

# 基于改进的深度残差网络的图像识别

汤 凯,何 庆,赵 群,王 旭

(贵州大学大数据与信息工程学院,贵州省公共大数据重点实验室,贵州 贵阳 550025)

[摘要] 随着大数据时代的发展,深度学习也渐渐变得更加实用,引领人工智能时代的发展.卷积神经网络在图像领域中发挥着非常重要的作用,是深度学习模型中重要组成部分之一.图像识别的关键攻破点在于如何提取图像的有效特征,从而有效地解决图像识别问题.针对这一难点,本文主要在残差网络(ResNet)的基础上引入空间变换网络.空间变换网络可以有效地提取目标区域特征,提高图像识别效率.同时由于 Softmax 分类器提取的特征区分并不明显,甚至存在类内间距大于类间间距弊端.但在图像识别任务中期望特征不仅可分,而且要求类间分别提取的特征区分差异大.针对这一问题,本文在软最大值(Softmax)分类器中引入中心损失函数(Center Loss).Center Loss 损失函数能够使得提取的特征类间距离大,类内距离小,从而提高提取的特征识别度.在公开的 CIFAR10 数据集上,该模型取得了不错的性能,识别准确率达到 89%.相同实验条件下,相对于未改善的残差网络模型,本文提出的模型在公开的 CIFAR10 数据集识别正确率提高了 6%.

[关键词] 图像分类,深度学习,深度残差网络,空间变换网络

[中图分类号] TP183 [文献标志码] A [文章编号] 1001-4616(2019)03-0115-07

## Image Recognition Based on Improved Deep Neural Network

Tang Kai, He Qing, Zhao Qun, Wang Xu

(College of Big Data and Information Engineering, Guizhou University, Guizhou Provincial Key Laboratory of Public Big Data, Guiyang 550025, China)

**Abstract:** With the development of big data era, deep learning has gradually become more practical, leading the development of the era of artificial intelligence. Convolution neural network plays a very important role in image recognition, and it is one of the important components of deep learning model. The key point of image recognition is how to extract the effective features of the image, so as to effectively solve the problem of image recognition. In view of this difficulty, the main work of this paper is to introduce spatial transformation network on the basis of residual network(ResNet). The spatial transformation network can effectively extract the region of interest and improve the efficiency of image recognition. At the same time for the feature extracted by Softmax classifier is not good. In many cases, the intra-class spacing is even larger than the inter-class spacing, but in the image recognition task, the expected features are not only divisible, and with require great differences. In order to solve this problem, this paper introduces the Center Loss function into the Softmax classifier. Center Loss function can make the distance between the extracted feature classes larger and the intra-class distance smaller, thus improving the recognition degree of the extracted features. In the open CIFAR10 dataset, the model has achieved good performance, and the correct recognition rate is up to 89%. Under the same experimental conditions, compared with the unmodified residual network model, the proposed model improves the recognition accuracy of open CIFAR10 dataset by 6%.

**Key words:** image classificaiton, deep learning, deep residual network, spatial transformation network

随着大数据时代的发展,深度学习也渐渐变得更加实用,引领人工智能时代的发展.在 1989 年,Le Cun 等人提出了卷积神经网络<sup>[1]</sup>,目前卷积神经网络被广泛使用,例如自动驾驶、人脸识别和语音识别.但深层的网络容易陷入梯度消失和梯度爆炸,反而会影响网络的性能.针对此问题,2015 年何凯明等人提出<sup>[2]</sup>深度残差网络模型,该模型更容易优化,关键通过解决增加深度带来的梯度弥散问题<sup>[2]</sup>,从而能

收稿日期:2019-07-05.

基金项目:块数据中多源异构数据关联识别理论模型研究、贵州省公共大数据重点实验室开放课题(2017BDKFJJ034).

