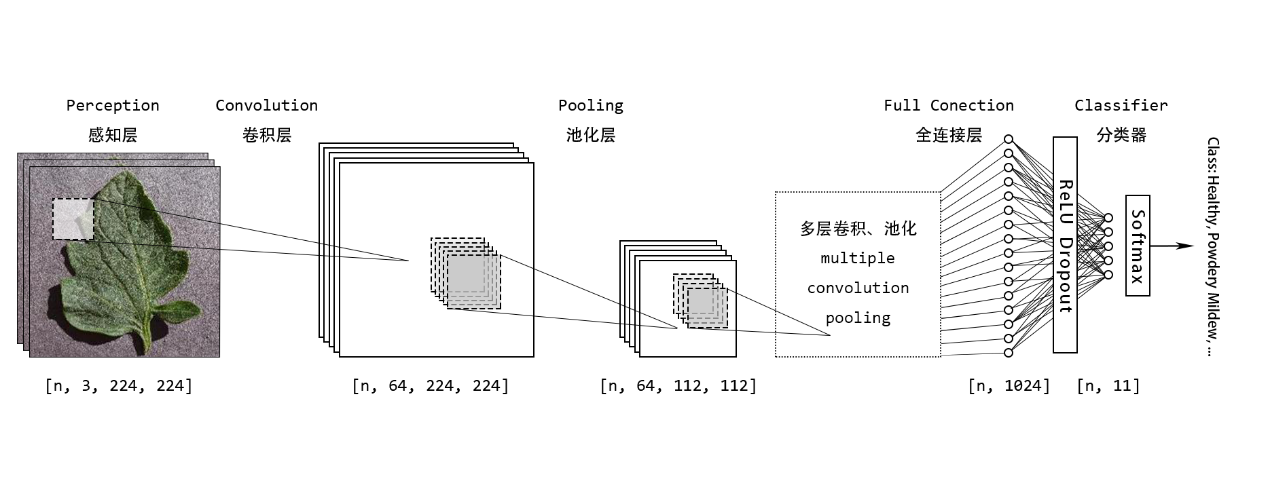
五、研究方法

**（一）卷积神经网络（Convolution Neural Network）**

**在过去的几年内，深度学习（DL）逐渐成为机器学习领域（ML）中应用最广泛的计算方法。深度学习模型在众多复杂的认知任务中取得了很好的效果，在很多时候其表现甚至超越人类，图像识别与处理便是深度学习模型的一个重要应用场景。卷积神经网络是计算机视觉领域最常用的深度学习模型，其在包括**图像分类和分割、目标检测、视频处理、自然语言处理和语音识别等领域应用广泛。本次研究主要使用用于图像分类的卷积神经网络，这种神经网络的基本结构如图所示：



1. **感知层（Perception Layer）**

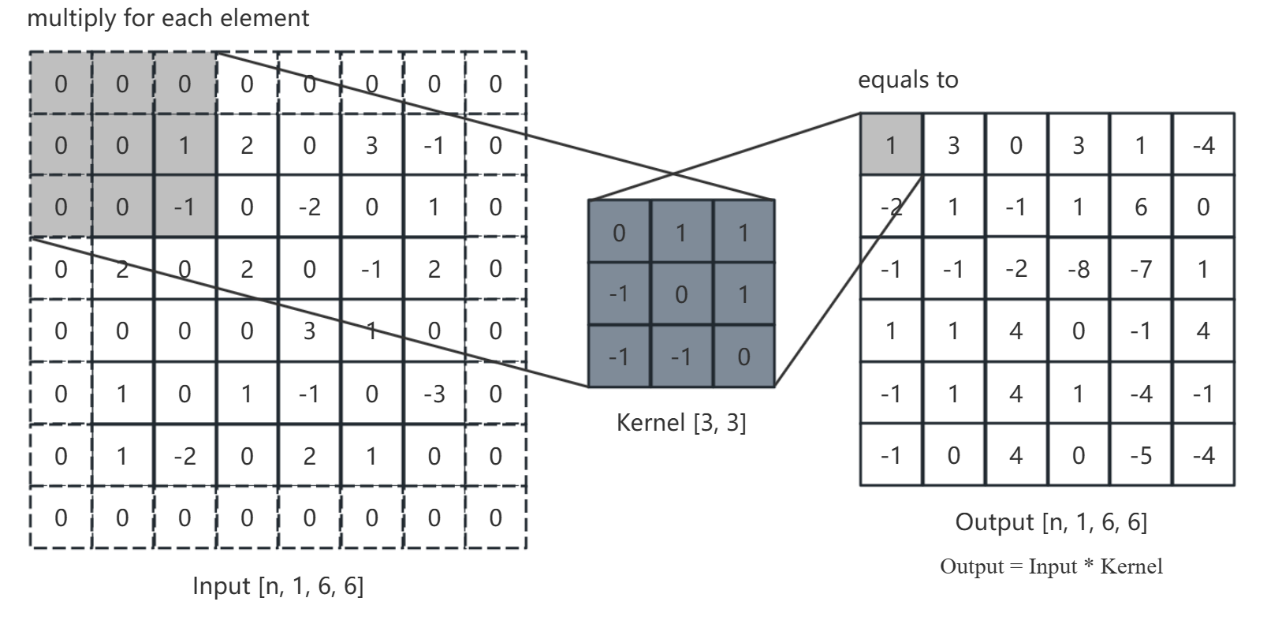
**在本模型中，感知层输入图像，使之可以用于模型训练的结构。这个结构通常会对图像进行随机旋转、平移、伸缩、翻转、对比度调整，并将像素数据按通道归一化，调整图像大小至目标尺寸，最终转化为张量形式。本模型图像输入尺寸是224×224像素，格式是有RGB三通道的.jpg格式图像。**

