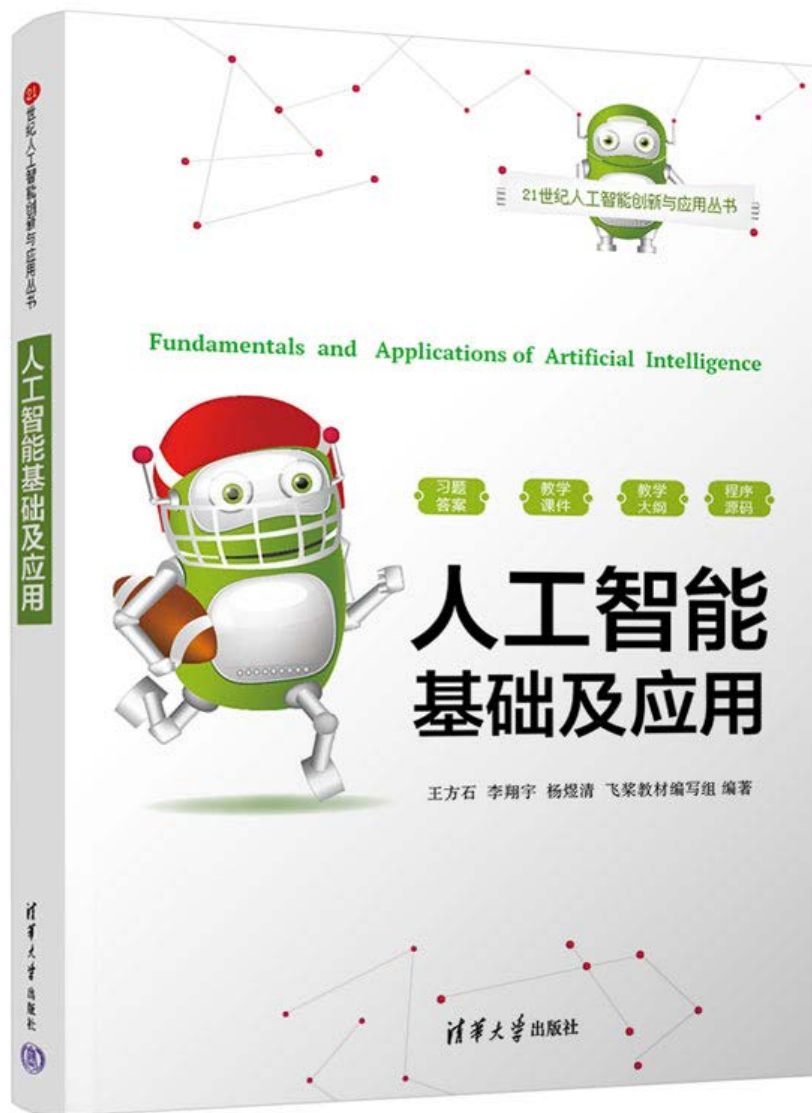


人工智能基础及应用



《人工智能基础及应用》第1版.

王方石, 李翔宇, 杨煜清, 飞桨教材编写组.

北京: 清华大学出版社, 2023年11月出版.

ISBN : 9 787 302 644244

第7章 智能图像处理

7.1 数字图像处理概述

7.2 传统的图像处理技术

7.3 基于深度学习的图像处理技术

本章学习目标

- ◆理解数字图像处理技术的基本概念和主要任务。
- ◆理解传统图像处理技术的原理、方法和评价指标。
- ◆理解基于深度学习的图像处理技术的方法和实现流程。

第7章 智能图像处理

◆ 人类的视觉

- 视觉是人类从大自然中获取信息、认识世界的最主要手段。
- 人类通过五官获取的外界信息中80%是视觉信息。
- 人能够记忆视觉感知信息的30%。

◆ 计算机视觉

- 希望计算机能够像人类的视觉那样具有“看”的智能。
- 可以感知、识别、理解客观世界中存在的各种目标。

◆ 计算机视觉与数字图像处理

- **计算机视觉**是一门研究对数字图像或视频进行高层理解的交叉学科
- **数字图像处理**则是计算机视觉的基础

7.1.1 数字图像处理的基本概念

◆ 图像

- **图像**是自然景物的客观反映，是人类认识世界的视觉基础，是人类获取、表达和传递信息的重要手段。
- “图”，就是物体透射光或反射光的分布，是客观存在的。
- “像”是人的视觉系统所接收的图在大脑中形成的印象或认识，是人的感觉。
- 图像则是两者的结合

7.1.1 数字图像处理的基本概念

◆ 数字图像

- **数字图像**是指用数字摄像机、扫描仪等成像设备经过采样和数字化得到的一个二维数组或矩阵。
- 该数组或矩阵的元素称为**像素**（Pixel），像素值均为整数，称为**灰度值**、**亮度值**或**强度值**。
- **数字图像**又称为**数码图像**或**数位图像**，通常表示为一个 $W \times H \times C$ 的数值矩阵，其中 W 为图像的宽度， H 为图像的高度， C 为每个像素点对应的信息维度。
- 矩阵中的元素值，即**灰度值**，记为 $f(w, h, c)$ ，其中 w 和 h 分别是该像素点的横坐标和纵坐标， c 表示该像素点的维度。
- 数字图像可以用数字计算机或数字电路存储和处理。

7.1.1 数字图像处理的基本概念

◆ 数字图像的种类

(1) 二值图像：又称为黑白图像。

$C=1$ ，即只有1个颜色通道；像素值 $f(w, h, 1)$ 只能取值0或1。

(2) 灰度图像： $C=1$ ， $f(w, h, 1) \in [0, 255]$ 。0为黑色，255为白色。

灰度图像非黑白图像，黑白图像只有黑白两种颜色，灰度图像则有256级颜色深度。

(3) 彩色图像， $C=3$ ，每个像素由R、G、B三个分量构成，取值范围为 $[0, 255]$ 。

(4) RGBD 图像： $C=4$ ，采用三维深度摄像头拍照时，每个像素点不仅具有R、G、B颜色值，还具有深度信息值，这种图像称为RGBD（RGB + Depth）图像。

7.1.1 数字图像处理的基本概念

◆ 数字图像处理

- **图像处理**，就是对图像信息进行加工，以满足人的视觉心理或应用需求的行为。
- 从20世纪60年代起，随着电子技术和计算机技术的不断发展和普及，电子学方法逐渐成为图像处理的主流手段，即数字图像处理。
- **数字图像处理**（Digital Image Processing）是指利用数学方法和计算机技术对数字图像或视频信息进行加工处理，以获取图像中的某些信息，提高图像的实用性。

◆ 数字图像处理技术的**主要内容**包括三大部分：

- 图像的编码与压缩；
- 图像的增强和复原；
- 图像的匹配、描述和识别。

7.1.1 数字图像处理的基本概念

进一步细分，图像处理的任务包括：

- 图像采集与获取
- 图像变换
- 图像去噪
- 图像增强
- 图像复原
- 图像重建
- 图像编解码
- 图像压缩
- 图像表示与描述
- 图像特征提取与分析
- 图像分类
- 图像目标检测
- 图像边缘检测
- 图像分割
- 图像质量评价
- 图像识别
- 图像理解

7.1.2 数字图像处理的主要任务

1. 图像复原 (Image Restoration)

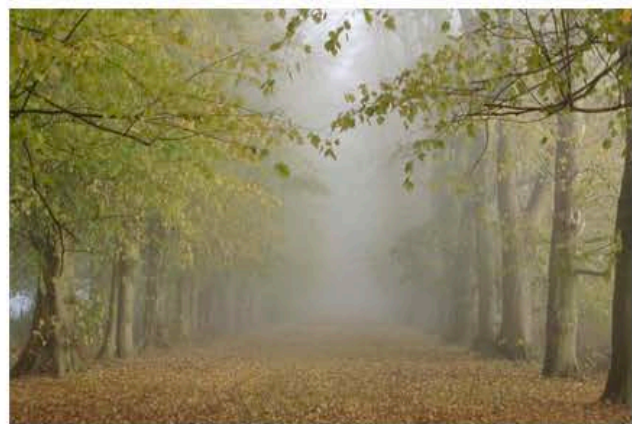
- ◆ 图像在形成、记录、处理和传输过程中，由于成像系统、记录设备、传输介质和处理方法的不完善，导致**图像质量下降**，这种现象称为**图像退化**。
- ◆ **图像复原**又称**图像恢复**：对退化的图像进行处理，使图像尽可能地恢复本来面目。
- ◆ **图像复原的目的**就是就是消除噪声、干扰和模糊，改善图像的质量。
- ◆ 图像复原认为图像是在某种情况下退化了，需要针对退化原因设法进行补偿，这就需要对图像的退化过程有一定的先验知识，再利用图像退化的逆过程去恢复原始图像。

