

文章编号: 2095-2163(2023)05-0070-06

中图分类号: TP317.4

文献标志码: A

基于 VGG 卷积神经网络的动物特性识别与优化路径

龚建伟, 孟博文, 童昱恒, 孔煜杰, 谭仪慧

(北京师范大学 香港浸会大学联合国际学院理工科技学院, 广东 珠海 519087)

摘要: 使用卷积神经网络进行图像识别, 可以大大降低图像辨识的成本, 在二分类问题中尤其如此。VGG 模型是一种相当流行的卷积神经网络, 其特性在于以小卷积核和“网络块”替代传统神经网络中的大卷积核与神经网络层, 这意味着其深度有所增加, 同时具有较强的迁移性与改进潜力。通过测试发现, 增加 VGG 块数的同时, 搭配图像增强是可靠的改进手段; 而增加 epoch 有利有弊, 对网络进行 dropout 的成效不太理想。针对于此, 实验构建了一个准确度为 83.3% 的轻量化 VGG 模型, 该模型相较 VGG-16 而言要轻量化许多, 表明根据合理的方向构建轻量化 VGG 模型用于动物识别是可行的。

关键词: 动物识别; VGG; 卷积神经网络; 优化路径

Study on the characteristics and optimization of VGG CNN for animal recognition

GONG Jianwei, MENG Bowen, TONG Yuheng, KONG Yujie, TAN Yihui

(Faculty of Science and Technology, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai Guangdong 519087, China)

【Abstract】 Using CNN for image recognition can greatly reduce the cost of image recognition, especially in classification problems. VGG model is a very popular convolution neural network. Its characteristic is that small convolution kernel and "network block" are used to replace the large convolution kernel and neural network layer in the traditional neural network, which means that its depth is increased and has strong improvement potential. Through the test, we found that increasing the number of VGG blocks and image enhancement are reliable means of improvement. Adding epoch has advantages and disadvantages, but the effect of dropout on the network is not ideal. Finally, this project has built a lightweight VGG model with an accuracy of 83.3%, which is much simpler than VGG-16, which shows that it is feasible to build a lightweight VGG model for animal recognition.

【Key words】 animal recognition; VGG; convolutional neural network; optimization

0 引言

随着图像识别技术的不断发展, 如今图像识别不仅可用于不同种类物体的识别, 对于较为相似的物体也可以进行准确度较高的区分。在诸多图像识别项目中, 动物识别具有较为特殊的意义, 在智能家居、智慧牧业等诸多场合都有用武之地。目前虽然已有的图像识别技术已经可以较好地地区分外貌差异较大的动物, 但是当面临长相相似的动物之时, 快速、高效地进行准确识别仍是一项具有挑战性的工作。猫、狗作为体型与外貌相对接近的动物, 对于图片识别来说还是较有挑战性的难题。如何在确保准确度的同时又能提升识别准确度, 在识别速度和准确度之间找到均衡, 是诸多动物图像识别项目需要面临的挑战。

目前, 利用卷积神经网络 (Convolutional Neural Network, CNN) 自动提取图像特征, 然后放入训练好的模型进行识别, 是图片识别项目最为常用的手段^[1-3]。Le Net 模型是非常经典的卷积神经网络模型, 不仅可用于识别颜色特征^[4], 也可用于识别人脸^[5]等彼此差别较小的图像。随着图像识别技术日益进步, 人们意识到传统的 Le Net 模型已不能满足图像识别的需求, 在精度等方面都有待提升。虽然 L2 正则化和 Dropout 优化网络等方法的引入, 可以减轻 Le Net 模型的过拟合问题^[6], 基于 Le Net 也可以搭建更好的混合激活函数卷积神经网络^[7], 但对其它模型的探索还是被人们提上了日程。在诸多进一步发展的 CNN 模型中, VGG 模型是较具代表性的优质模型, 其主要思路是通过增加卷积神经网络的深度和宽度来提升模型的性能^[8]。虽然现有