1. **卷积层（Convolution Layer）**

**卷积层是卷积神经网络的核心，也是整个网络学习图像形状、纹理的基础。卷积层的工作原理是二维数据的卷积操作，其过程可以用以下公式描述：**

**上述公式中，**代表当前被卷积图像第个通道的张量，是第个边长为的卷积核，就是卷积所得图像第个通道的张量。对于生成的每个像素数据，其值相当于以其为中心的范围内原图像数据与卷积核数据的元素积之和。随着卷积核在被卷积图像上滑动，原图像的局部特征便会以卷积形式映射到新产生的图像上。

通常，卷积操作不会改变输入张量的空间尺寸，但会扩展张量的通道数，使得模型能够学习到更多抽象的特征。对于正方形的输入图像，其空间尺寸满足以下公式：

**其中，是输出图像的边长，是输入边长，是边缘填充大小，是卷积核移动步长。下图展示了一个对通道数为1的图像做空间尺寸不变卷积的例子：**

**其中卷积核边长为3，移动步长为1，边缘填充大小为1（见Input四周虚线区域），填充数据为0；图中Output中的灰色像素对应着Input中大小的灰色区域，该区域称为这个像素的感受野。感受野的大小与卷积核空间尺寸匹配，决定了当前卷积层提取特征的局部性程度。**

**在卷积神经网络中，每一个卷积层输入张量的形状是，分别表示该批次样本个数、单个样本通道数、高度、宽度。在现代卷积神经网络的卷积层中，值也被称为batch size，往往被设置为2的整数次幂，是模型的重要超参数，直接关系到模型训练时表现和模型最终效果。对于每一个卷积层，卷积核上的权值是该层需要学习的参数。在模型训练过程中，经常发生模型过于复杂导致的过拟合现象，以及模型深度过大导致的梯度消失现象，这些现象往往会使得模型在陌生数据上表现不佳，或在训练时阻碍模型参数的进一步优化。为防止以上现象发生，我们在每一个卷积层后加入了批归一化（batch normalization）的正则化过程。**

