

深度卷积神经网络中激活函数的研究

李一波, 郭培宜, 张森悦

(沈阳航空航天大学 自动化学院, 辽宁 沈阳 110000)

摘要:针对深度卷积神经网络中经典的 AlexNet 网络模型中激活函数 ReLU 在网络模型训练时易产生神经元“死亡”和均值偏移的问题进行研究以及改进,通过结合反正切函数和对数函数的优势,在传统激活函数 ReLU 基础上提出了一种新的激活函数 sArcReLU,并在后续训练过程中进一步调参。并将文中改进后的激活函数 sArcReLU 用于 AlexNet 网络模型训练,将使用新激活函数训练的深度卷积神经网络模型应用于公开数据集进行分类实验以验证其性能。实验结果表明:利用 sArcReLU 激活函数训练的深度卷积神经网络比利用 ReLU 以及 ArcReLU 训练的网络模型在分类精度上分别提升了 1.7% 和 2.4%,证明了改进方式经过大量数据充分微调的深度卷积神经网络可有效地提高图像分类精度,该方法同时也提升了深度卷积神经网络的实际应用价值。

关键词:深度卷积神经网络;激活函数;反正切函数;对数函数;图像分类

中图分类号:TP183

文献标识码:A

文章编号:1673-629X(2021)09-0061-06

doi:10.3969/j.issn.1673-629X.2021.09.011

Research on Activation Function in Deep Convolutional Neural Network

LI Yi-bo, GUO Pei-yi, ZHANG Sen-yue

(School of Automation Institute, Shenyang Aerospace University, Shenyang 110000, China)

Abstract: The problem that the activation function ReLU in the classic AlexNet network model of deep convolutional neural network is prone to "death" of neurons and mean shift when the network model is trained is studied and improved. By combining the advantages of arctangent function and logarithmic function, we propose a new activation function sArcReLU based on the traditional activation function ReLU and further adjust the parameters in the subsequent training process. The improved activation function sArcReLU is used for AlexNet network model training, and the deep convolutional neural network model trained with the new activation function is applied to the public data set for classification experiments to verify its performance. The experiment shows that the deep convolutional neural network trained with the sArcReLU activation function improves the classification accuracy by 1.7% and 2.4%, respectively, than the network model trained with ReLU and ArcReLU, which proves that the proposed improved method is fully fine-tuned by a large amount of data. Neural network can effectively improve the accuracy of image classification, and this method also improves the practical application value of deep convolutional neural networks.

Key words: deep convolutional neural network; activation function; arctangent function; logarithmic function; image classification

0 引言

AlexNet 网络模型在图像识别分类、目标检测、语音文字识别等方面都表现得十分突出^[1]。AlexNet 虽然不是首个被创建的卷积神经网络模型(convolutional neural networks, CNN),可是它是首个引起众多研究者所关注的卷积神经网络,具有里程碑意义。AlexNet 模型是由多伦多大学教授 Geoffery Hinton 同他的学生 Krizhevsky 等人共同设计,2012 年创下 Imageclassification 比赛的新记录,并获得了 ImageNet Large Scale Visual Recognition Challenge(ILSVRC)挑

战赛的冠军^[2]。同年, Krizhevsky 等^[3]在大型图像数据库 ImageNet^[4]的图像分类竞赛中提出的 AlexNet 模型,以超越第二名 11% 的精确度获得了冠军,使得卷积神经网络成为学术界关注的焦点之一。随着机器学习研究领域的不断拓展, AlexNet 也被许多学者作为首选的网络模型,通过改进和优化,其分类精度不断提高。

