计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(1):14−21 [doi: 10.15888/j.cnki.csa.007243]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于深度残差网络图像分类算法研究综述①

赵志成[1,2], 罗　泽[1], 王鹏彦[3], 李　健[1]

[1](中国科学院 计算机网络信息中心, 北京 100190)
[2](中国科学院大学, 北京 100049)
[3](四川卧龙国家级自然保护区管理局, 卧龙 623006)
通讯作者: 罗　泽, E-mail: luoze@cnic.cn

摘　要: 近年来, 由于计算机技术的飞速迅猛发展, 特别是硬件条件的改善, 计算能力不断提高, 深层神经网络训练的时间大大缩短, 深度残差网络也迅速成为一个新的研究热点. 深度残差网络作为一种极深的网络架构, 在精度和收敛等方面都展现出了很好的特性. 研究者们深入研究其本质并在此基础上提出了很多关于深度残差网络的改进, 如宽残差网络, 金字塔型残差网络, 密集型残差网络, 注意力残差网络等等. 本文从残差网络的设计出发, 分析了不同残差单元的构造方式, 介绍了深度残差网络不同的变体. 从不同的角度比较了不同网络之间的差异以及这些网络架构在常用图像分类数据集上的性能表现. 最后我们对于这些网络进行了总结, 并讨论了未来深度残差网络在图像分类领域的一些研究方向.

关键词: 图像分类; 深度学习; 计算机视觉; 深度残差网络; 恒等变换; 卷积神经网络

引用格式: 赵志成,罗泽,王鹏彦,李健.基于深度残差网络图像分类算法研究综述.计算机系统应用,2020,29(1):14–21. http://www.c-s-a.org.cn/1003-3254/7243.html

Survey on Image Classification Algorithms Based on Deep Residual Network

ZHAO Zhi-Cheng[1,2], LUO Ze[1], WANG Peng-Yan[3], LI Jian[1]

[1](Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)
[2](University of Chinese Academy of Sciences, Beijing 100049, China)
[3](Sichuan Wolong National Natural Reseve Administration, Wolong 623006, China)

Abstract: Recently, the training time of deep neural network has been greatly shortened because of the rapid development of computer technology especially the improvement of hardware conditions. The deep residual network has rapidly become a new research hotspot. The architecture exhibits good features in terms of precision and convergence. Researchers have delved into its nature and proposed many improvements on deep residual networks, such as wide residual networks, deep pyramidal residual networks, densely residual networks, attention residual network, etc. This study analyzes the construction of different residual units from the design of residual network, and introduces different variants of deep residual network. From different aspects, We compare the differences between different networks and the performance of these network architectures on common image classification datasets. Finally, we summarize these networks and discuss some research directions of future deep residual networks in the field of image classification.

① 基金项目: 中国科学院科技服务网络计划区域重点项目 (Y82E01, KFJ-STS-QYZD-058); 国家研发基础设施和设施发展计划 (DKA2018-12-02-XX); 中国科学院战略性先导科技专项 (XDA19060205); 中科院信息化专项 (XXH13505-03-205, XXH13506-305, XXH13506-303); 中国科学院计算机网络信息中心一三五规划重点培育方向 (CNIC-PY-1408, CNIC_PY_1409); 专项大熊猫国际合作资金

Foundation item: CAS Science and Technology Service Network Initiative (STS Initiative) (Y82E01, KFJ-STS-QYZD-058); National Research and Development Infrastructure and Facility Development Program of China (DKA2018-12-02-XX); Strategy Priority Research Program of Chinese Academy of Sciences (XDA19060205); Special Project of Informatization of Chinese Academy of Sciences (XXH13505-03-205, XXH13506-305, XXH13506-303); Around Five Top Priorities of “One-Three-Five” Strategic Planning of Computer Network Information Center, Chinese Academy of Sciences (CNIC-PY-1408, CNIC_PY_1409); Giant Panda International Cooperation Fund

收稿时间: 2019-06-20; 修改时间: 2019-07-16; 采用时间: 2019-07-19; csa 在线出版时间: 2019-12-27

Key words: image classification; deep learning; computer vision; deep residual network; identity mapping; convolutional neural network