1. 图像复原

◆ 图像复原技术的**基本思路**是：

- 首先利用图像退化过程的先验知识建立图像退化的数学模型；
- 然后根据该模型对退化图像进行拟合，
- 再建立图像的复原模型。

◆ 图像复原最关键的是：对每种退化都建立一个合理的模型。图像复原技术就是要将图像退化的过程模型化，并且采用相反的过程恢复出原始图像。



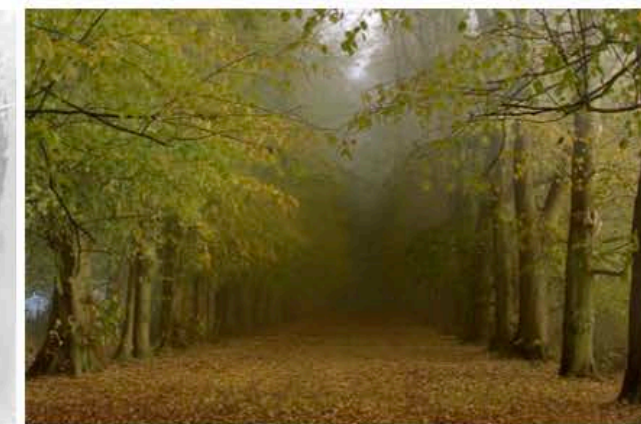
(a)



(b)



(c)



(d)

2. 图像增强（1）

- ◆ 图像增强（Image Enhancement）是一种增强摄影图像可读性的处理技术，即应用计算机或光学设备改善图像视觉效果。
- ◆ 图像增强的目的有两个：
 - 提高图像分辨率，使可见的细节更加清晰，使不易看清的细节呈现清楚，以便充分利用有用信息；
 - 增强图像对比度（即反差），突出感兴趣的目标或重要特征的细节，使人或计算机更易观察或检测到。
- ◆ 图像增强通常要完成的工作是：除去图像中的噪声，使边缘清晰；突出图像中的重要特征信息，同时减弱或去除不需要的信息。

2. 图像增强（2）

◆ 图像增强与图像复原的**相似之处**：两者的目的都是为了提高图像的整体质量。

◆ **不同之处**：

① 增强的图像不一定要逼近原始图像；而图像复原要尽可能地接近原图像。

② 图像增强更注重图像的对比度，它根据人类的视觉效果和喜好来处理图像，为观看者提供赏心悦目的图像，并不需要考虑图像退化的原因和过程；而图像复原技术则需要了解退化图像的某种先验知识，采用退化过程的逆过程进行补偿，去除图像中的模糊部分。

◆ 当无法获知图像退化的有关信息时，可使用图像增强技术，根据观看者的喜好改善图像的视感质量。

2. 图像增强（3）

- ◆ 图像增强的**思路**：根据某一指定的图像及其实际场景需求，借助特定的增强算法或算法集合来强化图像的有效信息或感兴趣信息，抑制不需要的信息或噪声。
- ◆ 现阶段，**比较流行的图像增强技术**有：灰度变换、同态滤波、直方图修正、频域滤波。
- ◆ 图像增强算法通常有其各自适用的范围，并不存在可适用于所有场景的、通用的图像增强算法。



3. 图像重建

- ◆ 图像重建（Image Reconstruction）是一种通过物体外部测量的数据、经过数字处理获得三维物体形状信息的技术。
- ◆ 图像重建技术起源于CT技术，最初采用CT设备显示人体各个部位鲜明清晰的断层图像。
- ◆ 后来，图像重建技术逐渐应用于许多领域，如工业自动化、机器人、地图测绘等。
- ◆ 图像重建与前述的图像增强、图像复原不同。
 - 图像增强和图像复原是从图像到图像的处理，即输入的是图像，处理后输出的也是图像；
 - 图像重建则是从数据到图像的处理，即输入的是某种数据，而处理后输出的是图像。
- ◆ 目前，图像重建与计算机图形学相结合，将多个二维图像合成三维图像，并加以光照模型和各种渲染技术，可生成各种具有强烈真实感的高质量图像。

4. 图像压缩编码

- ◆ 数字图像的特点之一是数据量庞大，会占据大量的存储空间和很宽的传输信道。
- ◆ 图像数据通常包含大量冗余信息，还有相当数量的不相干信息。
- ◆ **图像压缩的目的**：消除冗余和不相关的信息，降低数据量。
- ◆ **图像编码**主要是利用图像信号的统计特性和人类视觉特性，对图像进行高效**编码**，即**数据压缩技术**。
- ◆ 目前已制定了多种图像编码标准，如H.261、JPEG和 MPEG 等。
- ◆ 一般来说，**图像编码的目的**有三个：
 - ①减少需存储的数据量；
 - ② 降低数据率，减少传输带宽；
 - ③压缩信息量，便于特征提取，为后续的图像识别作准备。

5. 图像分类

- ◆ **图像分类**（Image Classification）是根据不同类别的目标在图像信息中所反映的特征不同，将不同类别的目标区分开来的图像处理方法。
- ◆ 它利用计算机对图像进行定量分析，“学习”图像中不同类别存在的规律或特征，将图像或图像中的每个像元或区域划归为若干个类别中的某一种，以代替人的视觉判读。
- ◆ 这里的类别可以是图像中不同种属的动物、不同品牌的车辆、人面部的不同表情、医学影像中不同的疾病或病灶等。
- ◆ 图像分类是给定一幅测试图像，利用训练好的分类器判定它所属的类别，而分类器则是利用带类别标签的训练数据集构建的图像分类模型。
- ◆ 图像分类可以对图像整体预测单个类别（**单标签**）或对图像中包含的多个物体类别进行预测（**多标签**）。

6. 图像目标检测

- ◆ **图像目标检测**（Object Detection）是指在给定图像中判断是否存在感兴趣的目标或物体（如人脸、汽车、猫等），并**确定目标的类别、位置及大小**。
- ◆ 可以用一个边界框（Bounding-Box, BBOX, 包含某个目标的最小矩形框）将一个目标定位出来，为下一步识别边界框中的目标是哪一个个体（如哪个人、哪辆车、哪只猫）作准备，即**目标识别**。
- ◆ **目标检测模型的输出形式通常是一个列表**，列表中的每一项对应一个检测出的目标，其中包括该目标的类别和位置，位置常用边界框的坐标表示。
- ◆ 图像分类只给出物体的类别，目标检测还需要输出目标的边界框，并且需要对众多对象进行分类和定位，而不仅仅是对个别主体目标进行分类和定位。
- ◆ **目标检测任务是在图像分类基础上又增加了定位的功能**。例如，人脸检测就是在一幅图像或视频中判断是否存在人脸；若存在，则返回人脸的位置和大小信息。

7. 图像分割（1）

- ◆ 图像中，人们感兴趣的那些部分常称为**目标或前景**，一般对应于图像中特定的、具有独特性质的区域，其他部分则称为**背景**。
- ◆ 一张图像通常包含多个目标，为了识别图像中的目标，需要按照一定的规则将图像分割成若干个区域，每个区域代表一个目标或目标的一部分。预先定义的目标可对应于单个区域，也可对应于多个区域。
- ◆ **图像分割**（Image Segmentation）是指根据一定的图像特征，将图像划分为多个不相交的区域，并提取出感兴趣目标的技术和过程。
- ◆ 图像特征可以是图像的灰度、颜色、亮度、纹理、形状等信息，也可以是更为抽象的语义信息。
- ◆ **划分的原则**是：使得同一区域内的图像像素具有特征一致性或相似性，而不同区域的图像像素之间表现出明显的特征差别。

7. 图像分割（2）

◆ 图像分割可分为3类：

➤ 语义分割（Semantic Segmentation）

是指对图像进行像素级分类，将同类目标的像素点合并成一个区域，以实现图像中的目标进行类别分割。

➤ 实例分割（Instance Segmentation）

是将图像中的每个目标分割成独立的实例，获得每个目标的轮廓，以实现图像中的每个目标进行个体分割。

➤ 全景分割（Panoptic Segmentation）。

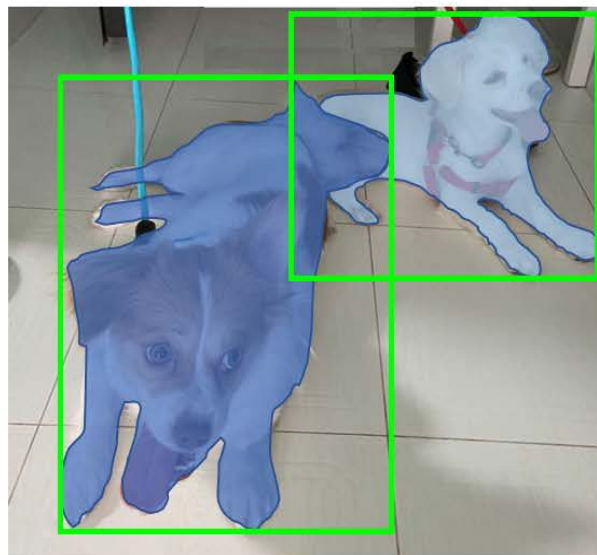
结合了语义分割和实例分割，不仅要对所有像素点进行分类，区分包括背景的所有类别，还要识别出图像中存在的所有目标个体。

