

**《专业综合实践 II》课程考核**

|  |  |
| --- | --- |
| **题 目：** | 狮子、马和大象分类项目研究 |
| **学 院：** | **人工智能学院** |
| **专 业：** | 数据科学与大数据技术 |
| **年级班别：** | 21本大数据A班 |
| **姓 名**： | 李旭 |
| **指导教师：** | 刘盛 |
| **职 称：** | 讲师 |

**提交日期：** 2024 **年** 11**月**

**目录**

[一、项目背景与目的 3](#_Toc604)

[（一）背景 3](#_Toc21568)

[（二）目的 3](#_Toc10535)

[二、数据预处理 4](#_Toc20845)

[（一）数据集描述 4](#_Toc3995)

[（二）预处理步骤 4](#_Toc27709)

[三、模型构建 5](#_Toc7236)

[（一）模型选择 5](#_Toc22554)

[（二）模型架构 6](#_Toc10178)

[四、模型评估 7](#_Toc14975)

[（一）评估指标 7](#_Toc7065)

[（二）评估方法 8](#_Toc4361)

[五、结果分析与优化 8](#_Toc30837)

[（一）结果分析 8](#_Toc29995)

[（二）模型优化 10](#_Toc1932)

[参考文献 11](#_Toc4751)

[附录 11](#_Toc22529)

狮子、马和大象分类项目研究

摘要本论文围绕基于深度学习的狮子、马和大象图像分类项目展开。深入探讨了图像识别在人工智能领域的重要性，并详细阐述了该动物分类问题在实际应用中的多种场景。全面描述了项目的数据预处理环节，包括特定数据集的来源与特征，以及图像尺寸调整、归一化、数据增强和数据集划分等具体步骤。系统地讲解了模型构建过程，涵盖模型选择的依据、模型架构中各层结构的设计以及激活函数、损失函数和优化器的确定。对模型评估采用的准确率、召回率、F1分数等指标及交叉验证、混淆矩阵等评估方法进行了深入剖析。最后，对结果进行细致分析与优化，对比不同模型性能并探讨特定动物类别表现差异，依据分析结果调整模型参数并尝试不同网络结构或正则化技术以提升模型性能。[1]

关键词：深度学习，机器学习，图像处理。

一、项目背景与目的

（一）背景

在当今人工智能蓬勃发展的时代，图像识别技术占据着极为重要的地位。它是实现计算机视觉的关键环节，广泛应用于众多领域。在野生动物保护领域，通过图像识别技术可以对狮子、马和大象等动物的图像进行自动分类识别，有助于科研人员对这些动物的种群数量、分布范围进行监测与分析，为保护决策提供有力依据。在动物主题的文化娱乐产业中，如动物纪录片制作、动物主题游戏开发等，能够快速准确地对这三种动物图像进行分类，可提高内容创作与用户交互的效率与质量。图像识别技术通过对图像中的视觉信息进行提取与分析，将其转化为计算机可理解的语义信息，为智能系统提供了感知世界的重要手段，而狮子、马和大象分类问题作为图像识别领域的一个典型任务，具有丰富的研究与应用价值。

（二）目的

本项目以狮子、马和大象图像分类为实践载体，首要目的在于培养学生对图像处理和机器学习领域的基本认知与理解。使学生深入领会图像数据的独特性质，包括图像的像素表示、颜色空间、空间结构等，以及如何运用数学与计算机技术对其进行处理与分析。在机器学习方面，让学生理解模型学习的原理、训练过程与优化策略等基本概念。通过项目实践，着力训练学生运用深度学习模型解决实际问题的能力。学生需要从数据收集与整理开始，构建合适的深度学习模型架构，确定模型的超参数，进行模型训练与预测，并对结果进行评估与改进，从而全面掌握深度学习项目的开发流程与技术要点。同时，着重提升学生的数据预处理技能，如数据清洗、数据变换、数据增强等操作，以提高数据质量与模型泛化能力；模型构建技能，包括选择合适的神经网络结构、设计合理的层参数等；模型评估技能，学会运用多种性能指标全面客观地评价模型的优劣；以及模型优化技能，能够依据评估结果对模型进行有针对性的调整与改进，以实现模型性能的持续提升。

