

DOI: 10.16781/j.0258-879x.2018.08.0897

· 论 著 ·

基于迁移学习的全连接神经网络舌象分类方法

杨晶东*, 张 朋

上海理工大学光电信息与计算机工程学院自主机器人实验室, 上海 200093

[摘要] **目的** 针对深度学习在舌象分类中训练数据量大、训练设备要求高、训练时间长等问题, 提出一种基于迁移学习的全连接神经网络小样本舌象分类方法。**方法** 应用经 ImageNet 海量数据集训练后的卷积 Inception_v3 网络提取舌象点、线等有效特征, 再使用全连接神经网络对特征进行训练分类, 将深度学习网络学习到的图像知识迁移到舌象识别任务中。利用舌象数据集进行训练、测试。**结果** 与典型舌象分类方法 K 最近邻 (KNN) 算法、支持向量机 (SVM) 算法和卷积神经网络 (CNN) 深度学习相比, 本实验使用的两种方法 (Inception_v3+2NN 和 Inception_v3+3NN) 具有较高的舌象分类识别率, 准确率分别达 90.30% 和 93.98%, 且样本训练时间明显缩短。**结论** 与 KNN 算法、SVM 算法和 CNN 深度学习相比, 基于迁移学习的全连接神经网络舌象分类方法可有效提高舌象分类的准确率、缩短网络的训练时间。

[关键词] 人工智能; 迁移学习; 深度学习; 舌象; 卷积神经网络

[中图分类号] R 241.25 **[文献标志码]** A **[文章编号]** 0258-879X(2018)08-0897-06

Tongue image classification method based on transfer learning and fully connected neural network

YANG Jing-dong*, ZHANG Peng

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

[Abstract] **Objective** To propose a classification method for small sample tongue images based on transfer learning and fully connected neural network, so as to solve the problems of large amount of data, high requirement of training equipment and long training time of deep learning in the classification of tongue images. **Methods** Effective features such as tongue points and lines of tongue images were extracted by the convolution Inception_v3 network after training on the massive data set of ImageNet. The above features were classified by the fully connected neural network, and the image knowledge acquired by the deep learning network was transferred to the tongue image recognition task, and then the tongue data set were used to train and test the efficiency of the network. **Results** Compared with the typical tongue image classification method such as K-nearest neighbor (KNN) algorithm, support vector machine (SVM) algorithm and convolutional neural network (CNN) deep learning method, the two methods (Inception_v3+2NN and Inception_v3+3NN) in our experiment had higher classification rates for tongue images, with the accuracy rates being 90.30% and 93.98%, respectively, and had shorter training time for the sample. **Conclusion** Compared with KNN algorithm, SVM algorithm and CNN deep learning method, the tongue image classification method based on transfer learning and fully connected neural network can effectively improve the accuracy rate of tongue image classification and shorten the training time.

[Key words] artificial intelligence; transfer learning; deep learning; tongue presentations; convolutional neural network

[Acad J Sec Mil Med Univ, 2018, 39(8): 897-902]

传统中医舌诊依赖医师肉眼观察、经验辨析, 其结果受诸多主观和外界因素影响, 缺乏客观评价标准, 制约了舌诊的应用与发展。随着图像处理、机器学习技术的发展, 舌诊客观化研究取得了

一定进展。有学者研究发现经平衡预处理后舌苔数据集训练 GoogLeNet 网络模型的结果优于传统舌象诊断方法^[1-4]。刘关松等^[5]和王昇等^[6]提出了一种基于神经网络集成舌苔分类方法, 发现该分类器具

[收稿日期] 2018-06-29 **[接受日期]** 2018-08-02

[基金项目] 国家自然科学基金(61374039), 上海市自然科学基金(15ZR1429100), 沪江基金(C14002)。Supported by National Natural Science Foundation of China (61374039), Natural Science Foundation of Shanghai (15ZR1429100), and Hujiang Foundation (C14002)。

[作者简介] 杨晶东, 博士, 副教授, 硕士生导师。

*通信作者(Corresponding author). Tel: 021-55271064, E-mail: eerfriend@yeah.net

有较好的精细分类效果,符合中医舌诊要求。传统舌象识别过程包括舌象图片采集、颜色校正、舌体分割、分类器训练等,需要人工寻找特征,这在一定程度上影响分类结果。随着深度学习技术的发展,利用深度学习技术来进行舌象分类成为一种趋势。然而,深度学习技术需要大量标注样本、设备配置要求高且训练时间长,难以在实际中运用。

