

基于VGG和ResNet的小样本人脸识别模型

黄嘉悦

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2023年5月9日; 录用日期: 2023年6月23日; 发布日期: 2023年6月30日

摘要

随着科学技术的发展和互联网信息的传播, 人们有了更加广泛的学习与娱乐平台, 可以通过互联网了解到世界各地的信息。人们的娱乐方式和以前相比也有了巨大提升。观众在观看国外比赛时, 由于不了解比赛人员信息, 很多时候会出现“脸盲”的情况。因此本文致力于通过分析国外明星的人脸信息, 帮助人们更快了解想要知道的明星情况。人脸识别技术在人们生活中许多方面都发挥着很大的作用。用来研究这项技术的方法也有很多。其中卷积神经网络(Convolutive Neural Networks, CNN)是当前推动机器学习的一项重要技术, 该项技术在图像分类中也表现出很好的效果。当数据集是小样本时, 使用预训练模型进行预测分析有比较好的效果, 是一种高效的识别方法。本文主要通过CNN和已有的预训练模型帮助进行人脸识别。通过几何变换、图像模糊化、调整亮度和对比度等方法, 进行随机搭配组合实现数据增强, 达到扩充数据集的效果。数据增强消除了样本数据的尺度、位置和视角差异等因素, 满足模型的平移不变性和尺度不变性, 增强了训练模型的鲁棒性, 提高了训练模型的识别准确率。此外在输入层加入Batch Normalization操作, 使每层神经网络输入是相同分布, 加快训练速度, 提高学习率, 并且使用自适应ReLU和RMSProp算法来提高收敛速度、降低错误率。最终该网络模型达到了76.6%的准确率。本文选择VGG16、VGG19和ResNet50预训练模型对样本数据进行拟合。通过大量试参数调整分析得到VGG19的模型效果最好, 达到了89.4%的准确率。

关键词

卷积神经网络, 人脸识别, 图像分类, 预训练模型

Small Sample Face Recognition Model Based on VGG and ResNet

Jiayue Huang

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: May 9th, 2023; accepted: Jun. 23rd, 2023; published: Jun. 30th, 2023

文章引用: 黄嘉悦. 基于 VGG 和 ResNet 的小样本人脸识别模型[J]. 运筹与模糊学, 2023, 13(3): 2474-2486.
DOI: 10.12677/orf.2023.133248

Abstract

With the development of science and technology and the spread of information on the Internet, people have a wider platform for learning and entertainment, and can learn about information from all over the world through the Internet. People's entertainment methods have also improved tremendously compared to the past. When watching foreign games, viewers are often "face-blind", because they do not know the information of the players. Therefore, this paper is dedicated to analyzing the face information of foreign celebrities to help people know more about the celebrities they want to know more quickly. Face recognition technology plays a big role in many aspects of people's lives. There are also many methods used to study this technology. One of them is Convolutional Neural Networks (CNN), an important technology that is currently driving machine learning, which has also shown good results in image classification. When the data set is a small sample, using pre-trained models for predictive analysis has better results and is an efficient recognition method. This paper focuses on face recognition with the help of CNN and existing pre-trained models. Data augmentation is achieved by random pairwise combinations of geometric transformation, image blurring, and adjustment of brightness and contrast to expand the data set. The data enhancement eliminates the scale, position and viewpoint differences of sample data, satisfies the translation invariance and scale invariance of the model, enhances the robustness of the training model, and improves the recognition accuracy of the training model. In addition, the Batch Normalization operation is added to the input layer so that the input of each neural network layer is identically distributed to speed up the training and improve the learning rate, and the adaptive ReLU and RMSProp algorithms are used to improve the convergence speed and reduce the error rate. The final network model achieves an accuracy of 76.6%. In this paper, VGG16, VGG19 and ResNet50 pre-training models are selected to fit the sample data. The best model of VGG19 was obtained through a large number of trial parameters adjustment analysis, and achieved an accuracy of 89.4%.

Keywords

Convolutional Neural Networks, Face Recognition, Image Classification, Pre-Trained Model

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

如今的社会，随着当代网络科技的不断发展和进步，人脸识别技术在无人超市、移动支付、门禁系统等领域都有应用。现阶段早已有许多技术能够用来进行人脸识别，其中卷积神经网络(Convolutional Neural Networks, CNN)在图像分类中早已获得了出色的实际效果。人脸识别验证既有串行识别也有并行实现，更符合人类的识别习惯，可交互性强[1]、带来了很大的推动力[2]。2015年庄福振等建立迁移学习模型并指出了迁移学习下一步可能的研究方向[3]。2018年尹红针对在小样本数据集的条件下CNN难以实现网络参数的最优化且易出现过拟合现象的问题，提出一种花卉图像分类算法。结果表明，在小样本花卉数据集的条件下，该方法能达到较好的效果[4]。2020年赵浩选用交叉验证的方式训练成功概括出一整套有效的实践策略[5]。左羽等将传统卷积神经网络VGG16与全卷积网络(FCN)结合在一起对图像进行分类，并取得了不错的效果[6]。梁雁等研究了CNN模型，实验证明了CNN模型在进行表情识别过程