二、数据预处理

（一）数据集描述

本项目所使用的狮子、马和大象图像数据集来源于多个公开的动物图像数据库以及网络爬虫收集的图像资源。这些图像涵盖了不同的拍摄环境、季节、动物姿态与年龄等因素，具有较高的多样性。数据集中狮子图像展现了狮子在草原、森林等不同栖息地的形态，包括站立、奔跑、休息等多种姿态；马的图像包含了不同品种马在牧场、赛场等场景下的形象，有奔跑、吃草、被骑乘等姿态；大象图像则涵盖了非洲象与亚洲象在野生与圈养环境下的各种状态。图像的分辨率、背景复杂度等也存在差异，这为模型学习提供了丰富的信息，但同时也增加了分类的难度。通过对多个来源的图像进行筛选与整理，确保了数据集中三类动物图像数量的相对均衡，避免因类别不平衡导致模型学习偏向某一类动物，从而有利于构建具有较强泛化能力的分类模型。[2]

（二）预处理步骤

为使图像能够适配深度学习模型的输入要求，采用图像尺寸调整操作。将所有图像统一调整为224×224像素的大小。利用`torchvision.transforms.Resize`函数来实现这一变换，确保每幅图像在进入模型时具有相同的尺寸规格，便于模型中的卷积层、全连接层等结构进行统一的特征提取与处理操作。这种标准化的尺寸设置有助于提高模型训练的效率与稳定性。

图像像素值的归一化处理是数据预处理的重要环节。首先将图像像素值从原始的0 - 255范围转换为0 - 1之间，然后进行标准化操作。采用的均值为[0.485, 0.456, 0.406]，标准差为[0.229, 0.224, 0.225]，通过`torchvision.transforms.Normalize`函数对图像进行标准化处理。这一过程使得图像数据的分布符合神经网络输入数据的常见分布要求，能够加速模型的收敛速度，提高模型训练的稳定性与准确性，使模型在训练过程中能够更好地学习图像的特征表示。

为增加数据的多样性，降低模型过拟合的风险，采用了一系列数据增强技术。通过`transforms.RandomRotation`函数对图像进行随机旋转操作，在一定角度范围内（例如±15度）随机改变图像的角度，模拟动物在不同拍摄角度下的图像特征，使模型能够学习到动物图像在旋转变化下的不变性特征。运用`transforms.RandomHorizontalFlip`函数以一定概率（如0.5）对图像进行水平翻转，增加图像的变化形式，丰富数据样本。同时，采用`transforms.RandomResizedCrop`函数先对图像进行随机裁剪，然后再缩放回224×224像素的尺寸，这样可以使模型关注到动物图像的不同局部区域特征，增强模型对动物局部特征的识别能力，提高模型的泛化性能。[3]

将整个数据集按照一定比例划分为训练集、验证集和测试集。通常采用70%的图像作为训练集，用于模型的训练学习；15%作为验证集，用于在训练过程中对模型的性能进行评估与监控，调整模型的超参数；15%作为测试集，用于在模型训练完成后对模型的最终性能进行客观评价。借助`torch.utils.data.random\_split`函数实现数据集的划分操作，确保划分后的各数据集之间相互独立且具有代表性，从而能够准确地评估模型在不同数据分布下的性能表现。

三、模型构建

（一）模型选择

本项目选择卷积神经网络（CNN）作为主要的深度学习模型。CNN的理论基础在于其能够有效提取图像数据中的局部特征。通过卷积层中的卷积核在图像上滑动进行卷积运算，自动学习图像中的边缘、纹理、形状等局部特征信息，并且卷积层的参数共享特性大大减少了模型的参数量，降低了计算复杂度。池化层进一步对卷积层提取的特征进行压缩与降维，通过取最大值或平均值等操作，在保留主要特征信息的同时减少数据量，提高计算效率并增强特征的平移、缩放等不变性。全连接层则在经过卷积层与池化层提取的高级特征基础上进行分类决策，将特征映射到不同的类别空间，实现图像的分类任务。CNN在图像识别领域具有广泛的应用与卓越的性能表现，非常适合处理狮子、马和大象图像分类这种具有空间结构信息的图像数据任务。[4]