深度神经网络的出现, 使得图像分类领域取得了一系列的突破. 深度神经网络可以通过多级表示, 学习到更加复杂的高级语义特征. 在一个端到端的多层模型中, 低、中、高不同级别的特征以及分类器被整合起来, 特征的等级随着通过所堆叠层的数量 (深度) 逐渐丰富[1–5]. 2012 年, Krizhevsky[6]用 5 个卷积层和 3 个全连接层的网络在 ILSVRC 竞赛上取得了 38.1% 的 Top-1 和 16.4% 的 Top-5 错误率. 2014 年, VGGNet[7]首次将网络的深度增加到了 19 层, 在 ILSVRC 竞赛上取得了 24.7% 的 Top-1 错误率和 7.3% 的 Top-5 错误率. 回顾深度卷积神经网络在图像识别领域的发展历史, 可以清楚地发现深度神经网络的表达能力和提取特征的能力随着网络的深度的增加而增加[8,9].

然而, 网络的深度并非越深越好. 在一定的深度范围内, 随着网络层数的增加, 模型可以拟合更加复杂的函数, 模型性能也可以得到提升. 但是在网络层数增加到一定的数目之后, 继续增加网络的层数, 训练精度和测试精度迅速下降. 研究人员发现, 随着网络深度的增加, 准确率达到饱和后再增加网络的深度分类的效果反而越来越差. 实验表明, 20 层以上的深度网络, 继续叠加增加网络的层数, 分类的精度反而会降低, 50 层网络的测试误差率大概是 20 层网络的一倍[10]. 这种网络的退化现象表明了直接增加深度的深度学习系统并不是很容易被优化.

为了解决由于深度增加带来的网络退化问题, 2015 年微软亚洲研究院的 He 等人[11]提出了深度残差网络 (deep residual network). 在残差学习 (residual learning) 的启发下, 网络中引入了恒等映射的设计, 巧妙的缓解了由于深度增加带来的梯度爆炸或梯度消失以及网络退化的问题, 提升了信息传递路径的数量, 使得网络的深度可以由几十层推到千层. 深度残差网络的出现极大的提高了系统的准确率, 使得训练极深的网络成为可能, 是图像分类领域具有重要意义的突破性进展.

由于深度残差网络的优良特性, 它被应用到多个领域, 例如人脸识别[12]、目标检测[13]、行人检测[14]、语义分割[15]、自然语言处理[16]等, 取得了很好的效果. 本文将以深度残差网络在图像分类领域的进展为线索, 简要阐述其成功的原因, 介绍深度残差网络在图像分类领域的一些研究进展, 比较这些不同的网络在图像分类数据集上的性能表现, 探索未来研究的一些方向.

1 深度残差网络概述

1.1 深度残差网络介绍

直接增加网络的深度, 会使得深度学习系统很难被优化. 假设我们现在有一个浅层的网络, 那么应该存在一个深层的网络: 它是在由浅层网络的基础上堆叠了多个$x \to x$(恒等映射) 的映射构成的, 那么该深层神经网络的性能至少应该不会比浅层网络的性能差. 然而实验却证明找不到这样一个我们理想当中的深层神经网络, 这种现象说明通过直接叠加深度的方式来拟合这样一个$x \to x$恒等映射是非常困难的.

在浅层网络已经达到了饱和之后, 通过在它后面再加上恒等映射层 (identity mapping), 不仅可以使得网络的深度增加使得模型有更强的表现能力, 而且保证了系统的误差不会随着深度增加而增加[11].

假设原始神经网络的一个残差单元要学习的目标映射为$H(x)$, 这个目标映可能很难学习. 残差神经网络让残差单元不直接学习目标映射, 而是学习一个残差$F(x) = H(x) - x$. 这样原始的映射变成了$F(x) + x$. 原始残差单元可以看做是由两部分构成, 一个线性的直接映射$x \to x$和一个非线性映射$F(x)$. 特别地, 如果$x \to x$是最优的学习策略, 那么相当于把非线性映射$F(x)$的权重参数设置为 0. 恒等映射使得非线性映射$F(x)$学习线性的 $x \to x$ 映射变得容易很多.