◆ 图像分割有时也用其他术语，如目标轮廓、目标检测、目标识别、目标跟踪等。

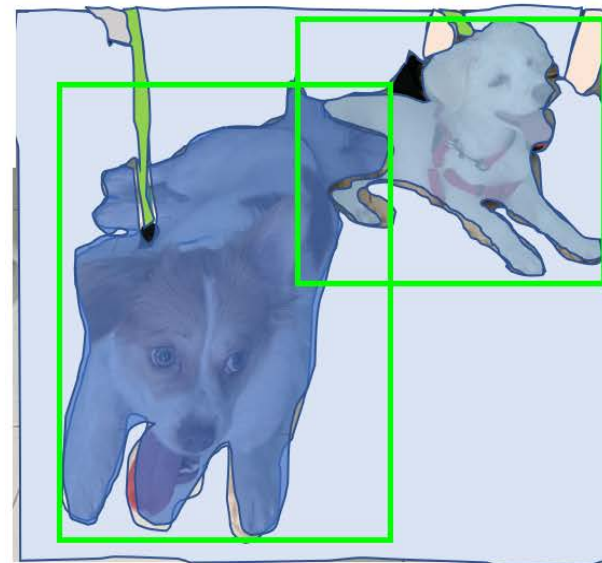
三种图像分割示例



(a) 语义分割



(b) 实例分割



(c) 全景分割

图像分类、目标检测 and 图像分割示意图



8. 图像识别

- ◆ 图像识别（Image Recognition）是指计算机对图像进行处理、分析和理解，以识别各种不同模式的目标和对象的技术。
- ◆ 其**目的**是通过一系列计算机算法或人工智能模型来识别图像中的各类目标（如商品、人脸、自然景观等）的类别、位置、轮廓等信息，从而更好地理解图像。
- ◆ 图像识别就是给定一幅测试图像，识别其中所包含目标的位置及其类别（即目标检测和分类）。
- ◆ 例如，人脸识别是在检测出图像中存在人脸的情况下，根据人脸的特征判断人的身份等信息。
- ◆ 与图像分类相比，图像识别是在同类别下更具体的细分。
- ◆ 图像分类、图像目标检测、图像识别并非完全不同的图像处理任务，三者的任务内容有时是交叠的。

7.2 传统的图像处理技术

- ◆ 图像分类、图像目标检测和图像分割是图像处理的三个核心任务。
- ◆ **图像分类**是对图像进行分类，通常根据图像中存在的目标，将图像归属于某个单一类别（单标签）或多个类别（多标签）。
- ◆ **图像目标检测**是在图像分类的基础上，将图像中所有同类或异类的目标用边界框定位出来。
- ◆ **图像分割**则在目标检测的基础上更进了一步，将整幅图像划分成若干个像素组，对应于各个目标区域，然后对其进行分类，它用目标边缘区分各个目标，而不是边界框。
- ◆ 可以理解为：**图像分类**只是对图像进行单标签标注或多标签标注；**图像目标检测**是在图像分类的基础上再进行粗糙的边界框定位；而**图像分割**则是对图像中所有目标进行分类，且给予精确的轮廓定位。

7.2.1 图像分类

1. 传统图像分类的流程包含4个步骤：

第1步：数据集构建与预处理。

- (1) 获取并整理图像数据集。
- (2) 若需要，对图像进行预处理，如去噪、增强、复原、裁剪、放缩、旋转等操作。
- (3) 将全部数据划分为训练集、验证集（可选）和测试集，保证3个集合互不交叠。

第2步：特征工程。

- (1) 根据应用场景和数据特点人工选择或设计图像特征，常用的经典图像特征有方向梯度直方图（HOG）和尺度不变特征转换（SIFT）等；
- (2) 提取图像特征，每一幅图像由相应的特征向量表示；
- (3) 必要时，还需进行特征变换（如归一化、标准化等），缺失值处理或异常值处理等操作。

7.2.1 图像分类

第3步：分类模型训练。

- (1) 选择适用于应用场景的机器学习分类算法，如支持向量机、随机森林和梯度提升树等。
- (2) 将表示图像的HOG特征向量与其所对应图像的类别标签组成一个个训练样本。
- (3) 将训练集中的HOG特征向量逐一输入到已选择的分类模型中，进行分类预测。
 - 当训练样本的实际输出结果（即预测的类别标签）与其预期结果（即真实的类别标签）不同时，则按照所选择的机器学习算法修正参数。
 - 重复模型预测和参数修正过程，直到满足算法的结束条件，
 - 如分类错误率低于事先设置的某个阈值，本轮训练过程才停止。

7.2.1 图像分类

第4步：分类模型测试与评估。

- (1) 若有验证集，则用验证集评估模型的性能，并根据需要调整超参数。
 - (2) 用测试集评估模型的性能，计算各种评价指标，如分类正确率等。若结果不令人满意，可继续执行第3步，直到达到停止条件，如最多训练1000轮。
- ◆ 至此，已确定分类模型的最终参数。当输入一张新图像时，已训练好的模型便可对其进行自动分类。

图像分类性能的评价指标

表 7.1 二分类问题的混淆矩阵

混淆矩阵		预测类别	
		Label = 0	Label = 1
真实类别	Label = 0	TN	FP
	Label = 1	FN	TP

$$\text{precision} = \frac{TP}{TP+FP}$$

$$\text{recall} = \text{sensitivity} = \frac{TP}{TP+FN}$$

$$\text{specificity} = \frac{TN}{TN+FP}$$

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

5个评价指标:

- ◆ 精度（precision）也称为查准率
- ◆ 召回率（recall）也称为查全率、敏感度（sensitivity）
- ◆ 特异度（specificity）
- ◆ 正确率（accuracy）
- ◆ F1值（F1-score）。

传统图像分类方法的局限性

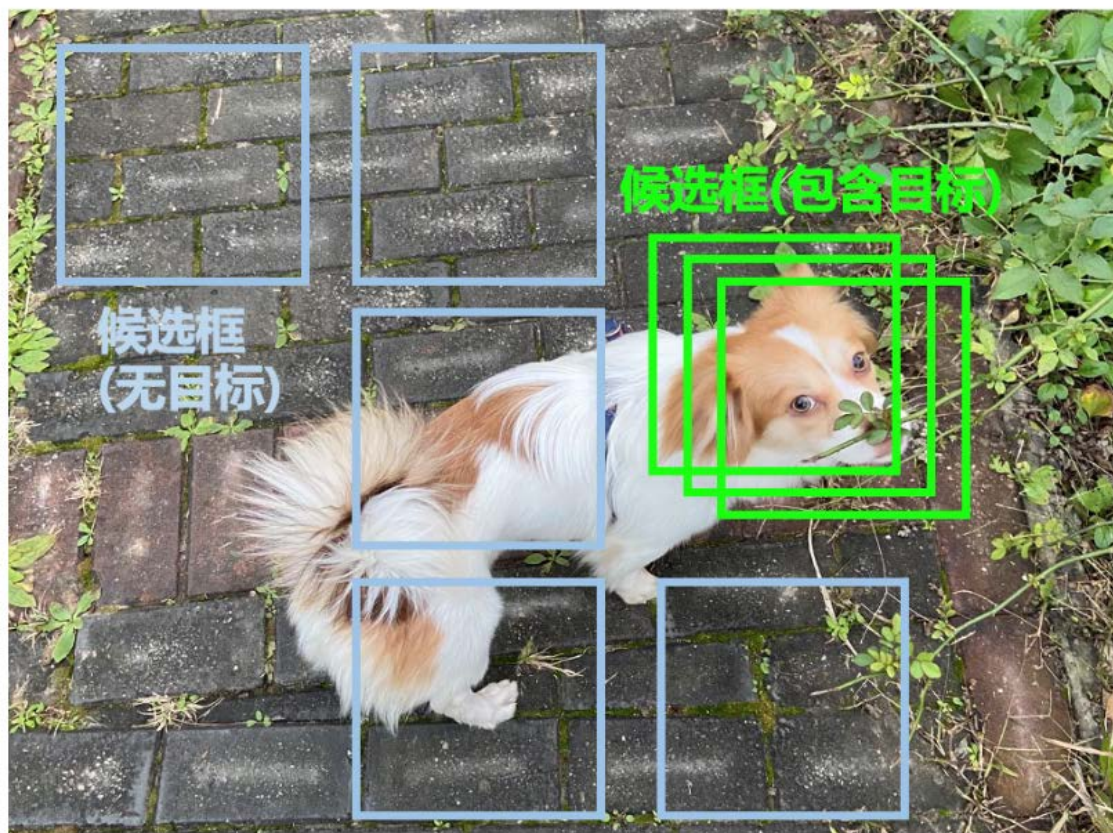
- (1) 图像分类之前，通常都需要耗费大量的时间和精力提取图像特征。
- (2) 需要人工选择或设计图像特征，若选择或设计的图像特征不合适，将导致分类效果不佳。
- (3) 图像特征的提取与分类模型的训练是互相独立的过程，导致在的过程中无法同步优化学习模型参数特征提取的过程，一定程度上限制了传统图像分类模型的性能。