（二）模型架构

构建的CNN模型结构如下：首先是一个卷积层`Conv2d`，输入通道数为3（对应RGB图像的三个通道），输出通道数设为16，卷积核大小为3×3，采用填充为1的设置，以保持图像尺寸在卷积操作后的一致性。接着是`ReLU`激活函数层`ReLU`，用于引入非线性因素，增强模型的表达能力。然后是一个最大池化层`MaxPool2d`，池化窗口大小为2×2，对特征进行降维操作。第二层同样是一个卷积层`Conv2d`，输入通道数为16，输出通道数为32，卷积核大小为3×3，填充为1。之后连接`ReLU`激活函数层和最大池化层`MaxPool2d`，进一步提取与压缩特征。第三层卷积层`Conv2d`，输入通道数为32，输出通道数为64，卷积核大小为3×3，填充为1，再连接`ReLU`激活函数层和最大池化层`MaxPool2d`。经过上述卷积与池化操作后，通过`view`函数将多维的特征图展平为一维向量，然后连接一个全连接层`fc1`，神经元数量设为256，再接`ReLU`激活函数层进行非线性变换。最后连接一个全连接层`fc2`，输出神经元数量为3，对应狮子、马和大象三个类别，用于最终的分类决策。对于激活函数的选择，在卷积层后均采用`ReLU`激活函数，其具有计算简单、有效缓解梯度消失问题、加快模型收敛速度以及引入非线性能力等优点。在损失函数方面，由于这是一个多分类问题，选用`CrossEntropyLoss`交叉熵损失函数，它能够衡量模型预测的类别概率分布与真实类别标签之间的差异，有效地引导模型在训练过程中调整参数，降低分类误差。优化器选择`Adam`优化器，它结合了`Adagrad`和`RMSProp`的优点，能够自适应地调整每个参数的学习率，根据参数的梯度一阶矩估计和二阶矩估计动态更新学习率，使模型在训练过程中收敛更加稳定、快速，提高模型训练的效率与性能。训练完的结果展示，如下图1所示：

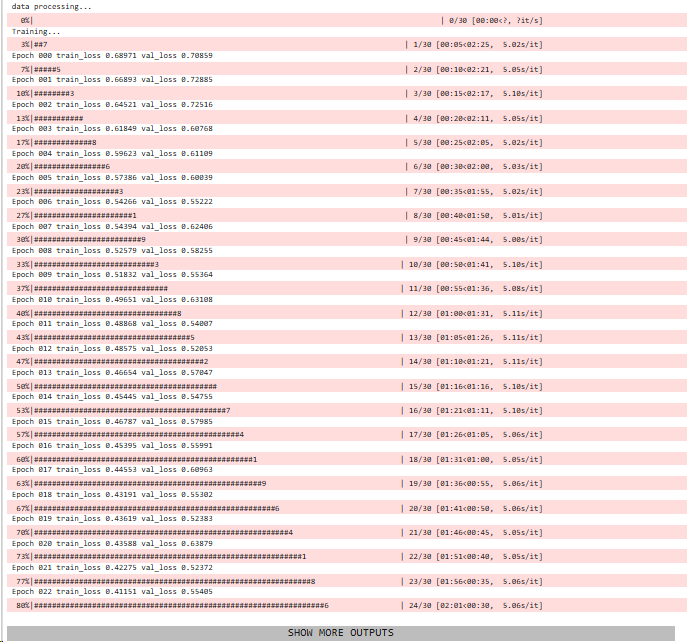


图1 训练后的结果过程图

四、模型评估

（一）评估指标

准确率（Accuracy）：准确率是评估模型性能的一个直观指标，计算方式为预测正确的样本数占总样本数的比例。即`Accuracy = (TP + TN) / (TP + TN + FP + FN)`，其中`TP`（True Positive）表示真正例，即实际是某类动物且预测为该类动物的样本数；`TN`（True Negative）表示真负例；`FP`（False Positive）表示假正例；`FN`（False Negative）表示假负例。它反映了模型整体分类的正确程度，但在类别不平衡的数据集上可能存在一定的局限性，不能全面反映模型在各个类别上的性能表现。