近年来,迁移学习获得了广泛认可和研究。Pan 和 Yang^[7]、Cheriyadat^[8]、Yang 和 Newsam^[9]对迁移学习的历史、分类和挑战进行了详尽解释;李冠东等^[10]利用迁移学习在深度卷积神经网络(convolutional neural network, CNN)模型下进行高分影像场景分类,提高了场景分类精度;刘晨等^[11]、Lin 等^[12]、Schmidhuber^[13]利用迁移学习先通过大样本对 CNN 进行预训练,再用超限学习机(extreme learning machine, ELM)代替全连接层,最后使用小样本对 CNN 模型进行训练,提高了分类准确率且缩短了训练时间;关胤^[14]提出了一种基于残差网络迁移学习的花卉识别方法,提高了网络在线识别花卉系统的准确率、实时反馈和泛化能力。迁移学习可协助人类解决小样本分类问题,使用在海量数据集上训练好的模型对小样本进行特征提取,再用特征微调全连接神经网络降低利用深度网络解决问题的难度系数。本实验针对传统机器学习与深度学习在小样本舌象分类中出现的问题,将迁移学习方法应用到舌象分类的任务中,发现迁移学习可在提高分类准确率的同时降低配置要求并减少训练时间,现报告如下。

1 材料和方法

1.1 数据集 为了使网络模型对移动互联网终端采集到舌象图片有较好的泛化能力,本研究主要收集了正常舌象、裂纹、厚苔、点刺、齿痕和剥落苔 6 种舌象图片,如图 1 所示。每种舌象图片有 30 张,共 180 张。简单处理使图片像素大小均为 256×256 ,且使用独热编码(one-hot-encoding)对样本类别进行标注。若正常舌象为第 1 类,裂纹为第 2 类,厚苔为第 3 类,则它们的标签分别为 $[1\ 0\ 0\ 0\ 0\ 0]$ 、 $[0\ 1\ 0\ 0\ 0\ 0]$ 和 $[0\ 0\ 1\ 0\ 0\ 0]$,其他类别以此类推。为拟合模型和评估模型泛化误差,将数据样本分为训练集和测试集,训练集和测试集尽可能互斥。训练集与测试集的划分方法有留出法

(hold-out)、交叉验证法(cross validation)和自助法(boot strapping)3 种。本研究采用留出法,直接将数据集 D 划分为 2 个互斥的集合,其中一个集合作为训练集 S ,另一个集合作为测试集 T ,即 $D=S \cup T$, $S \cap T=\varnothing$ 。将 $2/3 \sim 4/5$ 的数据用于训练,剩余样本用于测试。为评估结果稳定性,随机选取 90% 的数据集作为训练集,10% 的数据集作为测试集。

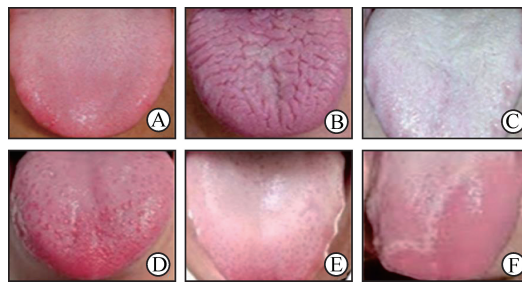


图 1 6 类舌象图片

Fig 1 Six tongue images

A: Normal tongue; B: Crack tongue; C: Thick tongue coating; D: Prick tongue; E: Indented tongue; F: Peeling coating

1.2 迁移学习与模型特征提取 迁移学习数学定义为:给定源域 D_s 和学习任务 T_s , 一个目标域 D_t 和学习任务 T_t , 迁移学习用 D_s 和 T_s 中的知识帮助提高 D_t 中目标预测函数 f_t 的学习, 并且 $D_s \neq D_t$ 或 $T_s \neq T_t$ 。迁移学习可以分为样本迁移、特征迁移、关系迁移和模型迁移 4 类, 研究如何将源任务上训练的知识迁移到仅有少量标签样本数据的目标任务上。深度网络的微调是一种深度网络迁移学习方法, 利用他人已训练好的网络再调整自己的任务, 节省时间和成本, 间接扩充了实验数据, 提升了模型鲁棒性。

