

基于改进 VGG16 的猴子图像分类方法

田佳鹭, 邓立国

(沈阳师范大学 数学与系统科学学院, 辽宁 沈阳 110034)

摘要:为提高对细粒度图像分类的准确性和分类速度,提出基于改进 VGG16 和迁移学习的图像分类方法。首先从 kaggle 平台中获取十种不同猴子数据集,并对数据集进行标准化处理,包含图片去椒盐噪声、将数据集转换为 TensorFlow 中提供的统一 TFRecord 数据格式。然后迁移学习改进的 VGG16 卷积神经网络,模型的优化包括利用 Swish 作为激活函数、将 softmax loss 与 center loss 相结合作为损失函数以实现更好的聚类效果、采用性能完善的 Adam 优化器。用训练集训练模型以确定微调参数信息,再用测试集检验模型准确性。结果表明,该方法对猴子图像分类的准确度可达到 98.875%,分类速度也得到了显著提升。与其他传统卷积神经网络模型相比,该方法具有更高的准确性和适用性。

关键词:迁移学习;VGG16;卷积神经网络;图像分类;TFRecord

中图分类号:TP391.41;TP18

文献标识码:A

DOI: 10.19358/j.issn.2096-5133.2020.05.002

引用格式: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.

Monkey image classification method based on improved VGG16

Tian Jialu, Deng Ligu

(School of Mathematics and Systems Science, Shenyang Normal University, Shenyang 110034, China)

Abstract: In order to improve the accuracy and speed of fine-grained image classification, an image classification method based on improved VGG16 and migration learning is proposed. Firstly, ten different monkey datasets are obtained from the kaggle platform, and the datasets are standardized, including image de pepper and salt noise, and the datasets are transformed into the unified TFRecord data format provided by TensorFlow. Then, the improved VGG16 convolutional neural network is migrated. The optimization of the model includes using Swish as the activation function, combining softmax loss with center loss as the loss function to achieve better clustering effect, and using Adam optimizer with perfect performance. The training set is used to train the model to determine the fine-tuning parameter information, and then the test set is used to test the accuracy of the model. The results show that the accuracy of monkey image classification can reach 98.875%, and the classification speed has been significantly improved. Compared with other traditional convolution neural network models, this method has higher accuracy and applicability.

Key words: transfer learning; VGG16; convolutional neural network; image classification; TFRecord

0 引言

随着深度学习技术的不断进步,对于通用对象的类别分析,卷积神经网络已经达到了很高的水平,但对于细粒度图像分类的稳定性目前还有待提升。所谓细粒度图像分类,通常用于描述对同一类事物进行细致的划分,所以待分类图像的区别要更加精准,类内区别大而类间区别小,粒度则更为精细。本文方法基于改进的 VGG16 深度卷积神经网络。尽管原始的 VGG16 模型已经拥有了良好的普适性和实用价值,并且在各种图像分类和目标检测

任务中都体现了极佳的效果,通过迁移学习改进的 VGG16 模型,能够将已习得的特征和性能应用到待解决的问题中,极大地节省了训练时间。此外在硬件上选取 GPU 进行训练,使得速度得到进一步提升。为了改善原始的交叉熵损失函数无法保证提取的特征具有识别度的缺点,在 VGG16 模型中引入将 center loss 损失函数与交叉熵损失函数相结合的办法。此外还运用了新型的 Swish 激活函数,以及拥有自适应学习率的 Adam 优化器。最后利用不同种类的猴子训练集对改进的模型重新训练,以获得

少量微调的参数信息。经验证该方法对猴子图像识别的精准度可提升到 98.875%,而原始的 VGG16 模型在该数据集上的分类仅能达到 90.210% 的准确率,可以证明改进后的模型具有更好的识别效果。

1 VGG16 模型及迁移学习

1.1 VGG16 模型

VGG 卷积神经网络的层次结构包括卷积层、激活层、池化层和全连接层,其中卷积层起到十分重要的作用,通过实现“局部感知”和“参数共享”的两种方式,达到了降维处理和提取特征的目的^[1]。卷积核是卷积层的核心,在卷积核的作用下可以提取位于图像中不同位置同一物体的形状,在起到降维作用的同时也减少了需要训练的参数^[2]。

池化层 pool 中用池化过滤器对输入数据进行降维^[3]。最后是全连接层,它的作用相当于“分类器”,由于此次待分类的数据与原先 VGG16 的分类数据不同,因此将全连接层的参数抹去,通过重新训练确定最后三层全连接层的参数信息,来实现分类目标^[4]。