7.2.2 图像目标检测

◆ 传统的图像目标检测方法多采用滑动窗口对整个图像进行遍历搜索。

◆ 这类算法的基本设计思路是：

- 首先使用滑动窗口在图像中生成目标候选区域，
- 然后采用分类算法判断目标候选区域是否包含需要的目标。



7.2.2 图像目标检测

1.传统图像目标检测的流程

第1步：构建数据集，并预处理图像。

- (1) 获取并整理图像数据集。
- (2) 对图像执行降噪、增强等预处理操作，主要目的是消除与检测目标无关的信息，增强相关信息的可检测性，从而提高后续特征提取、图像分割、分类的可靠性。常用的预处理方法有高斯滤波、均值滤波、图像腐蚀和膨胀、二值化等。
- (3) 将全部样本按照某种比例划分为训练集、验证集（可选）和测试集，保证三个集合互不交叠。

1. 传统图像目标检测的流程 (2)

第2步：生成目标候选区域。

- ◆ 采用滑动窗口，以一定的步长对全图进行遍历，生成若干目标候选区域，每个区域都用矩形框标出，记录其位置坐标及高宽。
- ◆ 通常采用穷举式滑窗的方式，即采用各种不同尺寸的滑窗，形成目标候选区域。
- ◆ 若已知某一特定应用领域图像的特点，可根据该领域图像的先验知识确定滑窗（或目标候选框）的大小。

例如，在行人检测中，将滑动窗口和目标候选框大小均设置为 64×128 像素。

第3步：提取目标候选区域的特征。（此为目标检测过程中最重要的步骤）

- ◆ 提取每个目标候选区域的图像特征，用特征向量表示该目标候选区域。
- ◆ 常用的图像特征有颜色特征、纹理特征、形状特征、空间关系特征等，如HOG或SIFT特征，也可以人工设计新的图像特征。

1. 传统图像目标检测的流程 (3)

第4步：训练目标分类器。

- (1) 选择适用于应用场景的机器学习分类算法，如支持向量机、随机森林、AdaBoost、朴素贝叶斯分类器和各种集成分类器等。
- (2) 将表示目标候选区域的特征向量与该区域对应的类别标签组成一个个样本。事先人工标注的包含目标的图像区域为正样本，不包含目标的图像区域为负样本。
- (3) 将训练集中的特征向量逐一输入已选择的分类模型中，进行模型预测。当训练样本的实际输出结果（即预测的类别标签）与其预期结果（即真实的类别标签）不同时，则按照所选择的机器学习算法修正参数。重复模型预测和参数修正过程，直到满足算法的结束条件，如分类错误率低于事先设置的某个阈值，本轮训练过程才停止。

1. 传统图像目标检测的流程 (4)

◆ 为了减少冗余候选框的数量，可采用非极大抑制算法（NMS）进行处理，将包含同一目标的多个候选框合并，仅保留其中可能性或置信度最高的候选框。

◆ NMS算法流程如下。

（1）按照置信度的降序排列候选框。

（2）计算置信度最高的候选框与其余所有候选框的重叠程度。

计算候选框之间的重叠程度时，可采用交并比（Intersection Over Union, IOU）方法，即计算两个候选框A和B的交集面积占并集面积的比例，其公式如下：

$$\text{IOU} = \frac{\text{area}(A) \cap \text{area}(B)}{\text{area}(A) \cup \text{area}(B)}$$

（3）删除重叠程度超过事先设定阈值的其余候选框。

（4）输出置信度最高的候选框，重复上述过程，直至所有候选框均被过滤或输出。

2. 图像目标检测性能的评价指标

(1) 交并比 (IOU)

IOU值越大，说明目标检测算法的定位越准确。设定一个置信度阈值（如0.5），若IOU值大于0.5，则认为目标被“检测成功”，否则被判定“检测错误”。

(2) precision和recall

$$\text{precision} = \text{TP} / (\text{TP} + \text{FP}), \quad \text{recall} = \text{TP} / (\text{TP} + \text{FN})$$

- TP表示 $\text{IOU} > 0.5$ 的检测边界框的数量（针对同一个标签框，只计算一次）；
- FP表示 $\text{IOU} \leq 0.5$ 的检测边界框数量，或是针对同一个标签框的多余检测边界框的数量；
- FN表示没有被检测到的标签框的数量。
- precision和recall是针对单张图像中某一目标类别的精度（准确率）和召回率。

2. 图像目标检测性能的评价指标

(3) AP

- AP是针对图像数据集中单类目标而言的，其值就是precision-recall (PR) 曲线下面的面积。
- 获得PR曲线的过程为：改变置信度阈值，分别计算单类目标的precision和recall值，随着阈值的变化，会得到不同的precision与recall值，从而得到PR曲线。

(4) mAP

mAP是针对整个图像数据集中所有目标类别而言的，其值为所有单类目标的AP值的平均值。mAP的取值为[0,1]，其值越大越好。

(5) 速度评价指标：有2个

- ① FPS (frames per second) 表示检测器每秒能处理图片的张数。
- ② 检测器处理每张图像所需要的时间。

3. 传统图像目标检测的局限性

- (1) 多采用滑动窗口进行遍历搜索，由于目标可能出现在图像的任何位置，且目标大小、长宽比例不确定，因此需要设置各种尺度和长宽比的滑动窗口，导致产生大量冗余的候选区域及其特征，使得算法的时间和空间复杂度都高，在实践中难以真正实现。
- (2) 一般只适用于一些特定场景，且效果表现一般。而现实生活中场景复杂多变，待检测的目标形状与大小不一。目前，尚未有通用的目标检测方法。
- (3) 在特征工程阶段，需要人工选择或设计图像中目标区域的特征。由于研究者的认知不全面或不准确，可能会采用不合适的特征，导致目标分类器的性能不佳。另外，在某种特定的目标检测任务中，针对不同目标（如人和狗）或同一目标的不同形态（如奔跑的狗和趴下的狗），需要选择不同的特征，若采用同一特征，会导致目标分类器的鲁棒性和可移植性较差。

7.2.3 图像分割

- ◆ 图像分割即根据灰度、颜色、亮度、纹理、形状等特征将图像划分为多个不相交的区域，这些区域对应不同类别的目标（物体）或目标的某一部分。
- ◆ 划分的原则是：使区域间呈差异性，区域内呈相似性。
- ◆ 传统的图像分割方法包括：基于阈值的分割方法、基于区域的分割方法、基于边缘的分割方法和基于图论的分割方法等。
- ◆ 这些方法一般都是根据相邻像素之间的相似性和不连续性来设计算法的。其中，相似性是指同类区域中像素的颜色、亮度、纹理等特征分布较接近，不连续性是指不同类别区域之间存在边界。各种分割方法又包括许多不同的算法。
- ◆ 最大类间方差阈值划分（OTSU）算法是一种基于阈值的分割方法。OTSU算法是由日本学者大津展之（Nobuyuki Otsu）于1979年提出的，故又称为大津算法。

1. OTSU算法 (1)

◆ OTSU算法的**设计思路**：通过计算确定一个合适的阈值，用于区分图像中每个像素点是属于目标区域（即前景）还是背景区域，从而产生二值图像，使得背景和前景之间的类间方差最大。

◆ OTSU算法的步骤如下。

第1步：将彩色图像转化为灰度图，图像尺寸为 $M \times N$ ，假设该灰度图中的最小灰度值为 min ，最大灰度值为 max 。

第2步：初始化，令前景和背景之间的类间方差 $V_0=0$ ；令灰度阈值 $t = min+1$ ，灰度值大于等于 t 的像素为前景，反之为背景。

1. OTSU算法 (2)

第3步： 若 $t < max$ ，则统计下列变量：

(1) 前景像素个数记为 N_T ，背景像素个数记为 N_B 。

(2) 计算属于前景和背景的像素在整个图像中所占的比例，分别记为 r_T 和 r_B 。

$$r_T = \frac{N_T}{N_T + N_B}, \quad r_B = 1 - r_T$$

(3) 计算前景像素的平均灰度值为 $\overline{u_T} = \frac{1}{N_T} \sum_{j=1}^{N_T} g_j$ ，背景像素的平均灰度值为

$$\overline{u_B} = \frac{1}{N_B} \sum_{j=1}^{N_B} g_j, \text{ 其中 } g_j \text{ 表示第 } j \text{ 个像素点的灰度值。}$$