通讯联系人:王旭,博士,副教授,研究方向:大数据应用、人工智能、量子通讯. E-mail: xuwang@ gzu.edu.cn

够通过增加网络深度,来提高网络性能. 刘万军等人<sup>[3]</sup>提出自适应增强卷积神经网络算法,在图像识别中有效地提高了收敛速率和图像识别准确率. 曾维亮等人<sup>[4]</sup>提出在卷积神经网络隐藏层中引入随机失活函数(Dropout),有效地避免了过拟合现象,减少网络计算复杂度,但采用 Dropout 随机失活方法在每次迭代的过程中会随机舍弃信息,容易造成重要信息的丢失,从而影响网络的性能. 但是由于图像分辨率低,图像正负样本数量不均衡,图像类间相似度高、类内差异性大等因素使得图像识别领域极具挑战性.

为进一步提升卷积神经网络对于图像分类的准确率,在残差网络模型的基础上引入空间变换网络,空间变换网络具有注意力机制,可以定位图像的感兴趣区域,获取有效特征,提高图像分类准确度. 同时在 Softmax 分类器中引入 Center-Loss 损失函数,Center-Loss 损失函数主要体现了“类间距离大,类内距离小”思想,使得获取的图像特征能够更有效地提高图像分类准确度. 并通过实验证明,本文提出的网络模型能够有效地提高图像分类准确度,识别准确率在基础网络模型上提高了 6%.

1 卷积神经网络

1.1 卷积层

卷积层是卷积核以类似滑动窗口的方式通过权重共享实现线性变换,同时保持平移不变性,缩减参数量<sup>[5-7]</sup>. 在每个卷积层后都会得到一个特征图, $n$  个卷积核得到  $n$  个特征图. 卷积层的计算公式如式(1)所示:

$$x^{(l+1)} = f(w^{(l+1)} x^{(l)} + b^{(l+1)}).$$
 (1)

式中, $l=1,2,\cdots,L$  表示网络层数; $w$  为共享卷积核的参数, $b$  为每一层的偏置参数, $f(\cdot)$  表示卷积层激活函数.

1.2 池化层

池化层也称为下采样,是图像特征的进一步提取,从而简化卷积层的输出,同时增强了解释性,保持平移不变性和避免丢失过多信息,一般都在卷积层得到的特征图上进行池化操作<sup>[8-11]</sup>. 在卷积神经网络中一般包括最大池化和均值池化,本文采用的是最大池化. 最大池化就是在得到的特征图上每个  $k * k$  ( $k$  取决于池化层参数)的区域内取最大值,特征图经过池化层则会特征维数降低,参数计算量减少. 池化层的计算公式如(2)所示:

$$x^{(l+1)} = f(pool(x^{(l)}) + b^{(l+1)}).$$
 (2)

式中, $l=1,2,\cdots,L$  表示网络层数; $pool(\cdot)$  表示池化函数; $f(\cdot)$  表示池化层激活函数; $b$  表示池化层的偏置参数.

1.3 全连接层

全连接层的作用就是将每个局部的特征全部结合起来得到全局的特征<sup>[13]</sup>. 全连接层通过聚合各个卷积层得到的特征来实现分类. 全连接层能够有效地提高模型拟合能力,但由于其特征数特别多,从而导致模型计算量较大,使得效率降低.

基本卷积神经网络模型如图 1 所示.

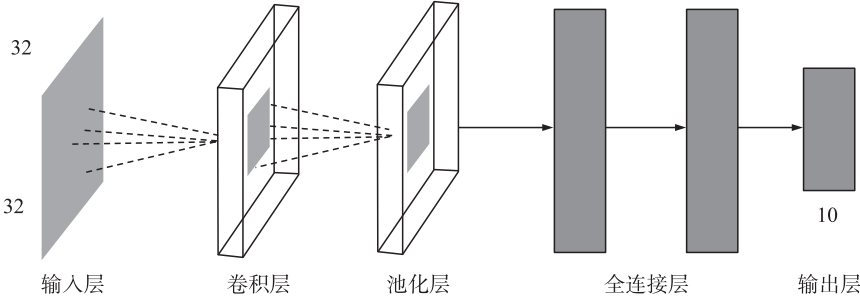


图 1 卷积神经网络基本模型  
Fig. 1 The basic model of convolution neural network

2 残差网络(ResNet)模型及其改进

2.1 ResNet-18 网络

ResNet 模型的初衷是为了解决单纯地把网络叠起来的深层网络的效果反而不如合适层数的较浅的

网络效果这一问题. 其关键之处在于普通网络的基础上增加了一个残差学习单元(residual unit), 构成一个残差模块(residual block). 假设某段神经网络的输入为  $x$ , 期望输出为  $H(x)$ , 此时学习的目标为  $F(x) = H(x) - x$ . ResNet 的残差学习模块如图 2 所示.