VGG16 卷积神经网络模型由 13 层卷积层和 3 层全连接层组成。该模型要求输入的图片数据大小为 $224 \times 224 \times 3$,初始卷积核的大小为 $3 \times 3 \times 3$,步幅 stride 的大小为 1,有效填充 padding 的大小为 1,池化层 pooling 采用 2×2 的最大池化函数 max pooling 的方式。模型中卷积的过程为:首先使用两次 64 个卷积核的卷积处理,接着进行一次池化层 pooling,完成后又经过两次 128 个卷积核的卷积,并采用一次池化层 pooling,再经过三次 256 个卷积核的卷积之后,采用一次池化层 pooling,最后重复两次三个 512 个卷积核卷积之后,进行一次池化层 pooling。在卷积层处理后是三次全连接层 Fc_layer,此时网络中需要处理的参数已经减少了很多^[5]。但传统 VGG16 的分类速度仍较慢,识别的准确度还有待提高。

1.2 迁移学习

对于新的问题往往需要重新搭建模型,并利用数据集再训练权重参数,成本相对较高且浪费时间,有时还难以实现。因此迁移学习的概念随之而来。在机器学习领域中,传统的机器学习方法局限于只能解决单一问题,无法一次性解决多个问题^[6]。所以迁移学习领域的任务,就是将在源域中解决任务时所学到的知识运用到目标域中,结合目标域中已有的训练数据,对模型进行微调从而构造

出泛化能力更好的模型。

本文方法中,将在 ImageNet 上训练好的 VGG16 模型的结构和参数,迁移到对猴子图形分类的问题中。通过迁移方法有效地减少大量训练参数,再次训练时只需确定微调参数即可,该过程不但减轻了工作量,基于以前的参数经验也可以实现更优化的分类目的。

2 对 VGG16 卷积神经网络模型的改进

2.1 图片进行去噪声处理

图片中通常会伴随着椒盐噪声,这是由于信号在图像传感器传输、解码处理等操作时,受到突如其来的强烈干扰而产生,属于一种因信号脉冲强度引起的噪声,可能在亮区域中出现黑色像素点或是在黑暗区域出现白色像素点^[7]。通常对于椒盐噪声处理的方法有异常值检测、中值滤波、伪中值滤波,其中常用的处理方式是中值滤波方法,它是一种非线性滤波技术,核心思想是利用该像素点邻域中强度值的中值来代替该像素点的灰度值(计算中值时包括该像素点的原始灰度值),并且能够做到完整地保留图像边缘信息,其利用 cv2.medianBlur 函数进行处理^[8]。

通常使用信噪比 SNR 来衡量图像噪声,但由于其中所需用到的功率谱难以计算,因此采用信号与噪声的方差之比来近似估计图像的 SNR。信噪比越小,图片中所混杂的噪声越多。图 1 所示是猴子数据库中一张伴随有椒盐噪声的图像,图 2 则是图像去噪处理后的结果。

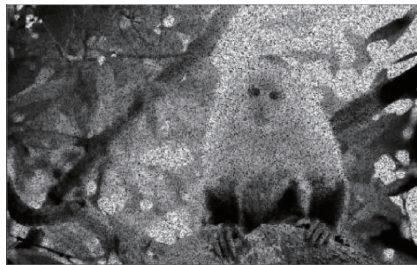


图 1 带有椒盐噪声的图像



图 2 去噪处理后的图像

2.2 数据标准化

运用数据集对模型进行训练时,通常需要将数据集分为训练集和测试集。前期利用训练集计算权重参数以及每个神经元的偏置,后期利用测试集对训练好的模型的准确率进行验证。为了实现对数据集的统一管理, TensorFlow 中提供了一种 TFRecord 的格式,能够对训练数据实现统一存储。TFRecord 格式可以将任意数据转化成 TensorFlow 所支持的格式,使得数据集更容易应用到相应的网络应用架构中^[9]。TFRecord 是一种将图片中重要的数据以及标签保存在一起的二进制文件,可以通过调用 tf.python_io.TFRecordWriter 类,将数据存放到 TFRecords 文件的 tf.train.Example 协议内存块中,Example 协议块中包含以字典形式存放的特征值和数据内容。

TensorFlow 标准化的建议是,一个 TFRecord 文件中最好仅存放 1 000 张左右的图片,若在一个包中存放太多的数据,则不利于多线程的读取,因此需要根据相应数据集的数量对图片进行打包。针对本文方法中用到的训练集和测试集的图片数量,将训练集数据打包为两个文件:traindata.tfrecords-000 和 traindata.tfrecords-001;因为测试集数据量不算多,因此打包为一个 testdata.tfrecords-000 文件即可。每种集合中不同类型的猴子图片分配情况如表 1 所示。

