

基于融合机制的多模型神经网络人物群体分类模型^①



郎波¹, 张娜², 段新新²

¹(北京师范大学珠海分校 工程技术学院, 珠海 519087)

²(北京师范大学研究生院珠海分院, 珠海 519087)

通讯作者: 郎波, E-mail: langbo666@126.com

摘要: 有效识别图像或视频中人物的不同群体, 是进行图像智能分析的重要环节, 归根结底是研究如何获取图像中的“有效特征”。本文以卷积神经网络模型为基础模型, 提出多模型融合卷积神经网络的方法, 利用 ImageNet 训练得到的模型参与本文神经网络模型的权值初始化, 在有效节省时间和计算资源成本的前提下获取更多有效的特征。实验结果证明, 本模型对于自然场景中的个体分类中成年男性、成年女性、儿童识别准确率可以保持在 85% 左右, 提高了人物群体分类的准确度和可靠度。

关键词: 图像智能分析; 有效特征; 卷积神经网络; 多模型融合; 群体分类

引用格式: 郎波, 张娜, 段新新. 基于融合机制的多模型神经网络人物群体分类模型. 计算机系统应用, 2020, 29(8): 127-134. <http://www.c-s-a.org.cn/1003-3254/7582.html>

Human Group Classification Model Based on Multi-Model-Integrated CNN

LANG Bo¹, ZHANG Na², DUAN Xin-Xin²

¹(School of Engineering Technology, Beijing Normal University, Zhuhai, Zhuhai 519087, China)

²(Zhuhai Branch of Graduate School, Beijing Normal University, Zhuhai 519087, China)

Abstract: Effectively identifying the different group of human in an image or video is an important part of intelligent image analysis. It is how to obtain “effective features” in the image. Based on the convolution neural network model, this study proposes a multi-model fusion convolution neural network method. The model trained by ImageNet participates in the initialization of the weights of the neural network model, achieves more effective features on the premise of effectively saving time and resource calculating costs. Experiments prove that the model can maintain the recognition accuracy of adult males, adult females, and children in natural scenes at about 85%, which improves the accuracy and reliability of group classification.

Key words: image analysis; efficient feature; Convolution Neural Network (CNN); multi-model-integrated; group classification

如何有效的识别图像或视频中人物的状态 (例如: 年龄、性别等), 归根结底是如何获取到图像中的“有效特征”。深度学习的出现解决了如何让机器自动学习“有效特征”的问题^[1,2], 它利用人脑处理信息的学习机制, 利用分级处理的过程, 逐层进行特征的变换, 将样

本在原空间的特征表示上升到一个新的特征空间表示, 从而大大提高了分类的效能, 可以更加丰富的表示出图像的内在信息, 即“有效特征”的表示问题^[3]。深度学习解决了如何自动学习“有效特征”的问题, 它通过逐层的特征变换, 将学习样本在原空间上的特征表示变

① 基金项目: 国家自然科学基金 (61375122); 广东省创新强校特色创新类项目 (201712009QX)

Foundation item: National Natural Science Foundation of China (61375122); Innovation Empowering School Characteristics Innovation Project of Guangdong Province (201712009QX)

收稿时间: 2020-02-08; 修改时间: 2020-03-03; 采用时间: 2020-03-20; csa 在线出版时间: 2020-07-29

换到一个新的特征空间,从而使得分类识别任务变得更加有效.目前主要的研究工作是通过对人脸图像信息来发现其性别属性.例如 Tian 等^[4]使用深度 LDA-Pruned Net, Edinger 等^[5]受到训练深度神经网络 Dropout 技术的启发,训练了一个随机丢失某些特征的支持向量机 (Support Vector Machine, SVM),通过它对人脸特征进行区分来完成性别分类, Brunelli 等^[6]使用两个径向基函数来分别提取男人和女人的脸部特征来进行性别分类,此外还有利用反向传播网络 (Back Propagation network, BP) 和极限学习机 (Extreme Learning Machine, ELM) 进行分类的模型^[7,8],张婷等^[9]提出的基于跨连卷积神经网络的性别分类模型.尽管这些研究实验在图像光照、遮挡等外部因素影响下保持较高的识别率,但其所采用的测试样本数据库普遍具有前景对象突出、背景简单的特点,而且必须是标准的人脸照片,并不具备自然图像的特点.后来 Levi 等^[10]利用 Adience 数据集使用卷积神经网络对年龄和性别进行分类,采用的图片为自然条件下光照多变、姿态各异、图片质量欠佳等现实场景,但在训练网络做预处理工作时仍然必须采用基于人脸特征来进行性别以及年龄的区分.

为了克服特征缺失的问题,还有研究者尝试使用局部特征来对人物个体进行分类^[11],局部特征的优势是能够详细描述图像的局部信息和细节,但是同时也存在缺点,例如,局部特征会忽略图像的色彩信息,而色彩信息对于识别人物的性别和年龄是至关重要的,另外过分强调局部特征也难以从整体构成和关系上来准确表达图像.