**在一个训练批次中，对形状为的张量。批归一化的计算过程如下，其中：**

为**单个通道内，所有样本的所有像素数据的均值，形状为；**

**为单个通道内，所有样本的所有像素数据的方差，形状同样为。**

**对于单个通道内的每一个像素数据，做以下线性变换：**

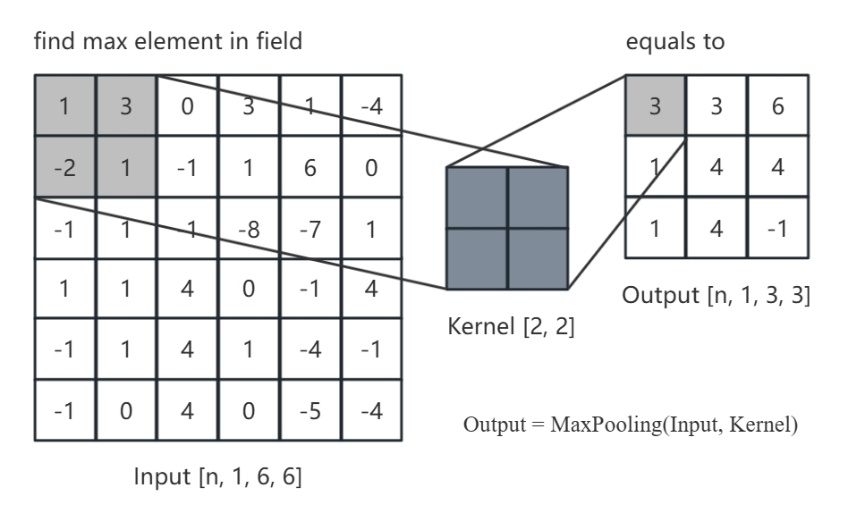
**其中**与为待学习参数，是一个小正数，防止分母为0。归一化结果。

**批归一化在通道层面使数据的分布接近标准正态分布，减少了样本内部的协变量偏移，使得卷积之后的模型更加稳定。**

1. **池化层（Pooling Layer）**

**池化层又称汇聚层、下采样层，用于在卷积层之后对图像的每个通道提取主要特征的同时，减小图像空间尺寸。池化层的存在减少了整个模型的参数量，简化模型结构，在加速模型训练的同时减小了过拟合风险。**

**池化的过程与卷积过程类似，即下采样核在原图像上滑动，生成下采样后图像。常见的池化方法有最大池化和自适应平均池化，其区别在于下采样函数不同。本模型主要采用最大池化法，其原理如下：**

****

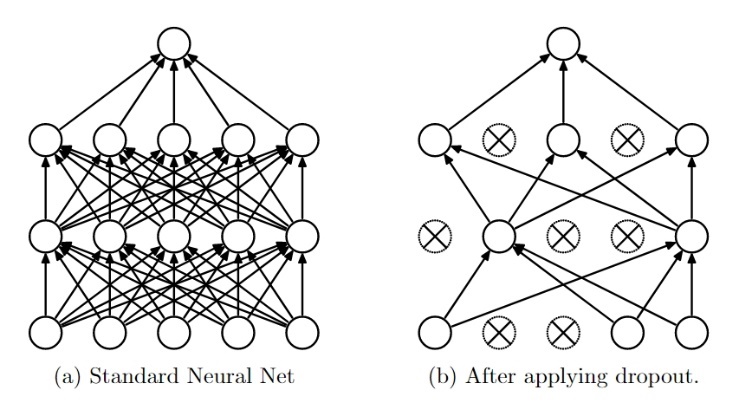
**其中下采样核边长为2，移动步长为2，无边缘填充。Output中每一个像素数据，都等于其Input感受野中所有元素的最大值。最大池化操作不会改变单个样本的通道数，但由于下采样核的移动步长为2，故图像线度会缩小为原来的一半。**

**现代卷积神经网络往往在池化层之后引入非线性的激活函数，以更好地拟合非线性数据、增强模型的特征提取能力。本模型使用的激活函数为ReLU函数，其定义为：**

1. **全连接层（Full Connection Layer）**

**全连接层用于收集前置卷积层、池化层传递过来的特征，并将这些数据进行处理、汇总、最终传递给分类器进行分类。在全连接层，我们需要将输入的张量展平为一个一维的向量并输入一个线性的神经元层。同上，为了强化模型的非线性拟合能力，输出的张量同样需要经过ReLU激活函数。**

**需要指出的是，全连接层中，相邻线性神经元层的任意两个神经元之间都有权重连接，这无疑会导致模型复杂化，从而提高模型过拟合风险。因此，我们在全连接层引入了Dropout正则化机制。Dropout机制可以将神经元按指定的几率关闭，从而减少了每次优化参数的数量，从而简化模型结构。其效果示意见下图：**

****

**在引入了关闭概率为的Dropout机制后，每一个全连接层的计算机制可表示为：**

**其中是一个元素为0或1的向量，对每一个元素而言，其为0的概率为。为输入全连接层的特征向量，则是输出向量。**

1. **分类器（Classifier）**

**分类器用于接受前置所有神经元的处理结果，并输出每个样本对应类别的概率。在本研究中，模型需要分辨源自11种不同类别的图片，因此分类器的输出形状为。为了满足概率归一原则，我们需要使用Softmax函数处理输出数据：**

**其中值是类别数。在经过Softmax函数归一化之后，输出的数据就是当前样本对应每个类别的概率。**

**（二）模型评价方法**

1. **交叉熵损失函数（Binary Cross Entropy Loss）**

**在神经网络中，损失函数用于量化表示当前模型对样本的预测与样本真实标签之间的差别大小，损失函数值越小，表示当前模型对样本的预测效果越好。在多分类问题中，现代神经网络模型常使用交叉熵损失函数以评估模型效果。其定义如下：**

**其中**是当前批中样本真实标签的独热编码向量，是模型预测值向量，上述公式中对向量的运算都是按元素运算，而不是张量运算。对模型进行优化的过程，就是调整模型参数，使得模型损失函数值越来越小的迭代过程。

1. 准确率（Accuracy）

对一批已知标签的数据，通过当前模型进行预测，其中预测类别正确的占比即模型在当前数据上的准确率。准确率是评估模型实际表现的最直观、最常用的数据。

1. **混淆矩阵（Confusion Matrix）**

**在对一批数据进行预测时，我们不仅想知道模型对其预测的准确率，还想知道预测错误的哪些数据被预测为了哪些其他的类别，而混淆矩阵便可以很直观地反映这一信息。**

**对于总类别数为的模型，其混淆矩阵的形状为，第i行第j列的数据表示实际标签为第i类、模型预测为第j类的样本数。由此可见，当一个模型的表现越出色，其混淆矩阵的对角线上元素和应该越接近预测样本总数。通过分析混淆矩阵，我们可以很清晰地得知当前模型容易混淆哪些类型的数据，从而给在数据层面进一步优化模型，如增加哪些类型图片的数量、图像增强程度，提供优化思路。**