深度卷积神经网络的观点起初来源于人工神经网络,而网络中激活函数则是人工神经网络模型训练过程和学习非线性函数过程中不能或缺的一部分。如果

网络模型不使用激活函数,神经网络每一层的输出都将等价于前一层输入的线性函数,那么无论神经网络将会有多少层,输出都将是输入的线性组合,这种情况使网络模型成为最基本的感知器,深度神经网络将变得失去本身意义^[5]。因此,使用激活函数作为神经元来引进非线性因素是十分必要的,可以使得神经网络能够任意逼近任何非线性函数,进而使神经网络能够应用到更多的非线性模型中。Krizhevsky 等人在 AlexNet 中提出了 ReLU 激活函数来训练网络^[6],其优点是前向区间为线性函数,加快了模型训练的收敛速度,解决了 softsign、Softsign、Tanh 等常见激活函数中的梯度消失问题,然而 ReLU 激活函数很可能导致一些神经元在模型训练中无法激活。为了解决这种神经元“死亡”现象,改进了 ReLU 激活函数,使其在 x 小于零的负轴区间由 Swish 函数来替代,使负半轴的 ReLU 激活函数称为非线性激活函数,有效地解决了 x 小于零的部分神经元不能激活的问题,并且在 x 大于零的区间部分范围内由 ReLU 激活函数来替代,提高收敛速度的同时也能够提升参数的利用率,并且降低了过

拟合现象的发生率,很好地提高了 AlexNet 的鲁棒性。

针对 AlexNet 网络中激活函数 ReLU 在网络训练中产生的神经元死亡和均值偏移问题,结合反正切函数和对数函数的优势,在传统激活函数 ReLU 基础上提出了一种新的激活函数 sArcReLU。将文中构建的激活函数训练网络模型并应用于公开数据集分类实验中以验证其建立的网络的性能。实验结果表明,利用 sArcReLU 激活函数训练的深度卷积神经网络在分类精度和适应性方面均有明显的改善。

1 AlexNet 模型结构

AlexNet 总共有 650 000 个神经元,63 000 万个神经连接,60 000 000 个网络参数。AlexNet 网络结构简单,引入了许多新方法来达到稳定的收敛速度^[7],网络结构如图 1 所示。网络模型结构共 8 层,其中分别含有五层卷积层和三层全连接层,包括了 LRN 局部响应归一化层和 Dropout 正则化。此网络模型在图像分类领域有着较为出色的优势^[8]。

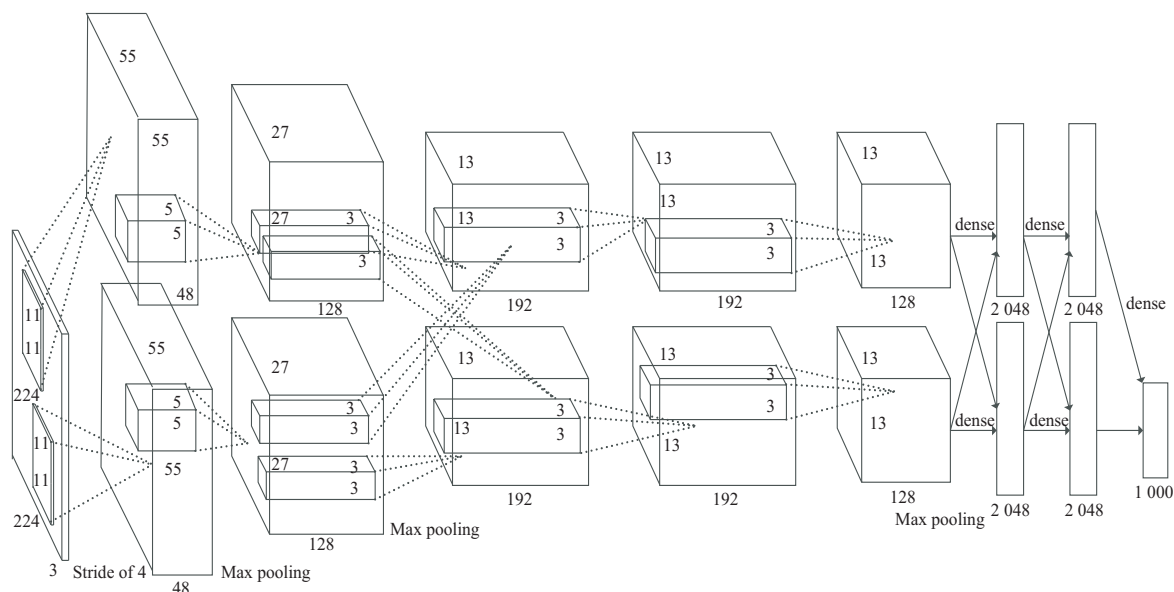


图 1 AlexNet 结构模型

图像数据输入格式是 $227 \times 227 \times 3$,其中 227 表示输入图像的宽度和高度,3 表示输入图像的三原色 R、G、B 通道模式,所以不需要对输入的数据集进行额外的格式裁剪。第一、二层计算均为卷积,ReLU,最大池化层和归一化,第二层的输出结果与 256 个特征图进行了卷积操作。网络中的第三、四层只需要进行卷积和 ReLU 操作。第五层的过程和第一层的过程类似,区别只在于没有经过归一化处理。网络最后将其第五层的输出转变为长向量,输入到三层全连接结构的卷积神经网络模型中,再运用 Softmax 回归函数即可计算其分类准确值。