深度残差网络的基本组成单元是残差单元, 残差单元一般由卷积 Conv 层, 批处理归一化 Batchnorm 层和非线性激活函数 Relu 共同构成. 图 1 给出了原始残差单元的示意图, 令第l个残差单元的输入为x_l, 那么下一层的输出为:

$$x_{l+1} = f(x_l + F(x_l, W_l)) \tag{1}$$

其中, $F(x_l, W_l)$是残差函数, W_l是该残差函数对应的权重参数; $f(*)$是非线性激活函数 Relu.

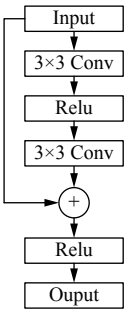

图 1 原始残差学习单元结构图

深度残差网络由多个残差学习单元堆叠而成. 给定输入的图像数据, 深度残差网络首先将输入数据依次送入卷积层 Conv、非线性激活函数层 Relu 和批处理归一化层 Batchnorm; 然后将处理的结果进一步送入到多个残差单元, 再经过批处理归一化层 BN 和多个全连接层; 最后得到输出结果.

在构建"超深"的网络比如 100 层以上的网络时, 将多个原始残差学习模块直接堆叠起来会造成参数的爆炸. 为了在不损失精度的情况下降低深层网络的的参数量, He 等人[10]提出了一种称为"瓶颈"(bottleneck) 的残差模块结构, 如图 2. 这种瓶颈结构的设计主要目的是为了减少参数量从而减少计算量, 使得深度残差网络的训练速度加快.

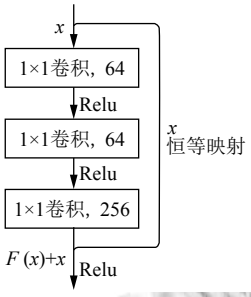

图 2 Bottleneck 结构

1.2 残差单元的不同构造方式

在原始残差单元的基础上, 研究者对于 Batchnorm[17] (批归一化) 和激活函数 Relu 的放置位置和组合方式进行了不同的探索. 研究发现, 残差单元不同的构造方式对于精度、网络的收敛以及训练的速度都有影响. He 等[10]在 2016 年提出了 Pre-activation (预激活) 的残差单元, 提高了模型的泛化能力, 减少了过拟合的影响. Han 等[18]在此基础上, 对于残差单元进行了更多的测试, 不同残差单元的结构图如图 3 所示, 不同残差单元在分类任务下的表现结果如表 1 所示.

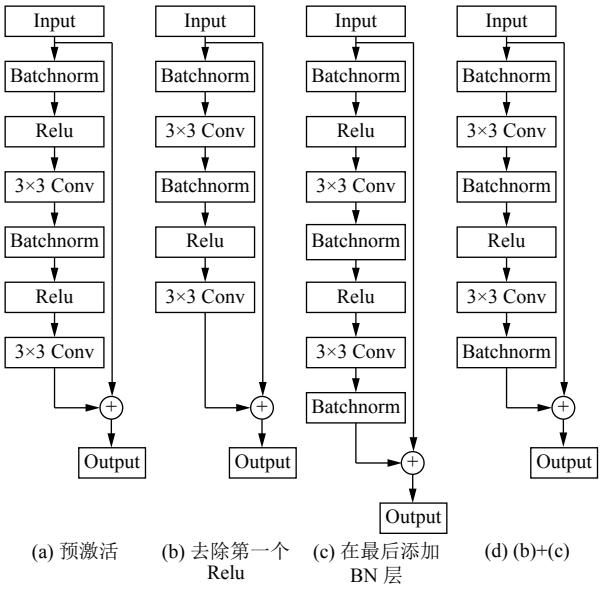

图 3 不同残差单元结构图

表 1 不同残差单元错误率对比 (单位: %)

不同残差单元	CIFAR-10	CIFAR-100
(a) 预激活	5.82	25.06
(b) 去除第一个 Relu	5.31	24.55
(c) 在最后添加 BN 层	5.74	24.54
(d) (b)+(c)	**5.29**	**23.74**

2 常用图像分类数据集介绍

CIFAR-10 和 CIFAR-100 是图像分类领域常用的比较性能的数据集. CIFAR-10 数据集由来自 10 个类的 60 000 个 32×32 彩色图像组成, 其中每个类包含 6000 个图像, 有 50 000 个训练图像和 10 000 个测试图像[19], 如图 4.