表 1 训练集和测试集中 10 种猴子的图片分类情况

Label	Common Name	Train Images	Test Images
n0	mantled_howler	131	26
n1	patas_monkey	139	28
n2	bald_uakari	137	27
n3	japanese_macaque	152	30
n4	pygmy_marmoset	131	26
n5	white_headed_capuchin	141	28
n6	silvery_marmoset	132	26
n7	common_squirrel_monkey	142	28
n8	black_headed_night_monkey	133	27
n9	nilgiri_langur	132	26

TFRecord 文件中存储的是图像经过压缩编码后的结果,因此后续将图片进行恢复时需要进行解码操作^[10]。通常训练模型时获取的图片数据集尺

寸大小并不统一,为了能够把 TFRecord 文件中的二进制的数重新构造为原来的图片,需要记录每张图片的 img_width 和 img_height 两个信息,并利用 TensorFlow 中的 reshape 函数实现图片重构。在本文方法中对重构的图片以 mun_Label_class id(数字_Label_标签)的方式进行命名。重构的标准化图片结果如图 3 所示。

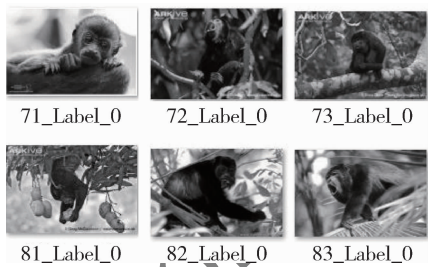


图 3 重构图片形式

2.3 损失函数的改进

损失函数 loss 描述的是真实值与预期值之间的偏离程度,有利于对模型的学习进行指导,损失函数越小代表该模型的鲁棒性能越好。在本文方法中使用的损失函数是将 softmax loss 与 center loss 相结合。

在实现多分类任务的情况下,人们通常会选用 softmax 激活函数与交叉熵损失函数 cross entropy 相结合的办法。交叉熵刻画的是两个概率分布的距离,交叉熵越大说明两者的距离越远。由于神经网络的输出结果并不一直都是概率分布,因此用 softmax 激活函数将多个神经元的输出归一到(0,1)区间内,将前馈神经网络的传播结果转换为概率分布的形式。这种方式对于不平衡的训练集十分有效(文中提到的 softmax loss 均指代交叉熵损失函数)^[11]。

首先在 softmax 函数的作用下将前馈神经网络的输出结果变为概率分布,假设卷积神经网络的输出为 $y_1, y_2, \dots, y_n$, 经过 softmax 函数处理后变为:

$$\text{softmax}(y_i) = e^{y_i} / \sum_{j=1}^n e^{y_j} \tag{1}$$

采用 softmax 函数的好处在于计算较为简单。交叉熵的计算公式如下:

$$f = -1/n \sum_x [y \ln a + (1 - y) \ln (1 - a)] \tag{2}$$

其中, y 为真实的输出值,而 a 为模型求得的模型值。

最终逻辑回归的损失函数定义如式(3)所示,在这个损失函数中 Y 是真实结果转化为 one-hot 的向量结果。

$$\text{Loss} = - \sum_i Y \log(\text{softmax}(y_i)) \quad (3)$$

通常希望训练出来的理想数据能够呈现出同类之间紧凑、异类之间分散的效果。上面介绍的 softmax loss 损失函数,可以有效地实现类间的区别较大,但是无法达到类内较为紧凑的效果,所以类内的间距还是比较大的。

利用经典的 MNIST 手写体数字识别数据集进行验证交叉熵损失函数,数据集中共有 10 个数字,利用二维空间展现训练到最后一层的特征分布情况。从图 4 可以看出使用 softmax loss 损失函数后能够呈现出较为清晰的类别界限。

