

基于残差的改进卷积神经网络图像分类算法^①



高磊, 范冰冰, 黄穗

(华南师范大学 计算机学院, 广州 510631)

通讯作者: 范冰冰, E-mail: lay771880416@163.com

摘要: 现有深度残差网络作为一种卷积神经网络的变种, 由于其良好的表现, 被应用于各个领域, 深度残差网络虽然通过增加神经网络深度获得了较高的准确率, 但是在相同深度情况下, 仍然有其他方式提升其准确率. 本文针对深度残差网络使用了三种优化方法: (1) 通过卷积网络进行映射实现维度填充; (2) 构建基于 SELU 激活函数的残差模块 (3) 学习率随迭代次数进行衰减. 在数据集 Fashion-MNIST 上测试改进后的网络, 实验结果表明: 所提出的网络模型在准确率上优于传统的深度残差网络.

关键词: 卷积神经网络; 残差网络; 学习率; SELU 激活函数

引用格式: 高磊, 范冰冰, 黄穗. 基于残差的改进卷积神经网络图像分类算法. 计算机系统应用, 2019, 28(7): 139-144. <http://www.c-s-a.org.cn/1003-3254/6997.html>

Improved CNN Image Classification Algorithm Based on Residuals

GAO Lei, FAN Bing-Bing, HUANG Sui

(School of Computer Science, South China Normal University, Guangzhou 510631, China)

Abstract: The existing deep residual network, as a variant of a convolutional neural network, is used in various fields due to its sound performance. Although the depth residual network obtains higher accuracy by increasing the depth of the neural network, there are still other ways to improve the accuracy at the same depth. In this study, three optimization methods are used to optimize the depth residual network. (1) Dimension filling by mapping through a convolutional network. (2) Building a residual module based on the SELU activation function. (3) Learning rate decays with the number of iterations. Testing the improved network on the dataset Fashion-MNIST, the experimental results show that the proposed network model is superior to the traditional deep residual network in accuracy.

Key words: convolutional neural network; residual network; learning rate; SELU activation function

1 引言

图像分类一直是深度学习的热门研究课题之一, 卷积神经网络是近些年发展起来, 并引起重视的一种高效识别算法. 卷积神经网络是人工神经网络的一种, 是深度学习的一个重要算法, 特别是在模式分类领域, 由于该网络避免了对图像的复杂前期预处理, 可以直接输入原始图像, 因此得到广泛的应用. 20 世纪 60 年代, Hubel 和 Wiesel 在研究猫脑皮层中用于局部敏感

和方向选择的神经元时, 发现其独特的网络结构可以有效的降低反馈神经网络的复杂性, 继而提出了卷积神经网络^[1-4](Convolutional Neural Networks, 简称 CNN).

与一般神经网络相比, 从构造上看, 卷积神经网络的基本结构包括两层, 其一为特征提取层, 其二为特征映射层. 从模型特征上看, 卷积神经网络有两个特殊性, 其一为神经元之间的连接是非全连接的, 其二为同一

^① 基金项目: 广东省重大科技专项 (2016B030305003)

Foundation item: Science and Technology Major Program of Guangdong Province (2016B030305003)

收稿时间: 2019-01-20; 修改时间: 2019-02-21; 采用时间: 2019-02-26; csa 在线出版时间: 2019-07-01

层中的某些神经元之间的连接的权重是共享的,从而降低了模型的复杂度.卷积神经网络的特殊结构在语音识别和图像处理方面有着独特的优越性,因此卷积神经网络主要用来识别位移,缩放以及其他形式扭曲不变性的二维图形.随着研究学者对卷积神经网络的研究和优化,提出了各种优秀的卷积神经网络的模型,并在图像处理上取得了不错的效果.

早在1998年,LEcun Yann, Bottou Leon就提出了LeNet-5模型^[5],这个模型是一种典型的用来识别数字的卷积网络,当年美国大多数银行就是用它识别支票上面的手写数字的. LeNet-5是早期的卷积神经网络模型,其结构一共有7层,每层都包含不同数量的训练参数,主要有2个卷积层,2个池化层,3个全连接层.

