



引用格式:徐静萍,王芳. 基于改进的 S-ReLU 激活函数的图像分类方法[J]. 科学技术与工程, 2022, 22(29): 12963-12968.

Xu Jingping, Wang Fang. Image classification method based on improved S-ReLU activation function[J]. Science Technology and Engineering, 2022, 22(29): 12963-12968.

基于改进的 S-ReLU 激活函数的图像分类方法

徐静萍, 王芳*

(燕山大学理学院, 秦皇岛 066004)

摘 要 为解决 ReLU 函数负区域取值为 0 而引发的对应权重无法更新的问题,提出了新的激活函数 S-ReLU。该函数在负区域具有软饱和性,增加了负样本数据的关注度。通过赋予负区域输出值较小的导数,促进了负输入值的反向传播,提高了模型的鲁棒性。通过与其他常见激活函数在数据集 MNIST、CIFAR-10 上使用 LeNet-5 模型的对比实验,探究基于 S-ReLU 激活函数的图像分类效果。实验结果表明,对于 MNIST 和 CIFAR-10 数据集,相比使用其他激活函数,S-ReLU 函数提高了模型分类精度。

关键词 激活函数; 图像分类; 卷积神经网络; 特征提取

中图分类号 TP391.41; **文献标志码** A

Image Classification Method Based on Improved S-ReLU Activation Function

XU Jing-ping, WANG Fang*

(School of Science, Yanshan University, Qinhuangdao 066004, China)

[Abstract] In order to solve the problem that the corresponding weight cannot be updated due to the value of 0 in the negative region of ReLU function, a new activation function S-ReLU was proposed. It has soft saturation in the negative region, which increases the attention of negative sample data. By giving a smaller derivative to the output value in the negative region, the back propagation of negative input value was promoted, the robustness of the model was improved. Through the comparative experiment of using LeNet-5 model on MNIST and CIFAR-10 datasets with other common activation functions, the image classification effect based on S-ReLU activation function was explored. The experimental results show that for MNIST and CIFAR-10 datasets, using S-ReLU function improves the classification accuracy of the model compared with other activation functions.

[Keywords] activation function; image classification; convolutional neural network; feature extraction

图像分类主要分为细粒度图像分类^[1]和图像分类两个方向。细粒度图像主要针对同一个大类下的不同子类,它们类间差异小而类内差异大,而图像分类是针对不同物种进行的识别,物种的类间差异性大,相对更容易些。对于图像分类任务,如何高效地提取物种间具有区分性区域的特征成为学术界讨论最为热烈的问题。在深度学习(deep learning, DL)^[2]没有提出之前,人们进行图像分类任务通常依靠的是人工设计的特征提取器,这需要研究者根据分类任务针对性的设计特征描述,人为设计局部不变性的特征描述需要大量的专业知识,这严重影响了模型的泛化能力。相较于传统手段图像识别,深度学习的出现使得图像分类的精度得到迅猛提升。随着人们对 DL 理解的加深,生成对

抗性网络^[3]、注意力机制^[4]、胶囊网络^[5]以及迁移学习^[6]等一系列优秀的深度学习模型被引入,极大地拓展了图像分类任务的手段,具有分类效果好、模型精度高、应用范围广泛等诸多优点。

卷积神经网络(convolutional neural network, CNN)的发展,使得人们在计算机视觉任务中取得了巨大进展。在 2012 年的 ImageNet 挑战赛中,He 等^[7]提出的 CNN 模型的识别性能首次超越了人类视觉能力,错误率仅有 4.94%。激活函数是 CNN 模型性能迅猛提升的重要原因,它使模型具备了非线性性质,增加了模型的复杂度,解决了以线性函数为激活函数时不能处理线性不可分问题的难题,极大地提升了模型处理问题的能力。

Sigmoid^[8]和 Tanh^[9]是最早应用于 CNN 的非线

收稿日期: 2022-01-11; 修订日期: 2022-07-08

基金项目: 国家自然科学基金(62073234); 河北省自然科学基金(F2020203105)