(4) 计算图像中全部像素的平均灰度值： $u = r_T \overline{u_T} + r_B \overline{u_B}$ (7.18)

(5) 计算前景和背景之间的类间方差： $V = r_T (\overline{u_T} - u)^2 + r_B (\overline{u_B} - u)^2$ (7.19)

将公式 (7.18) 代入公式 (7.19) 得到： $V = r_T r_B (\overline{u_T} - \overline{u_B})^2$

1. OTSU算法 (3)

第3步： 若 $t < max$ ，则统计下列变量：

(1) 前景像素个数记为 N_T ，背景像素个数记为 N_B 。

(2) 计算属于前景和背景的像素在整个图像中所占的比例，分别记为 r_T 和 r_B 。

$$r_T = \frac{N_T}{N_T + N_B}, \quad r_B = 1 - r_T$$

(3) 计算前景像素的平均灰度值为 $\overline{u_T} = \frac{1}{N_T} \sum_{j=1}^{N_T} g_j$ ，背景像素的平均灰度值为

$$\overline{u_B} = \frac{1}{N_B} \sum_{j=1}^{N_B} g_j, \text{ 其中 } g_j \text{ 表示第 } j \text{ 个像素点的灰度值。}$$

(4) 计算图像中全部像素的平均灰度值： $u = r_T \overline{u_T} + r_B \overline{u_B}$ (7.18)

(5) 计算前景和背景之间的类间方差： $V = r_T (\overline{u_T} - u)^2 + r_B (\overline{u_B} - u)^2$ (7.19)

将公式 (7.18) 代入公式 (7.19) 得到： $V = r_T r_B (\overline{u_T} - \overline{u_B})^2$

第4步： 若 $V > V_0$ ，则令 $V_0 = V$ ， $t = t + 1$ ，转至第3步；否则，算法停止，输出 t 。

2. 图像分割性能的评价指标

(1) 像素准确率。

像素准确率（Pixel Accuracy, PA），也称为**像素精度**，是指被正确分类的像素数占总像素数的比例，表示目标分割的准确度，判断标准为该像素点是否属于要检测

的目标，其计算公式可参见公式：
$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

(2) 类别像素准确率。

类别像素准确率（Class Pixel Accuracy, CPA）是指某一类的像素准确率。

(3) 类别平均像素准确率。

类别平均像素准确率（Mean Pixel Accuracy, mPA）是指各个类别的像素准确率的平均值。需要分别计算每个类中被正确分类的像素数的比例，即每个类别的像素准确率，然后累加，再求平均值。

2. 图像分割性能的评价指标

(4) 交并比。

- 交并比表示模型对某一类别的预测区域和真实区域的交集与并集的比值。
- 对于目标检测而言，是检测框和真实框之间的交并比；
- 对于图像分割而言，是计算预测掩码和真实掩码之间的交并比。
- 以二分类的正例（类别1）为例，预测掩码和真实掩码的交集为TP，并集为TP+FP+FN，计算交并比的公式为： $IOU = TP / (TP + FP + FN)$

(5) 平均交并比。

- 平均交并比（Mean Intersection Over Union, MIOU）是指所有类别交并比的平均值。
- 需要分别计算每个类别的IOU，再求它们的平均值。

7.3 基于深度学习的图像处理技术

- ◆ 采用传统机器学习方法进行数字图像处理，都有一个共同的局限性：需要人工设计与提取图像特征。
- ◆ 这不仅耗费大量的时间和精力提取图像特征，而且在人工设计或选择用于表示图像的特征时，还需要研究者充分了解图像处理任务的特性和应用领域的先验知识，否则会导致图像处理的效果不佳，任务模型的鲁棒性较差。
- ◆ 21世纪初，数字图像处理技术的发展陷入了瓶颈，在图像分类、目标检测、图像识别等实际应用中的效果一直不能令人满意。
- ◆ 直至2012年，基于深度学习的图像处理性能有了明显的提升，图像分类效果甚至超过了人眼的分辨能力。

7.3 基于深度学习的图像处理技术

- ◆ 基于深度学习的图像处理方法与传统的图像处理方法最大的区别在于：
 - 基于深度学习的方法不需要人工设计图像特征，能将图像特征提取与参数学习统一到模型训练的过程中。
 - 深度学习模型能自动提取图像特征，还能够直接通过调整模型网络结构的方式来优化图像特征提取方式。
- ◆ 另外，深度学习模型的训练对数据量和算力的要求比较高，网络结构复杂，训练时间较长；而传统机器学习模型的训练只需要较少的计算资源和较小训练数据量，模型结构相对简单，训练时间较短。

7.3.1 基于深度学习的图像分类

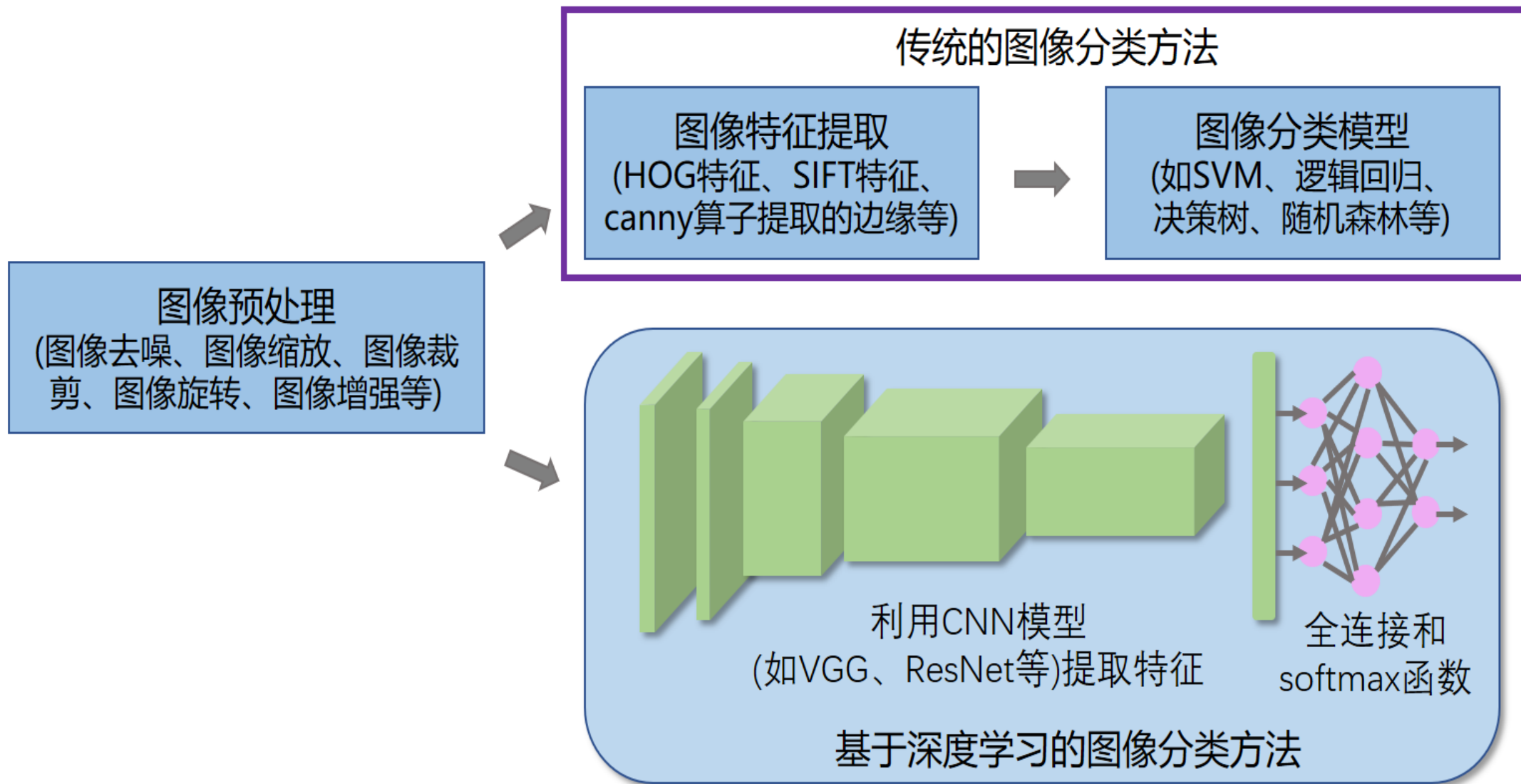


图7.6 基于深度学习的图像分类与传统的图像分类方法的过程对比

7.3.1 基于深度学习的图像分类

采用深度学习技术构建图像分类模型的过程如下。

第1步：数据集构建与预处理。

此过程与传统的图像分类方法的数据集构建与预处理过程一样。

- (1) 获取并整理标注完备的图像数据集。
- (2) 若需要，对图像进行必要的预处理，如去噪、增强、复原、裁剪、缩放、旋转等操作。
- (3) 将全部数据按照某种比例划分为训练集、验证集和测试集，3个集合互不交叠。