CIFAR-100 数据集是 CIFAR-10 的扩展数据集, 它包含的类别数更多 (100 个类别), 每个类别中有 600 张图像 (500 张训练图像和 100 张测试图像). 除此之外, CIFAR-100 中的 100 个小类还可以被分成 20 个大类. 数据集中的每个图像都带有一个"精细"的类别标签 (它所属的小类) 和一个"粗糙"标签 (它所属的大类).

3 不同变体在图像分类任务中的表现

在原始深度残差网络的启发下, 研究者们从不同的角度出发 (例如继续增加网络的深度、增加的网络的宽度、更细化的网络架构设计、引入注意力机制等等), 为了增加模型的表示能力和泛化能力, 提高模型提取特征的能力, 对于深度残差网络进行了不同方向

的改进, 同时在 CIFAR-10, CIFAR-100 这两个图像分类数据集上进行了性能验证.

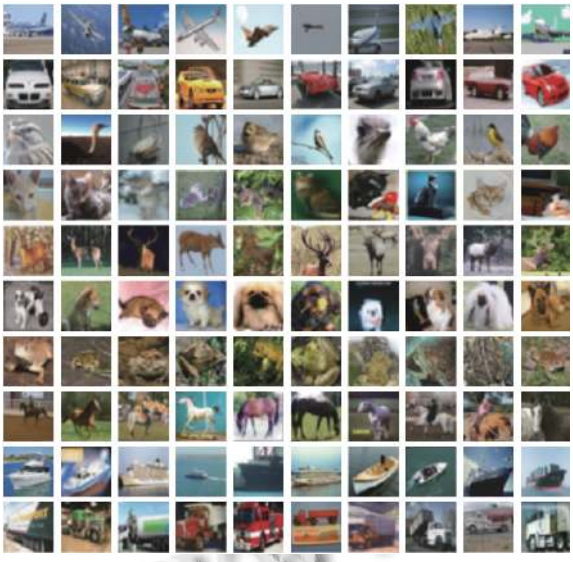

图 4 CIFAR10 数据集示例图[19]

3.1 更宽的残差网络

Zhang 等人[20]的研究表明, 不同于以往不断的叠加网络的深度, 网络的宽度作为一个重要的维度也需要被关注. 随着模型深度的加深, 在梯度进行反向传播时, 并不能保证能够流经每一个残差单元 (Residual unit) 的权重层, 绝大多数的残差单元只能提供很少的信息, 只有少数几个残差单元能够学到有用的表达提取到有用的特征[21,22]. 这里的“宽度”指的是特征映射的通道数, 在卷积层中指的式增加卷积核的个数. Sergey 等[23]从增加网络的“宽度”入手, 提出了 Wide Residual Netork (WRN), 使用一种较浅的但是宽度更宽的模型, 来更加有效的提升模型的性能.

宽残差网络的实验结果表明, 宽度可以提高特征的复用, 能够带来网络表现能力和泛化能力的提升, 同样的参数数量, 宽的网络训练训练速度也更快. WRN 的具体构造如表 2 所示, 不同深度的网络在数据集上的表现如表 3 所示.

3.2 “金字塔”型残差网络

Veit 等人[24]的实验证明深度残差网络其实可以看做是相对较浅网络的集合. 他们的研究表明, 从残差网络中删除一个单独的残差单元, 即只保留一个恒等映射, 不会对整体性能产生明显的影响, 经过不同的实验证明删除残差单元相当于删除集成的网络中的一些浅层的网络. 但是在普通网络如 VggNet 和 GoogleNet 中, 删除任意一个网络层都会导致网络的性能的骤降.