第一作者: 徐静萍(1997—),女,汉族,河南南阳人,硕士研究生。研究方向:图像分类。E-mail:xjp1631@163.com。

*通信作者: 王芳(1984—),女,汉族,安徽淮北人,博士,副教授。研究方向:复杂非线性系统控制。E-mail:wangfang@ysu.edu.cn。

性饱和和激活函数。但是 Sigmoid 和 Tanh 都存在软饱和和问题,即当参数在训练过程中处于饱和区时,会出现梯度消失的现象。基于以上问题,Glort 等^[10]提出了 ReLU 函数,它是非饱和和激活函数,可以缓解饱和函数存在梯度消失的问题。由于 ReLU 将所有负值取为 0,导致模型在训练过程中很容易出现“神经元死亡”^[11],从而致使模型分类效果下降。SoftPlus^[12]被认为是 ReLU 的平滑近似,避免了 ReLU 的“神经元死亡”的问题,但也导致了“均值偏移”的问题。Maas 等^[13]提出给所有样本负区域的值一个非零的斜率。Clevert 等^[14]解决了“均值偏移”的问题,提出了 ELU,但也引入了和 LeakyReLU 类似的问题——需要人为设定超参数。He 等^[7]则建议将 ReLU 负值的斜率设为一个可学习的参数,解决了人工调参的难题,但由于函数两部分都是线性的,模型的非线性能力下降。Misra^[15]提出了一个几乎处处光滑的非单调可导激活函数 Mish,提升了模型的分类能力,同时也增加了模型的计算量。

针对目前的激活函数存在的问题,现将 ReLU 和 SoftPlus 组合优化,构造新的激活函数 S-ReLU。在 MNIST 和 CIFAR-10 数据集上,利用 S-ReLU 与其他激活函数在 LeNet-5 模型下进行实验,以对比分析不同激活函数分类效果。

1 相关工作

主要介绍了非饱和的非线性激活函数 ReLU 和 SoftPlus,并分析了两者的优缺点。

1.1 ReLU 激活函数

Glort 等^[10]受到神经科学中人类脑神经元接受刺激后的表现的启发,首次将 ReLU 应用于神经网络中,通过让 CNN 模拟生物神经网络的工作流程,提升了模型的表达能力。ReLU 函数的形式为

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (1)$$

ReLU 函数与饱和激活函数 Sigmoid 和 Tanh 相比,将模型的负区域取值归为 0,使 CNN 具有了稀疏性,从而有效地解决了梯度消失的问题。但是,当训练阶段的学习率很大时,会出现“神经元死亡”现象,其图像如图 1 所示。

由图 1 可知,当 $x \leq 0$ 时,其输出值与导数恒为 0,一方面使模型输出具有了随机性,提升了模型的泛化能力;另一方面,使带有负样本信息的神经元无法工作,既造成了样本信息的缺失,又因其梯度常为 0,导致权重无法更新。当 $x > 0$ 时,输出值随着 x 的增大而增大,其导数为 1,数值小,计算简单,加快了模型的收敛速度。

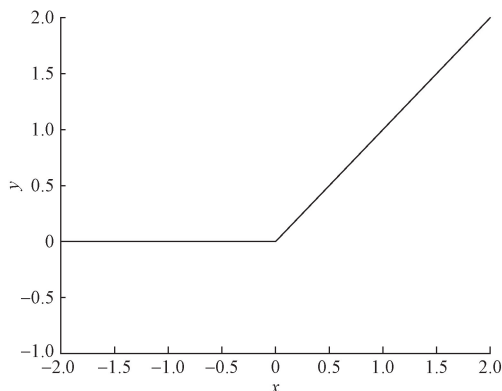


图 1 ReLU 函数

Fig. 1 ReLU function

1.2 SoftPlus 激活函数

SoftPlus 被认为是 ReLU 的平滑近似,它们具有许多相似性,SoftPlus 的表达式为

$$f(x) = \ln(1 + e^x) \quad (2)$$

SoftPlus 的图像如图 2 所示,由图 1 可知,当 $x < 0$ 时,ReLU 直接让所有负值等于 0,由图 2 可知,SoftPlus 则使用一种更加平滑的方式减少负区域数值的输出,即随着 x 的增大,结果不断减小直至趋于 0,避免了 ReLU 强制为 0 时导致的“神经元死亡”的问题,从而使网络模型变得更加稳定。