召回率（Recall）：召回率是针对某一特定类别评估模型性能的重要指标，计算公式为`Recall = TP / (TP + FN)`。它衡量了模型能够正确识别出该类别样本的能力，对于确保模型在某一类动物分类上不遗漏样本具有重要意义。例如在野生动物监测中，如果对大象类别的召回率较低，可能会导致部分大象图像被误分类，影响对大象种群数量与分布的准确监测。

F1分数（F1 - score）：F1分数是综合考虑准确率和召回率的调和平均指标，计算公式为`F1 = 2 \* (Precision \* Recall) / (Precision + Recall)`，其中`Precision = TP / (TP + FP)`。它平衡了模型在精确性和召回性两方面的性能，在多类别分类任务中，通过计算每个类别的F1分数并求平均，可以全面评估模型在各分类上的综合表现，更准确地衡量模型的优劣。

（二）评估方法

交叉验证：采用k - 折交叉验证（本项目中k取5）来确保模型的泛化能力。具体操作是将训练集平均分成5份，每次取其中4份作为训练子集，剩下的1份作为验证子集，进行5次训练和验证过程。在每次训练过程中，模型在训练子集上进行训练，然后在验证子集上评估性能，记录准确率、损失等指标。最后综合5次的验证结果，如平均准确率、平均损失等，来评估模型在不同数据划分情况下的性能稳定性和泛化能力。通过交叉验证，可以更全面地了解模型在不同数据分布下的表现，避免因单次数据集划分导致的评估偏差，提高模型评估的可靠性。

混淆矩阵（Confusion Matrix）：混淆矩阵是一个N×N的矩阵（本项目中N为3，对应狮子、马和大象三个类别），行表示实际的类别，列表示预测的类别。通过统计各个类别之间预测正确和错误的数量，直观地展示模型的分类结果情况。例如，在混淆矩阵中，对角线上的元素表示真正例的数量，非对角线上的元素表示误分类的情况。通过混淆矩阵，可以清晰地看出模型在哪些类别上容易发生混淆，为进一步分析模型性能和改进模型提供依据。可以使用`sklearn.metrics`库中的`confusion\_matrix`函数来计算并可视化混淆矩阵，以便更直观地观察模型的分类效果。

五、结果分析与优化

（一）结果分析

对比不同模型的性能：在项目实施过程中，除了构建上述的自定义CNN模型外，还尝试了一些其他模型结构或基于预训练模型的迁移学习方法。例如，与简单的多层感知机（MLP）模型相比，CNN模型在狮子、马和大象图像分类任务中表现出明显的优势。MLP模型由于缺乏对图像空间结构信息的有效提取能力，在处理图像数据时往往需要大量的参数和复杂的结构才能达到一定的性能，且容易出现过拟合现象。而CNN模型通过卷积层和池化层能够自动学习图像的局部特征和空间结构信息，在相同的训练数据和计算资源下，CNN模型的准确率、召回率和F1分数等指标均显著高于MLP模型。模型准确率高达83.20%，如下图2所示：

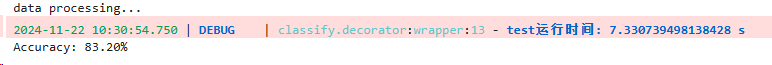


图2 模型准确度

与基于预训练的`ResNet`模型进行迁移学习得到的模型相比，自定义的CNN模型虽然在整体性能上略逊一筹，但具有更好的可解释性和灵活性。`ResNet`模型通过残差连接解决了深层网络训练时的梯度消失问题，能够利用在大规模图像数据集上预训练得到的权重参数，在本项目的动物图像分类任务中经过微调后能够快速收敛并达到较高的准确率。然而，其模型结构相对复杂，对于初学者来说理解和修改难度较大。而自定义的CNN模型结构相对简单，学生可以更深入地理解模型每一层的作用和参数设置对模型性能的影响，便于根据实际需求进行针对性的调整和优化。