2 模型的改进

AlexNet 网络采用 ReLU 非饱和线性函数,激活值的获取相对简单,只用一个阈值,省去了原本复杂的运算过程,相较于常见的非线性 S 型激活函数 Tanh、Sigmoid 等收敛速度更快^[9],改善了梯度消失和收敛不稳定的缺点。图 2 为 ReLU 与其他常见激活函数的曲线对比。

Tanh 函数换言之是 Sigmoid 函数的一种变换类型,Tanh 函数的输出值在 $[-1, 1]$ 区间内,而 Sigmoid 函数的输出值则在 $[0, 1]$ 区间^[10]。Tanh 和 Sigmoid 都

拥有饱和区,ReLU 函数在 x 大于零时导数一直是 1 且不变,十分有助于解决梯度消失、梯度爆炸等问题,促使加快训练速度。

一般地,当网络模型后向传递过程时,且使用梯度下降法求解网络参数,Sigmoid 激活函数向下传导的梯度中包含了有关自身输入的导数 $f'(x)$,当输入值进入饱和区时, $f(x)$ 的输出值逐渐接近于零。此特性十分容易导致梯度消失现象,所以深度神经网络一直很难得到有效的训练,也是阻碍神经网络发展的重要原因之一^[11]。

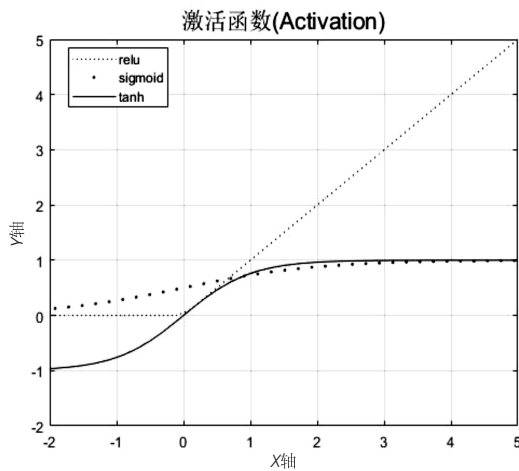


图2 ReLU 与常见的激活函数曲线对比

其中 ReLU 激活函数的数学表达式为:

$$f(x) = \max(0, x) \quad (1)$$

经过图 2 分析,ReLU 激活函数曲线也存在很多不足,例如当输出值恒大于或等于零时,易导致均值偏移的缺点,均值偏移会导致神经元将前一层网络的非零均值输出的信号作为输入信号,使参数的计算复杂度大幅度增加。处于网络模型训练过程的前向传播过程时,“强制”稀疏性作用会使网络中的某些神经元一直不会被激活,与其相对应的参数也一直不会被更新,所以会致使一些好的特征被屏蔽。众所周知,神经“死亡”和强制稀疏性的缺点对网络模型的收敛速度和网络性能影响很大^[12]。

文献[13]经过对几种经典激活函数的研究以及分析,结合这两种激活函数的优点,为了缓解 ReLU 函数神经元“死亡”的缺点,降低 ArcReLU 函数的计算复杂度,文中结合两者优点构造出了一种新的 ArcReLU 激活函数。函数原理是将 ReLU 函数负轴部分的输出更改为 $\frac{2}{\pi} \arctan x$,正轴部分替换为单调线性函数。网络训练中使用 ArcReLU 函数时,为了更好地控制输出范围,文中将算子扩增 $\frac{2}{\pi}$ 倍,因此网络模型遇到一些较大的输入值时依旧可以保持稳定。其定义如下:

$$f(x) = \begin{cases} x, & x > 0 \\ \frac{2}{\pi} \arctan x, & x \leq 0 \end{cases} \quad (2)$$