中的准确度最高、速度最快[7]。马俊、张荣福等人主要研究了基于迁移训练的 VGG16 模型对网络芯片进行分类[8]。田佳鹭等人将 VGG16 模型进行微调，从而提高准确率[9]。王文成等人将 ResNet50 预训练模型运用于动物的分类中，同样取得了很好的效果[10]。杨凡稳提出了一种基于改进猫群算法的图像目标分类的方法，运用该方法进行分类。结果表明该算法相比于原始的猫群算法运行速度有所提高[11]。俞显刚构建深度卷积神经网络与长短时记忆循环神经网络相结合的网络模型从而提高了模型准确率[12]。

早期的人脸识别方法是基于几何特征的方法，但该方法稳定性不高，识别效果差，典型的算法有活动轮廓模型[13]和可变形模板模型[14]等。2012 年勒尼德研究小组首次将深度学习运用进人脸图片中，取得了 87% 的准确率。1996 年 Dhananjay Thekkedath 等人基于不同的面部表情对比了 vgg16、resnet50 和 se-resnet50 三种预训练网络模型，得出了 resnet50 效果最好[15]。2021 年 Omiotek Zbigniew 等人结合 vgg16 模型将燃烧状态的准确率确定在 82%~98% 之间，实现了高精度和快速处理分类[16]。

以上的文章都是基于某特定人群的分类，如黄种人、白种人的分类或某组织中成员的分类。针对与国外明星人脸的分类并没有比较特定的文章。

2. 人脸识别的相关理论

2.1. 卷积神经网络基本概念

CNN 是有着一定深度构造且含有某种卷积运算的前馈神经网络。卷积神经网络有着表征学习功能，对比初期的人工神经网络，其最重要的特征是“局部连接”与“权值共享”。这两种特征的产生，不仅减少了权值的数量而且促使网络得到了提升，同时也相应的降低了过拟合的风险。是深度学习的代表算法之一。

CNN 在二维图像处理上也具备诸多优点，例如网络能自主提取图像特征如颜色、形状、纹路等。在处理二维图像问题上，尤其是在识别位移、放缩以及其它维持形式歪曲不变性的应用上具备优良的鲁棒性和高效的运算速率等。

2.1.1. 多层神经网络

卷积神经网络被广泛应用于自然语言处理、图像识别等众多领域。一个完整的卷积神经网络包含有输入层、卷积层、池化层、全连接层等。其中卷积层、池化层和全连接层被称作隐含层。每一层有很多个特征图，每个特征图利用一种卷积滤波器获取输入的一种特征，每个特征图又具有很多个神经元。该网络利用卷积层运算，提高特征信号。利用下采样层减少计算量并维持图像旋转的不变性。最终利用全连接层获得所提取到的图片特征。

2.1.2. 激活函数

激活函数就是在神经元中，将输入通过加权求和后要经过一个函数，这个函数就是激活函数。如果不加入激活函数那么整个输出就是输入的线性组合。如果加入激活函数后就可以拟合非线性函数，从而将模型适用于非线性模型。常见的激活函数有 Rectified Linear Unit (ReLU) 函数、Sigmoid 函数、tanh 函数和 Softmax 函数等。

ReLU 函数是最常见的激活函数。该函数的运算公式为：

$$f(x) = \max(0, x)$$

如果数据的输入值大于 0 就把输入的数据直接输出，否则输出为 0。ReLU 函数不仅在面对数据量大的情况下具有良好的性能，此外该函数的运算速度也比传统函数快。经过简单的阈值化就可以达到参数稀疏的效果。但是，该函数在神经元训练时会有出现“死亡”的情况，使得该单元无法激活数据。

Tanh 函数见图 1，也是一种较为常见的激活函数。但与上述 sigmoid 函数相比，其输出值均值为 0，收敛速度比 sigmoid 快、迭代次数少。tanh 函数的运算公式为：

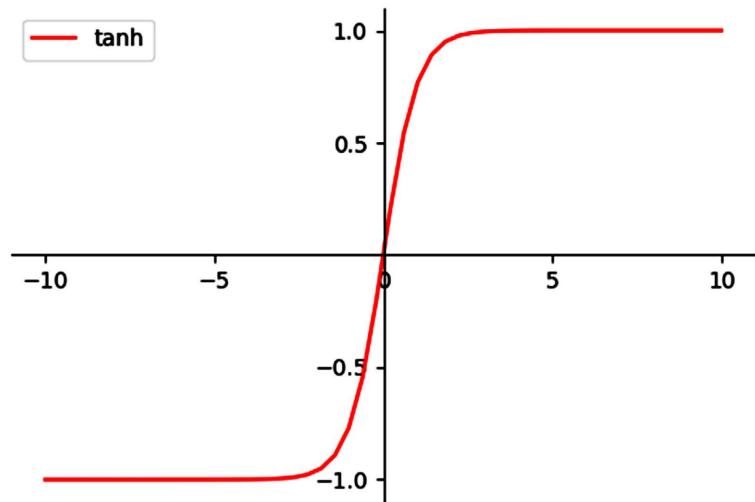


Figure 1. Tanh function
图 1. Tanh 函数

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Sigmoid 函数也被称为逻辑(Logistic)函数，取值范围为(0, 1)通常被用来做二分类。该函数比较平滑且易于求导。Sigmoid 函数的运算公式为：

$$f(x) = \frac{1}{1 + e^{-x}}.$$

Softmax 函数通常也被称为归一化指数函数，是 Logistic 函数的一种推广。通常用来作为多分类的激活函数。经过该函数可以使得每一元素的取值范围为(0, 1)，且所有元素之和为 1。在进行 CNN 分类时，可以通过 Softmax 将神经元的输出变成概率的模式。按照概率的大小确定到底是哪一类别。Softmax 函数的运算公式为：

$$\sigma(Z)_j = \frac{e^{Z_j}}{\sum_{k=1}^K e^{Z_k}} \quad (\text{for } j = 1, \dots, K).$$