作者简介: 龚建伟 (1998-), 男, 硕士, 主要研究方向: 人工智能与数码媒体。

收稿日期: 2022-05-31

的VGG-16等已经是相当优秀的模型,但由于VGG模型的基础架构,使其所包含的权重参数数目相当之大。因此,在此基础上寻找改进版的轻量级模型是许多图像识别项目的工作^[9]。设计合理的浅层VGG模型不但训练耗时少、泛化能力强,且准确度也有保障^[10]。目前已有研究者尝试在既有的VGG-16模型基础之上进一步改进以用于图像识别,取得了较好的成效,这证明VGG模型具有充分的改造潜力^[11]。但VGG-16模型对于动物识别来说可能过于臃肿。本研究试图构建多种不同的轻量级VGG模型并且验证准确度,最终构建出用于动物识别时准确度较高且足够轻量级的VGG模型。此外,在实验过程中亦可得到一些改进思路与方向,这对未来研究者进一步改进模型也有所帮助。

1 技术描述

1.1 图片处理

为了检查导入的图片是否正确,实验首先使用matplotlib库,绘制几张导入后的图片以进行检查。然后需要建立若干新目录,并将图片按一定比例随机分配至测试集与训练集当中。根据实验的设置,将有25%的数据被保留在测试集内。在完成这一步之后,实验选择利用Keras库中的ImageDataGenerator在模型运行时对图片进行预处理,以将其规范化为200×200像素的图片,有助于后面进一步开展模型训练与预测工作。

1.2 模型构建

在完成图片的导入、展示与预处理之后,进一步将图片划分为几部分,并使用卷积神经网络(CNN)进行图像识别。考虑到本次探究的目的,实验采用了传统VGG模型(全程使用3×3的卷积和2×2的max pooling)作为此次项目的基础,传统VGG模型结构简单且应用性强,而且有较为明确的改进方向。不仅如此,相较GoogLeNet等模型,VGG在迁移学习任务中的表现更好^[12],这意味着探究VGG的改进具有更深的意义。与其他模型相比,VGG的优势在于其采用连续的小卷积核代替较大卷积核,以获取更大的网络深度^[13]。

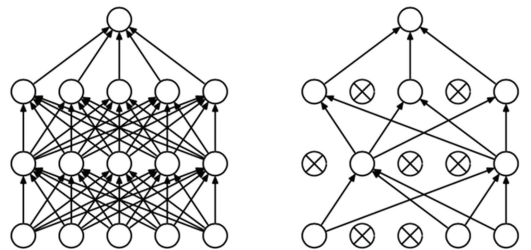
VGG模型的突出特点是使用卷积块和全链接块来代替传统的卷积层与全连接层,这里的块是由多个输出通道相同的层构成。max pooling被用于划分不同的块,而既有的VGG模型则根据网络层数不同被分为VGG-11、VGG-13、VGG-16、VGG19等。虽然既有模型层数较多且目前的诸多研究仍然侧

重以VGG-16为基础进行改进^[14],但直接使用更加轻量化的VGG模型进行图像识别仍是可能的^[15]。

为了节约资源,实验尝试了单块,以及两块、三块、四块与五块的VGG模型,不同块的过滤器数量分别为32、64、128、256、512个(均只有一个全连接层)。因为这是一个二分类问题,故class_mode被指定为binary。为了检查过拟合现象以及训练的成效,实验还绘制了损失曲线图以及准确率曲线图,其中蓝色曲线为训练集结果,橙色曲线为测试集结果。

1.3 图像数据增强与Dropout

测试过程中发现,在识别准确度提高的同时,也会存在较为明显的过拟合现象,这是卷积神经网络的通病。不仅如此,由于猫与狗确实较为相近,因此准确率也仍然有进一步提升的空间。为了解决这一问题,实验尝试了Dropout正则化和图像增强两种技术。Dropout正则化会丢弃一些网络中的神经元,以模拟不同结构的网络,其基础特性在于节点删除的完全随机性^[16];而图像增强则是改变图像的视觉效果,从而增加训练集的数据量。应用Dropout前后的神经网络示意如图1所示:



(a) Dropout应用前神经网络 (b) Dropout应用后的神经网络

图1 Dropout技术示意图

Fig. 1 Diagram of Dropout

2 实验与结果分析

2.1 实验环境

实验软硬件环境设定为:16 GB计算机内存,Windows10 64位1909版操作系统,搭载有Intel Xeon E3-1231v3 3.4 GHz处理器与GTX960显卡;基于Python3.6编程语言的TensorFlow深度学习框架中集成的Keras库。

