

图像分类

作者: Calvin

QQ: 179209347

Mail: 179209347@qq.com

介绍

笔记简介:

- 面向对象: 深度学习初学者
- 依赖课程: **线性代数, 统计概率**, 优化理论, 图论, 离散数学, 微积分, 信息论

知乎专栏:

<https://zhuanlan.zhihu.com/p/693738275>

Github & Gitee 地址:

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

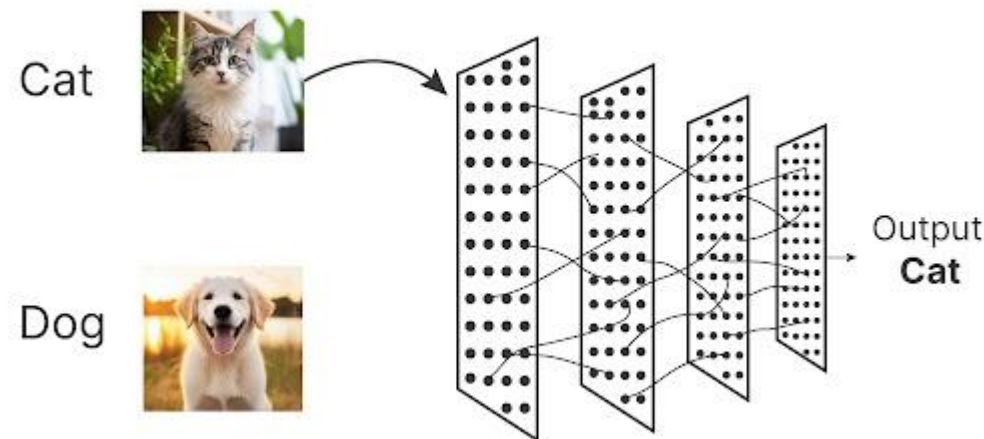
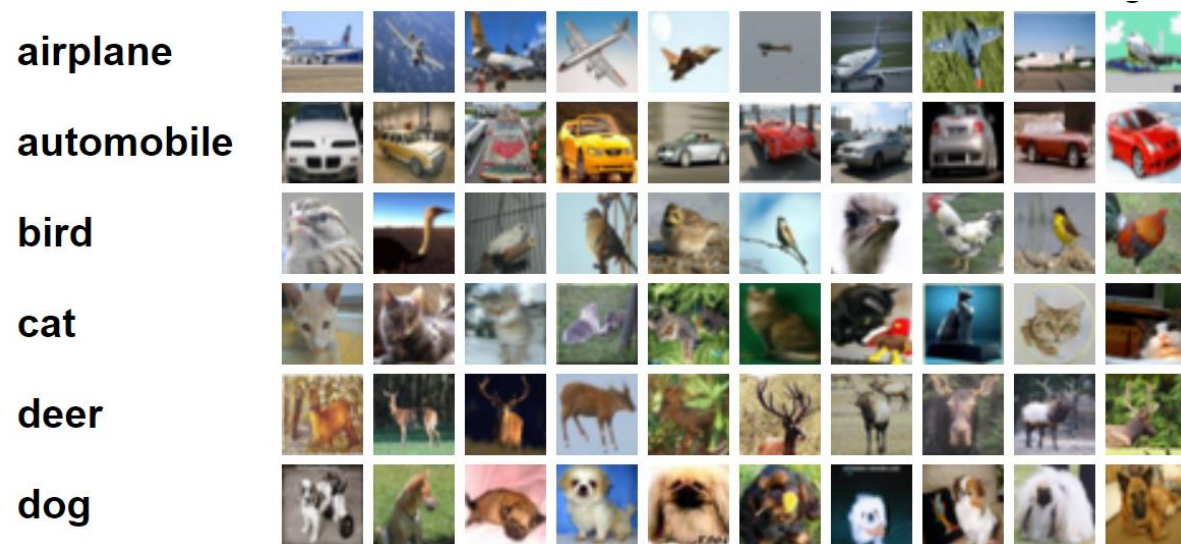
* 版权声明:

- 仅限用于个人学习
- 禁止用于任何商业用途

什么是图像分类

图像分类是计算机视觉领域中的一个重要任务，其目标是将输入的图像分配到预定义类别中。这个任务通常涉及使用机器学习和深度学习技术来训练模型，使其能够自动识别和分类图像。

在图像分类中，通常会使用大量带有标签的图像数据集来训练模型。这些标签指明了每张图像所属的类别，例如猫、狗、汽车等。训练过程中，模型会学习从输入图像中提取特征，并将这些特征映射到正确的类别标签上。

