

# 基于深度学习的工程图像识别与分类研究

金旭杨

浙江兆晟科技股份有限公司 浙江 杭州 311100

**摘要:** 本文探讨了基于深度学习的工程图像识别与分类方法,介绍了深度学习的基本原理、常用模型、框架与工具。重点分析了工程图像数据集的构建、预处理技术,以及基于深度学习的工程图像识别与分类模型设计,包括模型架构选择与优化、特征提取与表示学习、分类器设计与训练等方面。通过详细阐述深度学习在工程图像识别中的应用,本文旨在为提高工程图像识别与分类的准确性和效率提供理论支持和实践指导。

**关键词:** 深度学习; 图像识别; 分类算法

引言: 随着工程技术的发展,图像数据在工程领域中扮演着越来越重要的角色。传统的图像识别方法往往依赖于手工设计的特征,难以适应复杂多变的工程图像。而深度学习作为一种自动学习数据特征的技术,为工程图像的识别与分类提供了新的解决方案。本文综述了深度学习的基本原理和相关技术,特别是卷积神经网络、循环神经网络和生成对抗网络等模型在工程图像识别中的应用。同时,本文还探讨了工程图像数据集的构建与预处理技术,为后续基于深度学习的识别与分类模型设计提供基础。

## 1 深度学习基础理论与相关技术

### 1.1 深度学习原理概述

深度学习是机器学习的一个子领域,其核心在于通过模拟人脑神经网络的结构,从数据中自动提取并学习复杂特征,以解决各种复杂的任务。深度学习的基本构成单位是神经网络,它由多个层次组成,每个层次包含多个处理单元(神经元或节点)。这些层次可分为输入层、隐藏层和输出层。输入层接收原始数据,隐藏层负责学习和提取特征,而输出层则给出最终预测或分类结果。深度学习的训练过程通常涉及前向传播、损失计算、梯度计算和权重更新几个步骤。前向传播是指数据从输入层经过各隐藏层到达输出层的过程。损失计算评估输出层的预测值与真实值之间的差异。梯度计算通过反向传播算法,利用链式法则计算损失函数关于每个权重和偏置的梯度。最后,利用优化算法(如梯度下降、随机梯度下降、Adam等)更新网络中的权重和偏置,以最小化损失函数。

### 1.2 常用深度学习模型介绍

#### 1.2.1 卷积神经网络(CNN)及其变体

卷积神经网络(CNN)是一种专门用于处理具有网格结构数据(如图像、视频)的深度学习模型。其核心

在于卷积层和池化层的组合使用,能够自动学习图像的空间层次结构。卷积层通过一系列可学习的卷积核对输入图像进行卷积操作,以提取图像中的局部特征。池化层则对卷积后的特征图进行下采样,以减少数据维度并保留重要特征;CNN在图像识别领域表现出色,其变体如ResNet、VGG等通过增加网络深度和复杂度,进一步提高了性能。ResNet引入了残差学习,缓解了深层网络的训练难题;VGG则通过堆叠多个小卷积核,增强网络的特征提取能力。这些模型在图像分类、目标检测、图像分割等多个任务中取得显著成果。

#### 1.2.2 循环神经网络(RNN)及其改进型 LSTM、GRU

循环神经网络(RNN)是一类具有短期记忆能力的神经网络,特别适用于处理具有时间序列依赖性的数据,如语音、文本和视频等。在RNN中,神经元的输出不仅与当前时刻的输入相关,还与前一时间步的输出相关,从而形成一个循环结构。然而,传统的RNN在处理长序列时容易出现梯度消失或梯度爆炸的问题。为了解决这些问题,研究者们提出LSTM(长短期记忆网络)和GRU(门控循环单元)等改进型RNN。LSTM通过引入输入门、遗忘门和输出门三个门控机制,实现了对信息的长期记忆和选择性遗忘。GRU则是LSTM的简化版本,将遗忘门和输入门合并为更新门,减少计算量并保持较好的性能。这些改进型RNN在处理图像序列或具有时间关联的工程图像数据方面表现出巨大的潜力。

#### 1.2.3 生成对抗网络(GAN)

生成对抗网络(GAN)是一种通过生成器和判别器之间的博弈过程来生成逼真数据样本的深度学习模型。生成器尝试生成真实的样本,而判别器则试图区分真实数据和生成数据。通过不断对抗和优化,生成器能够生成越来越接近真实分布的数据样本。GAN在工程图像数