2.2 数据集检查

在正式实验开始之前,由于猫狗并非同时导入系统,因此预先使用matplotlib分别绘制数据集,以检查图像数据导入的正确行,如图2、图3所示。

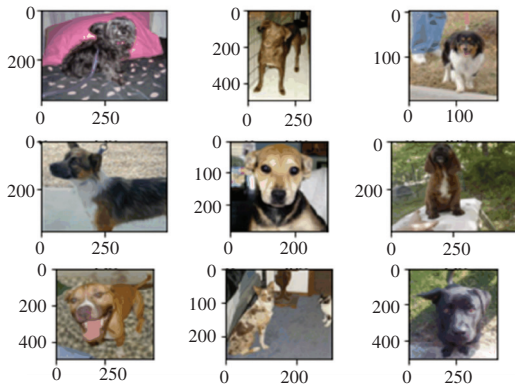


图2 使用 matplotlib 绘制的狗图像数据集

Fig. 2 Dog images dataset drawn with matplotlib

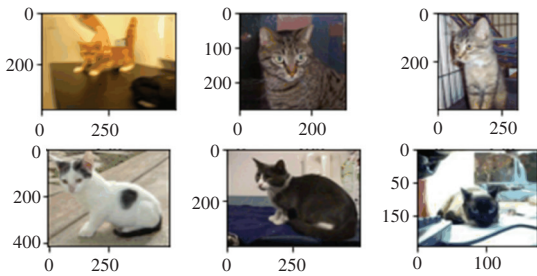


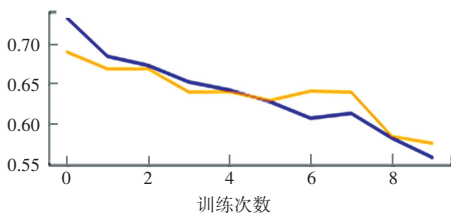
图3 使用 matplotlib 绘制的猫图像数据集

Fig. 3 Cat images dataset drawn with matplotlib

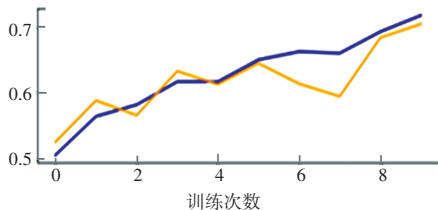
2.3 实验过程

2.3.1 Dropout 对比试验

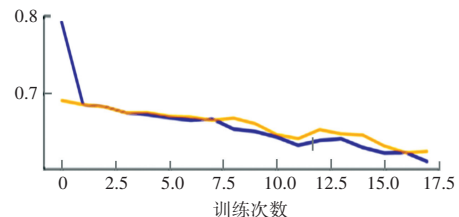
实验首先基于 5 000 张图片数据集尝试对双块



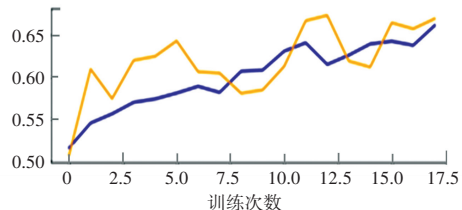
(a) 三块 VGG 且无 Dropout 的交叉熵损失函数



(b) 三块 VGG 且无 Dropout 的准确度



(c) 三块 VGG+Dropout 的交叉熵损失函数



(d) 三块 VGG+Dropout 的准确度

图4 三块 VGG 且无 Dropout 与三块 VGG+Dropout 的对比

Fig. 4 Comparison between 3 block VGG model without dropout and 3 block VGG model+dropout

2.3.3 VGG-16 模型实验

数据集样本中若仅有 5 000,在某些情境下或许显得有些少,因此需要探究数据集数量是否影响了模型的准确度。在与既有的 VGG-16 对比之后不

模型进行 Dropout 正则化以模拟不同的网络结构。实验在每一次卷积中设定 10% 的 Dropout 值,而在全连接层,这一数字被增加到 35%。经过测试可以发现,过拟合现象虽然有所减轻,但是准确率确有明显下降。如图 4 所示,对比三块 VGG 和三块 VGG+Dropout 的测试结果可以发现,虽然过拟合现象有所减轻,但是准确率始终保持在较低水平。