另外,由图 2 可知,SoftPlus 的输出范围为 $(0, +\infty)$,当 $x \rightarrow +\infty$ 时,其导数为 1,当 $x \rightarrow -\infty$ 时,其导数为 0,既具备了 ReLU 计算简单、收敛速度快的优点,又促进了负样本权重的更新。但当 $x < 0$ 时,其输出值恒为正值,导致了均值偏移。

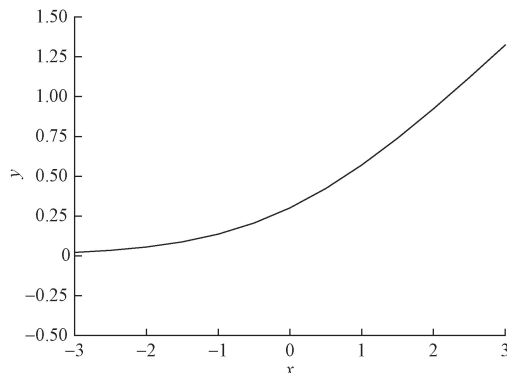


图 2 SoftPlus 函数

Fig. 2 SoftPlus function

2 S-ReLU 激活函数

基于 ReLU 函数和 SoftPlus 函数构造了一种新的函数 S-ReLU,分析了它的性质,并将其与 ReLU 和 SoftPlus 进行对比分析,体现了 S-ReLU 的优势,然后给出了 S-ReLU 如何在模型中进行正向传播与误差优化。

2.1 S-ReLU 函数

基于式(1)和式(2)所示两类激活函数构造如下的 S-ReLU 激活函数为

$$f(x) = \begin{cases} x, & x > 0 \\ x[e^x + x\ln(1 + e^x)], & x \leq 0 \end{cases} \quad (3)$$

从式(3)可知,当 $x > 0$ 时,输出结果是恒等函数,即输入等于输出,而当 $x \leq 0$ 时,输出结果是由 ReLU 和 SoftPlus 构成。

S-ReLU 激活函数的导数为

$$f'(x) = \begin{cases} 1, & x > 0 \\ e^x + x\left(e^x + 2\alpha + \frac{xe^x}{1 + e^x}\right), & x \leq 0 \end{cases} \quad (4)$$

式(4)中: $\alpha = \ln(1 + e^x)$ 。由 $f'_+(0) = f'_-(0) = 1$ 可知, $f(x)$ 是可微的,且从图 3 可以明显看到其导数并不是常数,因此, $f(x)$ 具有非线性,这将极大提高模型处理复杂问题的能力,增强模型的泛化能力。

首先,与 ReLU 相比, S-ReLU 增加了负样本信息,与 SoftPlus 在负区域取值总是恒大于 0 相比, S-ReLU 取值有正有负,可以有效减缓“均值偏移”问题。与同是增加负样本信息利用率的 ELU 函数相比, S-ReLU 更加贴合生物学上神经元响应的特点,提升了模型的抗干扰能力。

其次,当 $x \rightarrow +\infty$ 时, $f(x) \rightarrow \infty$, 函数没有上界,可以加快模型的收敛速度和避免函数饱和后训练速度的急剧下降;当 $x \rightarrow -\infty$ 时, $f(x) \rightarrow 0$, 说明函数是有下界的,使模型具有较强的正则性,增强了模型应对复杂问题的表达能力。

再次, S-ReLU 函数由指数函数和对数函数构成,与只有恒等函数构成的函数[式(1)]相比更加复杂,使得 S-ReLU 函数在实际应用中需要更多的时间,但 S-ReLU 函数在正向传播过程中可以返回一个有效的负值,防止了梯度消失。