2.2. 模型过拟合解决办法

但是由于神经网络是一种比较复杂的网络模型，需要建立在大样本的基础上才能比较好的体现它的精确度。如果数据样本比较小，分类比较多，很容易出现过拟合现象(Overfitting)。在进行迭代训练的过程中可能会出现训练集精度大于验证集精度，且测试集损失值并没有逐渐减少。这时候可以大体判断该模型出现过拟合状况。

Dropout 是一个很好的解决过拟合方式。模型中 dropout 掉不同的隐藏神经元就相当于在训练不同的网络，并随机删掉某些隐藏神经元，从而导致网络结构的不同。整个 dropout 过程就相当于对多个不同的神经网络取平均。而不同的网络产生不同的过拟合，一些互为“反向”的拟合相互抵消就可以达到整体上减少过拟合。

2.3. 深度学习模型

2.3.1. VGG 网络模型

VGG 卷积神经网络是牛津大学计算机视觉实验室参加 2014 年 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 比赛所用的网络结构, 目的是为了解决 ImageNet 中的 1000 类图像分类和定位问题。本文主要探讨的是 VGG 的两种模型, 分别是 VGG16 和 VGG19, 两者在网络深度不一样, 一个 16 层一个 19 层。

2.3.2. ResNet 网络模型

Resnet 分类网络模型, 可以说是目前应用最广泛的 CNN 特征提取网络。该网络模型由 He Kaiming 等人于 2015 年提出。并得到了当时 ImageNet 比赛等多个比赛的冠军。Resnet 网络作者则想到了在计算机视觉领域中常见的残差学习(residual representation)的概念, 并更进一步将它运用在了 CNN 模型的构建中, 因此就有了基本的残差网络模型。

Resnet 学习的是残差函数 $F(x) = H(x) - x$, 假如 $F(x) = 0$, 那便是之前提到的恒等映射。事实上, ResNet 是“shortcut connections”在恒等映射下的特殊情况, 它并没有引入额外的参数和计算复杂度。一个残差学习单元(Residual Unit)如图 2 所示, ResNet 相当于将学习目标改变了, 不再是学习一个完整的输出 $H(x)$, 只是输出和输入的差别 $H(x) - x$, 即残差。

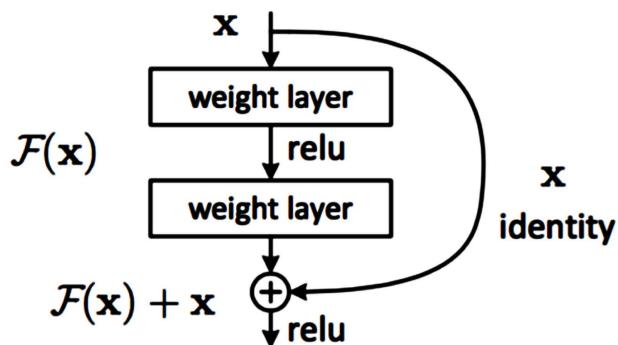


Figure 2. Residual learning unit
图 2. 残差学习单元

Resnet 网络模型是采用几个有参层来学习输入输出之间的残差表示, 并没有像通常的 CNN 网络那样可以直接采用有参层来尝试学习输入、输出之间的映射。很多的实验表明: 采用一般意义上的有参层可以直接学习残差, 比可以直接学习输入、输出之间的映射要容易得多(收敛速度更快), 也有效得多(可根据采用更多的层来达到更高的分类精度)。

ResNet 模型每个网络都包括三个主要部分: 输入部分、输出部分和中间卷积部分。尽管 ResNet 的变种形式丰富, 但都遵循上述的结构特点, 网络之间的不同主要在于中间卷积部分的 block 参数和个数存在差异。

3. 数据来源及处理

本次试验所选取的开发环境为 python3.6, 并选择在 TensorFlow2.0.0 平台上构建模型。本次试验的数据均来自百度图片(<https://image.baidu.com/>), 选择了 NBA 球队金州勇士队的现役 17 名球员的头像照片。该球队球员有白色人种、黑色人种等, 并且他们的发饰和五官轮廓也各不相同。每人的图片均为 40 张, 共计 680 张图片。部分球星图片如图 3 所示。



Figure 3. Some displayed samples

图 3. 部分展示样本

3.1. 数据预处理

3.1.1. 统一数据大小

由于网上的图片并没有规定的大小，所以本次找到的图片大小并不统一。为了使得试验更好的进行下去，首先将数据统一大小。通过 os 库和 cv2 库将所有图片的保持长宽比的方式将高度调整到 9 cm。

3.1.2. 背景填充处理

由于图片的长宽比不一致，经过上述操作后会发现某些图片的宽度不一致。图片的像素也就不一致，就不能保持同一的大小输入层。所以选择进行背景填充操作。首先找到宽度最大的图片，复刻一张相同大小的全黑图片，然后将其余图片居中放置到全黑图片中间。这样没有改变图片原有的比例同时解决了图片大小不一的情况。