表 2 k 倍宽残差网络具体架构设计

组名	输出大小	残差单元
Conv1	32×32	[3×3,16]
Conv2	32×32	$\begin{bmatrix} 3\times3,16\times k \\ 3\times3,16\times k \end{bmatrix}\times N$
Conv3	16×16	$\begin{bmatrix} 3\times3,32\times k \\ 3\times3,32\times k \end{bmatrix}\times N$
Conv4	8×8	$\begin{bmatrix} 3\times3,64\times k \\ 3\times3,64\times k \end{bmatrix}\times N$
Avg-Pool	1×1	[8×8]

表 3 不同配置的宽残差网络实验结果

深度	k	参数量 (M)	CIFAR-10 (%)	CIFAR-100 (%)
40	1	0.6	6.85	30.89
40	2	2.2	5.33	26.04
40	4	8.9	4.97	22.89
40	8	35.7	4.66	–
28	**10**	**36.5**	**4.17**	**20.50**
28	12	52.5	4.33	20.43
22	8	17.2	4.38	21.22
22	10	26.8	4.44	20.75
16	8	11.0	4.81	22.07
16	10	17.1	4.56	21.59

深度残差网络下一般在下采样残差单元中将通道数加倍. 实验发现, 在深度残差网络中分别删除下采样功能的残差单元和非下采样的残差单元, 删除下采样的残差单元会造成网络更多的性能下降. 为了更好的解决这个问题, Han 等人[18]设计了一个“金字塔”残差网络 (Pyramidal residual Network, PyramidNet). 如图 5 所示, 不同于以往在下采样单元时网络突然加倍, 该网络的宽度随深度的增加而逐渐增加, 这种形状类似于从顶部向下逐渐变宽的金字塔结构. 金字塔型残差网络具体结构如表 4, 分类错误率对比如表 5.

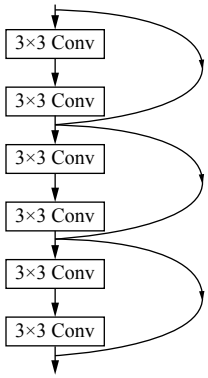

图 5 金字塔型网络残差单元

表 4 金字塔型残差网络具体架构

组名	输出大小	残差单元
Conv1	32×32	[3×3,16]
Conv2	32×32	$\begin{bmatrix} 3\times3,16+\alpha(k-1)/N \\ 3\times3,16+\alpha(k-1)/N \end{bmatrix}\times N_2$
Conv3	16×16	$\begin{bmatrix} 3\times3,16+\alpha(k-1)/N \\ 3\times3,16+\alpha(k-1)/N \end{bmatrix}\times N_3$
Conv4	8×8	$\begin{bmatrix} 3\times3,16+\alpha(k-1)/N \\ 3\times3,16+\alpha(k-1)/N \end{bmatrix}\times N_4$
Avg-Pool	1×1	[8×8,16+α]

表 5 金字塔型残差网络分类错误率对比

PyramidNet	参数量 (M)	CIFAR-10 (%)	CIFAR-100 (%)
PyramidNet(α=48)	1.7	4.58	23.12
PyramidNet(α=84)	3.8	4.26	20.66
PyramidNet (α=270)	**28.3**	**3.73**	**18.25**
PyramidNet (bottleneck,α=270)	27.0	3.48	17.01
PyramidNet (bottleneck,α=240)	26.6	3.44	16.51
PyramidNet (bottleneck,α=220)	26.8	3.40	16.37
PyramidNet (bottleneck,α=200)	**26.0**	**3.31**	**16.35**

金字塔残差网络的通道数的具体计算公式如式 (2), 其中 k 代表第 k 层, N 代表总的层数, D_k 代表第 k 层的通道数, α 代表最后一层输出通道数.

$$D_k=\begin{cases}16 & \text{if } k=1\\ D_{k-1}+\dfrac{\alpha}{N} & \text{if } 2\le k\le N+1\end{cases} \tag{2}$$

这种网络设计可以有效的改善过拟合的问题, 与其他网络架构设计相比, 它显示出很好的泛化能力; 而且在金字塔型的残差网络中, 删除具有下采样功能的残差单元不会降低性能.

3.3 密集型网络

受到 ResNet 将输入和输出相加形成残差结构的启发, Huang 等人[21]设计出一种将输出与输入并联到一起的网络架构, 实现了每一层都能直接得到之前所有层的输出的密集型卷积网络 (Densely convolutional Network, DenseNet). 该网络可以有效的缓解梯度消失的问题, 增加特征的重用性, 并大幅减少参数数量. 在这种新型网络架构中, 每层的输入由所有之前层的特征映射组成, 其输出将传输给每个后续层.