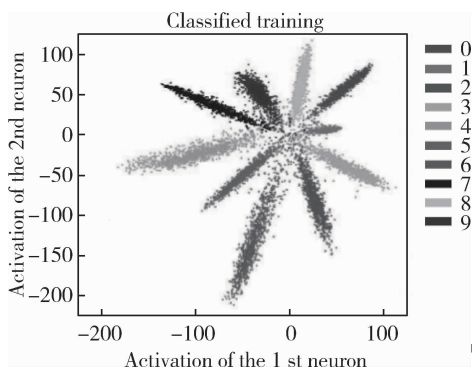


图 4 MNIST 数据集的特征在二维空间的分布情况

为缩小类内差距,本文方法运用辅助函数即中心损失函数 center loss。center loss 函数的核心意义是,对于每一个分类都需要维护一个类内中心 c ,即该类所有样本的特征平均值,如果输入样本距离类中心太远就要受到惩罚,从而起到缩小类内距离的作用^[12]。并且通过上述实验证明如果只用 center loss 函数,效果并不是很好,无法有效增加类间差异性。因此可以采用将 softmax loss 与 center loss 相结合的方法扬长避短,实现期望特征具有较高的识别度。

center loss 的定义为:

$$\text{CenterLoss} = 1/2N \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (4)$$

令 $\text{CenterLoss} = L_c$ 。

softmax loss 的定义为:

$$L_s = - \sum_{i=1}^m \log(e^{W_i^T x_i + b_i} / \sum_{j=1}^n e^{W_j^T x_i + b_j}) \quad (5)$$

其中 c_{y_i} 代表每类特征的中心。完整的损失函数为:

$$L = L_s + \lambda L_c = - \sum_{i=1}^m \log(e^{W_i^T x_i + b_i} / \sum_{j=1}^n e^{W_j^T x_i + b_j}) + \lambda L_c \quad (6)$$

对参数 λ 设置不同值,可以调节类内的密集程度, λ 越大类内越密集。将改进的损失函数应用到基于 MNIST 数据集的训练当中,从图 5 可以看出采用将两种损失函数相结合的办法,能够有效实现类间间距更大,而类内更加紧凑的预期效果。

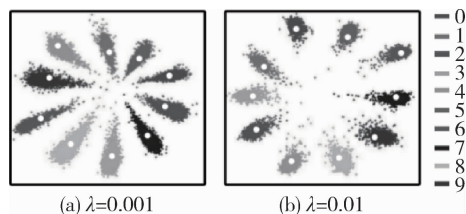


图 5 类间差距大、类内差距小的优化分类效果

2.4 Swish 激活函数

在神经网络中数据间的关系通常为非线性关系,因此激活层存在的最大目的在于引入非线性因素,来增加整个网络的表征能力。原始模型中使用的激活函数是修正线性单元 ReLU,尽管 ReLU 具有单侧抑制、相对宽阔的兴奋边界以及稀疏激活性等优点,但仍存在不足。当 $x > 0$ 时,它会一直无止境地“发飙”下去,ReLU 属于非饱和的线性函数,这样会导致数值不稳定,计算溢出,变成负值,这时 ReLU 的输出结果将永远为 0^[13]。

因此在该模型的改进中引用了 Swish 激活函数,Swish 函数有下界无上界,并且它拥有不饱和、光滑、非单调性的特征。此外在 Google Brain 的论文中提到在不同数据集上运用 Swish 激活函数时,准确率都有所提高,并且更适用于大的数据集和更深层的神经网络结构^[14]。Swish 的定义为:

$$f(x) = x \times \text{sigmoid}(\beta x) \quad (7)$$

为了测试 Swish 比 ReLU 的性能更优,利用三种不同的模型结构,分别为 VGG13、VGG16 和 AlexNet,同时在 MNIST 和 10 种不同猴子的两个数据集上进行测试。对于不同种类猴子的数据集,在其训练和测试中准确率、损失函数 loss 的变化程度上,也能够体现出 Swish 激活函数的性能比 ReLU 激活函数要好,因此在改进的模型中选用 Swish 激活函数。测试结果如表 2 所示。

表 2 检验结果

数据集	模型	训练	ReLU	Swish
	结构	次数	测试集/%	测试集/%
MNIST	VGG13	10 000	96.58	97.16
	VGG16	10 000	98.44	98.83
	AleNet	10 000	93.31	94.45
10 种不同猴子	VGG13	10 000	89.23	89.56
	VGG16	10 000	95.82	98.48
	AleNet	10 000	90.03	90.45

2.5 Adam 优化器

在该模型中引入 Adam 优化器,这是一种自适应学习率优化算法,是一种对随机目标函数执行一阶梯度优化的算法^[15]。Adam 算法与传统梯度下降优化器的区别在于,它能够为每一个参数设计独立的自适应性学习效率,通过对梯度的一阶矩估计和二阶矩均值计算来改变学习效率^[16]。