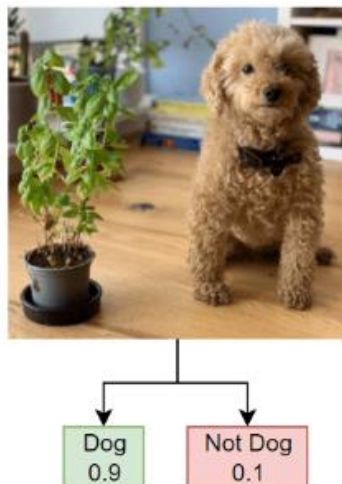


图像分类任务典型领域划分

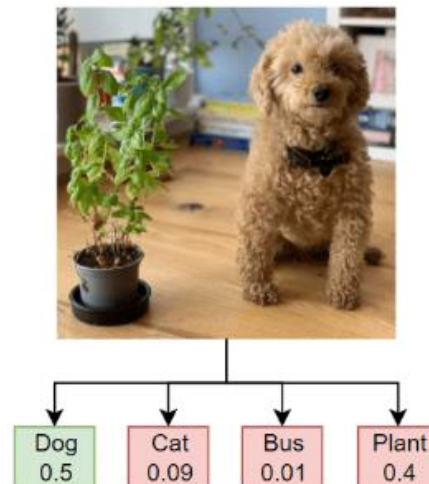
典型领域划分

- 二分类图像分类
- 多类别图像分类
- 多标签图像分类
- 细粒度图像分类

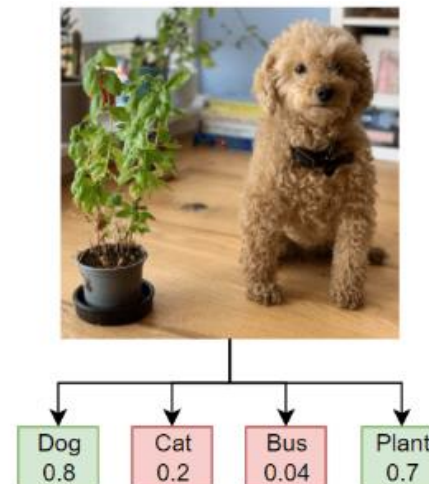
二分类



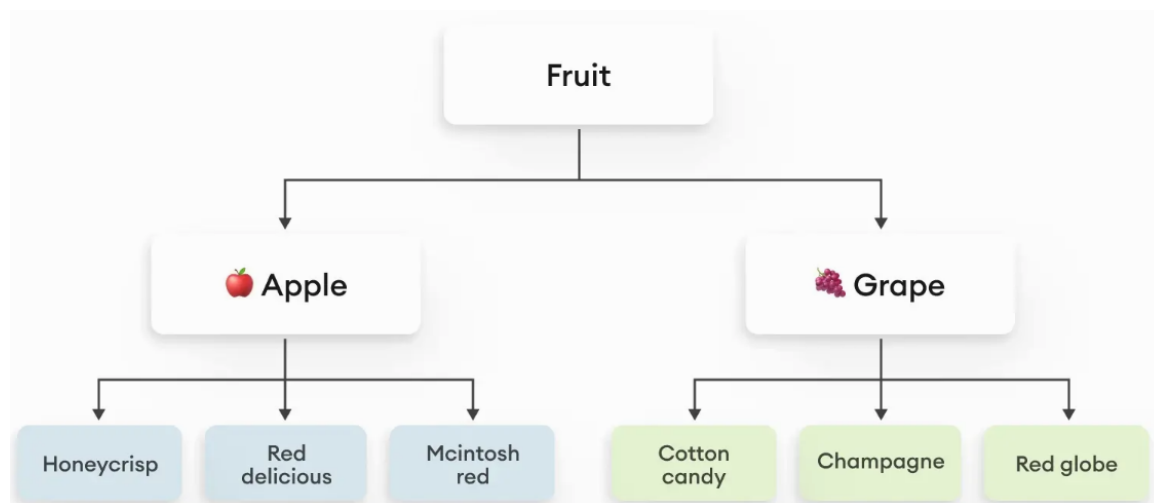
多类别分类



多标签分类



细粒度图像分类

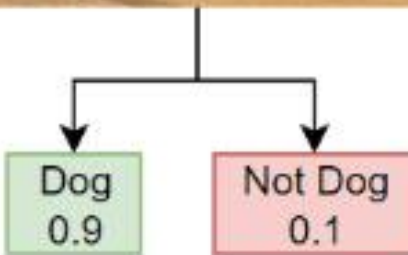


二分类图像分类

定义：二分类图像分类是指将图像分为两个类别的任务。。

示例：对于一组动物图像，模型需要将它们分为狗、非狗的两个类别。

二分类

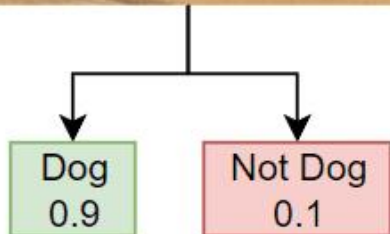


多类别图像分类

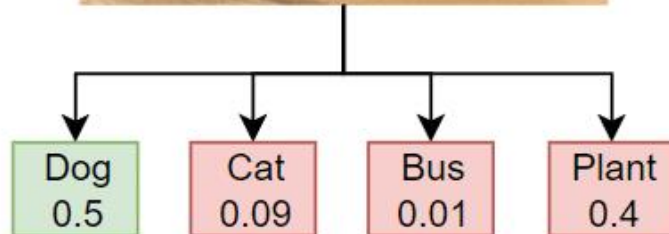
定义： 在多类别图像分类任务中，模型需要将输入的图像分为多个不同的类别中的一个。