在原始的深度残差网络中, 恒等映射的输出是通过加法结合起来的. 在这种情况下, 如果两个层的特征映射的分布差异性很大的话, 这有可能会影响特征的重用同时阻碍信息流的传播. 密集型网络 (DenseNet) (如图 6) 通过将特征映射级联而不是将特征映射直接相加, 可以在保留所有特征映射的同时增加输出的多样性, 促进特征被重用. 实验证明, 在相同的参数量下密集型网络具备更高的参数效率, 有更好的收敛效果. 表 6 是不同增长率 k 下的分类错误率.

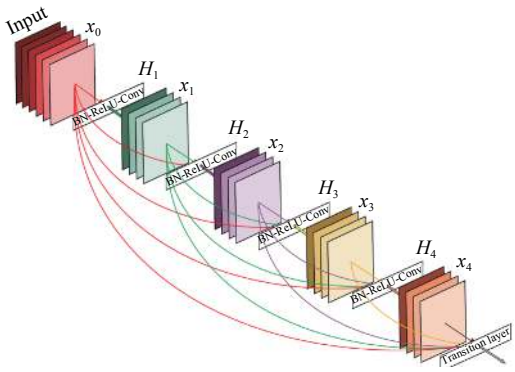

图 6 密集型网络结构图[21]

表 6 不同增长率 k 下分类错误率

DenseNet	参数量 (M)	CIFAR-10 (%)	CIFAR-100 (%)
Densenet (k=12)	0.8	4.51	22.27
Densenet (k=24)	15.3	3.62	17.60
Densenet (k=40)	**25.6**	**3.46**	**17.18**

3.4 深度注意力残差网络

注意力机制在计算机视觉中也起着重要的作用, 注意力机制不止能使得运算聚焦于特定的区域, 同时也可以使得该部分区域的特征重要性得到增强. 为了在深度残差网络中引入注意力的机制, Wang 等[22]提出了残差注意力网络 (Residual Attention Network, RAN).

一个注意力残差单元如图 7 所示, 分为两个分支, 右边的分支就是普通的卷积网络, 即主干分支, 叫做 Trunk Branch. 左边的分支是为了得到一个掩码 mask, 该掩码的作用是得到输入特征 x 的 attention map, 所以叫做 Mask Branch, 这个 Mask Branch 包含 down sample 和 upsample 的过程, 目的是为了保证和右边分支的输出大小一致.

注意单元的计算公式如式 (3), 其中 $M(x)$ 为 Mask Branch 的输出, $F(x)$ 为主分支的输出. 借鉴了 ResNet 中恒等映射的思想, 当掩码分支 $M(x)$=0 时, 该层的输入就等于 $F(x)$, 所以该层的效果至少不会比原始的 $F(x)$ 差, 残差单元更容易被优化. 同时掩码分支的设计, 使得特征图可以学习到不同大小的权重值, 进而让主干分支输出的特征图中显著的特征更加显著, 增加了特征的判别性.

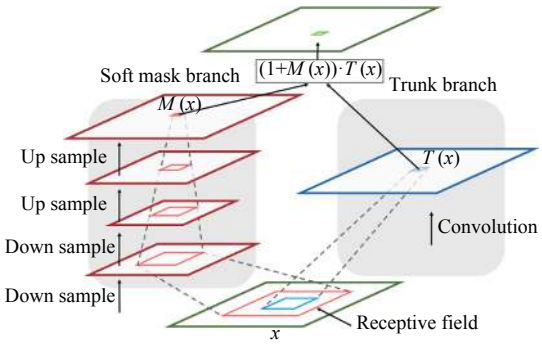

图 7 注意力残差单元结构图[22]

$$H(x) = (1 + M(x) * F(x)) \tag{3}$$

残差注意力模型不仅可以提升网络的性能, 还具有很强的扩展性, 可以结合到目前的大部分深层网络中, 做到端到端的训练. 因为残差结构的存在, 也可以很容易将网络扩展到百数层. 实验证明, 使用该种策略可以在达到其他大网络的分类准确率的同时显著降低计算量 (计算量基本上为原始 ResNet 深层网络的 69% 左右), 具体的实验结果如表 7 所示.