GoogLeNet 网络模型由 Szegedy 等^[15]提出, 其在 ILSVRC14 (ImageNet Large-Scale Visual Recognition Challenge 2014) 大赛上以 6.67% 的 top-5 错误率赢得冠军。该研究提出的 Inception_v1 模型在不增加计算量的同时扩大了网络的深度和宽度, 丰富了特征学习并提高了网络分类性能。Inception_v2 模型中加入了 BN 层, 减少了 Internal Covariate Shift; 同时模型使用 2 个 3×3 的卷积核代替 Inception 模块中 5×5 的卷积核, 降低了参数数量并加快了计算速度。Inception_v3 模型在上一模型的基础上进一步分解, 为了加速计算, 将 7×7 和 3×3 的卷积核分解为 2 个一维卷积

($1 \times 7, 7 \times 1$) 和 ($1 \times 3, 3 \times 1$) 形式^[16]; 又将 1 个卷积层分解为 2 个卷积层, 使网络更深并增加了网络的非线性。Inception_v3 模型经 2012 年的 ImageNet 数据集训练后, 在测试集上取得了 3.5% 的 top-5 错误率 (人类专家为 5%^[17])。

经 ImageNet 数据集训练好的 Inception_v3 模型具有抽取图像低层次特征 (如点、线、曲线和边缘等) 和抽象特征的能力。利用模型迁移和模型微调的思想, 本研究将在 ImageNet 数据集上预训练好的 Inception_v3 模型知识 (保存的所有卷积层参数) 迁移到舌象分类的任务中, 从而对舌象图像进行有效特征向量提取, 减少舌象训练样本量和模型训练时间, 提高分类的准确性。例如将图 1 中的每张图像输入训练好的 Inception_v3 模型得到的特征为 2 048 维的向量, 可视化像素大小为 32×64 的有效特征图, 如图 2 所示 (图 2 中的特征图与图 1 相应舌象的位置一一对应)。

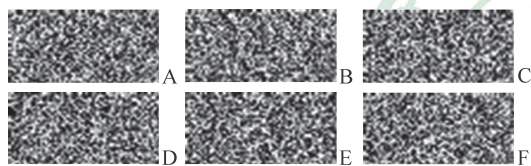


图 2 通过迁移学习得到的特征图

Fig 2 Feature images based on transfer learning

A: Normal tongue feature; B: Crack tongue feature; C: Thick tongue coating feature; D: Prick tongue feature; E: Indented tongue feature; F: Peeling coating feature

1.3 基于迁移学习的全连接神经网络分类模型 本实验采用的全连接网络为多层前馈神经网络, 如图 3 所示, 该网络有 d 个输入神经元、 q 个隐层单元和 l 个输出单元。隐含层第 h 个神经元的阈值为 γ_h , 输出层第 j 个神经元的阈值为 θ_j , 输入层第 i 个神经元与隐含层第 h 个神经元的连接权重为 v_{ih} , 隐含层第 h 个神经元与输出层第 j 个神经元的连接权重为 w_{hj} 。隐含层使用 ReLU 激活函数, 输出层使用 Softmax 函数, 分别记为 $f_1(v^T x + \gamma)$ 和 $f_2(w^T B + \theta)$ 。其中 v 为 $d \times q$ 大小的权值矩阵, x 为 $d \times 1$ 大小的样本输入向量, γ 为 $q \times 1$ 大小的阈值向量; w 是 $q \times l$ 大小的权值矩阵, $B = f_1(\cdot)$ 大小为 $q \times 1$, θ 为 $l \times 1$ 大小的阈值向量。网络前向传播输出为 $y = f_2(w^T B + \theta)$ (网络预测的概率值), 其中 y 为样本对应的标签向量, 大小为 $l \times 1$ 。用交叉熵计算损失函数 $Loss = Y_1 \log(y_1) + \dots + Y_l \log(y_l) + \dots +$

$Y_l \log(y_l)$, 使用反向传播 (back propagation, BP) 算法、梯度下降最小化损失函数, dropout 技术减少过拟合, 寻找最优模型分类参数。