Adam 算法将 AdaGrad 算法和 RMSProp 算法相结合,分别取二者之长。其中 AdaGrad 算法能够自动为不同参数适应不同的学习率,根据参数变化的频繁和稀疏程度来调整步长进行更新。RMSProp 算法利用梯度平方的指数移动平均数对学习率进行调节,能够有效加快目标函数的收敛速度。此外 Adam 算法的调参更为简单,通过默认参数就可以解决大部分问题,适应于拥有大规模数据及参数的情况。该算法具有良好的稳定性,能够让网络函数收敛得更快^[17]。定义公式如下:

$$m_t = \mu \times m_{t-1} + (1 - \mu) \times g_t \tag{8}$$

$$n_t = v \times n_{t-1} + (1 - v) \times g_t^2 \tag{9}$$

$$\hat{m}_t = m_t / (1 - \mu^t) \tag{10}$$

$$\hat{n}_t = n_t / (1 - v^t) \tag{11}$$

$$\Delta \theta_t = - \hat{m}_t / (\sqrt{\hat{n}_t} + \delta) \times \eta \tag{12}$$

式中, m_t 是对梯度的一阶矩估计, n_t 是对梯度的二阶矩估计,相当于是对两种期望 $E|g_t|$ 和 $E|g_t^2|$ 的估计; $\hat{m}_t$ 、 $\hat{n}_t$ 分别表示对 m_t 和 n_t 的矫正,相当于对期望的无偏估计^[18]。

3 实验结果与分析

本次试验基于迁移学习,为了实现对细粒度猴子图像的分类达到更优的效果,对传统的 VGG16 模型进行了改进。数据集采用从 kaggle 平台上获取的 10 种不同的猴子图像,首先运用训练集在改进的模

型上进行训练(统一将改进的模型称为 VGG16_2)。在训练速度方面,通过利用迁移学习已经节省了大量的训练时间,此外在硬件方面运用 NVIDIA 的 GPU 对模型进行训练,使得卷积神经网络的训练速度得到了大幅度的提升。原本需要花费一天才能训练完成,改进模型仅需要 1 个小时就可以达到预期效果。

模型训练的同时需要保存每一步的关键参数、当时的模型结构以及重新训练的权重信息,以便后续再次使用该模型时可以将有用信息直接恢复到所需要的模型当中。对于模型的保存使用 Saver 类中的 save 函数,保存时会产生四类文件,包括“.meta”文件用于保存图形的结构;“.data”文件用于保存变量的值;“.index”文件用于标识检查点信息以便于以后的恢复;“checkpoint”文件包含了最近检查点列表的协议缓冲区。恢复模型时使用 Saver 类中的 restore 函数,将这些文件加载到当前模型中即可^[19]。

最后对训练好的模型进行测试,运用测试集检验 VGG16_2 的准确率。通过验证得到该模型对于猴子图像分类的准确率可以达到 98.875%。图 6、图 7 分别描述了此次训练以及测试时损失和准确率的变化。

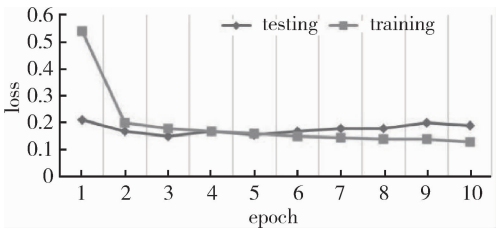


图 6 损失的变化

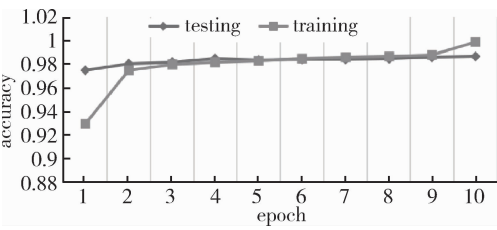


图 7 准确率的变化

为了验证改进的 VGG16_2 模型具有更好的识别性能,利用猴子数据集分别在 VGG11、VGG13、VGG16、alexnet、VGG19_bn、inceptionv3、resnet152_v2 以及改进的 VGG16_2 上进行训练。在不同模型上的准确率对比如表 3 所示,从表中可以看出经过

改进的 VGG16_2 的准确率较高。表 4 列出了对于不同种类猴子的识别率。

表 3 猴子数据集在不同模型上的准确率对比

试验模型	训练集准确率/%	测试集准确率/%
VGG11	80.23	78.21
VGG13	83.75	80.34
VGG16_2	98.875	95.62
alexnet	84.35	79.83
inceptionv3	85.75	80.62
resnet152_v2	84.63	80.86
VGG16	90.21	89.25
VGG19_6n	86.32	78.62