7.3.1 基于深度学习的图像分类

第2步：网络模型定义与设置。

设置网络模型有以下3种方法。

- (1) 可以选择已有的网络模型，如AlexNet、VGGNet、GoogLeNet和ResNet等，一旦选定模型，便已知其中的网络结构、损失函数、准确率计算公式和优化方法，但在训练的过程中可以自行修改。
- (2) 可以从零开始自行定义网络模型，包括网络结构、损失函数、准确率计算公式和优化方法。
- (3) 可以采用**基于模型的迁移学习方法**对预训练模型进行微调，构建出新的图像分类模型，即利用新的图像数据在已用其他图像数据训练好的分类模型上进行微调，而不是从零开始训练图像分类模型。

基于模型的迁移学习方法

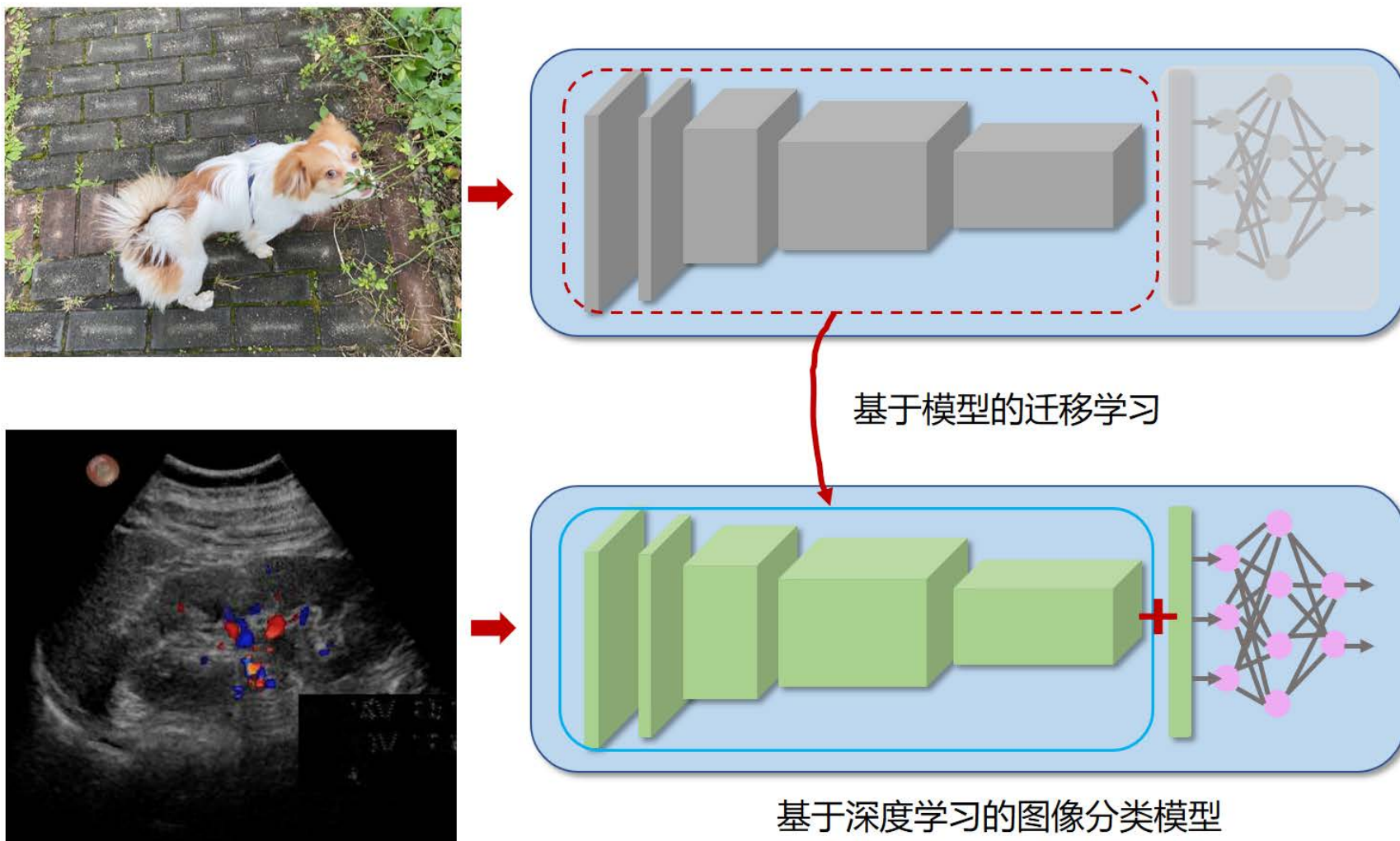


图7.7 采用迁移学习方法构建图像分类模型的过程示意图

7.3.1 基于深度学习的图像分类

第3步：网络模型训练。

采用第1步准备好的训练集训练第2步设置好的模型，还可以用验证集调试合适的超参数，并保存模型。

第4步：网络模型测试与评估。

- 训练好图像分类模型后，采用测试集中的图像数据对该模型进行评估。
- 输入一张测试图像，模型输出该图像属于每个类别的概率。
- 对于二分类（只有正类和负类）问题，分类模型输出一个 $[0,1]$ 的实数值，表示测试图像属于正类的概率，若该概率值大于事先设定的阈值（假设为0.5），则测试图像归入正类，否则归入负类。
- 对于多分类问题，假设类别数为 N ，则分类模型输出 N 个 $[0,1]$ 的实数值，表示测试图像属于每一类的概率，可以选择概率最高的类别作为预测结果。

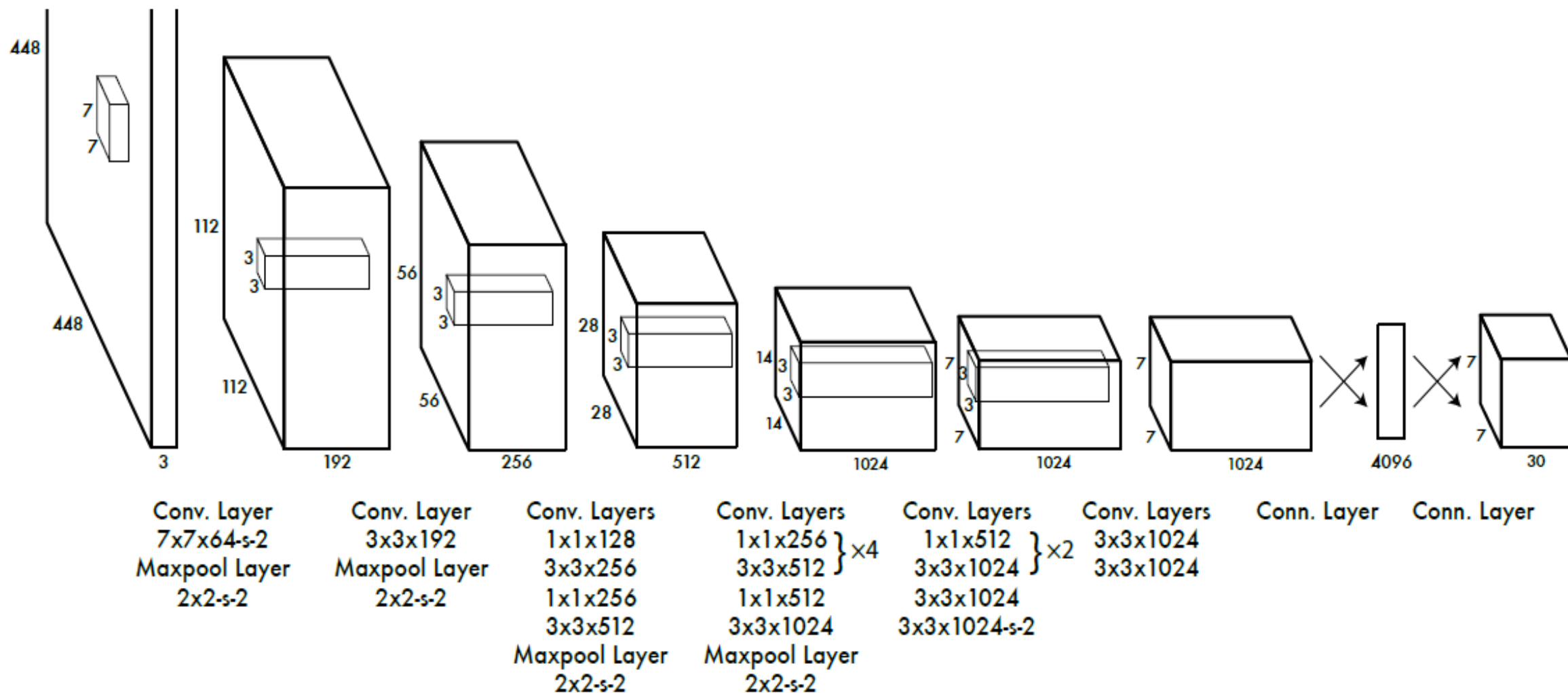
7.3.2 基于深度学习的图像目标检测

- ◆ 图像目标检测需要找到图像中感兴趣目标所在的位置，并判断其所属的类别。
- ◆ 传统的目标检测方法包含3个关键步骤，分别为：
 - 生成目标候选区域，
 - 提取目标候选区域的特征，
 - 训练目标分类器。
- ◆ 基于深度学习的目标检测则在上述3个步骤中采用卷积神经网络等深度学习模型改进和优化其性能，同时将原来独立的各个步骤整合到统一的训练过程中，实现端到端（End To End）的建模。
- ◆ 2016年，提出**YOLO 模型**，“You Only Look Once”（您只看一次）。
- ◆ YOLO的主要特点是将整个图像的目标定位和分类过程集成到一个CNN网络中，不需要产生候选区域，直接将目标边界框定位问题转化为回归（Regression）问题。