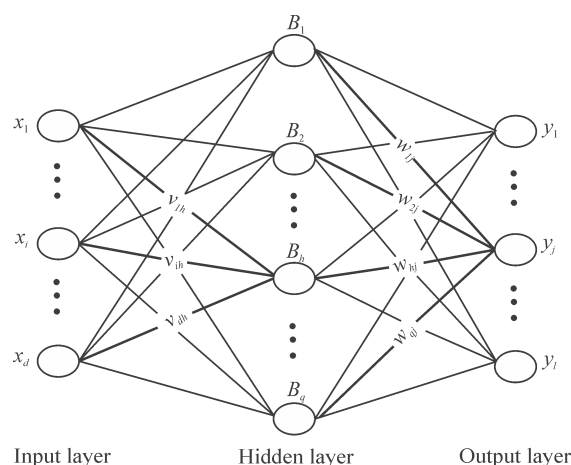


图 3 单隐层前馈神经网络

Fig 3 Single hidden layer feed forward neural network

基于迁移学习的全连接网络结构如图 4 所示。左侧为已经训练好的深度卷积 Inception_v3 模型 (即 Google 公开训练好的 Inception_v3 模型), 网络中保存了拟合网络参数, 具有提取图像强特征的能力, 使用该深度模型提取小样本舌象特征。右侧为全连接神经网络, 本实验使用了 2 层和 3 层 2 种全连接神经网络。Inception_v3+2NN 为 Inception_v3 模型+dense1(2 048, ReLU)+dense2(6, Softmax), Inception_v3+3NN 为 Inception_v3 模型+dense1(2 048, ReLU)+dense2(1 024, ReLU)+dense3(6, Softmax), 使用提取的强特征微调全连接神经网络进行小样本舌象分类。舌象分类流程如图 5 所示。

2 结果和讨论

2.1 实验环境 本实验使用了 python 和 Google 开源深度学习框架 TensorFlow 搭建神经网络, 训练笔记本电脑配置为处理器 Inter® Core™ i5-3230M CPU 2.60 GHz, 安装内存为 4.00 GB, 没有使用图形处理器, 全程单线程运行。180 张舌象图像中训练集为 162 张, 测试集为 18 张, 将数据转化为 TFRecord 格式备用。实验前已获取 Google 在 ImageNet 数据集上训练好的 Inception_v3 模型 BP 文件, 本实验的网络模型为 Inception_v3+2NN 和 Inception_v3+3NN。同时选取 CNN、改进的 CNN、K 最近邻 (K-nearest neighbor, KNN) 算

法、支持向量机 (support vector machine, SVM) 如表 1 和表 2 所示。
算法为对照, CNN 和改进的 CNN 的网络架构大纲

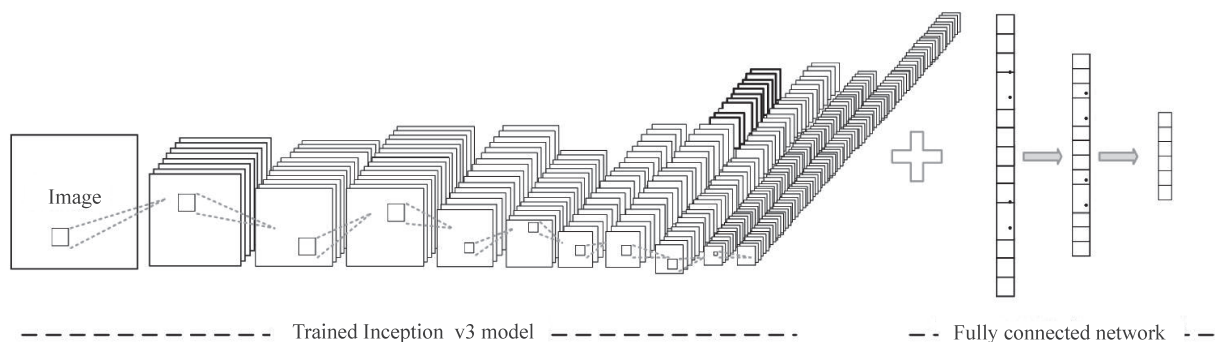


图 4 基于迁移学习的全连接网络结构图

Fig 4 Fully connected neural connection network structure diagram based on transfer learning