示例： 对于一组动物图像，模型需要将它们分为狗、猫、鸟等不同的类别。

二分类



多分类

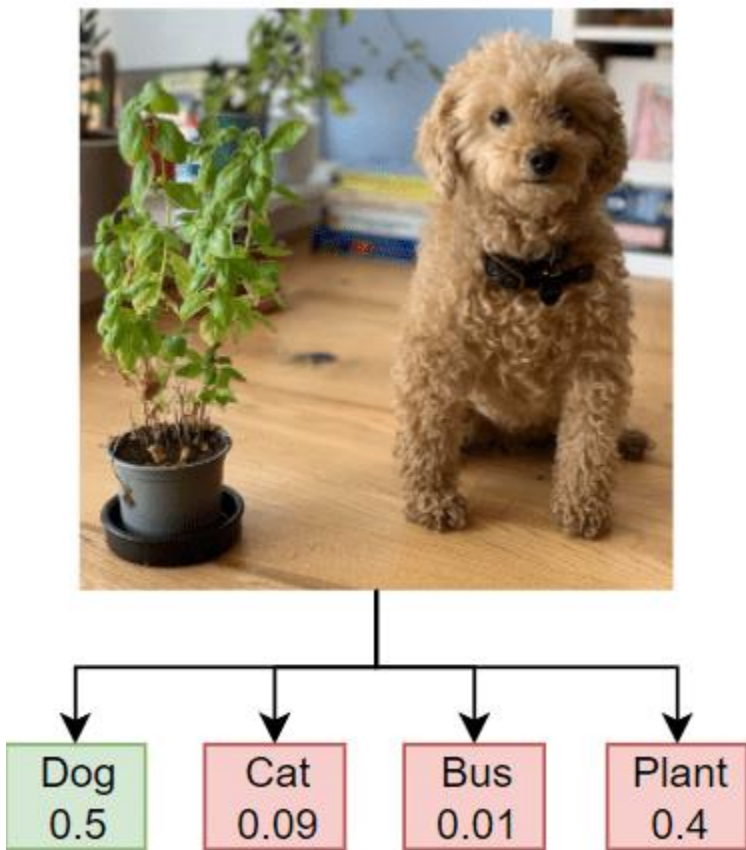


多标签图像分类

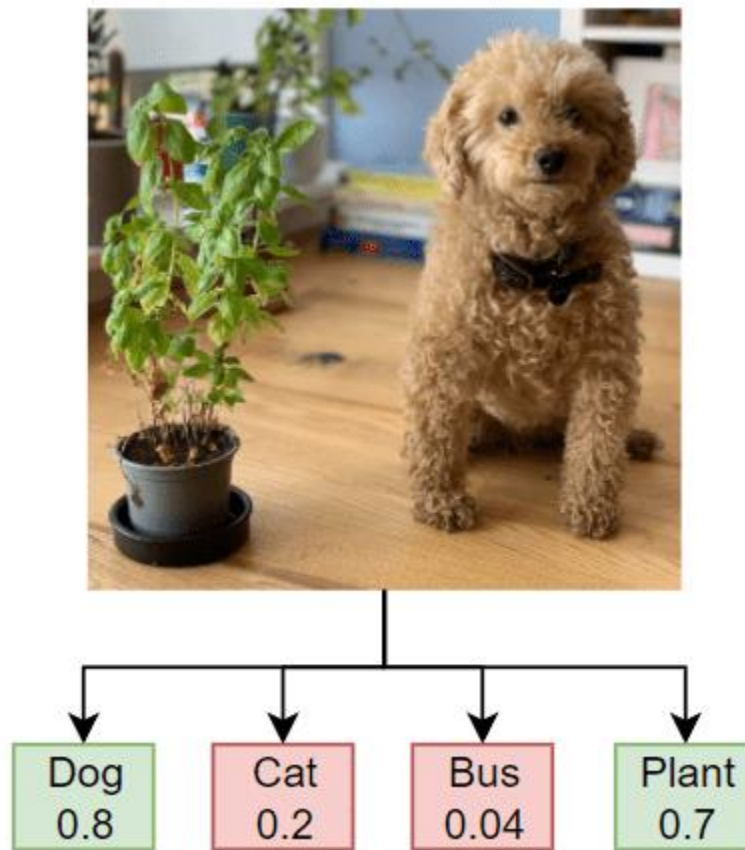
定义：在多标签图像分类任务中，每个图像可以被分配多个标签，而不仅仅是单个类别。

示例：对于一组包含多种物体的图像，模型需要为每个物体分配一个标签，例如在一张包含狗和树的图像中，模型需要识别出狗和树这两个标签。

多类别分类



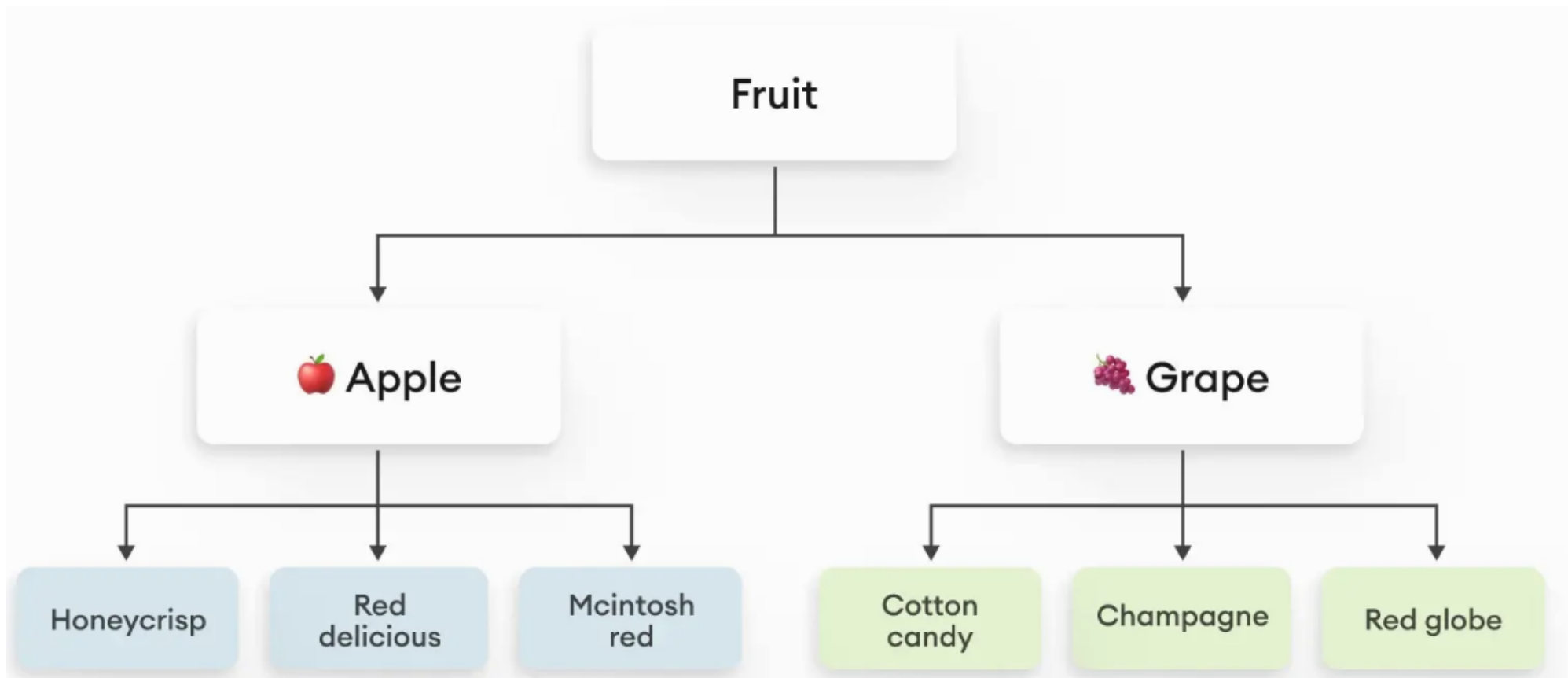
多标签分类