1 解决问题

本文尝试利用深度学习中卷积网络的学习方法来对场景中人物的年龄和性别进行区别,进而实现不同群体人物的分类.在设计学习模型时,主要考虑到以下 3 方面的问题:

(1) 选择卷积神经网络能否对人物特征 (年龄、性别) 的分类有促进作用?

卷积神经网络可以从原始图像中学习出包含丰富语义信息的特征矩阵,从目前的研究结果来看,利用卷积神经网络学习到的特征,要比手工特征或局部特征的分类准确率高^[12],所以,选取卷积神经网络来对人物特征进行分类是一个发展方向.

(2) 如何得到较为全面的人物特征来促进年龄和

性别的分类问题?

对于图像中人物特征的识别会受到很多因素的制约,例如颜色、构图、亮度、纹理等.如何从不同角度更加全面的挖掘图像信息、提取有效的图像特征,并且使得分类效果具有更广泛的适用性.卷积神经网络的学习效果依赖于输入源的质量.数字图像有不同的编码方式,将同一图像的不同表示矩阵作为输入源,可以获取更全面的图像特征.

实际生活中的自然图像背景复杂,人物着装发型各异,人物姿态多样,人物图像人物与背景关系复杂,因此如果要在这种自然图像集区分性别并判断年龄就比较困难.研究表明,综合有效利用高层次特征和低层次特征,对改善视觉系统的识别能力是很有作用的^[8].以此为出发点,为了能够使卷积神经网络 (Convolutional Neural Networks, CNN) 模型更为快捷的实现多特征信息的互补,本文以 CNN 模型为基础模型,提出一个多模型融合卷积神经网络的办法,利用 ImageNet^[13]训练得到的模型 (例如: Inception v3, Xception, Inception-Resnet v2, DenseNet, NASNetLarge),参与本文神经网络模型的权值初始化,在有效节省时间和计算资源成本的前提下获取更多有效的特征.将多模型融合卷积神经网络用于性别和年龄识别的优点在:对现有模型进行混淆矩阵的分析,利用迁移学习在现有模型得到的图像特征的基础上,采用无监督方式进行性别和年龄的分类,以构造性能更好的分类器.

2 算法模型

神经网络的结构主要由卷积层、采样层、全连接层组成.其中卷积层的作用是利用卷积核进行特征抽取,并对抽取到的特征进行过滤和强化,卷积的目的是将前一层输出的特征图与卷积核进行卷积操作,并通过激活函数输出特征图 (feature map),如式 (1) 所示:

$$P_j^t = f \left(\sum_{i \in P_j} K_{i,j}^t * P_j^{t-1} + q_j^t \right) \quad (1)$$

其中, f 是激活函数 Sigmoid, t 表示网络的层数, $K_{i,j}^t$ 表示卷积核, $*$ 表示卷积操作, q_j^t 表示偏移量, P_j 表示输入的特征图集合

采样层主要是用来降低特征维数,通过池化操作来获取抽样特征,抑制过拟合,提升特征的识别能力,如式 (2) 所示:

$$F = f \left(\text{simple} \left(y_i^{t-1} \right) \cdot \omega_j^t + b_j^t \right) \quad (2)$$

其中, ω 是权值函数, $simple()$ 是采样函数.

传统卷积神经网络是一种特殊的深层前向神经网络模型, 其结构一般由输入层、卷积层和池化层、全连接层及输出层构成, 它的一个缺点是难以有效的融合各种特征去构造性能更好的分类器^[3], 由于 CNN 模型主要利用灰度图像作为输入, 缺失了对颜色和亮度的理解, 为了更完善的表达图像, 因此采用多特征融合的方法来对图像进行处理, 针对该问题, 本文引入了多模型融合的思想, 提出了一个多模型卷积神经网络融合 (Multi-model-Integrated Convolution Neural Network, MICNN) 算法模型, 用于对自然场景下的人类性别和年龄进行分类. 该模型整体结构如图 1 所示.