Alex Krizhevsky提出的AlexNet模型^[6]是卷积网络进入大范围应用的经典网络模型,其结构包括5个卷积层和3个全连接层,其主要的优势在于:(1)使用ReLU激活函数,ReLU与Sigmoid相比,其优势是训练速度更快,因为Sigmoid的导数在饱和区会变得很小,导致权重几乎没有更新,从而造成梯度消失;(2)创造性的提出了Dropout层,Dropout层主要是防止过拟合的.每个全连接层后面使用一个Dropout层,从而减少过拟合.其主要的缺点在于当时人们的观念是卷积和越大,则感受野越大,从而看到图片的信息越多,因此获得的特征越好,因此AlexNet模型中使用了一些非常大的卷积核,比如 11×11 , 5×5 卷积核,造成了计算量的增加,降低了计算性能,并且不利于模型深度的增加.

VGGNet^[7]是由牛津大学的视觉几何组(Visual Geometry Group)提出的,VGGNet相比于AlexNet模型最大的改进就是采用连续多个 3×3 的卷积核替代了AlexNet模型中的大卷积核,同时增加非线性变换的能力,保证可以增加深度从而学习更复杂的模式,从而提升模型的效果.

GoogLeNet^[8]是2014年Christian Szegedy提出的一种卷积神经网络架构,其主要的贡献就是提出了inception模块^[8],多个inception模块构成了inception网络结构.其相对于AlexNet模型,GoogLeNet模型也采用了小卷积核,而相对于VGGNet模型,GoogLeNet模型采用了多个尺寸进行卷积再聚合的方式,其优势是获取了更多的特征,丰富的特征会让结果的判断更为准确.

ResNet^[9]是MSRA何凯明团队提出的,在深度网

络中存在以下问题:(1)深度学习的网络越深,常规的网络堆叠效果越不好;(2)网络越深,梯度消失现象越来越明显,网络的训练效果也不会很好.(3)而浅层的网络又无法明显提升网络的识别效果.因此ResNet模型引入了残差结构,其主要优点就在于解决了极深度条件下深度卷积神经网络的性能退化功能.

本文图像分类的模型,就是借鉴上述模型的优点,提出了一种基于残差的改进卷积神经网络,进一步提升训练效果.本文主要在一下方面进行改进:

(1)对ReLU激活函数,使用更新的SELU^[10]激活函数进行替换,形成新的残差结构.

(2)对传统深度残差网络的数据池化层进行改进.

(3)通过卷积的方式进行维度补偿.

(4)学习率随迭代次数进行减小.

2 模型相关技术

2.1 残差网络

ResNet的提出背景是解决或者缓解深层的神经网络训练中的梯度消失问题,假设有一个 N 层的深度神经网络,另外有一个 $N+K$ 层的神经网络,直观上来说,即使第二个深度神经网络效果比较差,但是也不会比第一个深度神经网络差.

根据神经网络的反向传播原理,先通过正向传播计算出结果output,然后与样本比较得出误差值 E ,根据误差结果,利用“链式法则”求偏导,使结果误差反向传播从而得出权重 w 调整的梯度.通过不断迭代对参数矩阵进行不断调整后,使得输出结果误差值更小,从而与输出结果与事实更加接近.从上面的过程可以看出,神经网络在反向传播的过程中要不断的传播梯度,当神经网络加深时,梯度会在传播过程中逐渐消失,导致无法对前面网络层的权重进行有效的调整.对于上述问题,研究学者给出了解决方法,正则初始化和中间的正则化层,但是这样也仅仅可以训练几十层网络.