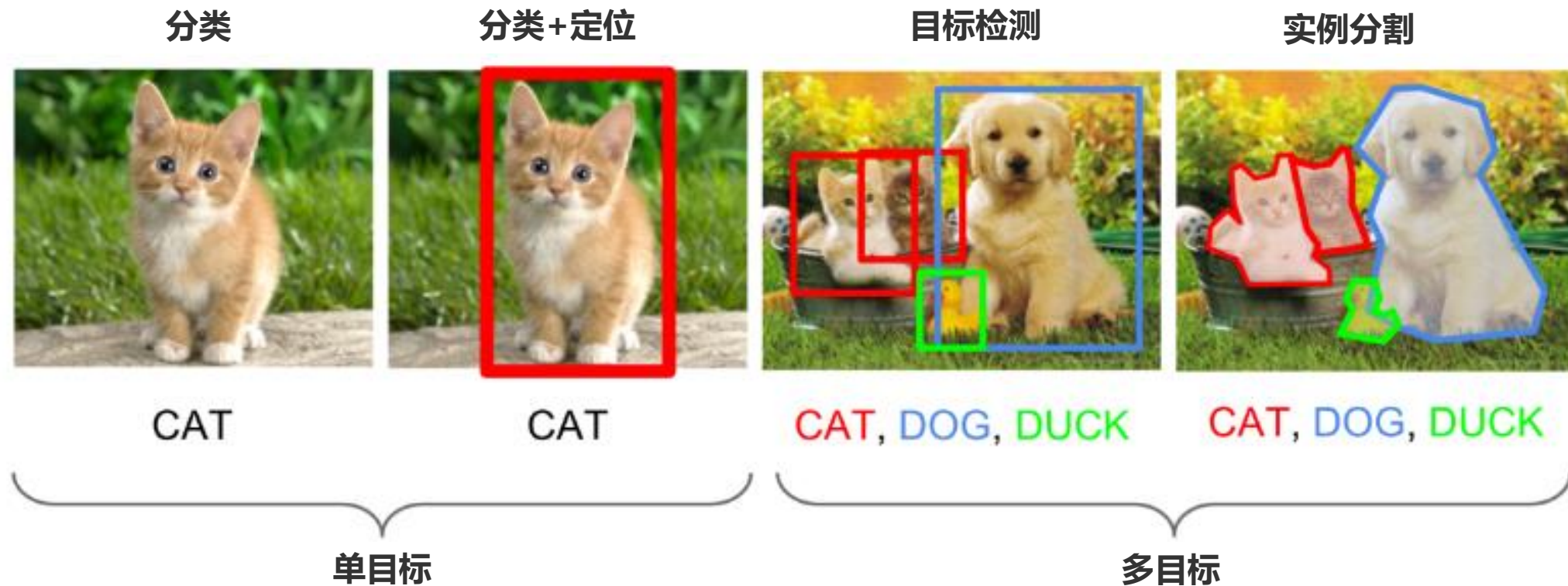
细粒度图像分类

定义：细粒度图像分类任务涉及区分属于同一类别但具有细微差别的图像，通常需要更高级的视觉辨识能力。

示例：在水果图像分类中，细粒度图像分类要求模型能够区分不同种类的水果，如区分不同种类的苹果和葡萄。



分类，目标检测，实例分割区别

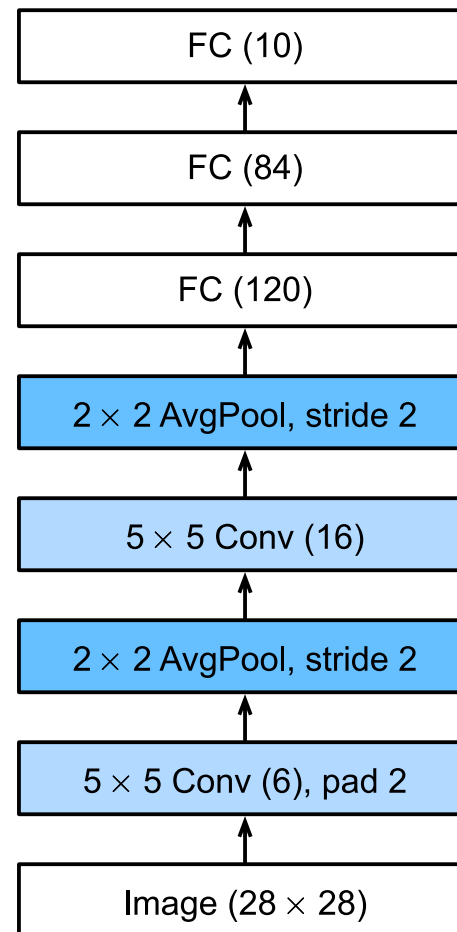


图像分类算法：LeNet

LeNet 是由Yann LeCun等人在1998年提出的经典卷积神经网络架构，是深度学习领域中的重要里程碑之一。LeNet-5主要用于手写数字识别任务，是最早用于数字识别的卷积神经网络之一。