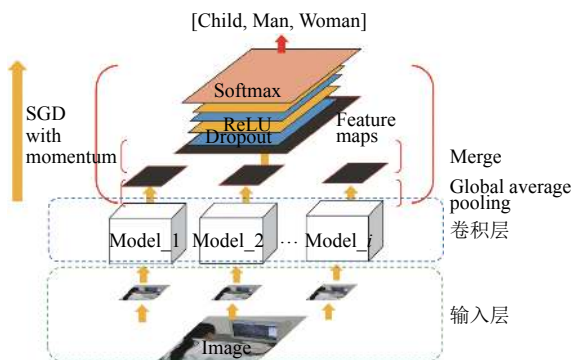


图 1 多模型融合卷积神经网络模型

2.1 特征抽取层

本层选取若干网络模型分别对训练数据集、验证数据集和测试数据集做特征提取的训练, 获得图像的特征表示, 目前选取的网络模型为 Inception v3^[9], Xception^[10], Inception-Resnet v2^[11], DenseNet^[12], NAS-NetLarge^[13]. 之所以选择这几个网络模型, 主要考虑到: Resnet 网络结构通过增加恒等映射, 用来解决深度网络中学习退化的问题, 从而优化训练结果; Inception 模块通过使用不同的滤波器对输入进行卷积并进行最大池化工作, 增加了网络宽度, 从而得到更为全面的特征表象; Inception-Resnet v2 可以有效防止模型梯度消失的问题; Xception 在减少模型参数的前提下增加了模型的表现能力; DenseNet 可以做到大量的特征被复用, 即使用少量的卷积核就可以生成大量的特征; NASNet 架构主要用于大规模的分类和识别, 并可以做到对模型规模的自动调整.

2.2 全局平均池化层

使用平均池化, 将从特征抽取层中得到的特征图进行全局平均化, 即把一个的张量平均化为一个 $1 \times 1 \times D$

的张量. 使用全局平均池化层的目的是为了将 N 个特征图降维成 $1 \times N$ 大小的特征图, 然后在全连接层用分类数 $class$ (本文设置为 3, 即男、女、年龄) 个 1×1 的卷积核将 $1 \times N$ 的特征图通过卷积得到 $1 \times class$ 的向量. 以 Xception 网络为例, ($nb_samples, 10, 10, 2048$) 在使用在全连接层 Softmax 之后的计算量为 $nb_samples \times 10 \times 10 \times 2048$, 而使用全局平均池化后计算量为 $nb_samples \times 1 \times 1 \times 2048$

全局平均池化通过加强特征图与类别的一致性, 让卷积结构更简单. 因为本文提出的融合训练的模型中每一个都是大型的网络结构, 同时本研究项目的数据集并不丰富, 所以训练过拟合是本项目重点考量的问题, 全局平均池化并且不需要进行参数优化, 所以在模型中设置这一层的目的是为了更加有效地避免过拟合, 而且, 全局平均池化层对空间信息进行了求和, 因对输入空间的变换更具有稳定性, 可以使计算网络对图像进行特征提取的同时也保留了图像的空间信息的特点.

2.3 Dropout+ReLU 层

全局平均池化层虽然在一定程度上避免了过拟合问题, 但是在研究过程中发现, 由于对人物个体的分类比较细化 (除了对性别进行分类, 还有对成人儿童进行分类), 所用到的训练数据集较小, 同时为了提高泛化性能, 数据集中的图片均是自然场景图片, 背景纹理复杂, 对提取有效特征有一定的难度. 此融合结构比较复杂, 所以在仍需要对解决过拟合问题采取一定措施.

神经网络的训练流程是将输入内容通过网络正向传播, 然后将训练误差反向传播进行权值更新, dropout 的作用就是在上述过程中临时随机删除部分神经元. 即在网络训练前向传播的时候, 让某个神经元的激活值以一定的概率 P 停止响应, 这样可以使训练模型不至于过分依赖某些局部特征, 从而提高模型的泛化能力. 如图 2 所示.

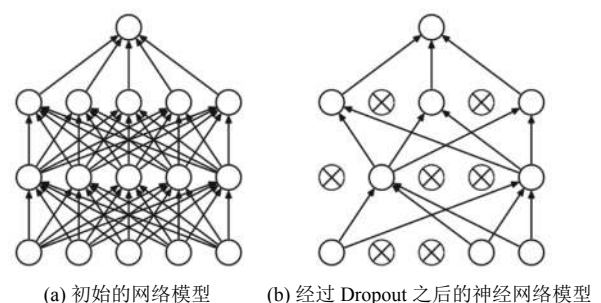


图 2 神经网络训练流程

Hinton 等的研究表明过拟合现象可以通过阻止某些特征的协同作用来进行缓解^[14]. 通过 dropout 过程的参与, 可以有效减少神经元之间的共适应性, 降低神经元的互相依赖性, 从而增强计算模型的鲁棒性. 经过实验验证, 为了减少特征之间的相关性, 本文中的 dropout 比率设置为 0.5, 假设网络中的神经元个数为 N , 其激活函数输出值个数也为 N , 则该层神经元经过 dropout 后, 大约 $0.5 \times N$ 的值被置为 0. 其各部分的数学定义为:

$$r_j^{(r)} \sim \text{Bernoulli}(p) \quad (3)$$

其中, $r_{0,1}^{(l)}$ 分布的向量通过 Bernoulli 函数以概率 P 随机生成一个服从 0,1 分布的向量, 其中 0 表示该神经元已经被 dropout, 即已被屏蔽, 不参与工作.

$$y^{*(l)} = r^{(l)} \times y^{(l)} \quad (4)$$

式中, y^l 为原本第一层的输出, $y^{*(l)}$ 为第一层经过 dropout 之后的输出.