3.1.3. 数据增强

由于本身试验的样本量不多，而在模型训练中样本量是一项非常重要的参数，样本量不足会导致模型拟合效果不好。数据增强是一种有效获取大量数据的方法，对于小样本案例非常有效。数据增强的使用有效的减少了模型的过拟合现象同时也提高了网络模型的鲁棒性。本文最开始采取对图像进行剪切、翻转、缩放、平移等方式扩大样本数量。用这样的处理后模型的精度达到了 61%。然后考虑到不同的图片光度有所不一样，所以在数据增强中加入了直方图均衡化和拉普拉斯算子(效果如图 4 所示)，运用同样的代码，最后得出的精度达到了 67%。所以本文采用最终的数据增强方法。

原始数据:

数据增强后:

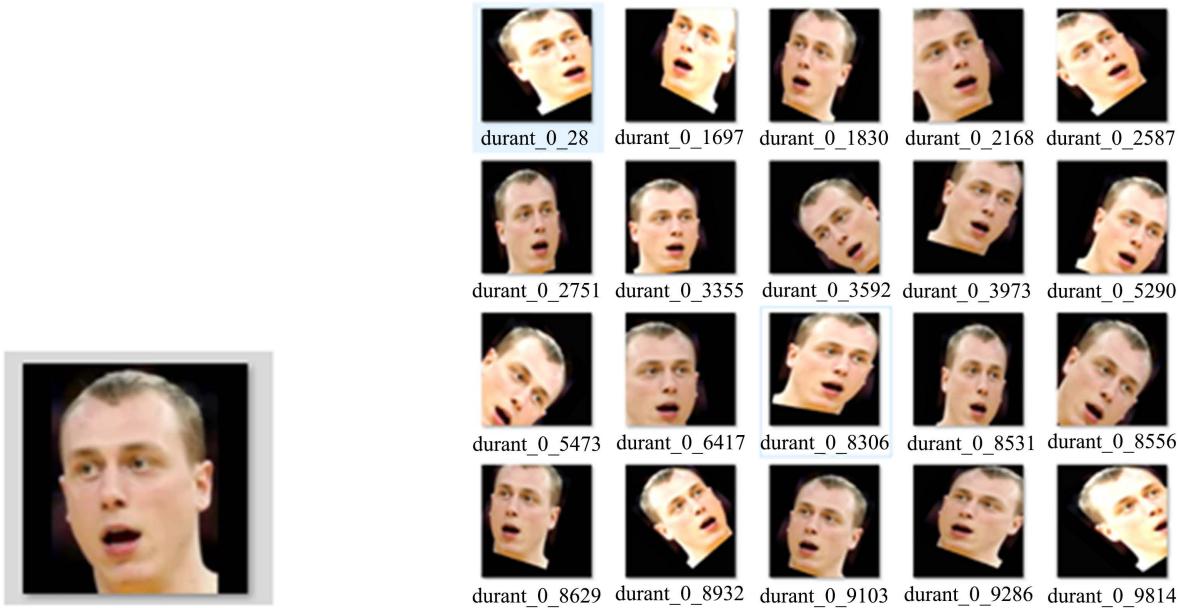


Figure 4. Data enhancement
图 4. 数据增强

3.1.4. 数据归一化处理

图像归一化就是对图像进行了一系列标准的处理变换，使之变换为一固定标准形式的过程，该标准图像称作归一化图像。图片归一化处理后，像素的取值范围就从(0, 255)变为了(0, 1)。会大大提升网络的训练效果。

3.1.5. 训练集测试集划分

为了更好的观察模型的训练效果，本文选择对样本数据进行划分。本次试验共 680 张图片，首先将数据进行随机打乱，选取 425 (62.5%)张作为训练集，选取 170 (25%)张作为验证集，选取 85 (12.5%)作为测试集。用训练集进行模型训练，模型的训练过程中用验证集验证，最后模型训练完成后用测试集测试模型的最终效果。

4. 明星人脸分类建模

4.1. CNN 模型的建立

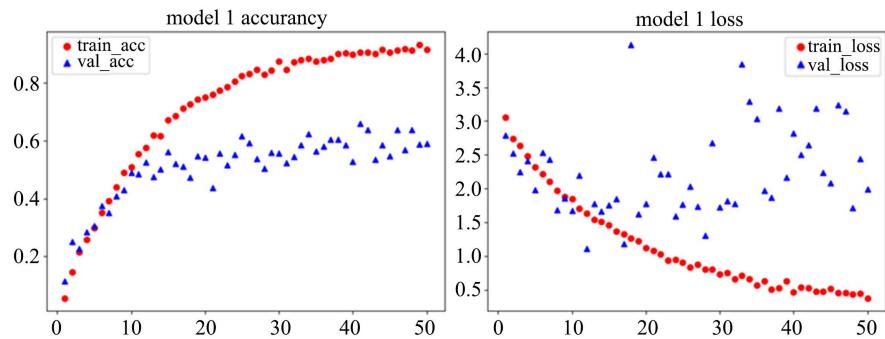
CNN 网络模型架构的搭建

本次的实验 CNN 模型采用的是 4 个卷积层 + 4 个池化层结构形式。

考虑到该模型为多分类模型，所以模型选择输出层的激活函数为 softmax 函数。损失函数选择交叉熵损失(categorical_crossentropy loss)函数。为了使得模型更好的拟合，自动调整学习率，在优化器的选择上模型选择自适应学习率优化算法 RMSProp 算法。