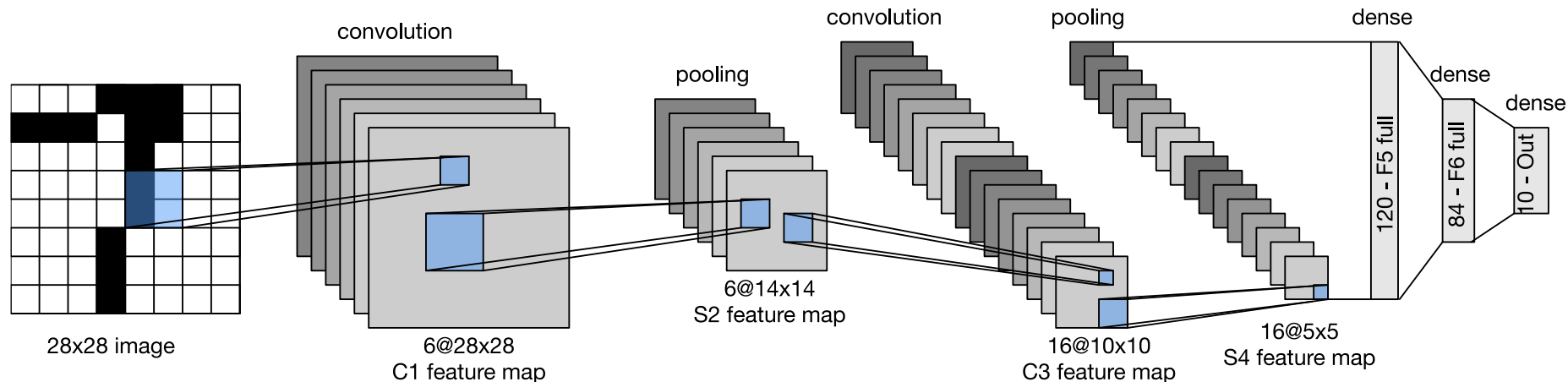


杨立昆
Yann LeCun



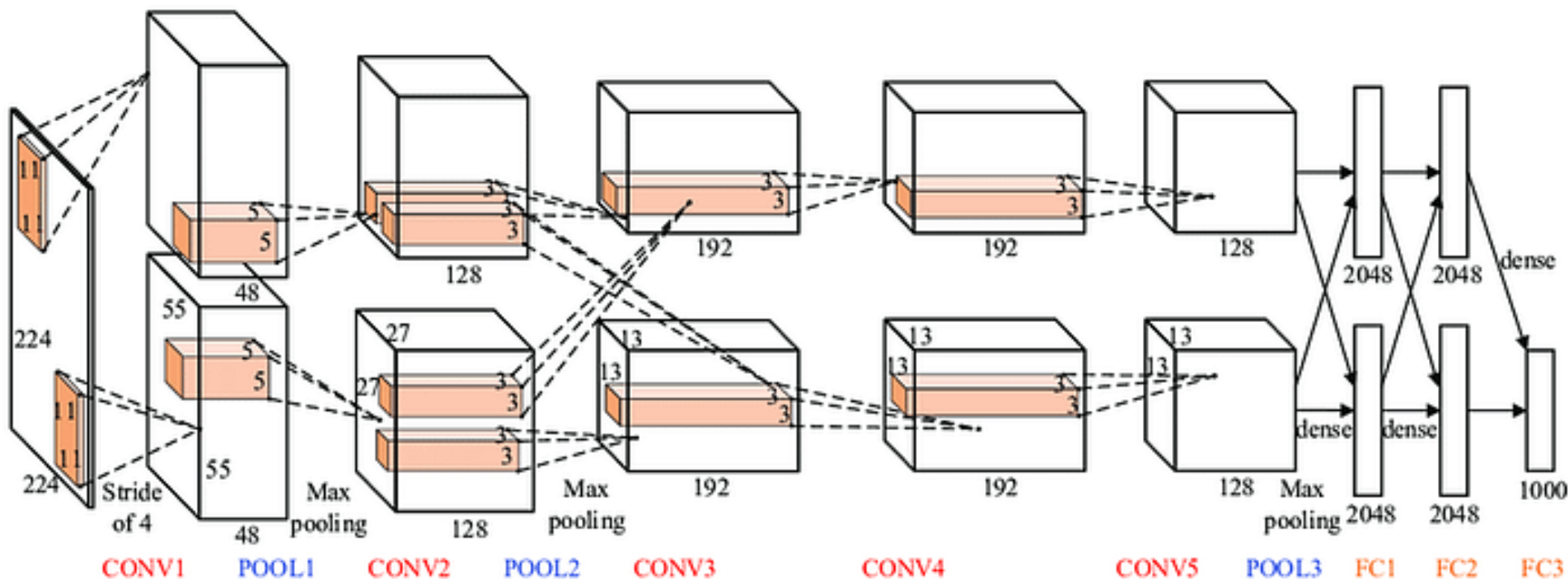
图像分类算法：LeNet 网络结构

输入层	输入为 32×32 的灰度图
卷积层C1	6个 5×5 的 VALID 卷积核-，填充为2，输出的特征图的尺寸为28×28，深度为6。
池化层S2	采用 2×2 最大池化模式，得到 6 个 14×14 的特征图。
卷积层C3	该层将输入与16个大小为5×5的卷积核进行VALID卷积运算，输出为 16 个 10×10 的特征图。
池化层S4	采用 2×2 最大池化模式，得到 16 个 5×5 的特征图。
卷积层C5	120 个 5×5 的 VALID 卷积核，输出为 120 个 1×1 的特征图。
全连接层F6	84 个神经元与 C5 层全连接。
输出层	由于手写数字识别处理的是 0~9 的 10 分类问题，因此，该层有10个输出。



图像分类算法： AlexNet