$$Z_i^{(l+1)} = w_i^{(l+1)} y^{*(l)} + b_i^{(l+1)} \quad (5)$$

其中, $Z_i^{(l+1)}$ 是第 $l+1$ 层未激活前的线性输出, $w_i^{(l+1)}$ 为第 $l+1$ 层的线性关系系数, $b_i^{(l+1)}$ 为第 $l+1$ 层的偏置.

$$y_i^{(l+1)} = f(z_i^{(l+1)}) \quad (6)$$

其中, $y_i^{(l+1)}$ 是第 $l+1$ 层的输出, $f(z_i^{(l+1)})$ 是激活函数.

2.4 全连接层

模型的最后一层是输出结果, 输出的结果用一个概率向量来表示 $[P_{\text{child}}, P_{\text{man}}, P_{\text{woman}}]$, 本文输出层激活函数使用 Softmax 对网络输出各权重值进行计算从而得到多分类下的先验概率. 若在 Softmax 层之前的隐藏层为 L 输出值为 M 维向量, 向量中的每个元素为待分类对象在各类别中的先验概率值 $P_j = P(y = j|x)$, $j = 1, \dots, M$, (即向量 x 属于 j 分类的概率), 公式表示为:

$$P_j = \frac{e^{x_j^L}}{\sum_{i=1}^M e^{x_i^L}} \quad (j = 1, \dots, M) \quad (7)$$

其中, x_j^L 是经过 L 层得到的 M 维向量的第 j 个分量.

本文选择交叉熵损失函数衡量真实值与网络的估计值的误差, 该方法可以克服某些损失函数在模型训练学习过程中权重和偏置的更新速度缓慢的问题, 如式 (8) 所示:

$$L = -\frac{1}{N} \sum_{j=1}^N \hat{p}_j \log(p_j) \quad (8)$$

其中, N 为测试集中样本的数量, \hat{p}_j 为第 j 个类别的真实值, p_j 是 Softmax 函数的第 j 个输出值.

2.5 误差修正

在神经网络训练过程中, 通常无法直接计算 loss 损失函数的全局最优解, 但可以通过损失函数求出网络中各参数的梯度 $\frac{\partial L}{\partial \theta}$, 目前使用梯度下降迭代的数值优化方法来降低损失函数, 即通过 $\theta \leftarrow \theta - \frac{\partial L}{\partial \theta}$ 来迭代更新参数. 本模型中使用 SGDM(伴随动量随机梯度下降) 来计算模型参数 $\theta \leftarrow \theta + V_t (V_t$ 表示每一次迭代后 θ 的更新量), 从而使模型更加逼近或者达到最优值并使 loss 损失函数最小, 具体表示如式 (9) 所示:

$$V_t = -\eta \times d\theta + V_{t-1} \times \text{momentum} \quad (9)$$

其中, η 表示学习率, 即当前梯度对最终更新方向的影响程度, $d\theta$ 是网络中更新参数相对于目标函数的梯度, V_{t-1} 表示上一次的更新量, momentum 的值是介于 $[0,1]$ 之间的一个随机量, 用来表示保留原来更新方向的程度. 该模型在训练过程中更新参数时会会在一定程度上保留之前的更新方向, 同时还会利用当前批次的训练结果微调下一轮迭代的更新方向, 可使得模型在训练过程中更加稳定, 并且有助于加快模型的收敛速度.

本模型在训练过程中, 初始训练梯度相对较大, 所以 momentum 初始值为 0.5, 训练一段时间后, 梯度变化相对减小时更改为 0.9, 考虑到学习率大小对梯度下降迭代更新参数的影响, 太小收敛速度慢, 太大学习率会阻碍收敛, 并会造成损失函数在最小值处的震荡, 甚至导致发散, 将 η 的初始值设为 0.0001, 在模型训练过程中监测在验证集的损失或者准确率, 当学习停滞时, 则减少学习率, 反之则进行调整.

3 实验方法

3.1 测试库的选择

本文在 Open Images Dataset V4 数据集提供的图像库的基础上, 经过人工筛选制作了实验所需的数据集. 该数据集中的图像源于普通真实的生活场景, 图像背景复杂, 人物姿态各异, 具有很强的实用性, 我们分别用 man、woman、child 表示不同类别的图像的标签, 得到的数据样本集为 $[\text{man}, \text{woman}, \text{child}] = [1504, 1504, 1504]$, 验证集为 1354 幅. 其数据集部分图像如图 3 所示.

3.2 数据预处理

数据集在输入模型训练之前对图像样本执行逐样本

均值消减的归一化的预处理,即分别对各个通道进行去中心化处理.归一化的目的是为了移除图像的平均亮度值,

很多情况下我们对图像的亮度并不感兴趣,而更多地关注其内容,这时对每个数据点移除像素的均值是有意义的.