首先构建上述的 CNN 网络模型(Model 1)。对归一化处理后的数据进行训练。模型训练 50 次，每次输入 50 张图片，训练次数达到后训练结束。

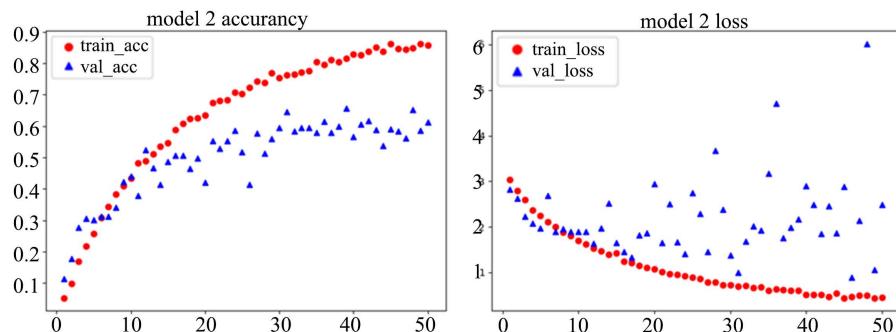
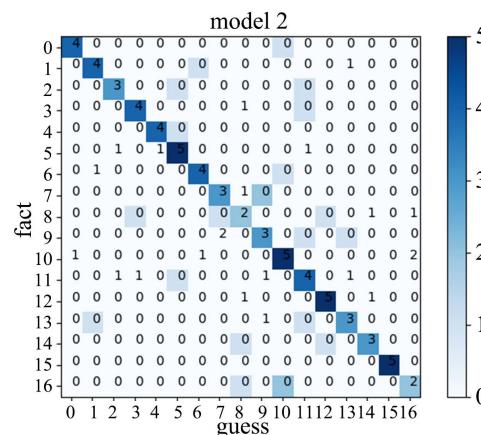
训练得到的结果如图 5 所示。

**Figure 5.** CNN model accuracy (left) and loss (right)**图 5.** CNN 模型精确度(左)和损失值(右)

左边红色线段代表的是训练集准确度，蓝色代表的是验证集准确度。通过上述图片可以看出训练集准确度达到了 90%，验证集准确度达到了 60%。由于训练集与验证集精度相差有 30% 左右，差别过大，所以我们判断该模型存在过拟合现象。下面我们在全连接层前加入 Dropout 函数，希望降低过拟合状况。

经过验证，隐含节点 Dropout 设置为 0.5 时网络模型的效果是最好的，此时生成的随机网络结构最多。所以最终选择加入设置为 0.5 的 dropout 函数建立模型(Model 2)。同样选择将模型训练 50 次，每次输入 50 张图片，训练次数达到后训练结束。

所得结果如图 6 和图 7 所示。

**Figure 6.** CNN_1 model accuracy (left) and loss (right)**图 6.** CNN_1 模型精确度(左)和损失值(右)**Figure 7.** CNN_1 model confusion matrix**图 7.** CNN_1 模型的混淆矩阵

但是通过上图可以看出，训练集准确度达到了 88.2%，验证集准确度达到了 76.6%。相比于 Model 1 验证集精度只提高了 6%，还是存在过拟合现象。这时我们选择使用预训练模型，在强大的图像基础上对数据进行分类。希望达到比较好的分类效果。

4.2. 预训练模型的建立

本文所采用的预训练模型为 VGG16 预训练模型、VGG19 预训练模型和 ResNet50 预训练模型进行建模。下面分别用这几个模型进行建模。

4.2.1. VGG16 预训练模型架构的搭建

搭建 VGG16 模型时，首先加载 VGG16 模型，将训练数据直接输入进 VGG16 模型中并改变输出层的大小，将预训练模型作为整体训练模型的一部分，训练整体模型。模型训练 50 次，每次输入 50 张图片，次数达到后训练结束，得到模型(model VGG16)。

得到的模型效果如图 8 所示。

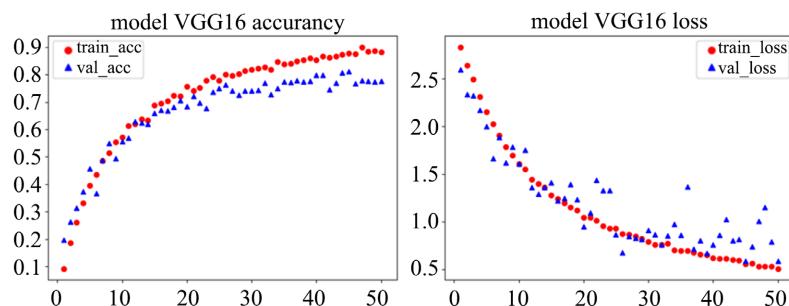


Figure 8. VGG16 model accuracy (left) loss of value (right)

图 8. VGG16 模型精确度(左)损失值(右)

通过图片可以看出模型的训练集准确率达到 90%，验证集准确率达到 78.37%，模型的精度提高了不少。这种情况说明在小样本的情况下，还是应该采用预训练模型对数据进行训练，从而达到较好的结果。