残差学习模块在反向传播时的梯度计算如式(3)所示:

$$\frac{\partial H(x)}{\partial x} = \frac{\partial F(x)}{\partial x} + 1. \quad (3)$$

在一般的网络模型中, 随着网络深度的增加, 在反向传播的过程中可能由于某一个导数值非常小或非常大, 在导数连乘后的梯度值则会越来越小或越来越大, 最终导致梯度弥散问题. 在残差网络模块中由于引入了残差学习模块, 其梯度计算表达式如(3)所示. 在其梯度计算中每一个导数值加 1, 从而能够保证网络有效地反向传播, 所以在一定程度上解决了梯度弥散问题.

ResNet-18 网络有 18 层网络, 其中包含 4 个残差模块, 每个残差模块都有 3 个卷积层, 每个卷积层后面都有一个 BatchNorm 层, 目的是为了降低网络对初始化权重不敏感, 同时一定程度可以避免过拟合问题.

## 2.2 ResNet-18 网络改进

### 2.2.1 空间变换网络

空间变换网络嵌入了注意力机制模块, 该模块可以根据任务本身的空间变换参数对输入的图片或空间中的学习特征进行对齐, 无需标注关键点, 从而减少物体由于空间中的旋转、平移、尺度、扭曲等几何变换对分类、定位等任务的影响, 同时可以定位图像的感兴趣区域, 获取有效特征, 提高图像分类准确度<sup>[14]</sup>. 空间变换网络由网络定位、网格生成和采样 3 个模块组成, 其结构如图 3 所示.

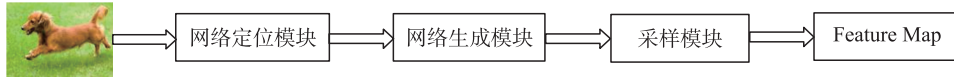


图 3 空间变换网络结构

Fig. 3 The spatial transformation network structure

#### (1) 网络定位模块

输入的特征图的维度为  $U \in R^{H \times W \times C}$ ,  $H$  为特征图的高度,  $W$  为特征图的宽度,  $C$  为通道数, 输出的结果为  $\theta$ ,  $\theta = f_{loc}(U)$ , 其中  $U$  为输入特征图,  $f_{loc}(\cdot)$  为网络定位模块. 本文中网络定位模块包含 4 个卷积层, 2 个全连接层, 其中  $\theta$  是一个 6 维的向量.

#### (2) 网格生成模块

网格生成模块根据预测的变换参数构造采样网格, 它是输入图像中一组点采样变换后的输出. 生成的图像的大小必须与输入图像的大小相同. 网格生成器其实得到的是一种映射关系  $\tau_\theta(G_i)$ , 假设网格生成模块的输入的图像每个像素点的坐标为  $\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix}$ , 输出的图像每个像素点的坐标为  $\begin{pmatrix} x_i^t \\ y_i^t \end{pmatrix}$ , 其中  $\tau_\theta(G_i)$  表示一种二维映射函数, 其仿射变换表达式如(4)所示:

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \tau_\theta(G_i) = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}. \quad (4)$$

#### (3) 采样模块

采样模块就是利用采样网格和输入的特征图同时作为输入产生输出, 这样就能通过采样核对输入特征图中的像素进行采样, 将采样的像素值放入输出特征图中对应的坐标位置. 具体表示如式(5)所示:

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c k(x_i^s - m; \Phi_x) k(y_i^s - n; \Phi_y), \forall i \in [1 \cdots H'W'], \quad \forall c \in [1 \cdots C]. \quad (5)$$

式中,  $V_i^c$  表示输出的特征图,  $U_{nm}^c$  表示输入特征图中坐标为  $(n, m)$  的像素值,  $k(\cdot)$  表示采样函数, 本文通过插值方式来获取输出的特征图每个像素点.