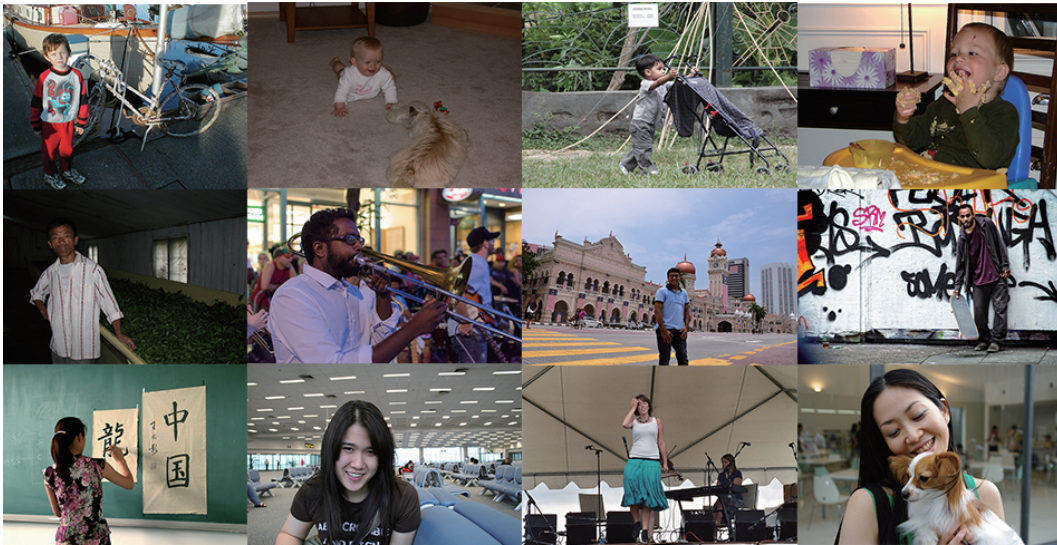


图3 本文采用的数据集部分样例

为了提高训练效率,本文采用已经在 ImageNet 数据集上预训练好的 CNN 模型作为特征提取器,通过这些预训练好的模型导出的特征向量可以快速概括图像上的大约内容,分别对训练集、验证集、测试集分别做特征提取的预训练.我们对目前比较流行的几种 CNN 模型进行了分析,如表 1 所示.

表 1 ImageNet 预训练模型分类性能比较

模型	Top-1 accuracy	Top-5 accuracy
VGG16	0.713	0.901
VGG19	0.713	0.900
Inception	0.779	0.937
Resnet50	0.749	0.921
DenseNet201	0.773	0.936
Xception	0.790	0.945
InceptionResnet v2	0.803	0.953
NASNetLarge	0.825	0.960

表 1 中, top-1 和 top-5 是指模型在 ImageNet 验证数据集上的性能. top-1 accuracy 为预测类别中概率向量中排名第一的类别与实际相符的准确率, top-5 accuracy 即为排名前 5 的类别包含实际结果的准确率.利用全局平均池化将卷积层输出的每个激活图直接求平均值,减少输出数据文件大小,在一定程度上减少过拟合问题.最后针对每一个预训练模型导出 5 个文件:

- (1) train (train_nb_samples,model_output).
- (2) val (val_nb_samples, model_output).
- (3) test(test_nb_samples, model_output).

(4) train_label (train_nb_samples).

(5) val_label(val_nb_samples).

其中, train_nb_samples, val_nb_samples, test_nb_samples, model_output 分别为训练集样本数、验证集样本数、测试集样本数、预训练模型的输出图像的特征表示长度.其中 VGG16, VGG19, Resnet50, InceptionV3, Xception, InceptionResnet v2, DenseNet201, NASNet-Large 的 model_output 分别为 512, 512, 2048, 2048, 2048, 1536, 1920, 4032.将多个不同的网络输出的特征向量进行保存,再利用这些特征向量训练、调参、优化模型,会大大减少训练时间和计算代价.

3.3 融合训练

融合的目的是为了尽可能的发挥各个单一模型的优势,从而提高整个模型的性能.通过多个模型融合训练,最终的特征表示是由多个不同的局部区域特征组合而成,这种“组合特征”的优越性能够在实际任务中显著的区分出个体的主要差别,其算法流程如算法 1.

