**对坐标进行数字化

**量化：**对幅度值进行数字化

所以，取样和量化都是对图像进行数字化的操作

1. 数字图像表示

连续图像函数通过取样和量化转换为数字图像，数字图像的每个像素值对应于离散坐标位置上的连续图像函数值。

表示数字图像的三种方法：

函数图表示：用两个坐标轴确定空间位置，用第三个坐标轴表示f的值，适用于处理以(x, y, z)形式表示的灰度集。

灰度值表示：显示器上每个点的灰度与对应点的f值成比例，通常使用有限的灰度值表示。

数值阵列（矩阵）表示：以数值f(x, y)组成的矩阵形式表示图像，每个元素对应图像中的一个像素，适用于计算机处理和分析图像数据。

1. 为什么数字图像可以用矩阵表示？

因为数字图像是由**像素（Picture Element）**组成的，每个像素包含了图像上的一个点的亮度或颜色信息。

对于**灰度图像**，每个像素的值表示该点的**亮度级别**，通常在**0到255**之间。这种情况下，图像可以表示为一个**二维矩阵**，其中每个元素对应一个像素的亮度值。

对于彩色图像，每个像素的值由红、绿、蓝三个分量组成，通常用RGB表示。在这种情况下，图像可以表示为一个三维矩阵，其中每个元素是一个包含三个分量值的向量。

1. 关于灰度的一些定义

动态范围：灰度跨越的值域，上限取决于饱和度，下限取决于噪声

对比度：一幅图像中最高和最低灰度级间的灰度差

反差比：一幅图像中最高和最低灰度级的比率

1. 线性索引和坐标索引

坐标索引：一个像素的位置由其二维坐标给出

线性索引：列扫描图像或者行扫描图像得到的

1. 空间分辨率和灰度分辨率

空间分辨率：图像中最小可辩别细节的测度，常见测度包括单位距离的线对数和单位距离的点数（像素数）。

灰度分辨率：在灰度级中可分辨的最小变化，通常以比特数来表示。

1. 图像内插

内插是**用已知数据来估计未知位置的值**的过程。

**最近邻内插**：当将一个500×500像素的图像放大到750×750像素时，可以创建一个大小为750×750像素的假想网格，使其像素间隔与原图像相同。然后将该网格收缩，使其与原图像完全重叠。在上覆网格中的每个点，找到最接近的原图像像素，并将其灰度值赋给上覆网格中的新像素。最后，展开上覆网格，得到放大后的图像。

**双线性内插：**使用4个最近邻的灰度来计算给定位置的灰度。V(x,y)=ax+by+cxy+d

**双三次内插：**包括16个最近邻点，v(x,y)=aijx^iy^i

1. 像素间的基本关系

坐标(x,y)处的像素p有2个水平的相邻像素和2个垂直的相邻像素，坐标为

(x+1,y)(x-1,y)(x,y+1)(x,y-1)

p的4个对角相邻像素的坐标是

(x +1,y+ 1), (x +l,y- 1),(x-1,y + 1), (x -1,y-1)

包含p则为闭邻域，不包含为开邻域

1. 邻接、连通、区域和边界

邻接（Adjacency）：邻接指的是图像中两个像素之间的关系。在二维图像中，通常使用四邻域或八邻域来定义像素之间的邻接关系。四邻域表示一个像素与其上下左右四个相邻像素之间的关系，而八邻域则表示一个像素与其周围八个相邻像素之间的关系。

连通（Connectivity）：连通指的是图像中像素之间的空间关系。连通性描述了图像中像素集合的拓扑结构。在二值图像中，连通性表示图像中连续的像素组成的对象或区域的性质。常见的连通性有4连通和8连通，分别表示像素可以通过四个方向或八个方向相邻的像素进行连接。

令S是图像中的一个像素子集。如果S的全部像素之间存在一个通路，则可以说两个像素p和q在S中是连通的。对于S中任何元素p，S中连通到该像素集称为S的连通分量。如果S仅有一个连通分量，则集合S称为连通集。

令R是图像中的一个像素子集。如果R是连通集，则称R为一个区域。在谈区域时，必须制定邻接的类型（4邻接或8邻接）。一个R的边界（也称为边缘或轮廓）是区域中像素的集合。

1. 距离测度

D：欧几里得距离

D4：城市街区距离（|x-u|+|y-v|）

D8：棋盘距离（max(|x-u|，|y-v|)）

1. 目标检测的定义

找到图像中所有感兴趣的目标，确定它们的类别和位置

1. 图像识别四大任务

分类定位检测分割

1. 理想的检测器要解决什么问题？

类内差异、环境干扰、类间相似、集群小目标、效率问题

1. 基于锚框的目标检测方法

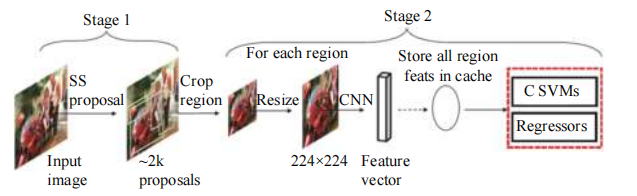
两阶段模型是指**有独立地、显示地提取候选区域(region proposal)的过程**，即先在输入图像上筛选出一些可能存在物体的候区域，然后针对每个候选区域，判断**其是否存在物体**，如果有就**给出物体的类别**和**位置修正信息**。一般而言，两阶段方法**具有较高的检测精度，**而单阶段方法**具有较快的推理速度**。