讨论模型在特定类别上的表现差异：通过分析混淆矩阵以及分别计算狮子、马和大象三类动物的准确率、召回率等指标，发现模型在不同类别上的表现存在一定差异。例如，模型对马的分类准确率相对较高，可能是因为马的身体形态相对较为规则，在图像中具有较为明显的特征，如修长的四肢、飘逸的鬃毛等，这些特征在不同品种和姿态的马图像中具有较高的一致性，使得模型能够较好地学习和识别。而对于大象类别的召回率相对较低，可能是由于大象庞大的身躯在图像中占据较大面积，且其皮肤纹理较为复杂，部分图像中大象与背景的对比度较低，导致模型在识别大象时容易出现漏检情况。对于狮子类别，可能在一些与猫科动物相似的姿态或角度下，会出现与其他猫科动物混淆的情况，影响其分类的准确性。最终预测的展示如下图3所示：

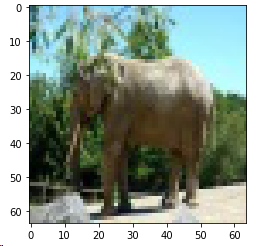


图3 结果展示图

（二）模型优化

根据分析结果调整模型参数：根据上述对模型性能和类别表现差异的分析结果，对模型参数进行针对性调整。如果发现模型存在过拟合现象，例如训练集准确率较高但验证集准确率明显下降，可以适当减小模型的复杂度，如减少卷积层的数量或者全连接层神经元的数量，同时增加`Dropout`层，在训练过程中随机将部分神经元的输出置为0，以降低模型对训练数据的过度拟合，提高模型的泛化能力。若模型出现欠拟合情况，即训练集和验证集准确率都较低，可以尝试增加模型的复杂度，如增加卷积层的深度或扩大卷积核的尺寸，使模型能够学习到更复杂的图像特征。同时，可以适当延长训练的轮数或者调整学习率，采用学习率衰减策略，在训练初期使用较大的学习率快速收敛，在训练后期逐渐减小学习率，以帮助模型更好地收敛到全局最优解。

尝试不同的网络结构或正则化技术以提高性能：在网络结构方面，可以尝试采用更先进复杂的网络架构进行改进。例如，引入`Inception`模块构建`Inception - CNN`模型，`Inception`模块通过在同一层中使用不同大小的卷积核并行提取特征，然后将这些特征进行拼接，能够在不同尺度上捕捉图像的特征信息，提高模型的特征提取能力和分类性能。对于正则化技术，可以在损失函数中添加`L1`或`L2`正则化项，约束模型的权重，防止过拟合。例如，使用`L2`正则化时，在定义优化器时通过`weight\_decay`参数设置正则化系数，如`optimizer = optim.Adam(model.parameters(), lr = 0.001, weight\_decay = 0.0001)`，这样可以使模型在训练过程中避免权重过大，提高模型的泛化性能。还可以采用`Batch Normalization`技术，对每一层的输入数据进行归一化处理，加速模型的收敛速度并提高模型的稳定性，进一步提升模型在狮子、马和大象图像分类任务中的性能表现。

参考文献

1. 曹彪,刘敏杰,余其鹰,等.基于PSO-TCN深度学习模型的台兰河流域洪水预报研究[J/OL].中国防汛抗旱,1-12[2024-12-13]
2. 曹玲玲.基于深度学习的信息系统异常检测与预测模型[J].信息系统工程,2024,(11):20-23.
3. 王聪慧,石崇东,张军,等.基于深度学习的金属表面腐蚀图像语义分割模型性能分析[J].石油管材与仪器,2024,10(06).2024.06.012.
4. 周峻宇,施水才,王洪俊.基于深度学习的图像字幕生成综述[J/OL].软件导刊,1-10[2024-12-13].

附录

大象，狮子，马的CNN模型深度机器学习代码

<https://github.com/wamgdaidaodedaoke/daoke.git>