算法 1. 多模型融合训练算法

输入: 训练集 $T = \{(x^1, y^1), (x^2, y^2), \dots, (x^n, y^n)\}$; 训练超参数 (学习率 η , 最大迭代次数 σ)

- ① 加载在 ImageNet 数据集上预先训练好的模型参数,进行基准模型权重迁移;
- ② 利用预训练的模型不同层结构,进行模型训练获取数据集(训练集,验证集,测试集)的特征向量;
- ③ 载入并融合不同模型训练的特征向量,完成对样本特征更全面提取;

- ④ 使用两层 ReLU+dropout 层和一层 Softmax 形成新的网络进行训练, 计算每个样本图像的实际输出类别;
- ⑤ 根据损失函数 loss function(衡量在单个训练样本上的表现) 及成本函数 cost function(衡量在全体样本训练样本上的表现), 计算出网络训练反馈误差;
- ⑥ 计算网络模型中的权重和偏置得到梯度值;
- ⑦ 使用 SGD with momentum 动量随机梯度下降法, 更新模型参数(权值和偏置) 以及模型参数的更新变量 v_i ;
- ⑧ 输出并保存网络模型的权重与偏置。

需要定义以下参数对模型的训练质量进行评估:

(1) 定义模型评估指标值:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

其中, TP 是被标记为正例的正例数据点, 称为真正例. FP 是被标记为正例的反例数据点, 称为假正例. TN 是被标记为反例的反例数据点, 称为真反例. FN 是被标记为反例的正例数据点, 称为假反例.

(2) 定义 $F-Score$ 是 $precision$ (精确率) 与 $recall$ (召回率) 的加权调和平均值:

$$F-score = (1 + \beta^2) \times \frac{precision \times recall}{(\beta^2 \times precision) + recall} \quad (11)$$

其中, $precision = \frac{TP}{TP + FP}$ 标识分类模型返回相关实例的能力, $recall = \frac{TP}{TP + FN}$ 表示分类模型识别所有相关实例的能力.

当 $\beta = 1$ 时, 将 $F-Score$ 设为:

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

理想中的期望是精度和召回率越高越好, 然而在工程实践中, 这两个指标是不能兼得的, 在 $\frac{F1}{F-Score}$ 较高时, 我们认为此时的特征检测结果最为理想.

计算各单一模型在个体识别分类混淆矩阵, 因为根据混淆矩阵可以分析单一预训练模型对男性、女性、儿童不同个体识别分类的精确率结果, 从而判断模型表现的优劣情况, 为选择合适的模型进行融合训练提供真实的数据依据. 表 2 表示从混淆矩阵中得到不同模型在个体识别任务中针对不同类别的表现能力.

根据表 2 提供的数据, 选择在 child 个体识别精确率最高的 Inception_ResNet_V2 模型, man 和 woman 个体识别精确率最高的 NASNetLarge 模型, 但 Inception_ResNet_V2 在 woman 模型上的精确率相对并不是很高, 所以选择在 child 精确率在 80% 以上, woman 精确率相对提高的模型 Xception, InceptionV3, DenseNet201. 最终融合模型训练选择: Inception_ResNet_V2, Xception, InceptionV3, DenseNet201, NASNetLarge. 在本研究中命名次新模型为: Merge_Model_V1.

在本研究中命名次新模型为: Merge_Model_V1.

表 2 各单一模型在识别不同个体类别的精确率 (%)

模型	Child	Man	Woman
VGG16	76.83	60.85	47.53
VGG19	72.10	49.00	53.57
MobileNetV2	85.97	64.04	79.70
Inception_ResNet_V2	90.23	73.30	79.42
ResNet50	79.95	71.94	79.29
Xception	83.17	72.98	81.13
InceptionV3	84.92	71.42	81.64
DenseNet201	87.41	71.91	81.66
NASNetLarge	87.50	73.86	82.96

4 实验结果与性能评估

在样本集上利用 Merge_Model_V1 进行训练学习并输出结果, 如图 4、图 5 所示的部分结果, 注意本模型对于输入的图片并未做任何人为标注, 仅使用样本集中本身具有的文件夹作为分类标签信息, 也未对图片中的目标主题做中心化处理, 从结果来看, 本模型网络对于个体性别的分类特征还是很准确的, 并能够在复杂场景下聚焦分辨人物性别. 另外分别计算出通过 Merge_Model_V1 训练网络得出在训练集与验证集上的交叉熵曲线、准确率曲线和 $F1$ 曲线, 如图 6. 从曲线的拟合度上可以看出, Merge_Model_V1 模型并没有发生“过拟合”的现象, 模型表现出了非常好的健壮性.

5 结论与展望

本文基于融合多模型预训练实现对自然场景中男性、女性、儿童实现群体识别分类. 本模型综合考虑识别全局图像信息和局部特征信息, 使用多模型预训练的方式, (利用全局平均最大池化的方式将卷积层输出的每个激活图直接求平均值, 减少输出的数据量, 在一定程度上减少过拟合问题.) 完成对数据集的特征向量提取, 融合模型阶段使用预训练阶段数据集的不同的特征图, 在网络训练阶段训练时间大大减少, 更加方便模型参数调优. 现阶段根据模型的实验结果表明, 本模型对于自然场景中的个体分类中成年男性、成年女性、儿童识别准确率相对平均可以保持在 85% 左右, 在考虑提高模型整体准确率的情况下, 我们下一步工作研究的重点将是使用本文提出的模型作为分类器, 应用与目标检测框架, 完成场景理解任务. 在场景安全隐患预估问题上, 可以做到针对特定不同个体的安全系数评估, 从而为更全面研究场景安全问题奠定一定基础.