而ResNet引入了残差网络结构^[11,12],通过残差网络,可以将网络层做的很深.残差网络结构如图1所示,残差网络的本质就是解决层次比较深的时候无法训练的问题,残差网络借鉴了Highway Network思想,通过对网络结构添加shortcut连接,使得输入可以直达输出,ResNet的思想是假设我们涉及一个网络层,存在最优化的网络层次,那么往往我们设计的深层次网络是有很多网络层为冗余层的.那么我们希望这些冗余层

能够完成恒等映射, 保证经过该恒等层的输入和输出完全相同. 图 1 中的残差模块的输入由原来的 $F(x)$ 变为 $H(x) = F(x) + x$, 如果这些网络层是冗余的, 那么只需要令 $F(x) = 0$, 即可完成恒等映射.

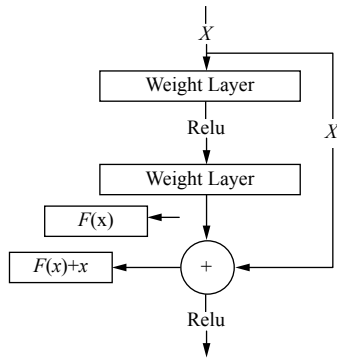


图 1 残差模块结构

2.2 激活函数

激活函数^[13 14]在神经网络中主要解决线性不可分问题, 通过对每一层线性变换后叠加一个非线性激活函数, 从而获得更强大的学习和拟合能力. 而 ReLU^[15,16] 激活函数则是其中的佼佼者. 从计算角度上来说 ReLU 计算简单, 只需要一个阈值就可以计算得到激活值, 其次 ReLU 的非饱和性可以有效的解决梯度消失的问题, 以提供相对宽的激活边界, 最后, ReLU 的单侧抑制提供了网络的稀疏表达能力. 但是 ReLU 仍然有其局限性, 根据 ReLU 公式 (1) 可看出, 由于其函数在负半轴上直接置 0, 由此导致负梯度经过 ReLU 单元时不被激活, 导致该神经元死亡. 其次 ReLU 函数在正半轴导数恒为 1, 导致梯度过小的时候, 经过该函数时, 容易导致梯度消失.

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

SELU 激活函数就是解决上述问题而产生的, 如式 (2), SELU 激活函数的正半轴导数是大于 1 的, 这样可以使方差过小的时候让它增大, 同时防止了梯度消失; 而在负半轴不再简单置 0, 从而解决神经元死亡问题, 又因为其坡度较为平缓, 以便尽量保证 ReLU 中的单侧抑制优势.

$$SELU(x) = \lambda \begin{cases} x, & \text{if } x > 0 \\ \alpha e^x - \alpha, & \text{if } x \leq 0 \end{cases} \quad \lambda > 1 \quad (2)$$

3 基于残差的改进卷积神经网络

3.1 改进后的残差网络结构

采用两层的残差结构, 通过 SELU 激活函数替代其 ReLU 激活函数, 并进行数据批量归一化. 改进后的残差网络结构如图 2

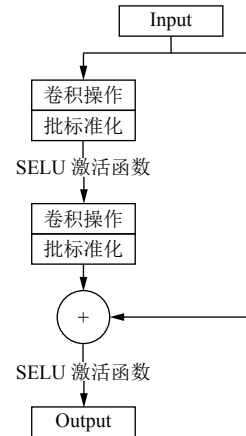


图 2 改进后的残差结构

3.2 基于 Inception 模块改进的数据池化层

Inception 模块是 GoogLeNet 神经网络中采用的优化模块, 其主要思想就是将一些大的卷积层替换成小卷积, 这样能够大大减少权值参数数量. 在传统深度残差网络中, 数据输入部分的池化层是通过的卷积层以及池化层实现的, 为了更快的提高训练速度, 减少权值参数, 因此在数据池化层引入 Inception 模块, 通过构建一个个更小的卷积网络然后将卷积结果进行批量正则化后, 作为层数进行维度上的合并. 将的卷积层拆分为三个的卷积层, 改进后的数据池化层如图 3.