本文对于 ResNet-18 引入空间变换网络, 使得新的网络模型既能够避免深层网络带来的梯度弥散问题的同时, 还能够根据任务本身学习图片或特征的空间变换参数, 对空间中的输入图像或学习特征进行对齐, 从而减少物体由于空间中的平移、旋转、扭曲、尺度等几何变换对分类、定位等任务的影响, 同时嵌入了注意力机制模块, 能够定位到图像目标区域, 得到有效的特征图, 从而提高图像分类准确度. 本文是将空间变换网络作为整个网络的第一部分, 从而直接对输入进行仿射变换. 得到后的图像作为 ResNet 的输入, 经过 ResNet 得到的输出特征向量通过分类器得到最终分类结果, 本文的模型如图 4 所示.

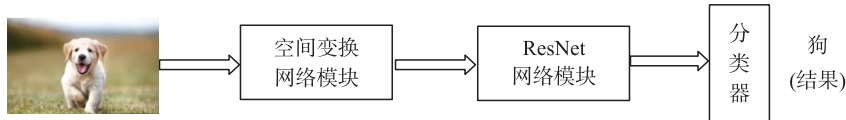


图 4 改进的网络模型结构  
Fig. 4 The improved network model structure

2.2.2 Center-Loss 损失函数

Center-Loss 函数的核心思想为“类间距离大, 类内距离小”, 更好地利用特征来高效地进行图像分类<sup>[15]</sup>. 其表达式如(6)所示:

$$L_c = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2. \quad (6)$$

式中,  $c_{y_i}$  表示第  $y_i$  个类别的特征中心,  $x_i$  表示全连接层之前的特征,  $m$  表示 mini-batch 的大小.

从聚类角度看, ResNet-18 网络中由于 Softmax 分类器提取的特征区分并不明显, 甚至存在类内间距大于类间间距弊端. 期望模型提取的特征不仅具有区分度, 而且类间差异大、类内差异小, 从而保证提取的特征具有很好的识别度. 所以本文将 Center-Loss 与 Softmax-Loss 结合得到一个新的损失函数, 其表达式如(7)所示:

$$L = - \sum_{i=1}^m \log \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} + \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2. \quad (7)$$

式中, 第一项为 Softmax-Loss 损失函数表达式, 第二项为 Center-Loss 损失函数表达式, 通过两个损失函数的结合, 能够有效地利用网络提取的特征来对图像进行分类.

3 实验结果及分析

3.1 数据集

本文实验采用的是公开的 CIFAR10 数据集, CIFAR10 数据集共有 60 000 张彩色图像, 其中训练集包含 50 000 张图像, 测试集包含 10 000 张图像. 每张图像的大小为  $32 \times 32$ . 此数据集包含 airplane、automobile、bird、cat、deer、dog、frog、horse、ship 和 truck 10 个类别, 每类 6 000 张图像. 部分数据集如图 5 所示.

3.2 实验内容及结果分析

3.2.1 实验环境配置

本次实验是在台式计算机运行, 其配置为 i7-4790 的 CPU, GeForce GTX 750 Ti 的 GPU, 使用 Python 语言, Pytorch 深度学习框架平台.