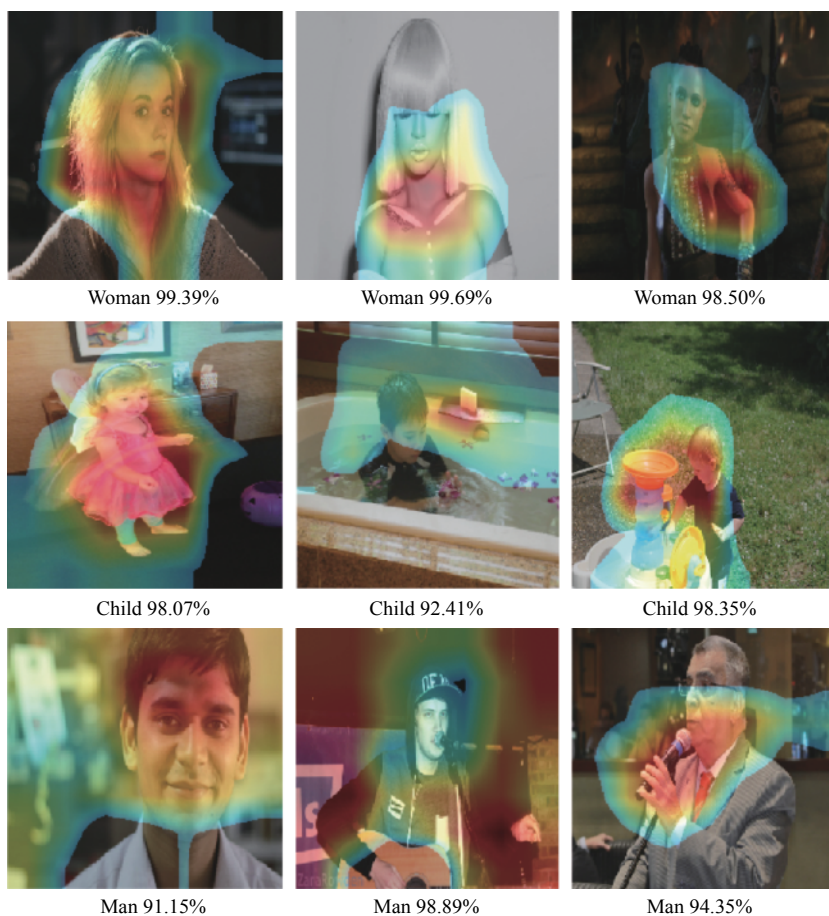


图4 融合模型对自然场景下的人物性别热力分析

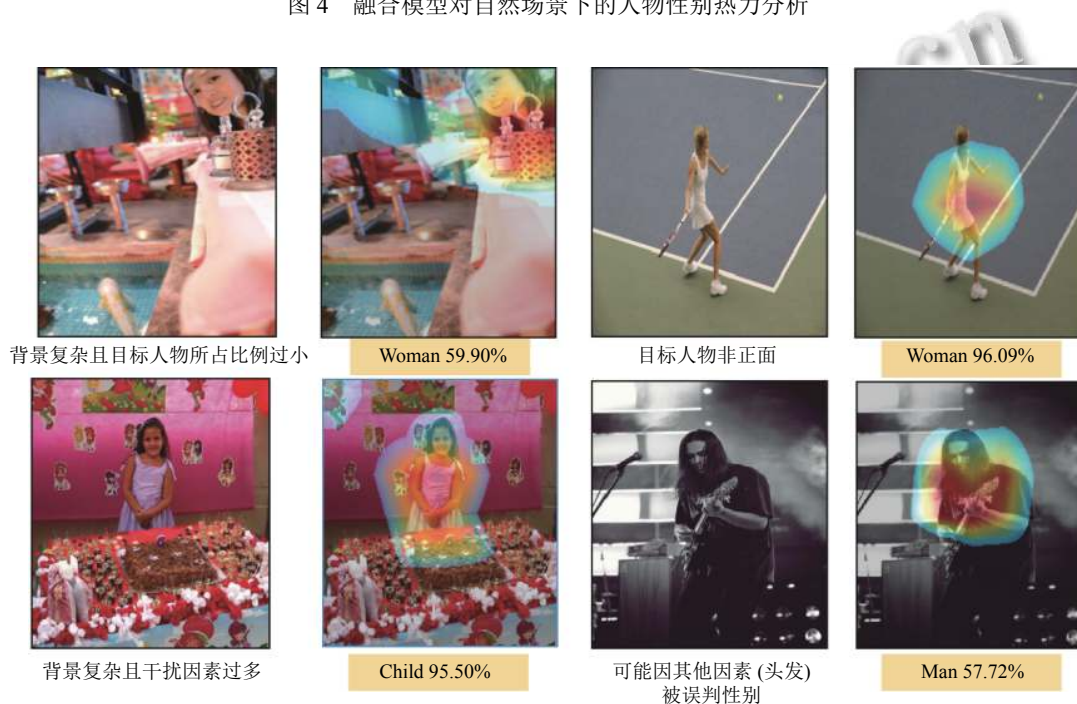


图5 融合模型对特殊条件下的人物性别热力分析

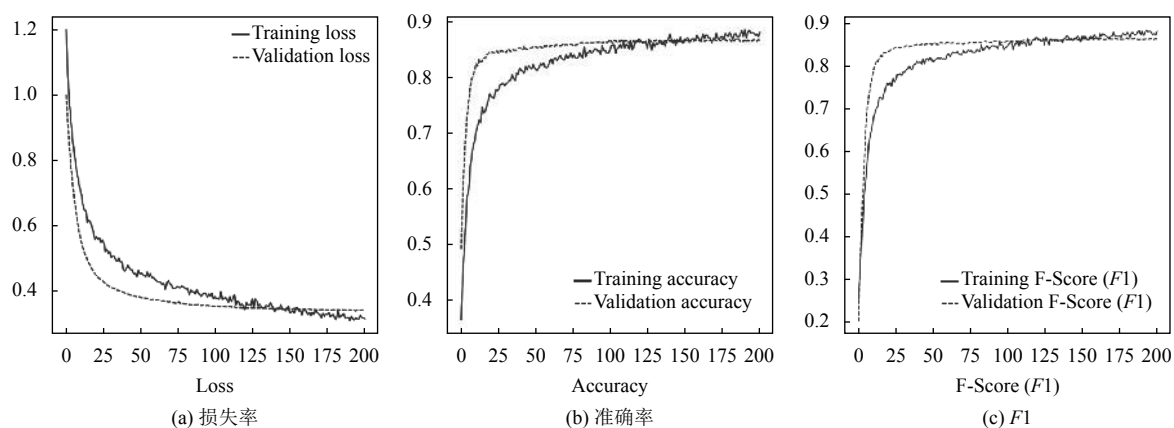


图6 损失率、准确率和F1曲线图

参考文献

- Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada. 2013. 6645–6649. [doi: 10.1109/ICASSP.2013.6638947]
- 刘建伟, 刘媛, 罗雄麟. 深度学习研究进展. 计算机应用研究, 2014, 31(7): 1921–1930, 1942. [doi: 10.3969/j.issn.1001-3695.2014.07.001]
- Najafabadi MM, Villanustre F, Khoshgoftaar TM, *et al.* Deep learning applications and challenges in big data analytics. Journal of Big Data, 2015, 2(1): 1. [doi: 10.1186/s40537-014-0007-7]
- Tian Q, Arbel T, Clark J. Deep LDA-pruned nets for efficient facial gender classification. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA. 2017. 512–521. [doi: 10.1109/CVPRW.2017.78]
- Eidinger E, Enbar R, Hassner T. Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security, 2014, 9(12): 2170–2179. [doi: 10.1109/TIFS.2014.2359646]
- Brunelli R, Poggio T. HyberBF networks for gender classification. Proceedings of DARPA Image Understanding Workshop. Detroit, MI, USA. 1995. 311–314.
