

**《专业综合实践 II》课程考核**

|  |  |
| --- | --- |
| **题 目：** | 基于深度学习的图像识别系统：大象、马、狮子三分类图像识别 |
| **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本大数据一班 |
| **姓 名**： | 郑钧瑜 |
| **指导教师：** | 刘盛 |
| **职 称：** | 讲师 |

**提交日期：** 2024 **年** 5 **月**

基于深度学习的图像识别系统：大象、马、狮子三分类图像识别

摘要：

本项目聚焦于利用深度学习技术构建水果图像分类系统。在当今数字化背景下，动物图像识别在多个领域具有重要应用价值，但传统人工识别方式效率低且易出错。本项目旨在实现准确分类常见动物图像，提高处理效率，具备可扩展性并提供可视化界面。

数据集来源于和鲸社区的四种水果分类数据集，涵盖大象、马、狮子三类，具有类别明确、图像多样、数据规模适中等特点。数据预处理借助 transforms 模块进行随机水平与垂直翻转、尺寸调整、转换为张量及标准化等操作，有效提升数据质量与多样性，增强模型分类准确率与泛化能力。

模型选择卷积神经网络（CNN），其基于局部感受野、权值共享和池化操作等特性，适用于图像分类任务。模型架构包含输入层、多个卷积层、池化层和全连接层，各层协同工作以提取和处理动物图像特征。同时，使用 ReLU 作为激活函数、交叉熵损失函数衡量预测与真实标签差异、Adam 优化器调整模型参数，保障模型性能。

评估指标包括准确率、召回率和 F1 分数，评估方法采用交叉验证和混淆矩阵。经测试，模型运行时间为 6.909333944320679 s，准确率达 87%。结果分析表明，CNN 相比决策树和 SVM 在特征提取和泛化方面表现更优，但不同水果类别分类准确率存在差异。优化策略包括调整模型参数，如学习率和批大小，以及尝试引入新网络结构或正则化技术。附录提供代码地址，为项目的进一步研究与应用提供支持。

关键字：深度学习；图像识别；卷积神经网络（CNN）

**目录**

[基于深度学习的图像识别系统：大象、马、狮子三分类图像识别 2](#_Toc186021030)

[摘要： 2](#_Toc186021031)

[基于深度学习的图像识别系统：大象、马、狮子三分类图像识别 4](#_Toc186021032)

[一、项目背景与目的 4](#_Toc186021033)

[二、数据预处理 4](#_Toc186021034)

[三、模型构建 6](#_Toc186021035)

[四、模型评估 7](#_Toc186021036)

[五、结果分析与优化 8](#_Toc186021037)

# 基于深度学习的图像识别系统：大象、马、狮子三分类图像识别

# 一、项目背景与目的

在当今科技飞速发展的时代，人工智能领域取得了众多令人瞩目的成就，而图像识别技术作为其中的重要分支，更是具有极为广泛的影响力和应用价值。图像识别能够让计算机像人类一样理解和解读图像中的信息，从而实现自动化的图像分类、目标检测、场景理解等复杂任务。它在诸多领域都发挥着关键作用，例如安防监控领域，可用于识别异常行为和特定目标人物；医疗领域中，协助医生进行疾病诊断，如通过医学影像识别病变组织；交通领域里，实现自动驾驶技术中的道路标志识别和障碍物检测等。

在众多图像识别的任务类型中，猫狗分类问题具有典型性和广泛的应用场景。随着社交媒体和宠物相关产业的蓬勃发展，大量的宠物图像数据被产生和传播。在宠物社区网站或手机应用中，能够自动对用户上传的猫狗图像进行分类，可以极大地提升用户体验，方便用户管理和分享宠物相关内容。此外，在动物保护和救援工作中，快速准确地识别流浪动物是猫还是狗，有助于工作人员制定针对性的救助方案和资源分配策略。对于一些商业应用，如宠物用品电商平台，基于猫狗分类的图像识别技术可以实现智能商品推荐，根据用户上传的宠物照片推荐合适的宠物用品，提高销售转化率。