AlexNet是由Alex Krizhevsky、Ilya Sutskever和Geoffrey Hinton设计的深度卷积神经网络架构，是在2012年ImageNet大规模视觉识别挑战赛（ILSVRC）中取得突破性成果模型。AlexNet在当时引领了深度学习领域的发展，证明了深度卷积神经网络在图像识别任务上的有效性。

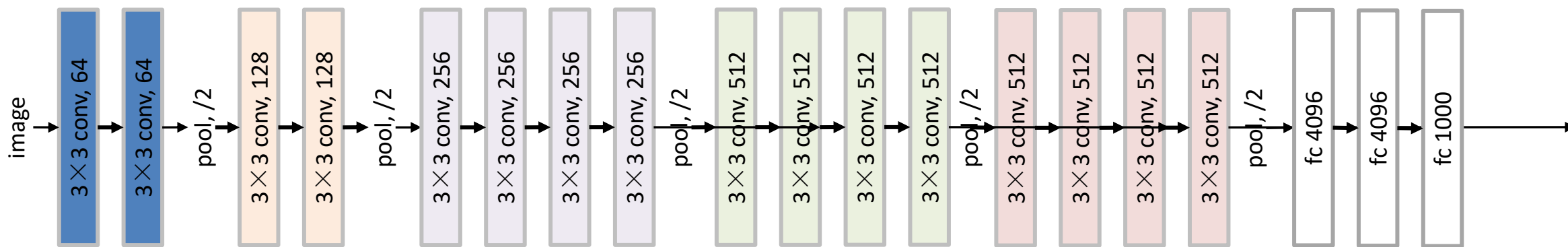


图像分类算法：VGGNet

VGGNet是由牛津大学计算机视觉组（Visual Geometry Group）提出的深度卷积神经网络架构，是2014年ImageNet大规模视觉识别挑战赛（ILSVRC）的冠军之一。VGGNet的主要贡献在于展示了通过增加网络深度（层数）可以提高模型性能的重要性。

VGGNet的一些关键特点：

- **深度**：VGGNet采用了较深的网络结构。
- **卷积层和池化层的堆叠**：卷积层用于提取图像特征，池化层用于降低特征图的维度。
- **卷积核大小为3x3**：有助于减少参数数量，同时保持有效的感受野。
- **全连接层**：全连接层用于最终的分类。
- **使用ReLU激活函数**：ReLU 作为激活函数，有助于加速模型的收敛。
- **模型简单且易复现**：VGGNet的结构相对简单，易于理解和复现。