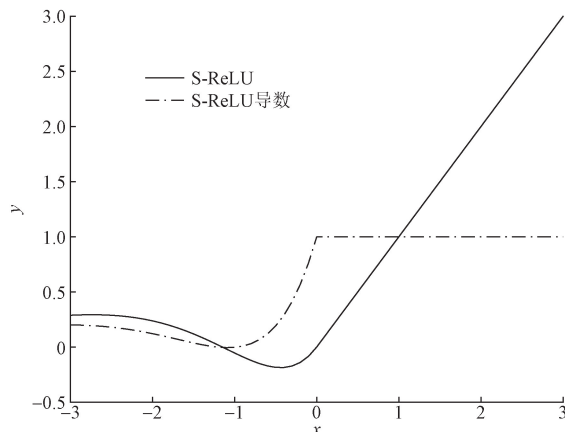


图3 S-ReLU 函数及导数

Fig. 3 S-ReLU function and derivative

S-ReLU 函数与 ReLU 函数的结构极度相似,使得 S-ReLU 函数也具备了 ReLU 函数的一些优点,如稀疏性,不仅可以减少模型的计算量,还能提升模型的表达能力,使网络更加关注任务本身。

2.2 基于 S-ReLU 的正向传播

假定每个神经元的输入值为 z , 经过激活函数后的输出结果为 a , 激活函数为 σ , 则

$$\begin{cases} z = wx + b \\ a = s(z) \end{cases} \quad (5)$$

现取 CNN 中的第 $i-1$ 、 i 和第 $i+1$ 层网络结构,设神经网络的第 $i-1$ 层的第 k 个神经元到第 i 层的第 j 个神经元的权重为 w_{ik}^i ,第 i 层的神经元到第 j 个神经元的偏置为 b_j^i ,第 i 层第 j 个神经元的输入为 z_j^i ,输出结果为 a_j^i ,有

$$a_j^i = s(z_j^i) = s\left[\sum_{k=1}^m (a_k^{i-1} w_{jk}^i + b_j^i)\right] \quad (6)$$

式(6)中: m 为第 $i-1$ 层的神经元个数。

当 σ 是 S-ReLU 函数时,有

$$a_j^i = \begin{cases} z_j^i, & z_j^i > 0 \\ z_j^i[e^{z_j^i} + z_j^i\ln(1 + e^{z_j^i})], & z_j^i \leq 0 \end{cases} \quad (7)$$

从式(7)可知, S-ReLU 与 ReLU 的不同之处是:在模型的正向传播过程中,输入值小于等于 0 时,输出结果不再是固定常数零,它使得网络在反向传播过程中产生了一个小的梯度流,促进了网络中权重矩和偏置更新,进而减小了模型的错误率。

此外,为了简化式(7),还可以用矩阵表示为

$$a^i = s(z^i) = s(a^{i-1}w^i + b^i) \quad (8)$$

2.3 权重和偏置的优化

神经网络在反向传播过程中利用梯度下降法^[16]更新模型中的权重 w 和偏置 b ,并以此减少模型的误差,在多分类问题中经常在模型最后一层使用 Softmax 函数作为激活函数,并用交叉熵损失函数作为 Loss 函数,公式为

$$L(y, \hat{y}) = - \sum_{k=1}^m y_k \ln a_k \quad (9)$$

式(9)中: y 为真实值; a_k 为预测结果。又因为在图像分类任务中常常会对数据标签进行 one-hot 处理,因此损失函数还可以化简为

$$L = - \ln a_k \quad (10)$$

接下来将说明如何使用交叉熵损失函数和梯度下降法更新权重 w 和偏置 b 。

假定 y_i 标签值为 1,其他标签值为 0,可得

$$L = - \ln a_i^l = - \ln \frac{e^{z_i^l}}{\sum_k e^{z_k^l}} = \ln \sum_k e^{z_k^l} - z_i^l \quad (11)$$

最后一层的激活函数是 Softmax,故为

$$\begin{cases} \frac{\partial L}{\partial z^l} = a^l - y \\ \frac{\partial L}{\partial w^l} = \frac{\partial L}{\partial z^l} \frac{\partial z^l}{\partial w^l} = (a^l - y) a^{l-1} \\ \frac{\partial L}{\partial b^l} = \frac{\partial L}{\partial z^l} \frac{\partial z^l}{\partial b^l} = a^l - y \end{cases} \quad (12)$$