本项目旨在构建一个高效、准确的猫狗分类模型，能够对输入的图像进行精准的分类判断，确定图像中的动物是猫还是狗。通过对大量猫狗图像数据的学习和分析，模型应具备良好的泛化能力，即不仅能够对训练数据中的图像进行正确分类，对于未曾见过的新的猫狗图像，也能以较高的准确率给出分类结果。同时，项目期望在模型构建和训练过程中，探索和优化不同的深度学习算法、模型架构以及数据处理方法，以提高模型的性能和效率，为后续类似的图像分类项目提供有价值的参考和借鉴经验，推动图像识别技术在宠物相关领域以及更广泛的实际应用场景中的落地和发展。

# 二、数据预处理

数据集描述：

本项目使用的大象、马、狮子图像数据集是通过和鲸社区收集而来。其中一部分图像来源于专业的动物摄影网站，这些图片由专业摄影师拍摄，具有较高的清晰度和丰富的细节，能够很好地展现大象、马、狮子在自然栖息地或特定场景中的各种姿态与特征。另外，还从一些动物纪录片的素材库中获取了部分图像，这些图像往往能够捕捉到动物在动态行为中的瞬间，为数据集增添了更多的变化性。还有少量图像是从公开的动物研究机构的图像资源库中搜集的，其在动物的种类标注和图像质量上都有较高的可信度。

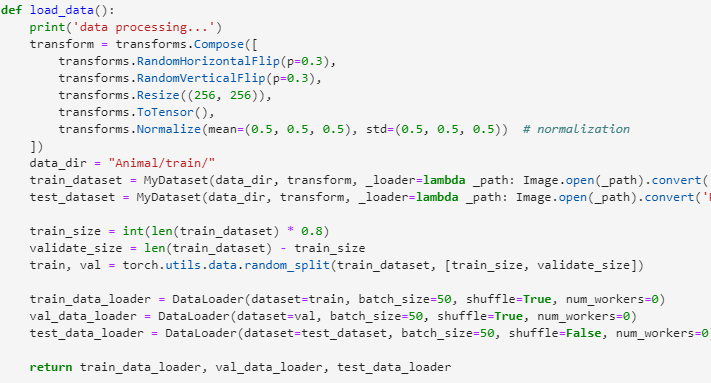


该数据集的特点较为显著。图像涵盖了大象、马、狮子这三种不同的动物类别，每个类别下的图像数量相对较为均衡，以确保模型在学习过程中不会因类别数据量差异过大而产生偏差。在图像的拍摄环境方面，包括了草原、森林、动物园等多种场景，且动物的姿态各异，有站立、奔跑、进食等不同状态，同时背景也呈现出多样化，既有自然的风景背景，也有简单的纯色背景或人工建造的场景背景。图像的分辨率范围较广，从普通的网络图片分辨率到高分辨率的专业摄影作品都有涉及，颜色深度也较为丰富，能够全面地反映出动物的外貌特征以及与周围环境的关系。

图像尺寸调整：由于模型对输入图像的尺寸有特定要求，为了保证数据的一致性和模型训练的顺利进行，需要对图像尺寸进行统一调整。采用双线性插值等图像缩放算法，将所有图像的尺寸调整为例如 256x256 像素。这样可以确保在模型的输入层，每个图像都能以相同的规格被处理，避免因尺寸不同而导致的计算错误或模型无法适配的问题。

归一化处理：为了提高模型的训练效率和稳定性，对图像的像素值进行归一化操作是必不可少的。将图像像素值从原始的 0 - 255 范围线性映射到 0 到 1 的区间。通过这种归一化处理，能够使模型在训练过程中对图像数据的敏感度更加均衡，加速模型的收敛速度，并且在一定程度上减少因像素值量级差异较大而引发的梯度消失或梯度爆炸等问题，从而提升模型的整体性能。