据扩充和样本生成方面具有广泛的应用前景。利用GAN可以生成大量的高质量图像数据，用于训练深度学习模型或进行数据增强。此外，GAN还可以用于图像修复、风格迁移和超分辨率等任务中，为图像处理领域带来了全新的解决方案<sup>[1]</sup>。

### 1.3 深度学习框架与工具

深度学习的发展离不开各种高效的框架和工具的支持。目前流行的深度学习框架包括TensorFlow、PyTorch、Keras等。这些框架提供了丰富的API和预训练模型，使得开发者和研究人员能够更方便地构建和训练深度学习模型。TensorFlow是一个由Google开发的开源机器学习框架，支持多种编程语言和平台。它提供了全面的机器学习工具集，包括张量计算、数据管道、模型训练和优化等功能。PyTorch则是一个由Facebook开发的开源机器学习库，特别注重灵活性和易用性。它允许开发者在Python中直接定义和执行计算图，大大简化了模型的构建和训练过程。Keras则是一个基于TensorFlow和Theano的高层神经网络API，提供了简洁易用的接口来构建和训练深度学习模型。随着技术的不断发展和完善，深度学习将在更多领域发挥重要作用，推动人工智能技术的进一步普及和应用。

## 2 工程图像数据集的构建与预处理

### 2.1 工程图像数据来源与特点分析

工程图像数据集的构建首先依赖于广泛而多样的数据源。这些数据通常来源于不同的工程领域，包括但不限于建筑工程、机械工程、电气工程等，每个领域都有其独特的图像类型和应用场景。例如，在建筑工程中，图像可能包括设计图纸（如蓝图）、施工现场照片、建筑物外观与内部结构照片等；在机械工程中，则可能涉及设备部件的照片、运行状态监测图像等；而电气工程则可能包括电路板布局、电线连接图等。分析这些数据时，需要关注其图像质量、分辨率以及标注信息的多样性。图像质量直接影响后续处理和分析的效果，低质量的图像可能包含模糊、噪声或失真等问题。分辨率则决定了图像中细节信息的丰富程度，对于需要精确分析的任务尤为重要；标注信息的准确性和一致性也是数据集质量的关键因素，包括目标对象的定位、分类或属性标注等<sup>[2]</sup>。

### 2.2 数据集构建方法

数据集的构建是一个系统性的过程，需要制定详细的采集方案。首先，确定数据采集设备，根据图像类型和需求选择合适的相机或扫描仪，确保图像的清晰度和细节保留。其次，设定采集环境，包括光照条件、拍

摄角度和背景设置等，以减少外界因素对图像质量的干扰。同时，确定采集参数，如曝光时间、ISO值等，以适应不同的光照条件和图像特征；数据标注是数据集构建中的关键环节。为了确保标注的准确性与一致性，需要制定详细的标注规范，明确标注对象的定义、分类标准和标注格式。在实际操作中，可以采用人工标注与自动标注相结合的策略。人工标注虽然耗时费力，但能够确保标注的准确性和灵活性；而自动标注则可以利用机器学习算法，在大量数据上进行快速标注，但可能存在一定的误差。因此，在实际应用中，可以根据任务需求和数据特点，灵活选择标注方法。按照一定比例将数据集划分为训练集、验证集和测试集。划分原则通常基于数据的多样性和分布特性，确保每个集合都能代表整体数据的特征。常用的划分方法包括随机划分、分层抽样等。划分过程中需要注意避免数据泄露和交叉污染，以保证模型训练与评估的可靠性。

### 2.3 图像预处理技术

图像预处理是深度学习模型训练前的重要步骤，旨在提高图像数据的质量和模型的训练效率。针对工程图像数据集的特点，可以采用多种预处理技术。第一，进行图像去噪和增强处理。由于工程图像可能受到环境噪声、光照不均等因素的影响，通过去噪算法（如高斯滤波、中值滤波等）可以有效减少噪声干扰。同时，利用图像增强技术（如直方图均衡化、对比度拉伸等）可以改善图像的视觉效果和细节信息。第二，进行图像归一化和标准化处理。归一化是指将图像的像素值映射到特定范围内（如0-1），以消除不同图像之间由于亮度和对比度差异造成的影响。标准化则是指根据图像的统计特性（如均值和标准差）对像素值进行调整，使得数据分布更加集中和稳定。这些处理有助于加速模型训练过程并提高模型的泛化能力<sup>[3]</sup>。第三，根据具体任务需求进行图像裁剪、缩放和旋转等操作。这些操作可以调整图像的大小和形状以符合模型输入的要求，同时增加数据的多样性以提高模型的鲁棒性。此外，对于具有时间关联性的图像序列或视频数据，还可以进行帧提取、时序对齐等预处理步骤以提取有效的时序特征。