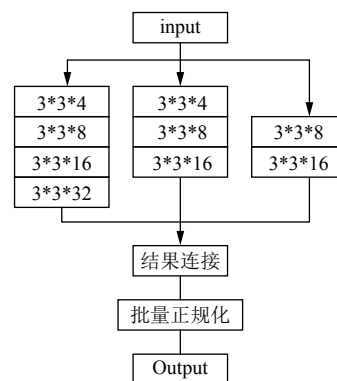


图 3 改进后的数据池化层结构

3.3 其他改进策略

改进一: 通过卷积网络将原始数据映射到新的空

间,实现维度拓展.

在设计残差网络时应遵循的规则有两点:

(1) 对于输出 feature map 相同的层,有相同数量的通道.

(2) 当 feature map 进行池化的时候, feature map 减半,通道数则会翻倍.

而对于残差网络来说,维度匹配的 shortcut 连接直接相加即可,对于维度不匹配的时候,需要将数据映射为相同维度的数据.而在传统残差网络中直接通过 0 进行填充.本文采用的是将数据乘以 W 矩阵投影到新的空间,实际实现是通过 1×1 卷积操作,直接改变 1×1 卷积的通道数据,从而实现通道数增加.

改进二:学习率优化方式^[17,18]

学习率决定了参数移动到最优值的速度快慢,如果学习率过大,很可能越过最优值,反而如果学习率过小,优化效率过低,会导致算法长时间无法收敛,因此学习率对于算法性能的表现至关重要,本文通过随迭代次数减小的优化方式,对学习率进行优化.

4 实验结果及分析

4.1 实验准备

本文通过在标准数据集上做实验来验证本文网络的性能. FashionMNIST^[19]是一个替代 MNIST 手写数字集的图像数据集,其涵盖了来自 10 种类别的共 7 万个不同商品的正面图片. FashionMNIST 数据集的大小以及划分和 MNIST 数据集一致,因此,可以直接测试算法性能,不需要改动源代码.

实验平台包括 PC 机, InterCore i7, 显卡型号 GTX1060, 6 GB 显存, Ubuntu 操作系统, PyTorch^[20]深度学习框架.在实验中,本文改进的神经网络结构为:首先采用改进后的数据池化层结构(图 3),替换传统深度残差网络中最开始的卷积池化层;其后为 4 个残差网络层,每个残差网络层都包含多个残差网络,第一个残差网络层包含 3 个残差块,第二个残差网络层包含 4 个残差块,第三个残差网络层包含 6 个残差块,第四个残差网络层包含 3 个残差块,每个残差块都使用 SELU 激活函数替代 ReLU 激活函数(图 2),同时进行批量归一化处理.最后通过均值池化和线性分类器得到输入结果.具体的神经网络结构见表 1.

在本文提出的模型训练中,损失函数采用交叉熵损失函数,采用 Adam 算法作为优化器进行优化.

表 1 改进后的神经网络结构

Layer name	模型
Inception	$3 \times 3, 32 + 16 + 16$
	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$
conv2	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$
conv3	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$
conv4	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$
conv5	池化, Softmax 分类器

4.2 实验结果及分析

本文通过控制变量,在同一台机器上分别运行算法,对每一个改进进行检测,最后将所有改进进行整合后检测,由于机器性能不足,因此只进行了 10 次迭代,但是为了防止结果的随机性,因此每个算法进行了十次实验,最后取得平均值.

(1) 优化学习率对准确率的影响

在仅仅优化学习率的情况下,在数据集 Fashion-MNIST 上测试分类结果,实验结果如表 2 所示.

表 2 学习率优化的结果比较 (%)

模型	ResNet34	ResNet34+LR 优化
迭代 10 次的平均准确率	93.12	93.22
迭代 20 次的平均准确率	94.21	94.82

从表中可以看到:在进行学习率优化后,迭代 10 次的平均准确率提高了 0.1%,但是迭代 20 次的平均准确率提高了 0.61%.可以看出之所以在迭代 10 次的效果不明显,在于迭代次数较少,原因是迭代次数会增加模型的细化程度,因此准确率会上升.其次学习率优化后的结果要比未优化的效果好,原因是学习率优化后,可以使模型更快收敛,损失函数波动更小,如图 4 和图 5 所示.