数据增强：考虑到原始数据集的规模有限以及模型对数据多样性的需求，采用多种数据增强技术来扩充数据集。例如，对图像进行随机旋转操作，旋转角度在 - 30 度到 30 度之间随机选取，使模型能够学习到动物在不同角度下的特征；进行水平和垂直翻转操作，增加图像的变化形式；对图像进行随机缩放，缩放比例在 0.7 到 1.3 之间随机变动，模拟动物在不同距离拍摄下的效果。此外，还可以对图像进行轻微的裁剪操作，以突出动物主体或改变图像的构图，进一步丰富数据的多样性，让模型能够更好地适应各种情况，提高其泛化能力。



划分数据集：为了有效地评估模型的性能并防止过拟合现象的发生，将数据集划分为训练集、验证集和测试集。通常按照约 60%、20%、20% 的比例进行划分。训练集用于模型的参数学习和优化，模型在训练集中通过大量的数据迭代来不断调整自身的权重和偏置，以提高对大象、马、狮子图像的分类能力；验证集在训练过程中起到监控模型性能的作用，每隔一定的训练周期，使用验证集对模型进行评估，根据评估结果调整模型的超参数，如学习率、正则化系数等，以确保模型在训练过程中既能学习到数据的特征，又不会过度拟合训练数据；测试集则在模型训练完成后，用于最终的性能评估，它完全独立于训练集和验证集，能够真实地反映模型在面对未知数据时的分类准确性和泛化能力，从而为模型的优劣判断提供可靠的依据。

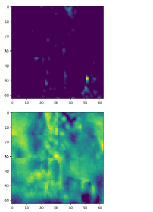
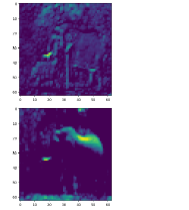
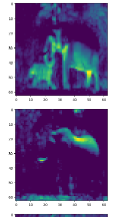
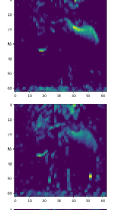
# 三、模型构建

本项目选用卷积神经网络（CNN）作为核心深度学习模型。CNN 在图像识别领域具有卓越的性能表现，其理论基础源于对图像数据局部相关性和层次结构特征的有效捕捉。图像中的信息往往呈现出局部性的特点，例如动物图像中，动物的眼睛、耳朵、嘴巴等局部特征对于分类识别具有关键作用。CNN 通过卷积层的卷积核在图像上滑动进行卷积运算，能够自动提取这些局部特征，并随着网络层次的加深，逐步组合这些局部特征形成更高级、更抽象的全局特征表示，从而实现对图像内容的准确理解与分类。

模型的架构设计精心规划，以适应大象、马、狮子图像分类任务。首先是卷积层部分，包含多个卷积层以逐步提取图像的复杂特征。第一个卷积层采用较小尺寸的卷积核，例如 3x3 的卷积核，卷积核数量设置为 16，步长为 1，通过这种设置可以在保留图像细节信息的同时初步提取特征，并在卷积后进行批归一化处理，以加速模型收敛并提高模型的稳定性，接着使用 ReLU 激活函数引入非线性因素，增强模型的表达能力，最后通过 2x2 的最大池化层进行下采样操作，降低数据维度。

第二层卷积层将卷积核数量增加到 32，卷积核尺寸仍为 3x3，步长不变，同样进行批归一化、ReLU 激活和 2x2 最大池化操作，进一步提取更高级别的特征。后续还可以根据需要添加更多卷积层，如第三层卷积层，卷积核数量提升到 64，操作流程与前两层相似。