在进行完上面步骤后，我们可以对模型进行微调。经过大量的试验我们最终选择解冻预训练模型结构中顶端的最后 3 个卷积层，再训练就是针对这部分解冻的层。模型训练 80 次，每次输入 50 张图片，次数达到后训练结束，并在输入层引入 BN 层之后的模型构建同理，BN 算法在解决网络训练收敛速度慢及模型精度低的情况下有很好的效果，同时也克服了内部协变量偏移和梯度消失等问题，得到模型(model VGG16_1)。

训练后得到的结果如图 9 和图 10 所示。

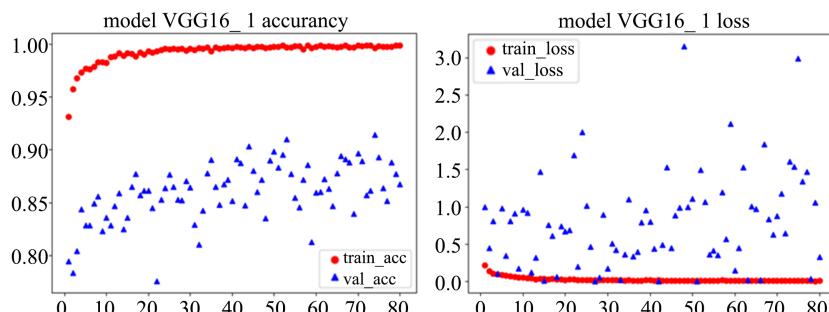


Figure 9. VGG16_1 model accuracy (left) loss of value (right)

图 9. VGG16_1 模型精确度(左)损失值(右)

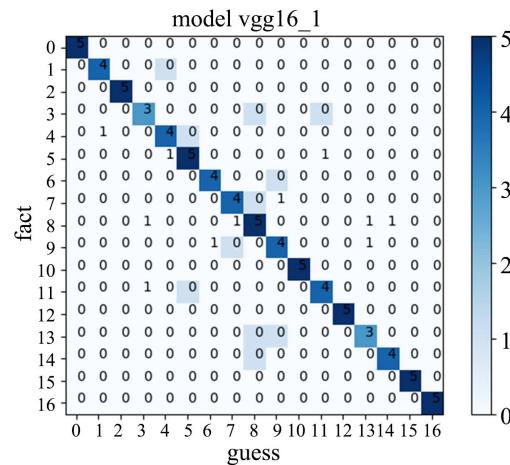


Figure 10. VGG16_1 model confusion matrix
图 10. VGG16_1 模型的混淆矩阵

通过图 9 可以看出微调后的 VGG16 模型训练集准确率几乎达到 100%，验证集准确率达到 87.8%。可以看出使用预训练模型后，模型准确率明显有所提升，这也说明在样本量少的情况下选择训练好的参数进行训练模型效果会大大提升。为了选择更好的预训练模型适应该项目，我们更换模型的参数不变继续加入不同模型与之对比。

4.2.2. VGG19 预训练模型架构的搭建

与构建 VGG16 模型一样，首先加载 VGG19 模型，将训练数据直接输入进 VGG19 模型中并改变输出层的大小，将预训练模型作为整体训练模型的一部分，训练整体模型。模型训练 50 次，每次输入 50 张图片，次数达到后训练结束，得到模型(Model 3)。

所得结果如图 11 所示。

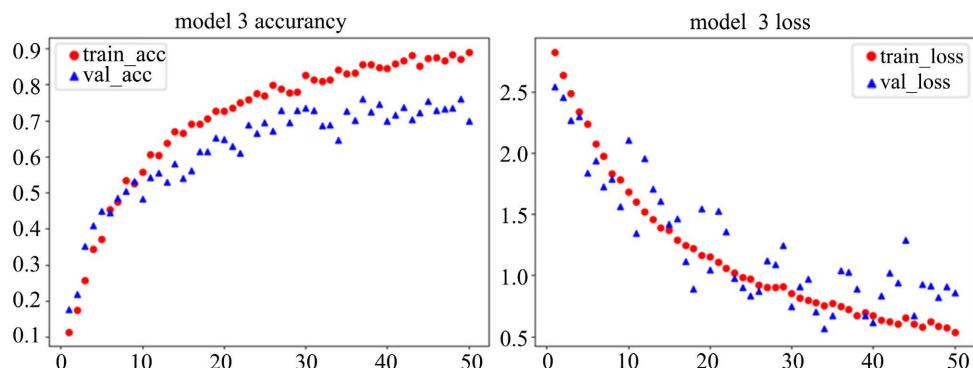
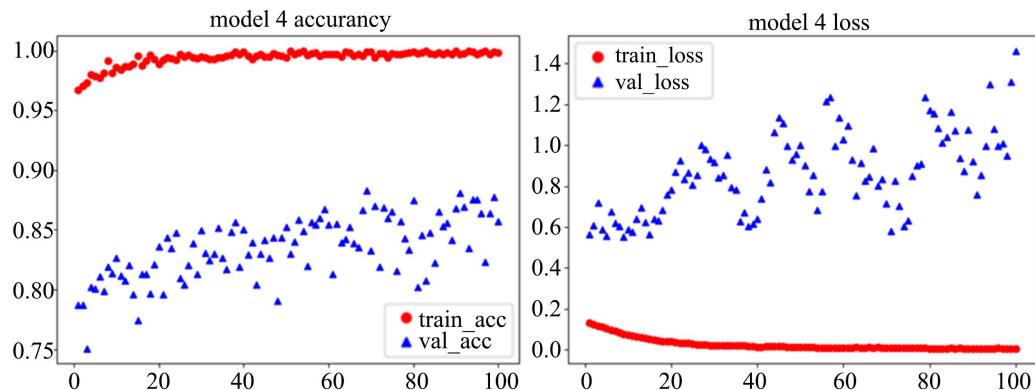
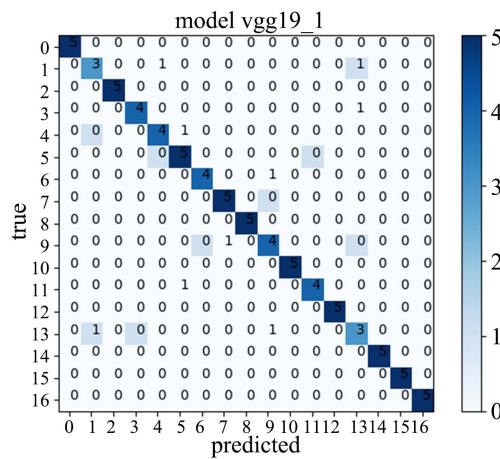