1. YOLO模型结构

- ◆ YOLO模型是受GoogLeNet模型的启发，采用包含4个分支的Inception模块，构建了拥有 24 个卷积层和 2 个全连接层的网络结构。
- ◆ 但不同于 GoogLeNet的是：YOLO的Inception模块中只有 3×3 卷积层，而没有 5×5 的卷积层。
- ◆ 3×3 卷积层之前的 1×1 卷积是降维层，目的是降低前一卷积层输出特征的维度，YOLO的输入图像尺寸为 224×224 ，预训练ImageNet分类任务的卷积层，然后使用尺寸为 448×448 的图像进行目标检测训练。
- ◆ 针对一张测试图像，YOLO模型最终输出的预测结果为一个 $7 \times 7 \times 30 = 1470$ 维的张量（Tensor）。

1. YOLO模型结构

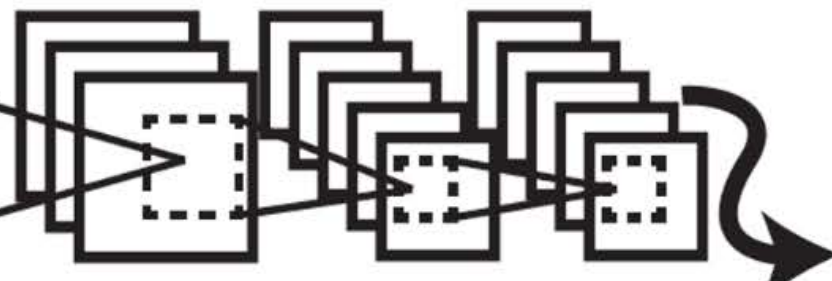
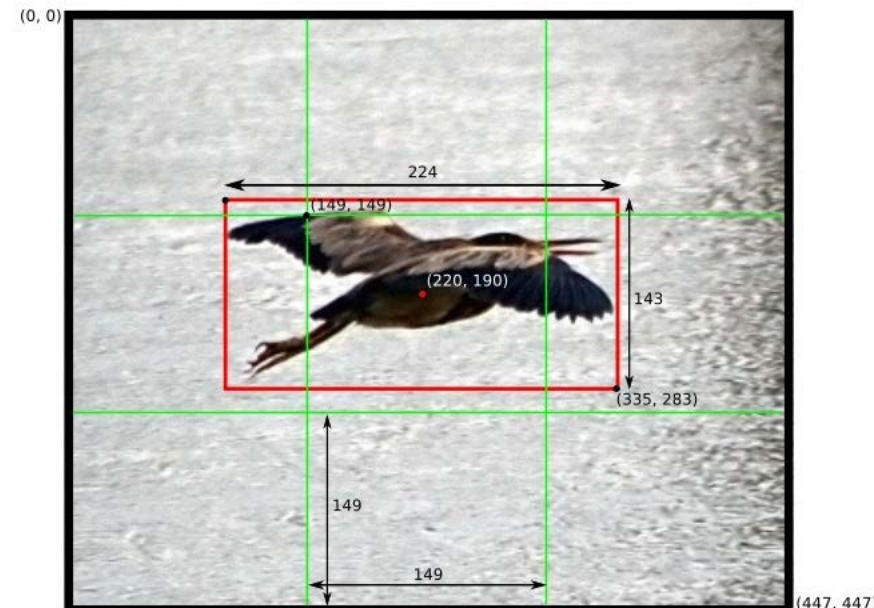


2. YOLO的工作流程

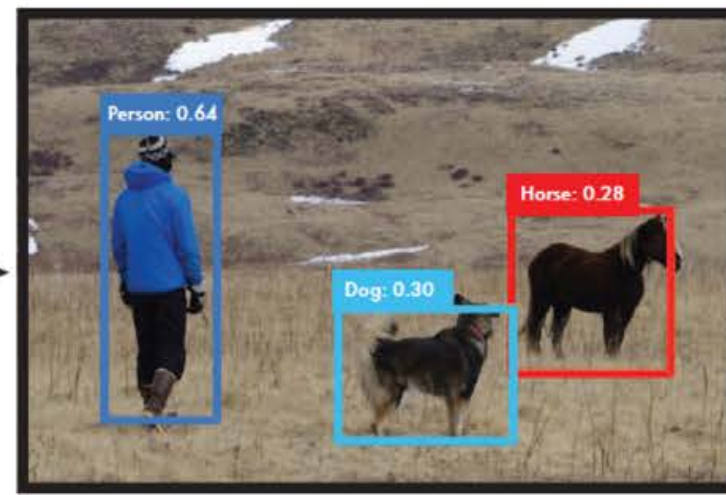
- (1) 构建数据集，并预处理输入图像；
- (2) 预测所有网格对应的边界框位置，并对网格中的目标进行分类。

- ① 将图像分割为 $S \times S$ 的网格。
- ② 计算预测边界框包含某类别目标的概率。
- ③ 归一化预测边界框的位置信息。
- ④ YOLO模型中超参数的设置与输出。

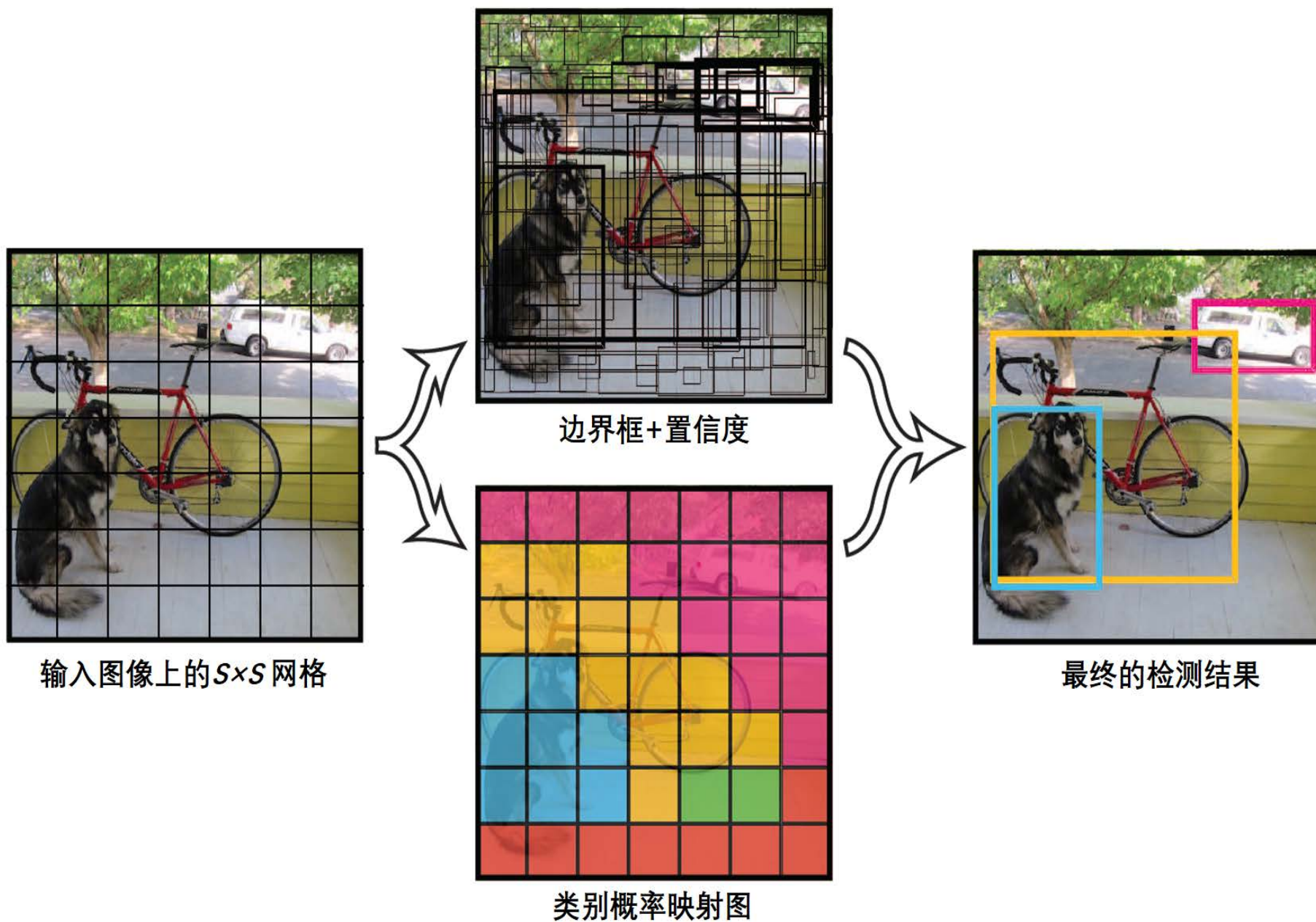
- (3) 非极大值抑制:执行NMS算法，去除重叠的边界框。



1. 调整图的尺寸
2. 运行卷积神经网络
3. 执行非极大值抑制算法



YOLO网络目标检测过程示意图



3. YOLO模型的损失函数

$$\begin{aligned} \text{loss} = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{\text{obj}} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\ & + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbf{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{S^2} \mathbf{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \end{aligned}$$