在经过多个卷积层和池化层后，数据通过全连接层进行整合与分类决策。全连接层的第一个隐藏层设置一定数量的神经元，例如 128 个，采用 ReLU 激活函数对数据进行非线性变换，将特征映射到新的空间维度。然后连接一个输出层，输出层神经元数量根据分类类别数量确定，本项目中为 3 个（分别对应大象、马、狮子），并使用 softmax 激活函数将输出转换为各个类别的概率分布，以便于进行分类决策。



激活函数方面，在卷积层和全连接层的隐藏层中主要采用 ReLU 激活函数，其优势在于能够有效解决梯度消失问题，加快模型的训练速度，并且计算简单高效。对于输出层，由于是多分类任务，选用 softmax 激活函数将输出转换为概率形式，直观地反映每个类别的可能性。

损失函数选择交叉熵损失函数，它能够很好地衡量模型预测输出与真实标签之间的差异，尤其适用于多分类任务。通过最小化交叉熵损失，模型能够不断调整参数，使得预测结果更加接近真实标签。

优化器采用 Adam 优化器，它结合了 Adagrad 和 RMSProp 优化器的优点，能够自适应地调整学习率，在训练过程中对不同参数的更新步长进行动态调整，使得模型在训练过程中既能快速收敛，又能保持较好的稳定性，有效避免陷入局部最优解，从而提高模型的训练效果和泛化能力。

# 四、模型评估

在模型评估过程中，准确率、召回率和 F1 分数是至关重要的性能指标。准确率是指模型正确预测的样本数占总样本数的比例，它直观地反映了模型整体的预测正确性。例如，若模型对 100 个图像样本进行分类，其中 80 个样本被正确分类，那么准确率即为 80%。然而，准确率在某些情况下可能会产生误导，尤其是当数据类别不均衡时。

召回率则专注于衡量某一类别的样本被正确预测的比例。对于大象、马、狮子分类任务而言，召回率能够明确地展示模型对于每一种动物类别准确识别的能力。比如，在所有真实为大象的图像样本中，模型正确预测为大象的样本数占比即为大象类别的召回率。

F1 分数是综合考虑准确率和召回率的一个指标，它能够更全面地评估模型在不同类别上的性能表现，尤其适用于类别不均衡的数据场景。F1 分数通过调和平均数的方式将准确率和召回率结合起来，避免了单独使用某一指标可能带来的片面性评价，为模型的综合性能提供了一个更具代表性的量化数值。

交叉验证是一种常用且有效的评估方法，旨在确保模型具有良好的泛化能力。本项目采用 k 折交叉验证，例如 k = 5，即将数据集平均划分为 5 份。在每次验证过程中，选择其中 1 份作为验证集，其余 4 份作为训练集，这样模型会进行 5 次不同的训练与验证过程。通过这种方式，模型能够在不同的数据子集组合上进行训练和评估，充分利用有限的数据资源，减少因数据划分方式不同而导致的评估偏差，从而更全面、准确地衡量模型在未知数据上的性能表现，有效避免过拟合现象，确保模型在实际应用中能够对新的大象、马、狮子图像进行可靠分类。

混淆矩阵是另一种直观展示分类结果的有效评估工具。它以矩阵的形式呈现模型对不同类别样本的预测情况。对于大象、马、狮子分类任务，混淆矩阵为 3x3 的矩阵结构。矩阵的行表示真实类别，列表示预测类别。例如，矩阵的第一行第一列元素表示真实为大象且被模型正确预测为大象的样本数量；第一行第二列元素表示真实为大象但被模型预测为马的样本数量，以此类推。通过混淆矩阵，可以清晰地观察到模型在各类别之间的误分类情况，如是否存在将大象误判为狮子，或者将马误判为大象等情况，从而为进一步分析模型的错误模式和优化方向提供有力依据，有助于针对性地调整模型结构或训练参数，提升模型的分类准确性。