Figure 11. Model 3 model accuracy (left) and loss value (right)
图 11. Model 3 模型精确度(左)和损失值(右)

通过图 11 可以看出模型的训练集准确率达到 90%，验证集准确率达到 78.37%。在进行完上面步骤后，我们同样对模型进行微调，与 VGG16 一致。选择冻结最后的 4 个卷积层，再训练就是针对这部分解冻的层。

模型训练 80 次，每次输入 50 张图片，次数达到后训练结束，得到模型(Model 4)结果见图 12 和图 13。

**Figure 12.** Model 4 model accuracy (left) and loss value (right)**图 12.** Model 4 模型精确度(左)和损失值(右)**Figure 13.** Model 4 model confusion matrix**图 13.** Model 4 模型的混淆矩阵

通过图 12 可以看出微调后的 VGG19 模型训练集准确率几乎达到 100%，验证集准确率达到 89.4% 结果表示随着训练次数的增加训练集的损失值几乎为 0，但是验证集还是存在一定的损失值，这与我们所选取的样本数据有关，本文选取的样本数据少，这也在情理当中。

4.2.3. ResNet50 预训练模型架构的搭建

搭建 ResNet 网络模型时与 VGG 相似，一开始也是选择将数据直接经过预训练模型，将预训练模型作为整体训练模型的一部分，训练的是整体模型。

但是在使用 ResNet 网络时出现了这样一个结果，训练集进度随着数据的训练在进行上升，但是验证集精度始终保持着比较小的值。经过大量的查找发现，出现这一现象的原因是因为在 ResNet 网络中出现了 VGG 网络中没有的 Batch Normalization 层(BN 层)，在 ResNet 网络中如果使用 pretrained 参数进行微调，这些 BN 层通常是使用了 K. learning_phase 的值作为 is_training 参数的默认值，所以导致训练的时候使用的一直是 mini batch 的平均值，而且由于 trainable 在进行微调时候一般设置为 false，从而导致整个 layer 不会更新，因此 moving_mean\variance 根本没有更新。导致在 test 时用的 moving_mean\variance 全是 imagenet 数据集上的值。

所以在初始化 ResNet 前 K.set_learning_phase(0)，在加入自己的 top layer 前改回 1 就能解决这个问题了。

在解决好 BN 层所带来的影响后直接将数据输入进 ResNet50 模型中，改变输出层的大小。

并对对 ResNet50 模型进行微调。解冻里面预训练模型顶端的 res5a_branch2c 层，然后针对这一网络模型进行训练。模型训练 60 次，每次输入 50 张图片，次数达到后训练结束，得到模型(modelResNet50_1)。得到的结果如图 14 和图 15 所示。

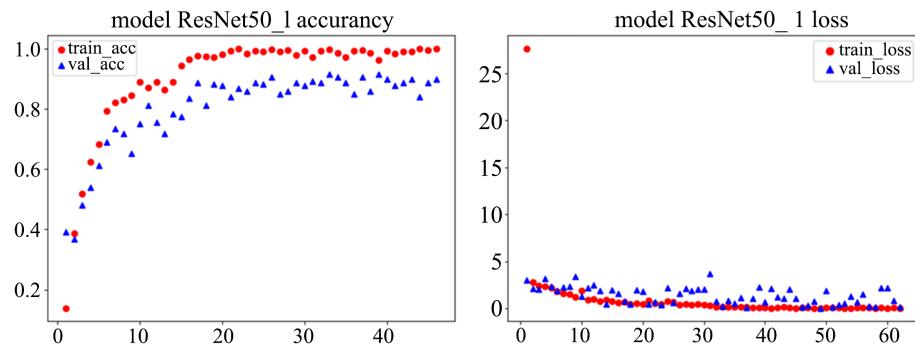


Figure 14. ResNet50 model accuracy (left) and loss value (right)

图 14. ResNet50 模型精确度(左)和损失值(右)