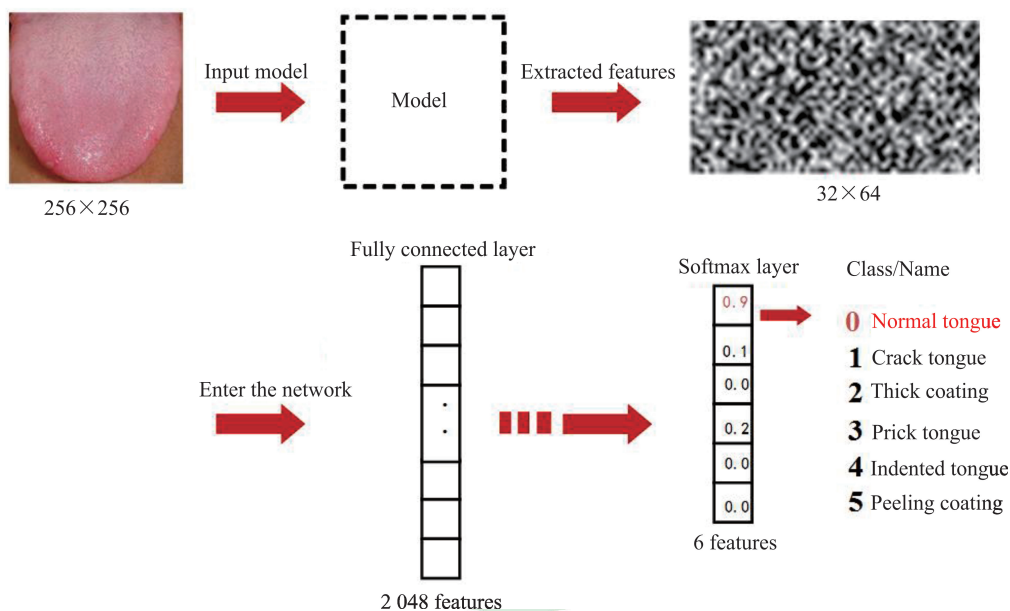


图 5 舌象图像分类流程图

Fig 5 Flow chart of tongue image classification

表 1 卷积神经网络架构大纲

Tab 1 Architecture outline of convolutional neural network

Type	Patch size/stride or remarks	Input size
Conv1	3×3/1	256×256×3
Max-pool1	3×3/2	256×256×64
Conv2	3×3/1	128×128×64
Max-pool2	3×3/2	128×128×128
Conv3	3×3/1	64×64×128
Conv4	3×3/1	64×64×64
Conv5	3×3/1	64×64×16
Dense layers	16 384/2 048/512/6	64×64×4
Softmax	Classifier	6

Conv: Convolutional layer

2.2 训练集的收敛性分析 分别采用 CNN、改进的 CNN、Inception_v3+2NN 和 Inception_v3+

3NN 基于训练集图像进行训练, 其中学习率为 0.001, 批样本数为 10, keep_prob 为 0.5, 迭代次数为 4 100, 每迭代训练 100 次评估 1 次网络模型的分类准确率。由于 KNN 算法没有训练过程, SVM 算法的训练过程是调优主要参数, 均无法记录准确率, 故只记录以上 4 种网络模型的准确率曲线, 如图 6 所示。

由图 6 分析可知, 4 种网络模型中 CNN 的收敛性最差, 其在测试集中的平均分类准确率较 Inception_v3+2NN 和 Inception_v3+3NN 分别低 73.63%、77.31%。分析其原因为 CNN 需要反复对训练集进行迭代以提取舌象特征, 而训练集样本数量较少, 池化层使用较多, 使 CNN 难以进行

特征学习, 故训练收敛性差。改进后 CNN 的平均准确率相比 Inception_v3+2NN 和 Inception_v3+3NN 分别低 62.52%、66.20%, 分析其原因为改进的 CNN 只使用了一个池化层, 并调小了网络结构, 使网络在小样本训练集上也可以提取到有用特征, 故收敛性较好。Inception_v3+3NN 收敛性最好, 这是由于 Inception_v3 模型能直接提取舌象的强表达基础特征, 然后将这些特征在全连接神经网络上进行迭代训练, 故其收敛性优于 CNN 和改进的 CNN。

表 2 改进的卷积神经网络架构大纲
Tab 2 Architecture outline of improved convolutional neural network