2.3.2 图像增强实验

除 Dropout 正则化之外,实验还考虑了图像增强。为避免破坏图片本体结构,实验基于 5 000 张图片数据集选择了一定的水平和垂直位移/水平翻转来获取更多的样本,用以提升模型的准确率。测试中尝试了多种位移值,三块 VGG 所对应的位移值为 0.12,两块 VGG 所设置的位移值为 0.22,翻转均被设置为 True。

通过对比发现,三块 VGG 且有图像增强时,过拟合现象显著小于单纯的三块 VGG 时的状态;无图像增强时,测试集准确率大多时候都低于训练集,但有图像增强之后有所不同。不仅如此,模型本身的准确率并未受到太大影响。但令人遗憾的是,虽然模型训练时 epoch 已被提升至 40,但最终的准确率仍然只有 72.1% 左右。两块 VGG+图像增强的最终准确度为 69.1%,较原先有明显提升,证明了图像增强的有效性,同时也可看到测试集曲线波动仍然较大,如图 5 所示。

难发现,模型实际上仍有较多改进空间。事实证明,VGG-16 在样本数为 5 000 的前提下依旧表现出色,仅仅两个 epoch 就可以达到很高的效能,但这与 VGG-16 模型的复杂度有关,如图 6 所示。

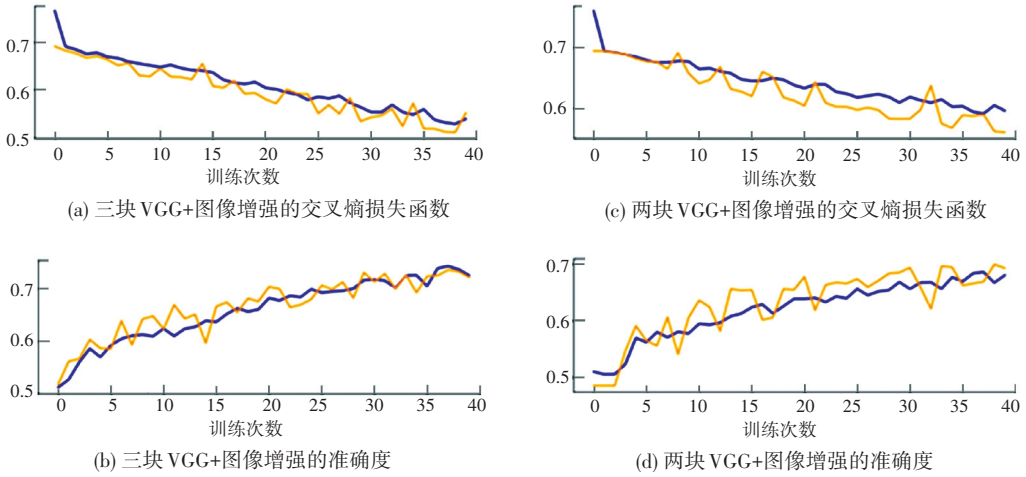


图 5 三块 VGG+图像增强与两块 VGG+图像增强的对比

Fig. 5 Comparison between 3 block VGG model+image data augmentation and 2 block VGG model+image data augmentation