ArcReLU 函数导数如下:

$$f'(x) = \begin{cases} 1, & x > 0 \\ \frac{1}{1+x^2}, & x \leq 0 \end{cases} \quad (3)$$

由图 3 分析,ArcReLU 激活函数导数值恒为正值。根据导数特征,可说明函数明显呈单调递增特性。当激活函数拥有单调特性时,单层网络能保证其为凸函数,由此推断出该函数在训练过程中更易收敛。

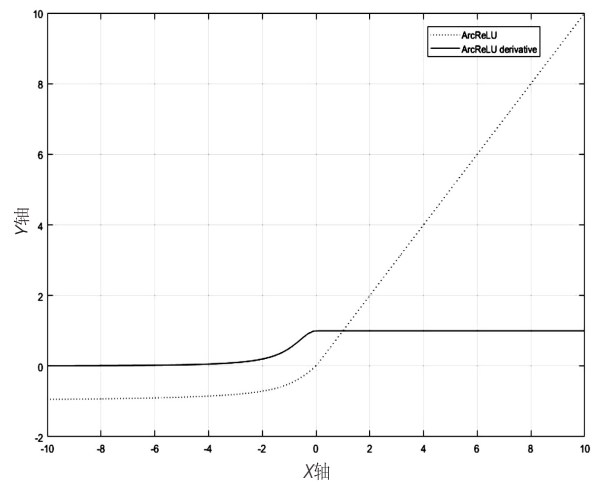


图3 ArcReLU 及其导数的图像

由于 ReLU 归类于分段线性非饱和的函数,其与传统的 S 型激活函数作对比发现,ReLU 函数随机梯度下降收敛速度更快,而且函数计算过程更加简洁明了。相较于 Sigmoid 激活函数,ReLU 稀疏特性更加明显。然而稀疏性也将会带来更高的错误识别率并且降低了网络模型的有效容量。通过分析激活函数,不仅计算过程更加简洁,而且负半轴的输出会得以更好的保存。在负轴使用了反正切函数进行替换,不仅可以使均值更趋向于零,缓解了函数均值易偏移的缺点,并且其负半轴部分具有软饱和性,使其不会轻易出现神经元“死亡”的现象,同时具备单调递增的特性,进一步提高了收敛速度。

文中结合 ArcReLU 激活函数设计了一种新的激活函数,记为 sArcReLU,表达式如公式(4)所示,函数图像如图 3 所示。

$$f(x) = \begin{cases} s \cdot \log\left(\frac{1}{1+e^{-x}} + 1\right), & x \geq 0 \\ \frac{2}{\pi} \arctan x, & x < 0 \end{cases} \quad (4)$$

上述公式中,系数 s 表示超参数,取值范围为 $[0, 1)$ 。当输入为负值时,函数的梯度与参数 s 不相关;当输入为正值时,函数值取决于超参数。

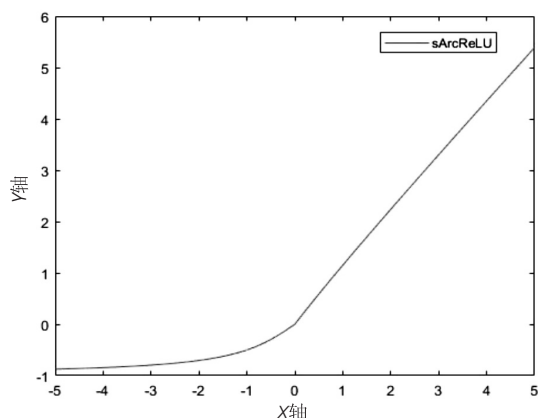


图 4 sArcReLU 函数图像

改进后激活函数的图像如图 4 所示。由函数图像分析可以得到初步结论,该函数在其定义范围内可导并且单调递增,现只需证明该函数在其原点处的可导特性。以下证明过程中将 x 大于零的部分称为 f_1 , x 小于等于零的部分称为 f_2 ,过程证明如下:

$$f(0) = f(0^-) = f(0^+) = 0 \quad (5)$$

$$\lim_{x \rightarrow 0^+} \left[\frac{f_1(x) - f_1(x_0)}{\Delta x} \right] = \lim_{x \rightarrow 0^+} \left[\frac{s \cdot \log\left(\frac{1}{1 + e^{-x}} + 1\right) - 0}{x - 0} \right] = 1 \quad (6)$$

$$\lim_{x \rightarrow 0^-} \left(\frac{f_2(x) - f_2(x_0)}{\Delta x} \right) = \lim_{x \rightarrow 0^-} \left(\frac{\frac{2}{\pi} \arctan x - 0}{x - 0} \right) = 1 \quad (7)$$