(2) 使用 SELU 激活函数替换 ReLU 激活函数

在仅仅使用 SELU 激活函数替换 ReLU 激活函数的情况下,在数据集 FashionMNIST 上测试分类结果,实验结果如表 3 所示.

从表中可以看到:在仅优化激活函数的条件下,模型的准确率提高了 0.23%.但是可以从图 6 中可以看到,SELU 比 ReLU 的收敛速度要快很多.

(3) 基于 Inception 模块优化改进的神经网络

在仅仅使用 Inception 模块替换第一层的卷积池化

层,在数据集 FashionMNIST 上测试分类结果,实验结果如表 4 所示。

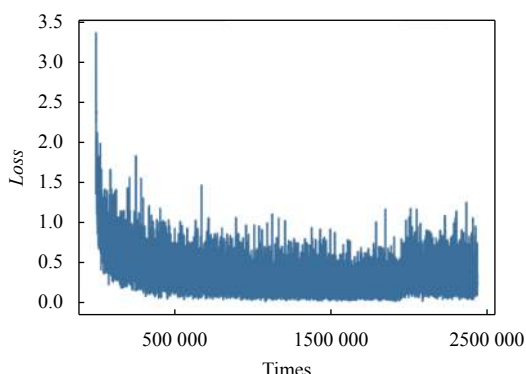


图 4 没有经过学习率优化的损失变化曲线

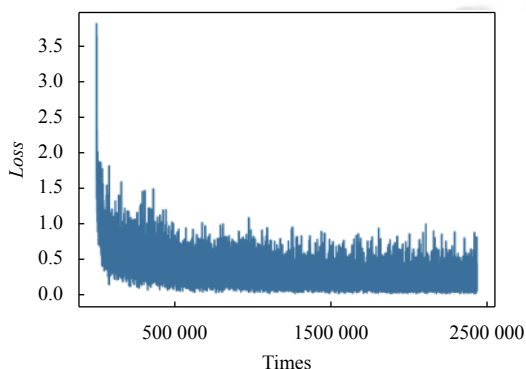


图 5 经过学习率优化的损失变化曲线

表 3 激活函数优化结果比较 (%)

模型	平均准确率
ResNet34	93.12
ResNet34+SELU	93.35

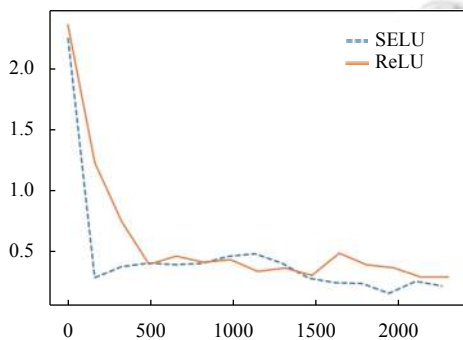


图 6 SELU 和 ReLU 激活函数收敛速度

表 4 Inception 模块优化结果比较 (%)

模型	平均准确率
ResNet34	93.12
ResNet34+Inception	94.88

从表中可以看到:相比传统残差网络,准确率提高了 1.76%,原因在于,通过多个小的卷积核,可以获得更丰富的特征,从而使结果更为准确。

(4) 基于所有优化方式的结合的神经网络

综上所述所有优化方式,最终的训练结果,在数据集 FashionMNIST 上如表 5 所示。

表 5 最终的模型优化后的结果比较 (%)

模型	平均准确率
ResNet34	93.12
整体改进后的模型	96.04

从表中可以看到:准确率比传统深度残差网络准确率提高了 2.92%,因此,多种优化方式对于提高深度残差网络的准确性是有积极意义的。

5 结论与展望

针对卷积神经网络,本文在传统残差网络的基础上进行了进一步的优化,通过对数据池化层,激活函数和学习率的优化,提升网络的准确率.实验结果表明,改进后的深度残差网络比传统深度残差网络表现要好,因此,本文在深度残差网络优化方面有一定的参考价值,当然,深度残差网络依然还有其他改进方式,例如其他方式的维度填充,其他损失函数或者优化器对训练模型的影响等方面,这将是下一步的研究方向。