- Tamura S, Kawai H, Mitsumoto H. Male/female identification from 8×6 very low resolution face images by neural network. Pattern Recognition, 1996, 29(2): 331–335. [doi: 10.1016/0031-3203(95)00073-9]
- Jiao YB, Yang JC, Fang ZJ, *et al.* Comparing studies of learning methods for human face gender recognition. Proceedings of the 7th Chinese Conference on Biometric Recognition. Guangzhou, China. 2012. 67–74, doi: 10.1007/978-3-642-35136-5_9.
- 张婷, 李玉鑑, 胡海鹤, 等. 基于跨连卷积神经网络的性别分类模型. 自动化学报, 2016, 42(6): 858–865. [doi: 10.16383/j.aas.2016.c150658]
- Levi G, Hassner T. Age and gender classification using convolutional neural networks. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Boston, MA, USA. 2015. 34–42. [doi: 10.1109/CVPRW.2015.7301352]
- 郎波, 黄静, 危辉. 利用多层视觉网络模型进行图像局部特征表征的方法. 计算机辅助设计与图形学学报, 2015, 27(4): 703–712.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.
- Deng J, Dong W, Socher R, *et al.* ImageNet: A large-scale hierarchical image database. Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA. 2009. 248–255. [doi: 10.1109/CVPR.2009.5206848]
- Hinton GE, Srivastava N, Krizhevsky A, *et al.* Improving neural networks by preventing co-adaptation of feature detectors. arXiv: 1207.0580, 2012.