式(5)说明 sArcReLU 在原点连续且具有定义。由式(6)与式(7)的结果分析,函数都存在且相同,依据导数的定义,该函数在零点处可导。

sArcReLU 的偏导数为:

$$f' = \begin{cases} \frac{1 + e^{-x}}{\ln 10(2 + e^{-x})}, & x \geq 0 \\ \frac{1}{1 + x^2}, & x < 0 \end{cases} \quad (8)$$

从上述公式可看出,sArcReLU 激活函数的导数值始终大于零。根据导数定义,证明函数具有单调递增特性。当激活函数始终具有单调特性时,单层网络可以保证其为凸函数^[14]。当激活函数中 x 大于等于 0,即神经元处于兴奋激活状态时,梯度随着 x 的增加而不断下降,并且最终收敛于 1。当 x 小于零,即神经元处于抑制区域时,具有接近于零的梯度 s 。

sArcReLU 函数具有以下优势:

①负轴的输出值保持非零值,因此激活函数拥有保留梯度变化的特性,即可以有效解决由 ReLU 强制置零特性导致的神经元“死亡”缺点。

②超参数 s 使得激活函数输出值可变,消除了当 ReLU 取正值时的线性特性,使得网络更新过程更接近于生物神经元接收刺激信号时展现的生物特性,并且应用于不同的数据集将会得到不同的最优值,也使改进后的激活函数更具备适应性。

③由于超参数 s 的存在,正半轴的值保持可变性,可以实时修正数据的分布稀疏性,保留了网络快速收敛的特性。

3 实验与分析

UC Merced Land Use (UCM) 数据集为公开的遥感数据集。数据集中的遥感图像由工作人员从美国地质调查局“市区图像”集合中的大图像中手动提取^[15]。其中包括美国不同城市地区的 21 种地物的遥感图像,数据集中每一类有一百幅,共 2 100 幅,空间分辨率分别为 1 英尺,分别具有 R、G、B 三个颜色通道。21 类场景分别为农田、机场、棒球场、沙滩、建筑、丛林、密集住宅区、森林、高速公路、高尔夫球场、港口、十字路口、普通住宅区、房车停车场、立交桥、停车场、河流、跑道、高级住宅区、储油罐、网球场。其中部分场景之间有部分重叠,比如住宅区分了密集住宅区、普通住宅区、高级住宅区三类。21 类场景图像的部分样例如图 5 所示。

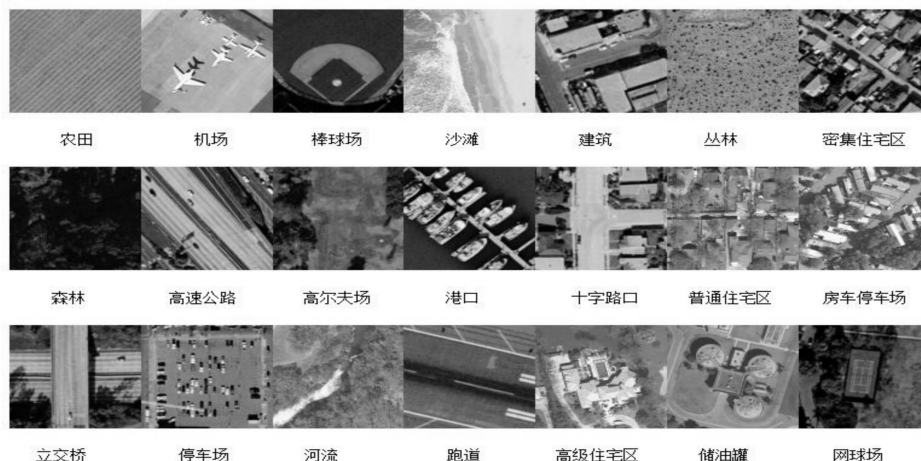


图 5 UCM 数据集样例图

实验采用开源框架 Tensorflow-1.10.0, 编程软件 python3.6.6 实现模型结构构建, 采用的 CPU 是 Intel (R) Core (TM) i5-8500 CPU @ 3.00 GHz, 8G 内存, 64 位 windows10 操作系统。