以此类推,第 $l-1$ 层有

$$\begin{cases} \frac{\partial L}{\partial a^{l-1}} = \frac{\partial L}{\partial z^l} \frac{\partial z^l}{\partial a^{l-1}} = (a^l - y) w^l \\ \frac{\partial L}{\partial z^{l-1}} = \frac{\partial L}{\partial a^{l-1}} \frac{\partial a^{l-1}}{\partial z^{l-1}} = (a^l - y) w^l g'(z^{l-1}) \\ \frac{\partial L}{\partial w^{l-1}} = (a^l - y) w^l g'(z^{l-1}) a^{l-2} \\ \frac{\partial L}{\partial b^{l-1}} = (a^l - y) w^l g'(z^{l-1}) \end{cases} \quad (13)$$

式(13)中:

$$\begin{cases} g'(z^{l-1}) = \begin{cases} 1, & z^{l-1} > 0 \\ e^{z^{l-1}} + z^{l-1}(e^{z^{l-1}} + 2\beta + \gamma), & z^{l-1} \leq 0 \end{cases} \\ \beta = \ln(1 + e^{z^{l-1}}) \\ \gamma = \frac{z^{l-1} e^{z^{l-1}}}{1 + e^{z^{l-1}}} \end{cases} \quad (14)$$

而权重 w 和偏置 b 的更新方法为

$$\begin{cases} w^{l_new} = w^{l_old} - \eta \frac{\partial L}{\partial w^l} \\ b^{l_new} = b^{l_old} - \eta \frac{\partial L}{\partial b^l} \end{cases} \quad (15)$$

式(15)中: η 为学习率,是后期设置的超参数。CNN 通过多次正向、反向传播更新模型中的权重和偏置,从而缩小了精度误差,达到提升分类准确率的目的。

3 实验结果与分析

在 MNIST 和 CIFAR-10 数据集上将 S-ReLU 与其他激活函数应用 LeNet-5^[17] 模型进行对比实验,并分析不同激活函数的分类效果。

3.1 实验条件

实验是在硬件参数为 i5-7500 CPU, GTX960 GPU 的 tensorflow-gpu-1.1.0 深度学习框架下进行的。在模型开始训练之前,对所有的数据集进行了数据预处理,且对除 MNIST 数据集之外的数据集使用平移、水平翻转、缩放等数据增强技术^[18] 以扩充数据。出于节约实验时间和成本的考虑,使用 LeNet-5 深度学习模型,为了更客观地对比不同激活函数的性能,除改变模型的激活函数外,相同数据

集下的其他设置均一致。实验中,参与对比的激活函数为 Tanh、ReLU、SoftPlus、ELU、Mish、Swish^[19] 和 S-ReLU,所需参数设置如表 1 所示。

实验中采用的数据集是图像分类中最常用的数据集:MNIST^[17] 和 CIFAR-10^[20],其详细信息如表 2 所示。

表 1 实验超参数及指标设置

Table 1 Experimental hyperparameters and index settings

参数名称	参数值
优化器	梯度下降法
学习率初始值	0.001
权重衰减	0.000 5
动能	0.9
遍历次数(epoch)	两个数据集分别为 30, 50
批数据大小(batch size)	128
损失函数	交叉熵损失函数
评价指标	准确率

表 2 数据集信息

Table 2 Dataset information

数据集	类别数	训练集/幅	测试集/幅
MNIST	10	60 000	10 000
CIFAR-10	10	50 000	10 000

MNIST 数据集有 48 000 张训练图片、12 000 张验证图片和 10 000 张测试图片,每张图片是大小为 28×28 的灰度图,类别为 0~9 的 10 个数字。

CIFAR-10 是由大小为 32×32 的 60 000 张 RGB 图片组成,包含鸟类、汽车、飞机等 10 个不同类别属性的物种。与 MNIST 数据集相比,尽管 CIFAR-10 也是 10 个类别,但由于其图像包含了更多额外的信息,因此识别难度会更大。

3.2 实验分析

3.2.1 MNIST 数据集

MNIST 数据集选择 LeNet-5 网络作为基准框架,其中批数据大小为 100,遍历次数为 30 次。LeNet-5 框架下,各个激活函数的分类效果如表 3 所示。

由表 3 可知,在遍历次数为 30 的情况下,7 个激

表 3 LeNet-5 框架下的各个激活函数对比

Table 3 Comparison of activation functions under the LeNet-5 framework

激活函数	准确率/%	最低误差
Tanh	97.67	0.077 2
ELU	97.89	0.066 0
ReLU	97.97	0.069 1
Mish	97.93	0.063 3
Swish	97.87	0.070 1
SoftPlus	96.98	0.091 6
S-ReLU	98.16	0.058 5