表 7 注意力残差网络错误率对比

网络架构	参数量 (M)	CIFAR-10 (%)	CIFAR-100 (%)
Attention-92	1.9	4.99	21.71
Attention-236	5.1	4.14	21.16
Attention-452	**8.6**	**3.90**	**24.54**

3.5 随机深度残差网络

深度残差网络的由于网络更深或者更宽, 网络的参数量很大, 往往非常容易造成过拟合, 模型在训练集上表现很好, 在测试集上却表现很差. 为了解决过拟合的问题, 同时受到 Dropout 思想的启发, 随机深度残差网络 (ResDrop)[25]在训练时使用伯努利随机变量, 随机使得一部分的残差单元"失活", 使得网络不依赖于某个特定的残差单元, 起到一部分正则化的效果. 和 Dropout 类似, 在进行测试时使用整个网络进行预测.

在训练期间, 随机深度残差网络的深度会减小, 进而会导致前向传播和反向传播的深度变短, 所以其训练时间不会随着深度残差网络的深度而线性地增加.

此外, 训练期间网络深度的减少会增强前边层参数的梯度更加有利于梯度的传播, 这将使得 1000 层以上的随机深度残差网络能够正常训练. 随机深度的残差网络可以被看做不同深度网络的集成[24],与恒定深度的深度残差网络相比不易过拟合. 随机深度残差网络在 CIFAR-10 和 CIFAR-100 上分别取得了 5.25% 和 24.98% 的错误率.

4 深度残差网络改进总结

深度残差网络一直是图像分类领域研究的热点. 自从深度残差网络被提出以来, 研究者们为了提升深度残差网络的表征能力和泛化能力, 提高在分类任务上的表现, 研究出了多个改进的版本[26–28]. 这些改进或变体可以大体可以分成基于残差单元的优化改进, 基于整体网络结构的设计的改进、加入 attention 机制 3 种.

基于残差单元的改进主要是通过修改残差单元的不同层的摆放位置和修改残差单元的残差函数. Zhang 在残差单元中加入 Dropout[29]层取得了更好的表现, Xie 等人[30]引入了一个"基数"的超参数通过增加残差单元独立路径的数量提高了准确率, 在此基础上 Gastaldi 提出 Shake-Shake 正则化残差网络[31], 采用随机仿射组合替换并行分支的标准求和来提高多分支网络的泛化能力.

基于整体网络结构设计的改进的研究是指改变网络结构的整体框架. 通过改进深度残差网络的架构使得梯度更加容易传播, 模型的表示能力更强, 残差网络更容易优化. Zhang 等人[32]在原始残差网络的基础上增加了一个层级的快捷连接构建了一个多级网络, Yamada 等人[33]进一步把随机深度引入到"金字塔"残差网络框架中, 提出了 PyramidSepDrop 网络模型.

将 Attention 机制引入深度残差网络是目前研究的热点方向之一. Squeeze and excitation networks[34]认为不同的特征映射通道的重要性不同, 在他们的压缩和激励模块中, 他们使用全局平均池化 (Global Average Pooling) 来计算通道的注意力 (权重值). Woo 等人[35]在此基础上, 提出了卷积注意力模块 CBAM (Convolutional Block Attention Module), 利用一个有效的结构设计来结合空间 (feature map) 和通道的注意力, 通过将空间注意力和通道注意力结合取得了在不同的数据集上取得了更好的性能.

此外, 还有一些研究者将 3 种方法混合也取得了很好的效果, 例如 Tan 等人[36]通过混合改进在 CIFAR-10 和 CIFAR-100 上分别取得了 1.1% 和 8.3% 的错误率, 不同的深度残差网络性能表现如表 8 所示.