VGG-19

图像分类算法： VGGNet 网络结构

根据卷积核大小与卷积层数目不同，VGG可以分为6种子模型，分别对应的模型为：

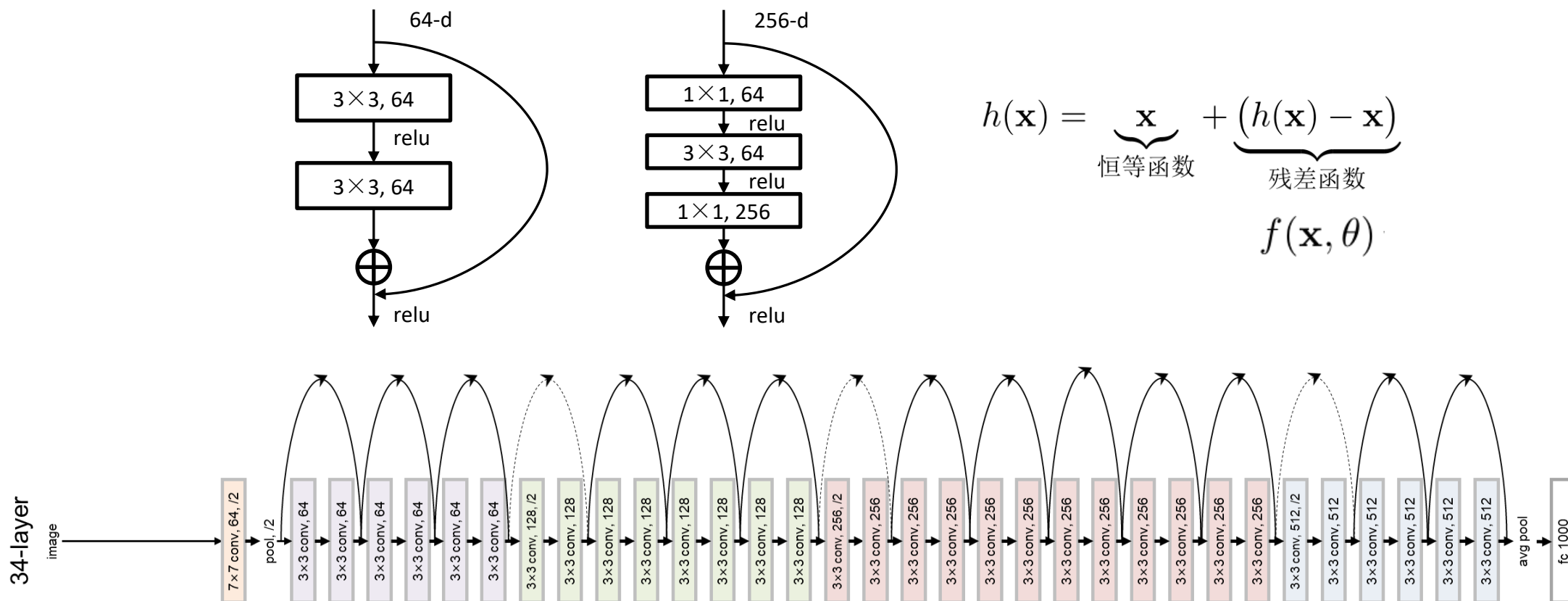
- VGG11
- VGG11-LRN
- VGG13
- VGG16-1
- VGG16-3
- VGG19

不同的后缀代表不同的网络层数。
VGG16-1表示后三组卷积块中最后一层卷积采用卷积核尺寸为1*1。VGG19位后三组每组多一层卷积，VGG19为3*3的卷积。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

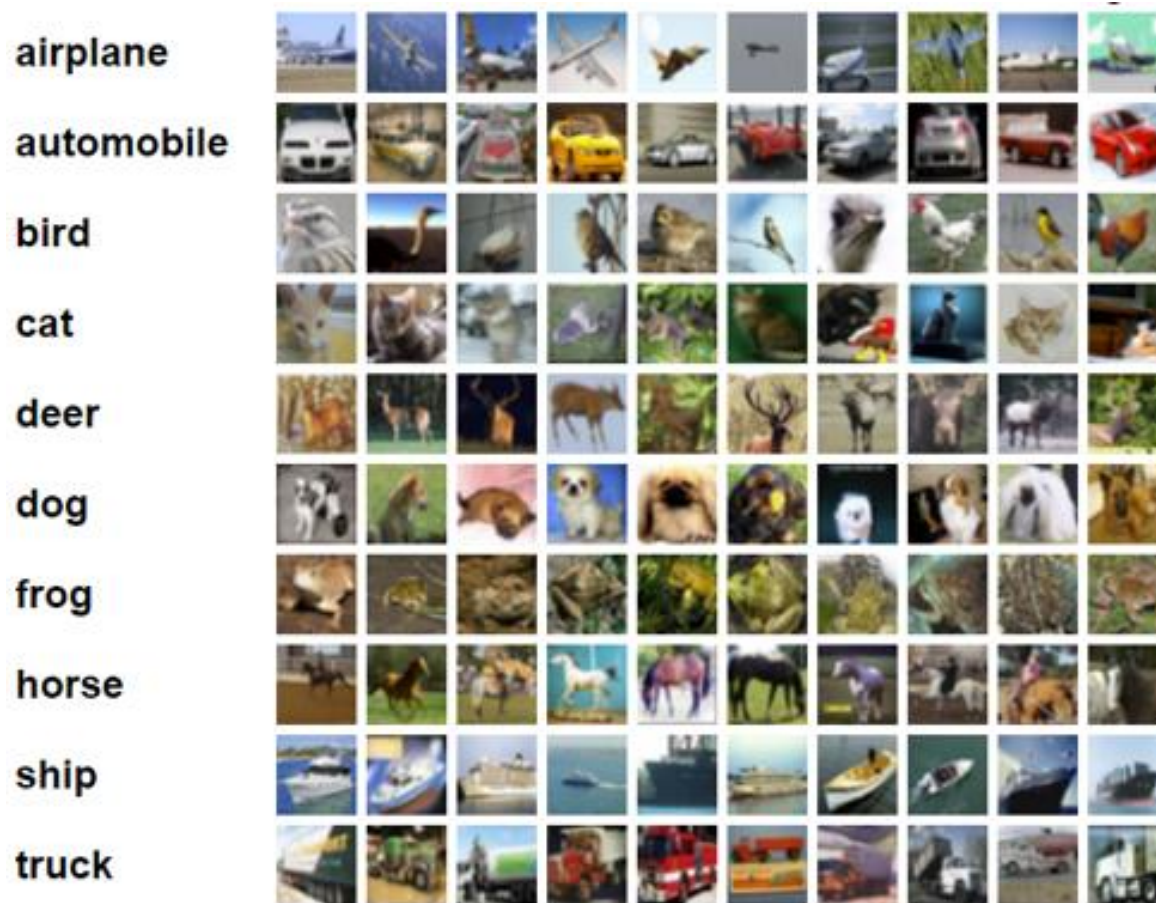
图像分类算法：ResNet

ResNet (Residual Network) 是由微软亚洲研究院的研究员提出的一种深度神经网络架构，旨在解决深度神经网络训练过程中的梯度消失和梯度爆炸问题。ResNet通过引入残差学习的概念，使得网络可以更轻松地训练非常深的层次，甚至超过1000层。