1. R-CNN->Fast RCNN->Faster RCNN？

RCNN：

首先使用无监督的选择性搜索(Selective Search , ss)方法将输入图像中具有相似颜色直方图特征的区域进行递归合并，产生约2000个候选区域;然后从输入图像中截取这些候选区域对应的图像，将其裁剪缩放至合适的尺寸，并相继送入一个CNIN特征提取网络进行高层次的特征提取，提取出的特征再被送入一个 **SVM分类器**进行**物体分类**，以及一个**线性回归器**进行**边界框位置和大小的修正**;最后对**检测结果进行非极大值抑制操作**，得到最终的检测结果。

也就是先提取proposal，然后将proposal输入CNN提取特征，使用SVM分类，最后做bbox reg。



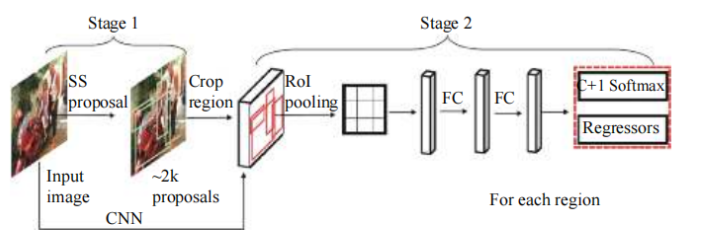
Fast RCNN：

1. 将任意size的图片输入CNN，得到特征图。在RCNN中，先生成region proposals再做卷积，相当于做了多次卷积，浪费时间。
2. 对原始图片使用selective search算法得到约2k region proposals（相当于RCNN的第一步）
3. 在特征图中找到每一个region proposals对应的特征框。采用RoI池化将**每个候选区域分成固定大小的网格，然后在每个网格内进行最大池化，转为固定大小的特征图**
4. 统一大小的特征框经过全连接层得到固定大小的特征向量，分别进行softmax分类（使用softmax代替了RCNN里面的多个SVM分类器）和bbox回归

使用了共享卷积特征图加速检测过程，从而提高检测速度

第二阶段

Fast R-CNN组合了classification和regression, 做成single Network，实现了端到端的训练，实际上它相对RCNN最大的改进是抛弃了多个SVM分类器和bounding box回归器的做法，一起输出bbox和label



**ROI Pooling（Region of Interest）考点**

它的输入是特征图，输出则是大小固定的channel x H x W的vector。ROI Pooling是将一个个大小不同的region proposals，映射成大小固定的（W x H）的矩形框。它的作用是**根据region proposals的位置坐标**在**特征图中将相应区域池化为固定尺寸的特征图**，以便进行后续的分类和输出回归框操作。它可以加速处理速度。这个ROI Pooling在下面的Faster RCNN也会出现，它的作用是类似的，就是将region proposals池化成同样大小的vector，便于传入后续分类网络。

ROI Pooling有两个输入，一个是图片进入CNN后的特征图，另一个是区域的边框。ROI 的输出是一个region\_nums x channels x W x H的向量。

Faster RCNN：

1. 提取特征：输入固定大小的图片，进过卷积层提取特征图feature maps
2. 生成region proposals: 然后经过Region Proposal Networks(RPN)生成region proposals。该层通过softmax判断anchors属于foreground或者background，再利用bounding box 回归修正anchors获得精确的proposals（候选区域）。
3. ROI Pooling: 该层的输入是feature maps和proposals，综合这些信息后提取proposal feature maps
4. Classification: 将Roi pooling生成的proposal feature maps分别传入softmax分类和bounding box regression获得检测物体类别和检测框最终的精确位置
5. SPPNet

空间金字塔池化，该层将输入特征图分为不同大小的子网格，并对每个子网格进行池化操作，最终输出一个固定大小的特征向量。这个固定大小的特征向量可以作为输入到全连接层中进行分类。

1. 首先，将输入特征图按照不同的比例进行分割，例如将特征图分割为1x1、2×2、4x4等多个子网格。
2. 对于每个子网格，进行池化操作，例如最大池化或平均池化。这样可以得到每个子网格的一个固定大小的特征向量。
3. 将所有子网格的特征向量拼接在一起，得到一个固定大小的特征向量作为输出。
4. FPN

FPN的结构主要包括两个部分：自底向上的特征提取和自顶向下的特征融合。

自底向上的特征提取：FPN首先通过一个基础的卷积神经网络（如ResNet）从输入图像中提取特征。这个过程是一个逐层下采样的过程，每层的特征图分辨率逐渐降低，但语义信息逐渐增强。

自顶向下的特征融合：在自底向上的特征提取过程中，FPN从每个下采样层中得到一组特征金字塔。为了将不同尺度的特征进行融合，FPN引入了自顶向下的连接。具体地，高分辨率的特征图通过上采样操作得到更低分辨率的特征图，然后与低分辨率的特征图进行逐元素相加，得到融合后的特征图。这样的操作可以保留高层特征的语义信息同时融合低层特征的细节信息。

通过自底向上的特征提取和自顶向下的特征融合，FPN能够生成具有多尺度信息的特征金字塔。这使得网络可以在不同尺度上进行目标检测或语义分割，并且对于处理不同大小的目标具有较好的适应性。

1. Mask-RCNN