表 4 10 类猴子分别在 VGG16_2 上的识别率

猴子类别	mantled_ howler	Patasmonkey	bald_ uakari	japanese_ macaque	black_headed_ night_monkey	nilgiri_ langur	pygmy_ marmoset	white_headed_ capuchin	silvery_ marmoset	common_ squirrel_monkey
识别率/%	96.452	97.432	98.532	97.856	97.789	98.112	98.424	98.525	97.425	96.633

经最后的测试集验证,改进的 VGG16_2 模型的总体准确率高达 98.875%。尽管如此,模型还有待提升的空间,并且针对这种细粒度图像的识别还需对模型进行完善,以达到最好效果。

参考文献

- [1] 刘建伯. 基于卷积神经网络图像分类方法研究及应用[D]. 大庆:东北石油大学,2019.
- [2] 李妮. 基于深度信念网络的时间序列预测研究[D]. 西安:西安理工大学,2019.
- [3] 岳清清. 深度学习在岩石薄片图像检索中的应用研究[D]. 西安:西安石油大学,2019.
- [4] 黎式南. 基于双通道卷积神经网络的立体匹配算法研究[D]. 岳阳:湖南理工学院,2019.
- [5] 赵嘉. 基于迁移学习的遥感图像分类研究[D]. 长春:吉林大学,2019.
- [6] 徐旭东,马立乾. 基于迁移学习和卷积神经网络的控制图识别[J]. 计算机应用,2018,38(S2):290-295.
- [7] 黄晶晶,张明海. 基于偏微分中值滤波的巡检图像去噪研究[J]. 光电技术应用,2019,34(5):42-47.
- [8] 黄友亮,孙广形,战胜,等. 基于自适应中值滤波高炉渣颗粒小波去噪[J]. 矿业研究与开发,2019,39(9):141-145.
- [9] ALKASHASH A M, AWAD S, COLLINGWOOD R, et al. Classification of pap test images using TensorFlow as an open source artificial intelligence software[J]. Journal of the American Society of Cytopathology,2019,8(5):29-30.
- [10] LIU T J, GUO Q Q, LIAN C F, et al. Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural net-

通过实验可以证明经过改进的 VGG16_2 模型对猴子图像的分类具有更好的性能。

4 结论

本文基于迁移学习实现了对 10 种不同猴子图像的识别,并为了适应该数据集使模型达到更好的识别效果,对传统 VGG16 模型进行了改进,分别从激活函数、损失函数和优化器的角度入手,生成 VGG16_2 模型。结合各种已有模型,以及在各种不同数据集上的测试对比结果,选取较该数据集来说性能最优的 Swish 激活函数、center loss 与 softmax loss 相结合的损失函数、Adam 优化器对模型进行完善。

works[J]. Medical Image Analysis,2019,58:10-12.

- [11] 崔义新. 基于交叉熵的随机赋权网络[D]. 保定:河北大学,2017.
- [12] 杨青,陆安江,高进. 基于深度学习的 Center Loss 算法研究[J]. 通信技术,2019,52(8):1878-1883.
- [13] 刘华. 深度卷积网络中的自适应激活函数研究[D]. 广州:华南理工大学,2018.
- [14] 米硕,田丰收,孙瑞彬,等. Swish 激活函数在中小规模数据集上的性能表现[J]. 科技创新与应用,2018(1):4-5.
- [15] 杨观赐,杨静,李少波,等. 基于 Dropout 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(7):122-127.
- [16] 常子汉. 基于小波变换与 Adam 优化的 LSTM 电价预测研究[D]. 兰州:兰州大学,2019.
- [17] 苏越. 基于深度学习卷积神经网络图像识别技术的研究分析探讨[J]. 信息通信,2019(7):7-8.
- [18] 高子翔. 基于多尺度池化的卷积神经网络图像分类方法研究[D]. 包头:内蒙古科技大学,2019.
- [19] 王曼. 轻量级卷积网络的构建及其在图像分类中的应用[D]. 西安:西安理工大学,2019.

(收稿日期:2020-03-30)

作者简介:

田加鹭(1996-),女,硕士研究生,主要研究方向:人工智能。

邓立国(1971-),男,博士,副教授,主要研究方向:人工智能、数据库。

版权声明

经作者授权，本论文版权和信息网络传播权归属于《信息技术与网络安全》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《信息技术与网络安全》编辑部
中国电子信息产业集团有限公司第六研究所