Type	Patch size/stride or remark	Input size
Conv1	$3 \times 3/1$	$256 \times 256 \times 3$
Conv2	$3 \times 3/1$	$256 \times 256 \times 32$
Conv3	$3 \times 3/1$	$256 \times 256 \times 16$
Max-pool1	$3 \times 3/2$	$256 \times 256 \times 4$
Conv4	$3 \times 3/1$	$128 \times 128 \times 4$
Dense layers	16 384/2 048/512/6	$128 \times 128 \times 1$
Softmax	Classifier	6

Conv: Convolutional layer

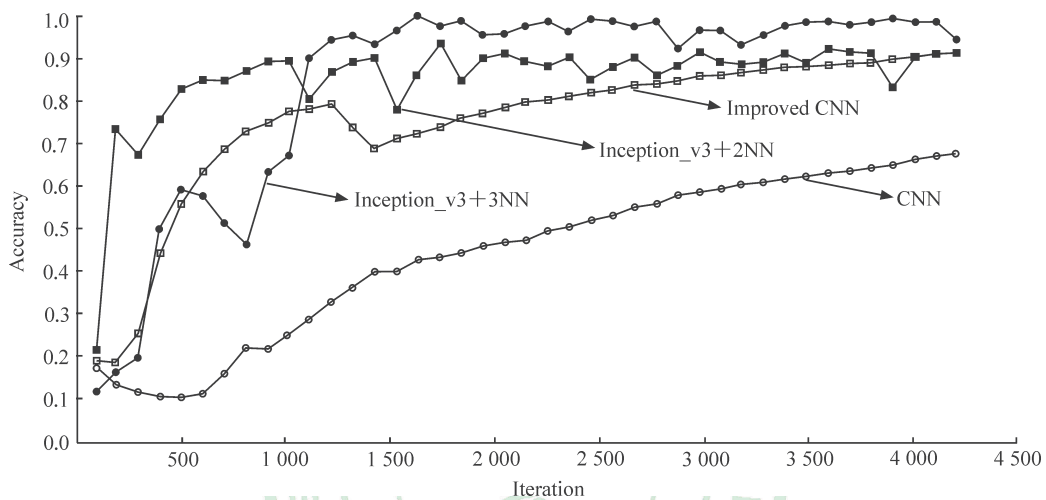


图 6 原始数据集训练正确率曲线

Fig 6 Accuracy rate curve of four networks trained by origin tongue sets

CNN: Convolutional neural network; Inception_v3+2NN: Inception_v3 model plus two-layer neural network; Inception_v3+3NN: Inception_v3 model plus three-layer neural network

2.3 测试集的泛化性分析 使用迁移学习将训练集输入 Inception_v3 模型, 提取具有强表达能力的特征向量, 再分别使用 2NN (两层全连接神经网络) 和 3NN (三层全连接神经网络) 对特征向量进行训练, 训练完成后使用测试集分别进行测试, 记录算法模型的平均分类结果和运行时间。本实验方法与传统 KNN、SVM 和 CNN 方法的比较结果如表 3 所示。

3 结论

KNN 算法没有训练过程, 其在利用测试集进行测试时, 将图像转化为 196 608 维, 根据离测试集最近的 K 个训练集类别出现频率进行分类, 分类结果易受 K 值影响, 样本维度较高时分类效果受限, 其分类准确率为 50%; SVM 算法需要在一个最大边缘超平面进行分类, SVM 算法的分类准

确率低于 KNN 算法; CNN 内有大量参数用于学习知识表示, CNN 训练时需要大量数据拟合参数用于分类过程。为了减少计算量和训练时间, 其采用了池化技术, 这将造成原始图像信息丢失, 致使 CNN 在小数据集上迭代训练时提取不到具有强代表性的舌象特征, 易产生过拟合, 因而表现较差。经减少 CNN 的池化次数并调节网络结构改进后 CNN 的分类准确率由 16.66% 提高至 27.78%, 但仍然低于其他方法, 表明 CNN 不适用于小样本数据分类。

本实验将迁移学习与全连接神经网络结合, 使用训练好的 Inception_v3 从小样本舌象训练集中提取特征, 再利用特征微调全连接神经网络, 较好地解决了传统机器学习和深度学习在小样本、多分类中的问题。本实验中 Inception_v3+2NN 和 Inception_v3+3NN 两种方法的分类准确率分别为