活函数都达到了 96.9% 以上的准确率,其中 S-ReLU 函数表现最为突出,达到了 98.16% 的准确率,误差仅为 0.058 5。在相同的遍历次数下,与 SoftPlus 函数相比,准确率提升了 1.18%,误差降低了 0.0331。ReLU 函数达到了 97.97% 的准确率。Mish 与 Swish 都达到了 97.7% 以上。饱和激活函数 Tanh 的准确率为 97.67%,仅高于非饱和激活函数 SoftPlus 函数,准确率仅提高了 0.69%。因此,从以上的实验结果上可知,非饱和激活函数整体上要优于饱和激活函数。

另外,需要说明的是,由于 S-ReLU 函数同时含有指数和对数函数,表达式相对复杂,所以在实验中,在同样的遍历次数下,S-ReLU 的计算时间会比 ReLU 等函数更长些。

3.2.2 CIFAR-10 数据集

为了进一步探索不同激活函数在同一深度学习框架下的图像分类效果,将 7 个激活函数应用到 CIFAR-10 数据集,同样采用 LeNet-5 框架进行对比实验。

表 4 可知,构造的 S-ReLU 函数分类效果最好,其准确率为 63.64%,与 SoftPlus 相比,准确率提高了 8.13%。激活函数 Tanh 的准确率与非饱和激活函数 Swish 和 SoftPlus 相比分别提升了 1.18% 和 3.29%。Mish 的准确率为 59.90%,仅次于 S-ReLU 函数。与此同时,Swish 的准确率仅有 57.62%。与 MNIST 数据集上的分类结果相比,模型表达能力有所下降,说明同一个激活函数在不同数据集会有不同的表现,因此在以后的实验中可以通过对不同的数据集选择恰当的激活函数的方式提高模型的表达能力。对比表 3 和表 4 可以得到,同样是 10 个类别,CIFAR-10 数据集的准确率远低于 MNIST 数据集,这是由于 CIFAR-10 数据集是一个三通道的 RGB 图像,图像更容易受到光照、物体姿态、遮挡等因素的影响,从而导致模型的分性能有所下降。

表 4 CIFAR-10 数据集中不同激活函数的结果比较

Table 4 Comparison of the results of different activation functions in the CIFAR-10 dataset

激活函数	准确率/%	最低误差
Tanh	58.80	1.206 0
ELU	59.86	1.179 8
ReLU	59.11	1.192 6
Mish	59.90	1.176 3
Swish	57.62	1.231 6
SoftPlus	55.51	1.282 5
S-ReLU	63.64	1.085 3

4 结论

针对目前函数存在的“神经元死亡”“均值偏

移”和人工调参难的问题。提出了新的激活函数,并通过实验分析得到如下结论。

(1) S-ReLU 在 MNIST、CIFAR-10 上均达到了现有激活函数的最好水平,且准确率分别提升了 1.18% 和 8.13%。

(2) 解决了 ReLU 函数的“神经元死亡”的问题,使模型能在为负区域数值引入负激活值和非零导数值的同时,维持了负区域的稀疏性。

(3) S-ReLU 函数兼具了 ReLU 和 SoftPlus 的优点,负区域取值有正有负,更加贴合生物学上神经元响应的特点,提升了模型的抗干扰能力。

(4) S-ReLU 不含有任何自定义的超参数,解决了 ELU 和 LeakyReLU 等人工调参的难题,且样本负区域函数是非线性的,提升了模型的非线性表达能力。

为了更好地利用 S-ReLU 函数的性能,今后还可以研究该激活函数在不同的学习率、优化器的表现,不断探索模型的网络结构,以寻求最优超参数值和模型设计。此外,还能进一步探索 S-ReLU 函数在目标检测、人脸识别、语义分割等应用场景的表现,拓展 S-ReLU 的使用范围。