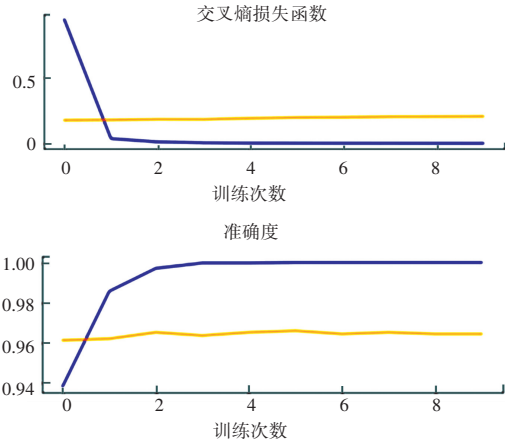


图 6 VGG-16 测试结果

Fig. 6 Test results based on VGG-16 model

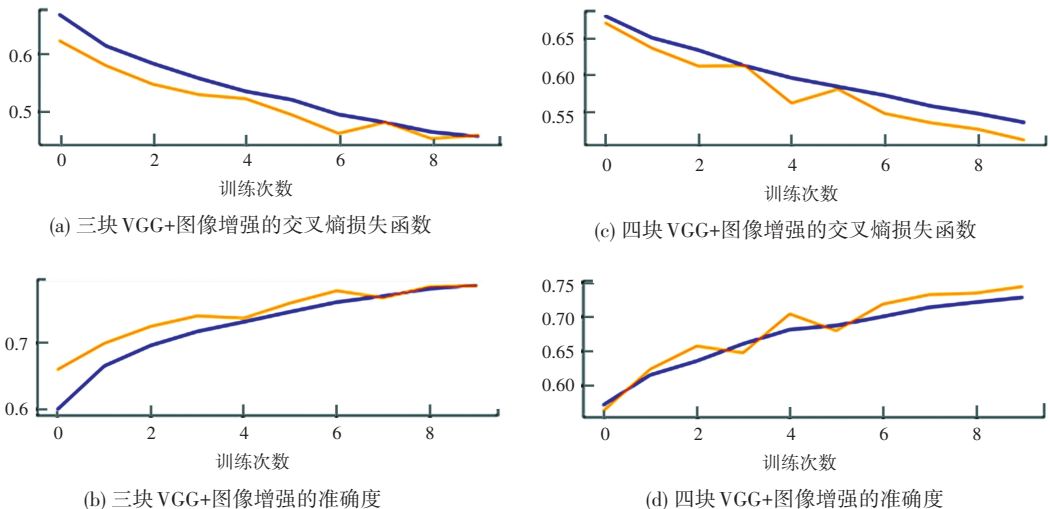


图 7 三块 VGG+图像增强和四块 VGG+图像增强的对比

Fig. 7 Comparison between 3 block VGG model+image data augmentation and 4 block VGG model+image data augmentation

2.3.4 大样本实验

对实验所设计的轻量级模型而言,采用包含 2.5 万张图片的数据集,结果如图 7 所示。在针对 2.5 万张图片的完整数据集进行测试时可以发现,测试集曲线要显著比 5 000 张图片时平缓。此外,三块 VGG+图像增强和四块 VGG+图像增强的准确率都有所提升,分别达到了 74.2%和 78.5%的水平,这说明有更多样本用于训练对提升准确率很有帮助。

2.3.5 单块多卷积层模型实验

由于 2.5 万张图片的数据集测试起来比较费时,所以后续实验基于 1 万张图片的数据集进行测试。图 8 分别为四块 VGG 和三块 VGG,但是最后一个块有两个卷积层的测试结果。