经典数据集：CIFAR-10

CIFAR-10数据集包含10个类别的60000张32x32彩色图像，每个类别有6000张图像。其中有50000张训练图像和10000张测试图像。



Accuracy

Model	Acc.
VGG16	92.64%
ResNet18	93.02%
ResNet50	93.62%
ResNet101	93.75%
ResNeXt29(32x4d)	94.73%
ResNeXt29(2x64d)	94.82%
DenseNet121	95.04%
PreActResNet18	95.11%
DPN92	95.16%

经典数据集： ImageNet

ImageNet是一个大型视觉识别数据库，它包含超过1400万张手动标记的图像，涵盖了超过2万个类别。ImageNet数据库最初由斯坦福大学的李飞飞教授团队创建，旨在推动计算机视觉领域的发展。

ImageNet数据集的特点：

- **大规模性**： ImageNet包含超过1400万张图像，涵盖了超过2万个类别。
- **手动标记**： 每张图像都经过了人工标记，确保了数据集的准确性和可靠性。
- **多样性**： 包含各种各样的图像，涵盖了不同类别、场景和视角，使得模型在各种情况下都能进行有效的识别。
- **用途广泛**： ImageNet数据集不仅仅用于ImageNet挑战赛，还被广泛应用于训练和评估各种图像识别算法，成为计算机视觉领域的重要基准数据集之一。
- **推动研究**： 推动了深度学习和计算机视觉领域的发展，为研究人员提供了一个共同的平台，促进了算法的创新和进步。

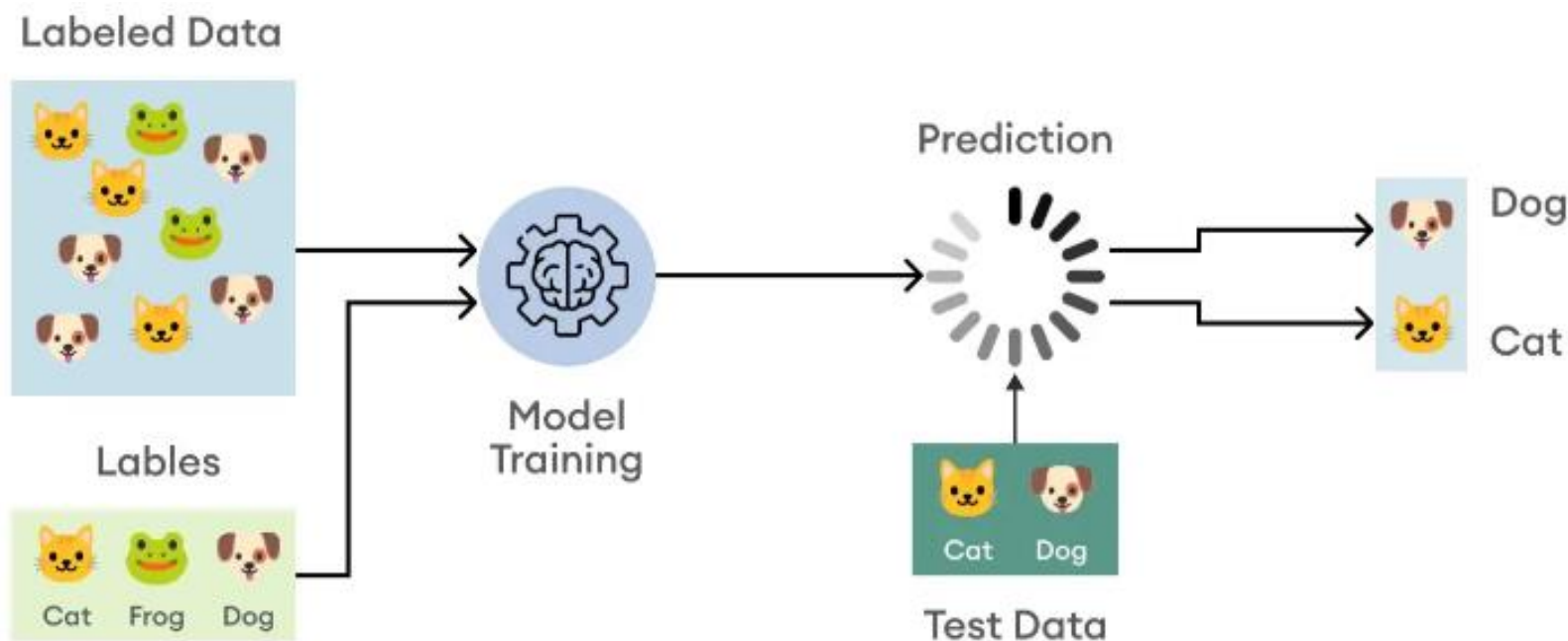


图像分类模型训练推理流程

图像分类模型的训练和推理流程通常包括以下步骤：

训练阶段：

- 数据收集与预处理
- 选择模型架构
- 模型训练
- 模型评估
- 模型优化
- 模型保存



推理阶段：

- 数据预处理
- 加载模型
- 模型推理
- 后处理
- 结果输出



Thank

You