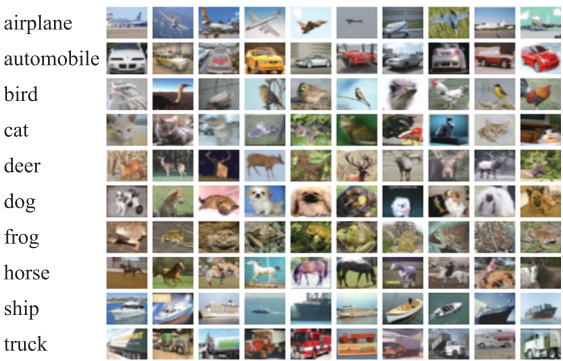


图 5 CIFAR10 数据集  
Fig. 5 The CIFAR10 dataset

### 3.2.2 实验参数设置

实验中初始学习率设置为 0.01,同时学习率使用指数衰减方式依次迭代递减. 优化器采用的是 SGD 随机梯度下降法,其中动量设置为 0.9,并采用 L2 正则化来实现权重衰减,目的是为了避免网络陷入过拟合.

### 3.2.3 实验结果及分析

#### (1) ResNet-18 网络模型实验结果

其轮次 (Epoch) 设置为 300, Batch size 设置为 128, 训练集损失迭代、训练集准确率和测试集准确率如图 6、7 和 8 所示.

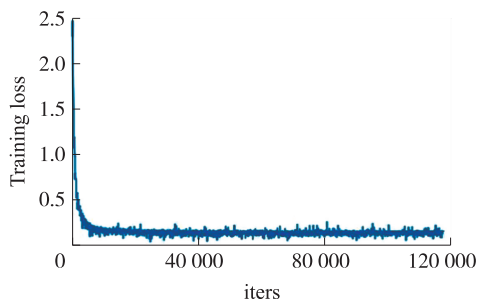


图 6 训练集损失迭代

Fig. 6 The training set loss iteration

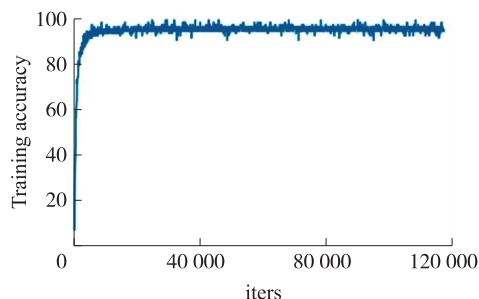


图 7 训练集准确率

Fig. 7 The training set accuracy

#### (2) ResNet-18 网络+Center-Loss 损失函数

其轮次 (Epoch) 设置为 150, Batch Size 设置为 1024, 训练集损失迭代、训练集准确率和测试集准确率如图 9、10 和 11 所示.