1. **Kappa系数（Cohen’s Kappa）**

**在混淆矩阵的基础上，我们引入了Kappa系数来进一步评估模型表现。Kappa系数的计算基于混淆矩阵，并修正了随机猜测对准确率的影响，在多分类模型的评估上应用广泛。其计算方法由以下公式给出：**

**其中被称为观测一致性，**即分类器的总体准确率；**为期望一致性，**即随机猜测时的一致性概率。这两个一致性系数的计算依靠混淆矩阵：

**Kappa系数越接近1，表示模型预测效果越好。**

**（三）模型训练方法**

**本模型使用Adam梯度下降优化器，对模型参数进行优化。Adam算法源于随机梯度下降算法（Stochastic Gradient Descent），同时融合了Momentum、RMS Prop等改进算法，在深度学习神经网络中应用广泛。其优化目标是使得损失函数的值尽可能小，基本优化策略是对损失函数关于权重求梯度，并将权重向梯度的反方向小幅移动，从而试图找到使得损失函数值尽可能小的权重值。其基本迭代方法如下：**

**其中、、分别为学习率、动量衰减系数、二阶矩衰减系数，是整个模型训练过程中的重要超参数，其取值直接影响到最终模型效果。**

**Adam梯度下降优化器可以根据梯度的变化自适应地调节学习率，以避免模型训练过程中的震荡与不稳定，并相较传统优化器显著减少了模型的训练数据。**

**（四）模型超参数搜索方法**

**本研究使用了麻雀搜索算法（Sparrow Search Algorithm）、TPE搜索算法（**Tree-structured Parzen Estimator）两种搜索算法进行对比实验论证，并优化模型超参数，使得模型能在训练过程中更快收敛，并取得更高的准确率。

1. **麻雀搜索算法**

**作为一种新兴的元启发搜索算法，麻雀搜索算法于2020年被提出，并被应用于深度学习模型超参数优化、NP难问题求解等领域。其核心思想是模拟麻雀在觅食过程中的分工行为，通过动态调整发现者、跟随者、警戒者三种角色的位置，有效平衡了全局搜索与局部优化的性能。**

**该算法将一个麻雀种群分为三种角色：发现者、跟随者、警戒者。算法开始时，种群中的个体被随即放置在搜索空间中。在每一次迭代时，随机生成预警值，该预警值将决定发现者的行为；种群中位置最优的前20%的个体为发现者，当预警值小于安全阈值时，发现者扩大搜索范围，即趋向全局搜索，否则向附近安全区域移动，即趋向局部优化；剩余80%为追随者，对于每一个追随者个体，若其当前位置较优，则会向全局最优个体的方向靠近，否则会移动至更远的区域搜索；警戒者是种群中随机挑选的20%个体，当警戒者距离当前全局最优位置过远时，表示其被天敌捕食的概率很大，故其向最优位置靠近。重复上述迭代过程，直到达到最大迭代数，或当前已满足收敛条件。**

**麻雀搜索算法很好地平衡了全局搜索和局部优化，并创新性地引入了随机扰动和警戒者的身份，避免过早陷入局部最优。相较粒子群算法、模拟退火算法等传统元启发搜索算法，麻雀搜索算法的收敛速度更快，寻优能力更强，因此也在近些年得到广泛应用。**

1. **TPE搜索算法**

**作为贝叶斯优化算法的变体，TPE搜索算法通过智能采样，用尽可能少的采样次数搜寻最优的参数组合。区别于传统贝叶斯优化方法，TPE算法采用了Parzen窗估计器对参数的分布进行建模，并使用树结构处理各参数之间的依赖关系。**

**算法开始时，在搜索空间中随机采样若干个点并进行评估，并保存评估结果。对于每一次迭代，将历史采样观测集，按评估结果排序，并按照一定的比例分为性能优良和较差的两个集合和，并按照以下公式对这两个集合建立核密度估计模型：**

**其中是核函数；接着在中选择新的采样点，使得的值尽可能大，以保证新采样的观测在下一次迭代时被划到性能优良的集合之中。而在计算、的值时，树状结构能够较好地处理参数之间的层级依赖，并保证了高效地在高维解空间中进行搜索。**

**TPE搜索算法能够通过较少的采样次数获得更优的参数组合，而且在高维搜索之中表现优异，而树状结构又保证其可以处理离散/连续混合参数空间，因此非常适合深度学习模型的超参数调优任务。**