按照 9 : 1 的比例将 UCM 数据集中每一个场景的高分辨率图像随机分为 90 张训练数据、10 张测试数据。初始学习率选取 0.000 1、0.001、0.01、0.1、1, 实验结果如图 6 所示。

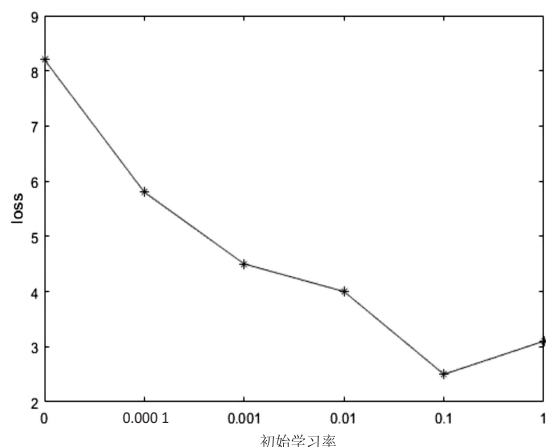


图6 不同初始学习率对应的 loss 值

随着学习率不断扩大, 网络的损失值逐渐减小, 随后 loss 值又开始小幅度增大。由图 6 分析选取 0.1 作为初始学习率最为合适。

为了验证激活函数参数 s 取值对分类结果的影响, s 分别取值 0、0.01、0.1、0.4、0.7、1 进行测试, 实验结果如表 1 所示。

表1 超参数 s 分析

s	迭代次数/K	分类准确率/%
0	0.3	73.4
0.01	0.3	76.9
0.1	0.2	78.8
0.4	0.2	76.7
0.7	0.2	75.9
1	0.2	72.3

实验结果表明, 当 $s = 0$ 时, 激活函数退化为 ReLU, 准确率就是 AlexNet 经典网络模型训练得出的结果, 即 73.4%。当 $s = 0.1$ 时, 网络会在 1 000 次时最先达到收敛状态, 而且分类效果最好, 网络准确率将会达到 78.8%。

在数据集 UCM 和同等条件下对三种激活函数 ReLU、ArcReLU、sArcReLU 做了训练测试。实验结果如表 2 所示。

通过结果分析, 用 sArcReLU 激活函数训练的 AlexNet 在分类准确率上比受其他两个激活函数的 AlexNet 更胜一筹。

表2 数据集 UCM 下的准确率

激活函数	准确率/%
ReLU	76.4
ArcReLU	78.1
sArcReLU	78.8

为了验证提出的激活函数的适应性, 新建了一个数据集 UCM - NWPU, 选取 NWPU - RESISC45 和 UC Merced_LandUse 中相同 15 种场景, 每种场景有 900 张, 总有 13 500 张, 对图像尺寸归一化, 调整输入图像的尺寸统一为 $256 \times 256 \times 3$ 。训练集是在每类图像场景中随机挑选出其中 90% 图像, 因此原有训练集共有 12 150 张图像, 剩下的 1 350 张图像就将作为测试集。

若选取的学习率过大, 十分有可能会直接越过最优值; 若选取的学习率过小, 优化的效率可能过低, 长时间就无法收敛。同上一个实验一致, 经过测试, 本实验的学习率为 0.1 时测试结果最好, 激活函数中超参数 s 选取 0.1 时效果最好。实验结果如下所示:

图 7 和图 8 分别是 sArcReLU 作为激活函数在数据集 UCM-NWPU 上训练验证的分类准确率和损失函数曲线。

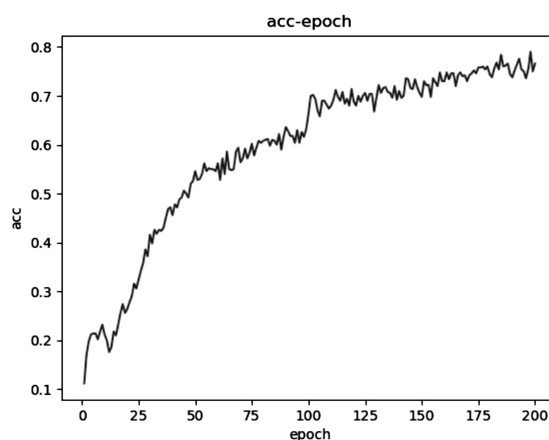


图7 准确率

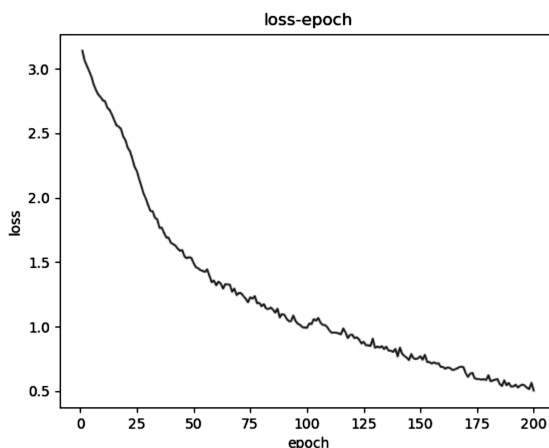


图8 损失函数曲线

4 结束语

文中设计了一种新激活函数 sArcReLU, 并且与 ReLU 和 ArcReLU 激活函数进行了实验分析和比较。实验证明:sArcReLU 激活函数能够明显提高网络迭代速度并且有效降低训练的误差率。表明 sArcReLU 中超参数 s 对迭代次数和准确率有一定的影响。下一步研究工作的重点将放在激活函数中超参数 s 的优化上, 以进一步提高激活函数 sArcReLU 的性能。

参考文献:

- [1] KRIZHEVSKY A, SUTSKEVE R I, HINTON G E. Image Net classification with deep convolutional neural networks [C]//Proceedings of advances in neural information processing systems. Cambridge, MA: MIT Press, 2012: 1106–1114.
- [2] CHENG Jian, WANG Peisong, LI Gang, et al. Recent advances in efficient computation of deep convolutional neural networks[J]. Frontiers of Information Technology & Electronic Engineering, 2018, 19(1): 64–77.
- [3] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527–1554.
- [4] DENG J, DONG W, SOCHER R, et al. Image Net: a large-scale hierarchical image database[C]//IEEE conference on computer vision & pattern recognition. Kyoto, Japan: IEEE, 2009: 248–255.
- [5] 刘建伟, 刘媛, 罗雄麟. 深度学习研究进展[J]. 计算机应用研究, 2014, 31(7): 1921–1930.
- [6] YUAN Zhenlong, LU Yongqiang, XUE Yibo. Droid detector: Android malware characterization and detection using deep learning[J]. Tsinghua Science and Technology, 2016, 21(1): 114–123.
- [7] 党宇, 张继贤, 邓喀中, 等. 基于深度学习 AlexNet 的遥感影像地表覆盖分类评价研究[J]. 地球信息科学学报, 2017, 19(11): 1530–1537.
- [8] PAN Guangliang, LI Jun, LIN Fei, et al. A combined activation function for learning performance improvement of CNN image classification [C]//Proceedings of 5th international conference on vehicle, mechanical and electrical engineering (ICVMEE 2019). Wuhan: Wuhan Zhicheng Times Cultural Development Co., Ltd., 2019: 365–371.
- [9] 陈兆佳. 基于三元组深度哈希的跨模态检索方法[D]. 西安: 西安电子科技大学, 2018.
- [10] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229–1251.
- [11] 陈先昌. 基于卷积神经网络的深度学习算法与应用研究[D]. 杭州: 浙江工商大学, 2014.
- [12] 郭敏钢, 宫鹤. AlexNet 改进及优化方法的研究[J]. 计算机工程与应用, 2020, 56(20): 124–131.
- [13] 许赞杰, 徐菲菲. 基于 ArcReLU 函数的神经网络激活函数优化研究[J]. 数据采集与处理, 2019, 34(3): 517–529.
- [14] WU Huaiqin. Global stability analysis of a general class of discontinuous neural networks with linear growth activation functions[J]. Information Sciences, 2009, 179(19): 3432–3441.
- [15] 何婧媛, 阿茹罕. 基于卷积神经网络的遥感图像分类[J]. 电子设计工程, 2020, 28(12): 109–113.