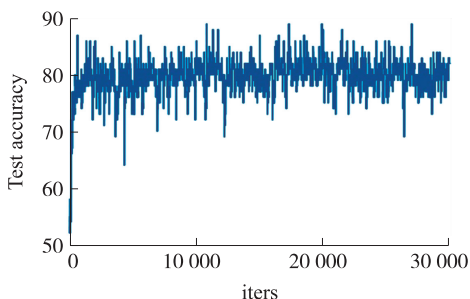


图 8 测试集准确率

Fig. 8 The test set accuracy

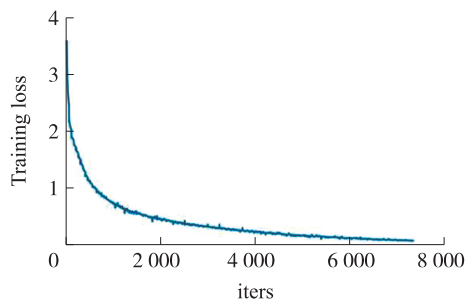


图 9 训练集损失迭代

Fig. 9 The training set loss iteration

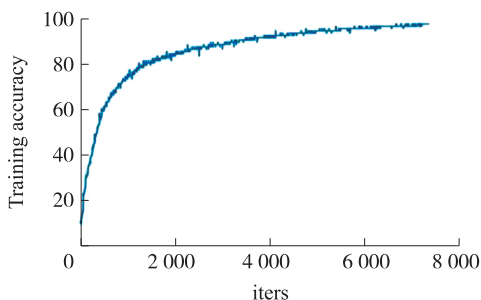


图 10 训练集准确率

Fig. 10 The training set accuracy

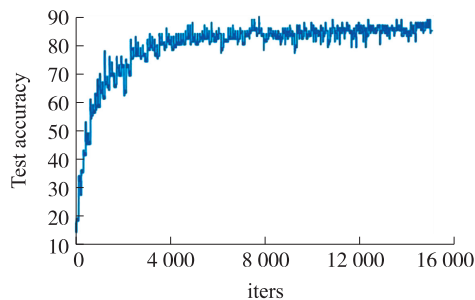


图 11 测试集准确率

Fig. 11 The test set accuracy

#### (3) ResNet-18 网络+空间变换网络+Center-Loss 损失函数

其轮次 (Epoch) 设置为 250, Batch size 设置为 128, 训练集损失迭代、训练集准确率和测试集准确率如图 12、13 和 14 所示.

通过对比图 8 和图 14, 可以看出本文提出的模型在测试集准确率迭代过程中波动明显少于 ResNet-18 网络模型, 从而说明本文提出的模型有效提高了网络收敛的稳定性.

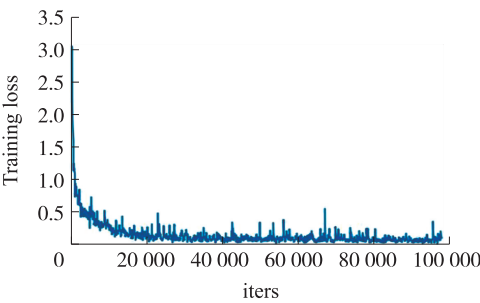


图 12 训练集损失迭代

Fig. 12 The training set loss iteration

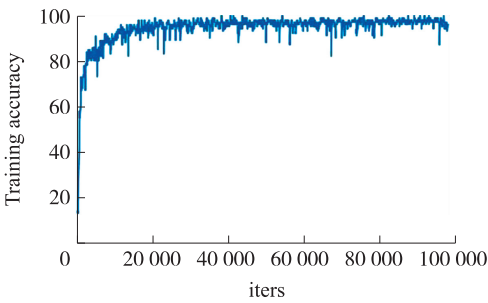


图 13 训练集准确率

Fig. 13 The training set accuracy

实验结果对比如表 1 所示.

表 1 实验结果对比

Table 1 Comparison of experimental results

方法	正确率(最高)/%
ResNet18 网络	83
ResNet18 网络+Center-Loss 损失函数	87
ResNet18 网络+空间变换网络+Center-Loss 损失函数	89
改进激活函数的卷积神经网络 <sup>[16]</sup>	78.59

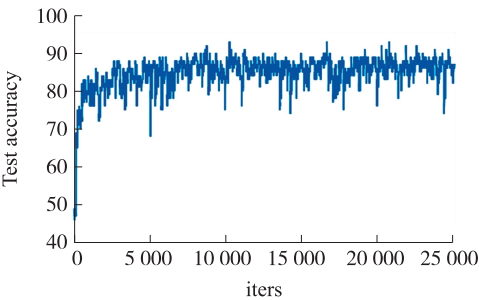


图 14 测试集准确率

Fig. 14 The test set accuracy

通过表 1 可知,只用 ResNet-18 网络做图像分类测试集最高准确率为 83%,“ResNet-18 网络+Center-Loss 损失函数”网络模型测试集最高准确率为 87%，“ResNet-18 网络+空间变换网络+Center-Loss 损失函数”网络模型在公开 CIFAR10 数据集图像分类的测试集准确率最高达到了 89%,曲之琳等人<sup>[16]</sup>提出设计一个新的激活函数嵌入到卷积神经网络模型,其模型在测试集准确率为 78.59%. 通过实验数据表明,本文提出的方法能够有效地提高图像分类的准确率,并且在 ResNet 网络的基础上提高了 6%,相对于文献[16]提高了 10.41%.