参考文献

- 1 白琮,黄玲,陈佳楠,等.面向大规模图像分类的深度卷积神经网络优化.软件学报,2018,29(4):1029-1038.[doi:10.13328/j.cnki.jos.005404]
- 2 刘万军,梁雪剑,曲海成.不同池化模型的卷积神经网络学习性能研究.中国图象图形学报,2016,21(9):1178-1190.[doi:10.11834/jig.20160907]
- 3 王民,刘可心,刘利,等.基于优化卷积神经网络的图像超分辨率重建.激光与光电子学进展,2017,54(11):111005.
- 4 李鸣,张鸿.基于卷积神经网络迭代优化的图像分类算法.计算机工程与设计,2017,38(1):198-202.
- 5 Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324. [doi: 10.1109/5.726791]
- 6 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA.

2012. 1097–1105.
- 7 Guo ST, Luo YX, Song YZ. Random forests and VGG-NET: An algorithm for the ISIC 2017 skin lesion classification challenge. arXiv preprint arXiv: 1703.05148, 2017.
 - 8 Ballester P, Araujo RM. On the performance of GoogLeNet and AlexNet applied to sketches. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, AZ, USA. 2016. 1124–1128.
 - 9 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2015. 770–778.
 - 10 Zhang C, Woodland PC. DNN speaker adaptation using parameterised sigmoid and ReLU hidden activation functions. Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China. 2016. 5300–5304.
 - 11 张振焕, 周彩兰, 梁媛. 基于残差的优化卷积神经网络服装分类算法. 计算机工程与科学, 2018, 40(2): 354–360. [doi: [10.3969/j.issn.1007-130X.2018.02.023](https://doi.org/10.3969/j.issn.1007-130X.2018.02.023)]
 - 12 王一宁, 秦品乐, 李传朋, 等. 基于残差神经网络的图像超分辨率改进算法. 计算机应用, 2018, 38(1): 246–254. [doi: [10.11772/j.issn.1001-9081.2017061461](https://doi.org/10.11772/j.issn.1001-9081.2017061461)]
 - 13 Scardapane S, Van Vaerenbergh S, Hussain A, *et al.* Complex-valued neural networks with nonparametric activation functions. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018. [doi: [10.1109/TETCI.2018.2872600](https://doi.org/10.1109/TETCI.2018.2872600).]
 - 14 Nasser A, Njikam S, Zhao H. A novel activation function for multilayer feed-forward neural networks. Applied Intelligence, 2016, 45(1): 75–82. [doi: [10.1007/s10489-015-0744-0](https://doi.org/10.1007/s10489-015-0744-0)]
 - 15 Yarotsky D. Error bounds for approximations with deep ReLU networks. Neural Networks, 2017, 94: 103–114. [doi: [10.1016/j.neunet.2017.07.002](https://doi.org/10.1016/j.neunet.2017.07.002)]
 - 16 Arpit D, Zhou YB, Ngo H, *et al.* Why regularized auto-encoders learn sparse representation? arXiv: 1505.05561, 2015.
 - 17 Chandra B, Sharma RK. Deep learning with adaptive learning rate using Laplacian score. Expert Systems with Applications, 2016, 63: 1–7. [doi: [10.1016/j.eswa.2016.05.022](https://doi.org/10.1016/j.eswa.2016.05.022)]
 - 18 乔俊飞, 王功明, 李晓理, 等. 基于自适应学习率的深度信念网设计与应用. 自动化学报, 2017, 43(8): 1339–1349.
 - 19 Xiao H, Rasul K, Vollgraf R. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv: 1708.07747, 2017.
 - 20 Neubig G, Goldberg Y, Dyer C. On-the-fly operation batching in dynamic computation graphs. arXiv preprint arXiv: 1705.07860, 2017.