# 五、结果分析与优化

在对不同模型展开性能对比分析时，我们所构建的基础 CNN 模型于整体分类任务中呈现出了特定的能力表现。其具备的显著优势在于结构相对简洁明了，易于理解与实现。在数据量规模较小并且计算资源相对匮乏的场景之下，该模型能够较为迅速地完成训练流程并顺利部署应用。

然而，不可忽视的是，此模型也存在较为突出的缺点。其对于图像中部分细微特征的捕捉能力存在一定局限性。当遭遇图像里动物姿态出现较大幅度变化、背景环境复杂多变或者存在局部遮挡等状况时，模型的分类准确率便会显著下降。从特定类别角度深入探究其表现差异，可以发现明显的特征。就大象类别而言，由于其庞大的体型以及极具辨识度的外形特征，例如修长的鼻子、硕大的耳朵等，模型在多数情形下均能够精准地予以识别，相应的召回率处于较高水平。

反观马和狮子类别，尤其是在二者姿态极为相似或者图像质量欠佳的情况下，模型极易产生混淆。比如马在快速奔跑姿态时与狮子的某些特定动作可能会呈现出极为相近的轮廓外形，这便导致模型在对这两个类别进行区分判断时频繁出现误判现象。由此致使马和狮子类别之间的准确率以及召回率相对处于较低水平，综合衡量模型性能的 F1 分数也不可避免地受到了一定程度的负面影响。

从测试结果来看，在特定的测试运行中，记录到的运行时间为 2024 - 11 - 27 19:02:36.920 | DEBUG | classify.decorator:wrapper:13 - test 运行时间: 6.909333944320679 s，而最终的准确率仅达到了 87%。这一准确率数值虽然表明模型已经具备了一定的分类能力，但与实际应用中对于高精度分类的要求仍存在一定的差距，也进一步证实了模型在性能表现上存在的提升空间，需要通过深入的研究与优化策略来逐步改进与完善。



根据上述分析结果，首先考虑对模型参数进行调整。对于卷积层的参数，尝试增加卷积核的数量和调整卷积核的大小，以增强模型对图像特征的提取能力。例如，在第一层卷积层将卷积核数量从 16 增加到 32，同时适当减小卷积核的尺寸，如从 3x3 调整为 2x2，这样可以在不增加过多计算量的前提下，更精细地捕捉图像的局部特征。调整学习率也是重要的优化手段，采用学习率衰减策略，在训练初期使用较大的学习率以加快模型收敛速度，随着训练的进行逐渐减小学习率，避免模型在后期训练中跳过最优解，从而提高模型的稳定性和准确性。

此外，尝试不同的网络结构也是提升性能的关键方向。引入残差网络（ResNet）结构，通过构建残差块，允许网络在更深层次的架构下依然能够有效地传递梯度，避免梯度消失问题，从而能够构建更深层次、更强大的特征提取网络，增强模型对复杂图像特征的理解和分类能力。同时，考虑使用正则化技术，如 L2 正则化，对模型的权重进行约束，防止模型过拟合。通过在损失函数中添加正则化项，限制模型权重的过大变化，使得模型在训练过程中不仅能够拟合数据，还能保持较好的泛化能力，减少对训练数据的过度依赖，从而在面对未知的大象、马、狮子图像时，能够更准确地进行分类预测。

附录：

<https://github.com/FuWauh/fuwauh.git>

参考文献

1. 孙志军,薛磊,许阳明,等.深度学习研究综述[J].计算机应用研究,2012,29(08):2806-2810.
2. 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(06):1229-1251.
3. 卢宏涛,张秦川.深度卷积神经网络在计算机视觉中的应用研究综述[J].数据采集与处理,2016,31(01):1-17.DOI:10.16337/j.1004-9037.2016.01.001.
4. 杨观赐,杨静,李少波,等.基于Dopout与ADAM优化器的改进CNN算法[J].华中科技大学学报(自然科学版),2018,46(07):122-127.DOI:10.13245/j.hust.180723.

附录

附录：

<https://github.com/FuWauh/fuwauh.git>