90.30% 和 93.98%, 表明对小样本数据进行分类时, 迁移学习方法具有较好的特征提取效果。

表 3 本研究分类方法与其他方法效果对比

Tab 3 Comparison of our methods with other methods

Method	Accuracy (%)	Time <i>t</i> /min
KNN	50.00	1.25
SVM	33.33	1.36
CNN	16.67	170
Improved CNN	27.78	139
Inception_v3+2NN	90.30	15
Inception_v3+3NN	93.98	18

KNN: K-nearest neighbor; SVM: Support vector machine; CNN: Convolutional neural network; Inception_v3+2NN: Inception_v3 model plus two-layer neural network; Inception_v3+3NN: Inception_v3 model plus three-layer neural network

综上, KNN 和 SVM 算法由于没有大量迭代过程, 因此分类时间较短, 但在面对高维数据时容易受自身分类能力限制; CNN 每次迭代都需大量的卷积运算、梯度计算进行参数拟合, 且图片尺寸较大, 故处理时间较长, 难以使用小样本数据拟合大量参数并提取有用特征, 泛化性较差。本实验将迁移学习与全连接神经网络结合并应用到小样本舌象分类领域, 采用预训练好的 Inception_v3 模型, 在实验中无需训练即可直接输入图像进行特征提取, 有效缩短了训练时间。综合分析发现, 与 KNN、SVM 和 CNN 方法相比, 基于迁移学习的全连接神经网络舌象分类方法可有效提高舌象分类准确率、缩短网络训练时间。

[参考文献]

- [1] FU S, ZHENG H, YANG Z, LIU Y. Computerized tongue coating nature diagnosis using convolutional neural network[C]//IEEE, International Conference on Big Data Analysis. IEEE, 2017: 730-734.
- [2] MARMANIS D, DATCU M, ESCH T, STILLA U. Deep learning earth observation classification using ImageNet pretrained networks[J]. IEEE Geosci Remote S, 2016, 13: 105-109.
- [3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[Z/OL]. arXiv: 1409.1556, 2014. <https://arxiv.org/pdf/1409.1556.pdf>.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems. Curran Associates Inc., 2012: 1097-1105.
- [5] 刘关松,徐建国,高敦岳. 基于神经网络集成的舌苔分类方法[J]. 计算机工程,2003,29:100-102.
- [6] 王昇,刘开华,王丽婷. 舌诊图像点刺和瘀点的识别与提取[J]. 计算机工程与科学,2017,39:1126-1132.
- [7] PAN S J, YANG Q. A survey on transfer learning[J]. IEEE T Knowl Data En, 2010, 22: 1345-1359.
- [8] CHERIYADAT A M. Unsupervised feature learning for aerial scene classification[J]. IEEE T Geosci Remote, 2013, 52: 439-451.
- [9] YANG Y, NEWSAM S. Spatial pyramid co-occurrence for image classification[C]//IEEE International Conference on Computer Vision. IEEE, 2011: 1465-1472.
- [10] 李冠东,张春菊,王铭恺,张雪英. 卷积神经网络迁移的高分影像场景分类学习[J]. 测绘科学,2019(6):1-13.
- [11] 刘晨,曲长文,周强,李智,李健伟. 基于卷积神经网络迁移学习的 SAR 图像目标分类[J]. 现代雷达,2018,40: 38-42.
- [12] LIN K, YANG H F, CHEN C S. Flower classification with few training examples via recalling visual patterns from deep CNN[C]//IPPR Conference on Computer Vision, Graphics, and Image Processing (CVGIP), 2015.
- [13] SCHMIDHUBER J. Deep learning in neural networks: an overview[J]. Neural Netw, 2014, 61: 85-117.
- [14] 关胤. 基于残差网络迁移学习的花卉识别系统[J/OL]. 计算机工程与应用. [2018-05-15]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20180404.1703.020.html>.
- [15] SZEGEDY C, LIU W, JIA Y, SERMANET P, REED S E, ANGUELOV D, et al. Going deeper with convolutions[C/OL]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2014. doi: 10.1109/CVPR.2015.7298594.
- [16] SZEGEDY C, VANHOUCKE V, IOFFE S, SHLENS J, WOJNA Z. Rethinking the inception architecture for computer vision[C]//Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV. IEEE, 2016: 2818-2826.
- [17] RUSSAKOVSKY O, DENG J, SU H, KRAUSE J, SATHEESH S, MA S, et al. ImageNet large scale visual recognition challenge[J]. Int J Comput Vision, 2015, 115: 211-252.

[本文编辑] 杨亚红