4 总结

本文提出将空间变换网络嵌入到 ResNet-18 网络中构造一个新的网络模型,通过实验表明本文提出的新的网络模型能够有效避免由于深度网络带来的梯度弥散问题,根据任务本身学习图片或特征的空间变换参数,对空间中的输入图像或学习特征进行对齐,从而减少物体由于空间中的平移、旋转、扭曲、尺度等几何变换对分类、定位等任务的影响,嵌入了注意力机制模块,能够定位到图像目标区域,得到有效的特征图,从而提高图像分类准确度. 其次,本文在 Softmax 分类器中引入 Center-Loss 损失函数,解决 Softmax 分类器提取的特征区分并不明显,甚至存在类内间距大于类间间距弊端. 本文的不足之处在于尚未解决只引入空间变换网络模型未收敛问题,从而影响了只引入空间变换网络模型最终效果. 但同时引入空间变换网络和 Center-Loss 损失函数增强了只引入 Center-Loss 模型的效果,从而说明空间变换网络在一定程度上可以提高图像分类准确率. 通过与 ResNet-18 网络模型对比,本文提出的 ResNet-18 网络+空间变换网络+Center-Loss 损失函数模型在公开 CIFAR10 数据集图像分类准确率提高了 6%.

[参考文献]

[1] LECUN Y,BOTTOU L,BENGIO Y,et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE,1998,86(11):2278-2324.

[2] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision & Pattern Recognition. USA:IEEE,2016.

[3] 刘万军,梁雪剑,曲海成. 自适应增强卷积神经网络图像识别[J]. 中国图象图形学报,2017,22(12):1723-1736.

[4] 曾维亮,林志贤,陈永洒. 基于卷积神经网络的智能冰箱果蔬图像识别的研究[J]. 微型机与应用,2017,36(8):56-59.

- [ 5 ] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[ J ]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39( 6 ): 1137–1149.
- [ 6 ] LIU M, SHI J, LI Z, et al. Towards better analysis of deep convolutional neural networks[ J ]. IEEE transactions on visualization and computer graphics, 2017, 23( 1 ): 91–100.
- [ 7 ] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of go without human knowledge[ J ]. Nature, 2017, 550( 7676 ): 354.
- [ 8 ] ZHU P, ISAACS J, BO F, et al. Deep learning feature extraction for target recognition and classification in underwater sonar images[ C ]//2017 IEEE 56th Annual Conference on Decision and Control( CDC ). Melbourne, Australia: IEEE, 2018.
- [ 9 ] WEI W, ZHAO M, WANG J. Effective android malware detection with a hybrid model based on deep autoencoder and convolutional neural network[ J ]. Journal of ambient intelligence & humanized computing, 2018( 1 ): 1–9.
- [ 10 ] 刘晨, 曲长文, 周强, 等. 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( 3 ): 38–42.
- [ 11 ] WANG Y, FEI L, ZHANG K, et al. LFNet: a novel bidirectional recurrent convolutional neural network for light-field image super-resolution[ J ]. IEEE transactions on image processing, 2018, 27( 9 ): 19–26.
- [ 12 ] CHEN B, ZHANG Y, XIN P. An effective time-domain microwave image reconstruction algorithm for loss-y layered media utilizing the ADI-FDTD method[ J ]. Journal of China universities of posts & telecommunications, 2018, 25( 2 ): 93–99.
- [ 13 ] MALLAHI M E, ZOUHRI A, QJIDAA H. Radial meixner moment invariants for 2D and 3D image recognition[ J ]. International journal of automation & computing, 2018, 28( 2 ): 207–216.
- [ 14 ] ARCOS G A Ã, ÃLVAREZ G A J A, SORIA M L M. Deep neural network for traffic sign recognition systems: an analysis of spatial transformers and stochastic optimisation methods[ J ]. Neural Netw, 2018, 99( 12 ): 158–165.
- [ 15 ] JIN X B, XIE G S, HUANG K, et al. Discriminant zero-shot learning with center loss[ J ]. Cognitive computation, 2019( 7 ): 1–10.
- [ 16 ] 曲之琳, 胡晓飞. 基于改进激活函数的卷积神经网络研究[ J ]. 计算机技术与发展, 2017( 12 ): 83–86.

[ 责任编辑: 黄 敏 ]