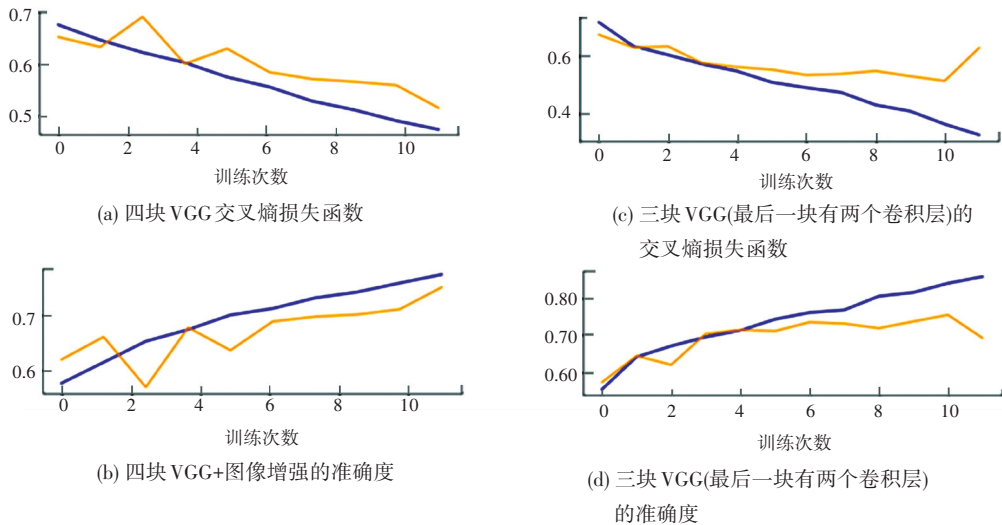


图 8 四块 VGG 和三块 VGG 测试结果

Fig. 8 Comparison between 4 block VGG model and 3 block VGG model

两个模型均出现了相当明显的过拟合现象。其中四块 VGG 的最终准确率为 75%，三块 VGG 的准确率则为 69.2%。虽然两个模型均有 4 个卷积层，但池化层的分布(用于划分块数)仍然对结果产生了巨大的影响。

2.3.6 四块 VGG+图像增强模型实验

由上述结果可以看出,四块 VGG 的准确率较 5 000 张图片时(70.9%)有明显提升,这说明其具有进一步改进的潜力。因此,实验尝试了四块 VGG+图像增强(偏移值设置为 0.09)的模型,最终在经过 40 epochs 的训练后,得到了 83.3% 的准确度,如图 9 所示。这一结果已经超过了 2007 年的 SVM 分类器的水平^[17]。

3 结束语

从模型构建结果来看,实验最终构建得到的四块 VGG+图像增强(偏移值设置为 0.09)的模型准确度在 80% 以上,对于轻量级模型而言,这一结果比较令人满意。模型相对传统的 VGG-16 而言可算相当轻量级的模型,在训练速度等方面都有显著优势。因此,这一模型可以较好地满足智慧牧业、社区监控等应用的需求。除此之外,实验所得到的方向性结论也可为试图构建轻量级 VGG 模型的研究者提供一定帮助。

参考文献

- [1] Brahim Achour. Image analysis for individual identification and feeding behaviour monitoring of dairy cows based on Convolutional Neural Networks (CNN) [J]. Biosystems Engineering, 2020, 198 : 31-49.
- [2] 杜紫薇,周恒,李承阳,等. 面向深度卷积神经网络的小目标检测算法综述[J]. 计算机科学, 2022, 49(12) : 205-218.
- [3] 李炳臻,姜文志,顾佼佼,等. 基于卷积神经网络的目标检测算法综述[J]. 计算机与数字工程, 2022, 50(5) : 1010-1017.
- [4] 王济民,魏怡,周宇,等. 基于 LeNet-5 卷积神经网络和颜色特征的限速标志识别[J]. 计算机科学, 2021, 48(S2) : 345-350.
- [5] 赵彩敏,刘国红. 基于改进的 LeNet-5 卷积神经网络的人脸表情识别[J]. 许昌学院学报, 2021, 40(2) : 113-116.
- [6] 赵小强,罗维兰. 改进卷积 Lenet-5 神经网络的轴承故障诊断方法[J]. 电子测量与仪器学报, 2022, 36(6) : 113-125.
- [7] 刘国柱,赵鹏程,于超,等. 基于混合激活函数的改进卷积神经网络算法[J]. 青岛科技大学学报(自然科学版), 2021, 42(1) : 114-118.
- [8] 苟鹏. 基于 VGG 图像分类算法研究[D]. 四川:西南交通大学, 2021.

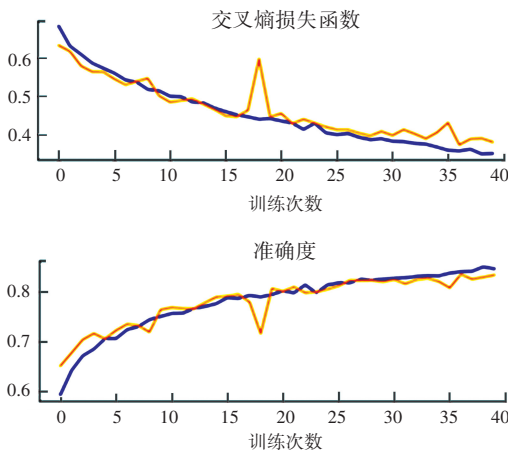


图 9 四块 VGG+图像增强的测试结果

Fig. 9 Test results of 4 block VGG model + image data augmentation