参 考 文 献

- [1] 蓝洁,周欣,何小海,等. 基于跨层精简双线性网络的细粒度鸟类识别[J]. 科学技术与工程, 2019, 19(36): 240-246.
Lan Jie, Zhou Xin, He Xiaohai, et al. Fine-grained bird recognition based on cross-layer simplified bilinear network[J]. Science Technology and Engineering, 2019, 19(36): 240-246.
- [2] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313: 504-507.
- [3] 李志欣,周韬,张灿龙,等. 利用对抗网络改进多标记图像分类[J]. 计算机辅助设计与图形学学报, 2020, 32(1): 16-26.
Li Zhixin, Zhou Tao, Zhang Canlong, et al. Using adversarial networks to improve multi-label image classification[J]. Journal of Computer Aided Design and Graphics, 2020, 32(1): 16-26.
- [4] 李庆盛,缪楠,张鑫,等. 基于注意力机制非对称残差网络和迁移学习的玉米病害图像识别[J]. 科学技术与工程, 2021, 21(15): 6249-6256.
Li Qingsheng, Miao Nan, Zhang Xin, et al. Image recognition of corn diseases based on attention mechanism asymmetric residual network and transfer learning[J]. Science Technology and Engineering, 2021, 21(15): 6249-6256.
- [5] 冯伟业,廖可非,欧阳缙,等. 基于胶囊神经网络的合成孔径雷达图像分类方法[J]. 科学技术与工程, 2019, 19(28): 203-207.
Feng Weiye, Liao Kefei, Ouyang Shan, et al. Synthetic aperture radar image classification method based on capsule neural network[J]. Science Technology and Engineering, 2019, 19(28): 203-207.
- [6] 谢小红,陆建波,李文韬,等. 基于迁移学习的服装图像分类模型研究[J]. 计算机应用与软件, 2020, 37(9): 88-93.
Xie Xiaohong, Lu Jianbo, Li Wentao, et al. Research on clothing image classification model based on transfer learning[J]. Computer

- Applications and Software, 2020, 37(9): 88-93.
- [7] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification [C]// Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015, 1: 1026-1034.
- [8] Han J, Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning[J]. International Workshop on Artificial Neural Networks: from Natural to Artificial Neural Computation, 1995, 930: 195-201.
- [9] 蒋昂波, 王维维. ReLU 激活函数优化研究[J]. 传感器与微系统, 2018, 37(2): 56-58.
- Jiang Angbo, Wang Weiwei. Research on relu activation function optimization[J]. Sensors and Microsystems, 2018, 37(2): 56-58.
- [10] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[J]. Journal of Machine Learning Research, 2011, 15: 315-323.
- [11] Bengio Y, Glorot X. Understanding the difficulty of training deep feed forward neural networks[J]. International Conference on Artificial Intelligence and Statistics, 2010, 9: 249-256.
- [12] Hahnloser R H R, Seung H S, Slotine J J. Permitted and forbidden sets in symmetric threshold-linear networks[J]. Neural Computation, 2003, 15(3): 621-638.
- [13] Maas A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models[J]. International Conference on Machine Learning, 2013, 30(1): 3.
- [14] Clevert D A, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs) [J]. Computer Science, 2015: 1511.07289.
- [15] Misra D. Mish: a self regularized non-monotonic neural activation function[J]. ArXiv, 2020: 1908.08681v3.
- [16] 朱延栓, 戴晓强, 兰武, 等. 基于动量梯度下降法的姿态数据采集系统[J]. 软件导刊, 2021, 20(10): 181-185.
- Zhu Yanshuan, Dai Xiaoqiang, Lan Wu, et al. Attitude data acquisition system based on momentum gradient descent[J]. Software Guide, 2021, 20(10): 181-185.
- [17] 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.
- [18] 高友文, 周本君, 胡晓飞. 基于数据增强的卷积神经网络图像识别研究[J]. 计算机技术与发展, 2018, 28(8): 62-65.
- Gao Youwen, Zhou Benjun, Hu Xiaofei. Research on convolutional neural network image recognition based on data augmentation[J]. Computer Technology and Development, 2018, 28(8): 62-65.
- [19] Ramachandran P, Zoph B, Le Q V. Searching for activation functions[J]. Computer Science, 2017: 1-13.
- [20] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images [J]. Handbook of Systemic Autoimmune Diseases, 2009, 1(4): 1-60.