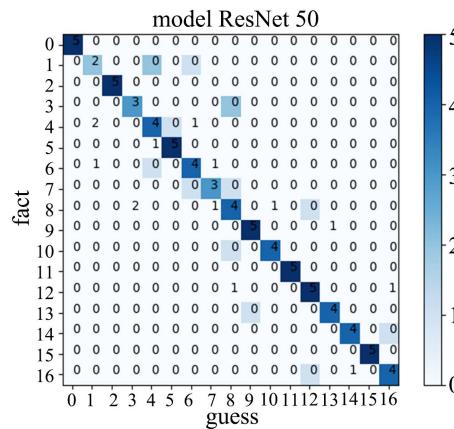


Figure 15. ResNet50 model confusion matrix

图 15. ResNet50 模型混淆矩阵

从上图可以看出模型的损失值基本趋于零，模型收敛。模型训练集的准确率接近 100%，但是模型验证集准确度只达到了 83% 左右，测试集准确率为 83.5% 没有 VGG19 模型的效果好。

4.3. 模型对比与分析

模型对比

上述我们共建立了 4 个模型，分别为 CNN 模型，VGG16、VGG19 和 ResNet50 模型，结果如表 1 所示。

Table 1. Model comparison
表 1. 模型对比

	CNN 模型	VGG16 模型	VGG19 模型	ResNet50
训练集精度	88.2%	98.2%	99.4%	97.4%
测试集精度	76.6%	87.8%	89.4%	83.5%

通过上述的试验可以看出，模型 VGG19 的效果最好，准确率达到 89.4%。我们选择该模型作为本次试验的最终模型。

5. 总结与展望

5.1. 总结

本课题主要是对某国外球队明星进行分类，并分别利用自己构建的 CNN 模型、VGG16、VGG19、ResNet50 几个重要模型作对比。利用 TensorFlow 深度学习框架对数据进行预处理，利用模型训练调节各参数，最终得出 VGG19 模型的效果最好。使测试数据集的识别正确率达到 89.4%，其效果最佳。

在之后的研究中可以对这方面进行更好的处理从而获得更好的模型。该实践策略为搭建神经网络模型处理图像分类问题提供了一些构模思路，具有一定的参考意义。

5.2. 展望

本论文基于以上的内容还有许多不足之处，有以下内容需要进一步完善：

- 1) 在数据集的选取上应该加大数据样本，提取的效果才能更加明显。
- 2) 除了本文所选用的模型，应该继续利用其他模型进行训练，以此得到最优分类算法。
- 3) 本文对黑皮肤人群的分类效果比较差，要对该情况进行更高效的区分需要引入更加精细的特征提取方法。

参考文献

- [1] Lades, M., Vorbruggen, J.C., Buhmann, J., et al. (1993) Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers*, **42**, 300-311. <https://doi.org/10.1109/12.210173>
- [2] Wiskott, L., Fellous, J.M., Kruger, N., et al. (1997) Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, 775-779. <https://doi.org/10.1109/34.598235>
- [3] 庄福振, 罗平, 何清, 史忠植. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26-39.
- [4] 尹红. 基于深度学习的花卉图像分类算法研究[D]: [硕士学位论文]. 南昌: 南昌航空大学, 2018.
- [5] 赵浩. 基于 TensorFlow 的卷积神经网络图像分类实践策略研究[J]. 价值工程, 2020, 39(9): 205-207.
- [6] 左羽, 陶倩, 吴恋, 王永金. 基于卷积神经网络的植物图像分类方法研究[J]. 物联网技术, 2020, 10(3): 72-75.
- [7] 梁雁, 刘广峰. 基于卷积神经网络的人脸识别研究[J]. 数字通信世界, 2021(1): 101-102.
- [8] 马俊, 张荣福, 郭天茹, 张喆嫣, 李卿, 王蓉, 李子莹. 基于迁移学习的 VGG-16 网络芯片图像分类[J]. 光学仪器, 2020, 42(3): 21-27.
- [9] 田佳鹭, 邓立国. 基于改进 VGG16 的猴子图像分类方法[J]. 信息技术与网络安全, 2020, 39(5): 6-11.
- [10] 王文成, 蒋慧, 乔倩, 祝捍皓, 郑红. 基于 ResNet50 网络的十种鱼类图像分类识别研究[J]. 农村经济与科技, 2019, 30(19): 60-62.
- [11] 杨凡稳. 基于猫群算法的图像分割与分类[D]: [硕士学位论文]. 株洲: 湖南工业大学, 2015.
- [12] 俞显刚. 基于神经网络的视频手语识别[D]: [硕士学位论文]. 合肥: 安徽大学, 2018.
- [13] Olstad, B. and Torp, A.H. (1996) Encoding of a Priori Information in Active Contour Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 863-872. <https://doi.org/10.1109/34.537341>
- [14] Jain, A.K. and Yu, Z. (1998) Deformable Template Models: A Review. *Signal Processing*, **71**, 109-129. [https://doi.org/10.1016/S0165-1684\(98\)00139-X](https://doi.org/10.1016/S0165-1684(98)00139-X)
- [15] Theckedath, D. and Sedamkar, R.R. (2020) Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. *SN Computer Science*, **1**, Article No. 79. <https://doi.org/10.1007/s42979-020-0114-9>
- [16] Zbigniew, O. and Andrzej, K. (2021) Flame Image Processing and Classification Using a Pre-Trained VGG16 Model in Combustion Diagnosis. *Sensors (Basel, Switzerland)*, **21**, Article No. 500. <https://doi.org/10.3390/s21020500>