## 3 基于深度学习的工程图像识别与分类模型设计

### 3.1 模型架构选择与优化

在工程图像识别与分类任务中，选择合适的深度学习模型架构是基础且关键的一步。根据工程图像数据的特点，如图像复杂度、分辨率、目标对象的多样性等，可以选择卷积神经网络（CNN）作为基础架构，因为其在处理图像数据方面表现出色。另外，还可以考虑使用

如ResNet、VGG等经过优化和验证的CNN变体，以利用它们在大型图像数据集上的成功经验。模型的优化是提高性能与效率的重要途径；这包括调整网络层数、卷积核大小与数量、池化方式等超参数。网络层数的增加可以增强模型的特征提取能力，但同时也会增加计算复杂度和过拟合的风险。因此，需要通过实验找到性能与复杂度的平衡点<sup>[4]</sup>。卷积核的大小和数量则决定了特征提取的精细程度，较小的卷积核和较多的数量可以捕获更丰富的局部特征，但也会增加计算量。池化方式的选择（如最大池化、平均池化）则影响特征图的降维方式和模型的平移不变性。通过对比实验，可以针对特定任务优化这些超参数。

### 3.2 特征提取与表示学习

深度学习模型在工程图像识别中的核心优势在于其自动提取和表示学习特征的能力。这些特征不仅比手工设计的特征更加复杂和抽象，而且能够自动适应不同的数据分布和任务需求。为了提升模型的特征提取能力，可以进一步优化网络结构，如引入残差连接（如ResNet中的残差块）、注意力机制等，以更好地捕捉图像中的关键信息。通过可视化技术展示模型学习到的特征，可以帮助我们理解模型的决策机制并发现潜在的问题。特征可视化通常包括卷积核权重可视化、特征图可视化和类激活映射（CAM）等。这些技术可以帮助我们观察模型在不同层次上的特征表示，并识别出哪些特征是模型做出分类决策的重要依据。

### 3.3 分类器设计与训练

分类器的设计是基于深度学习模型进行工程图像识别与分类的最后一步。通常，分类器位于网络的最后几层，负责将学习到的特征映射到目标类别上。常用的分类器包括全连接层加softmax激活函数的形式，它可以输出每个类别的预测概率。在训练过程中，需要选择合适的损失函数来评估模型的性能并指导模型的优化。对

于多分类任务，交叉熵损失函数是一个常见的选择。此外，还需要选择合适的优化算法来更新模型的参数，如随机梯度下降（SGD）、Adam等。训练过程中还需要注意避免过拟合和欠拟合的问题，可以通过数据增强、正则化、早停等方法来提高模型的泛化能力<sup>[5]</sup>。基于深度学习的工程图像识别与分类模型设计涉及模型架构选择与优化、特征提取与表示学习以及分类器设计与训练等多个方面。通过不断实验和优化这些关键环节，可以构建出高性能的模型来满足实际工程应用的需求。

### 结束语

综上所述，深度学习在工程图像识别与分类中展现出强大的潜力和优势。通过合理选择和优化深度学习模型，结合高质量的工程图像数据集和有效的预处理技术，可以显著提高图像识别的准确性和效率。未来，随着深度学习技术的不断发展和完善，相信基于深度学习的工程图像识别与分类方法将在更多工程领域得到广泛应用，为智能工程的实现提供有力支持。

### 参考文献

- [1]李大湘,张玥.融合全局与局部特征的国画图像分类算法[J].现代计算机.2019,(34).DOI:10.3969/j.issn.1007-1423.2019.34.010.
- [2]李清扬,赵晔.苏绣文化保护与传承研究[J].开封教育学院学报.2019,(1).DOI:10.3969/j.issn.1008-9640.2019.01.098.
- [3]邱晨,罗璟,赵朝文,等.基于BP神经网络的空气质量模型分类预测研究[J].软件.2019,(2).DOI:10.3969/j.issn.1003-6970.2019.02.025.
- [4]曾伟辉,李森,李增,等.基于高阶残差和参数共享反馈卷积神经网络的农作物病害识别[J].电子学报.2019,(9).DOI:10.3969/j.issn.0372-2112.2019.09.023.
- [5]吴杰,施磊,张志安.基于深度学习的害虫图像识别与分类方法研究[J].计算技术与自动化,2023,42(1):166-173. DOI:10.16339/j.cnki.jsjsyzdh.202301029.