边界框中心位置
与宽高的误差

边界框包含目标与不包
含目标的置信度误差

网格单元的分类误差

4. YOLO模型的优势与不足

(1) 优势

- ① YOLO误识背景的数量远远少于其他系统，不到Fast R-CNN的一半。
- ② 端到端的预测：只需经过一个CNN处理一次图像，便可同时预测目标位置和类别。
- ③ 目标检测速度快，适合在资源紧张、实时性要求高的场景中使用。
- ④ 泛化能力强，容易进行迁移学习。

(2) 不足

- ① 每个网格仅预测了两个边界框，YOLO对小目标的检测效果不好。
- ② YOLO划分网格的方式较粗糙，目标边界框的定位不是很准确。
- ③ 当目标的长宽比非常大（如超过105）时，YOLO无法检测出这样的目标。

7.3.3 基于深度学习的图像分割

- ◆ 2015年有学者提出了全卷积网络（Fully Convolutional Network, FCN）使用反卷积实现了图像像素级分类。
 - 该模型在PASCAL VOC图像分割数据集上平均交并比为62.2%，比之前方法的性能提升了20%
- ◆ 同年，德国弗莱堡大学的奥拉夫·龙内贝格（Olaf Ronneberger）等科研人员基于FCN 提出了U-Net网络
 - U-Net网络是一款专门为生物医学图像分割而开发的CNN，即使在医学图像分割数据量相对较少的情况下，也取得了不俗的成绩。
 - 在IEEE国际生物医学影像研讨会的细胞追踪挑战比赛中取得了第一名。

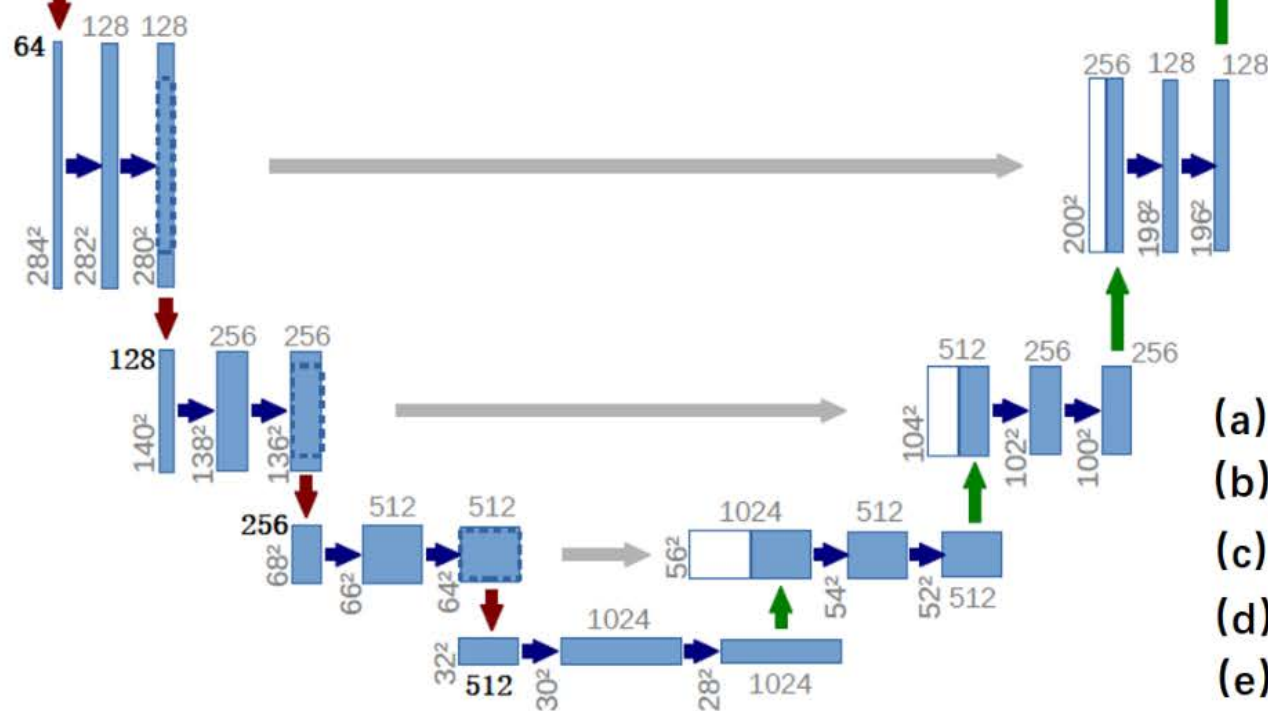
7.3.3 基于深度学习的图像分割



包含两部分：

◆ 编码——收缩路径

◆ 解码——扩展路径



不同的操作

(a) ➡ conv 3x3, ReLU

(b) → copy and crop (复制和裁剪)

(c) max pool 2x2

(d)  up-conv 2x2 (上卷积)

(e) \Rightarrow conv 1x1

U-Net模型

◆ 编码——收缩路径

- 包含5个卷积块，每个卷积块都包括两个操作(a)，即两个（Conv 3×3 + ReLU）的组合运算。
- 相邻两个卷积块之间都执行一次最大池化操作(c)，相当于进行下采样，使得特征图的宽和高均变为前一层尺寸的1/2。
- 为了避免特征信息丢失，在下一个卷积块中加倍特征图的通道数。

U-Net模型

◆ 解码——扩展路径

- 包含5个卷积块，只是与收缩路径共用U型底部的卷积块，即扩展路径上的第一个卷积块就是收缩路径上的第五个卷积块。
- 扩展路径上每个卷积块的卷积运算和激活函数操作同收缩路径上的卷积块完全一样。
- 相邻两个卷积块之间都执行一次 2×2 的上采样操作(d)，使得特征图的宽和高均变为前一层尺寸的2倍，但通道数减半。
- 增加了一个输出层，执行2个 1×1 卷积操作(e)，得到像素分类图像。

◆ **扩展路径与收缩路径不同**的是：在扩展路径中，执行每个卷积块之前，需要先执行(b)操作，即将两部分特征图拼接（Concatenate）在一起。

基于深度学习图像分割的过程

第1步：数据集构建与预处理

- 需要有标注完备的图像分割数据集，并按比例划分为不相交的训练集、验证集和测试集，以作为训练、测试、评估模型的实验载体。
- 必要时，还需要进行图像增强操作，以提高图像质量和增加数据量。

第2步：网络模型定义与设置：设置网络模型有以下3种方法。

- (1) 选择已有的网络模型，如FCN、U-Net等，一旦选定模型，便已知其中的网络结构、损失函数、准确率计算公式和优化方法，但在训练的过程中可以自行修改。
- (2) 从零开始自行定义网络模型。
- (3) 采用模型迁移学习方法对预训练模型进行微调，从而构建出新的图像分割模型。

基于深度学习图像分割的过程

第3步：网络模型训练

采用第1步准备好的训练数据集，训练第2步设置好的网络模型，还可以用验证集调试合适的超参数，并保存模型。

第4步：模型测试与评估

- 训练好图像分割模型后，采用测试集中的图像数据对该模型进行评估。
- 计算出类别平均像素准确率、交并比和平均交并比等评价指标值，用于指导模型结构和优化的改进。

7.4 本章小结

◆ 数字图像处理概述

图像、数字图像、数字图像的种类、数字图像处理、图像处理的主要任务

◆ 传统的数字图像处理技术

- 图像分类、图像目标检测、图像分割等不同图像处理任务的步骤
- 图像分类、图像目标检测、图像分割等不同图像处理任务的评价指标

◆ 基于深度学习的图像处理技术

- 基于深度学习的图像处理技术的步骤
- 基于深度学习的图像处理技术的优势