表 8 深度残差网络性能对比

网络架构	参数量 (M)	CIFAR-10 (%)	CIFAR-100 (%)
WRN-28-10	36.5	4.17	20.5
PyramidNet(bottleneck,α=270)	26.0	3.31	16.35
Densenet (k=40)	25.6	3.46	17.18
Attention-452	8.6	3.90	24.54
Stoc-depth-110	1.7	5.25	24.98
EfficientNet-B7	**64**	**1.1**	**8.3**

5 结论

深度残差网络的出现, 极大的提高了深度学习的表征能力和学习能力, 成为图像分类领域研究的热点方向.

本文分析了深度残差网络和其变体, 比较不同模型在常用图像分类数据集上的性能表现, 通过分析可见在图像分类领域深度残差网络已有一定的研究成果. 鉴于目前深度残差网络和其变体还存在收敛速度慢、训练时间长、网络参数冗余、网络设计复杂、对于数据需求量大依赖人为标注等缺点, 未来的研究方向在于:

(1) 减少深度残差网络的参数, 在不损失精度的情况下对于深度残差网络进行有效的压缩. 深度残差网络由于在宽度和深度上增加了很多, 会产生很多的冗余参数, 如何在不损失性能的情况下减少深度残差网络的参数量从而提高深度残差网络的计算性能是个具有现实意义的问题.

(2) 在数据量较小的情况下, 获得更好的性能. 目前在图像分类领域深度残差网络的精度仍然依赖于数据集样本的多少, 数据增强的策略等. 在数据标注不足的情况下, 如何获得相同的性能也是一个值得关注的问题. 更少的依赖有监督学习和人类的先验标注信息, 将无监督学习或者强化学习和深度残差网络结合值得我们不断的探索.

(3) 增强深度残差网络的学习能力和泛化能力. 深度残差网络的参数量往往很大, 模型往往在训练集上效果效果很好, 在测试集上效果很差, 如何防止过拟合使得模型可以很好地泛化是一个值得研究的问题. 另外, 在现有基础上, 改进残差单元和残差网络的结构、引入注意力机制以及混合改进等, 使得深度残差网络在分类任务上取得更高的准确率是值得深入探索的核心问题.

参考文献

1 温煌璐. 基于卷积神经网络的图像分类算法研究[硕士学位论文]. 乌鲁木齐: 新疆大学, 2018.

2 郭玥秀, 杨伟, 刘琦, 等. 残差网络研究综述. 计算机应用研究, 1–8. https://doi.org/10.19734/j.issn.1001-3695.2018.12.0922. [2019-10-27].

3 张帆, 张良, 刘星, 等. 基于深度残差网络的脱机手写汉字识别研究. 计算机测量与控制, 2017, 25(12): 259–262.

4 陈晨, 刘明明, 刘兵, 等. 基于残差网络的图像超分辨率重建算法. 计算机工程与应用: 1–8. http://kns.cnki.net/kcms/detail/11.2127.TP.20190527.1720.009.html. [2019-10-27].

5 马慧. 基于卷积神经网络的图像分类技术及应用[硕士学位论文]. 江门: 五邑大学, 2018.

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 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556v1, 2014.

8 Wu ZF, Shen CH, Van Den Hengel A. Wider or deeper: Revisiting the ResNet model for visual recognition. Pattern Recognition, 2019, 90: 119–133. [doi: 10.1016/j.patcog.2019.01.006]

9 Szegedy C, Liu W, Jia YQ, *et al*. Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.

10 He KM, Zhang XY, Ren SQ, *et al*. Identity mappings in deep residual networks. In: Leibe B, Matas J, Sebe N, *et al*., eds. Computer Vision – ECCV 2016. Cham: Springer, 2016. 630–645.

11 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. 2016. 770–778.

12 杨雨浓. 基于深度学习的人脸表情识别方法研究[博士学位论文]. 杨凌: 西北大学, 2017.

13 宋伟, 纳鹏宇, 刘宁宁. 基于双目视觉系统的运动目标跟踪算法研究. 北京交通大学学报, 2013, 37(6): 13–17, 26. [doi: 10.3969/j.issn.1673-0291.2013.06.003]

14 苏松志, 李绍滋, 陈淑媛, 等. 行人检测技术综述. 电子学报, 2012, 40(4): 814–820. [doi: 10.3969/j.issn.0372-2112.2012.04.031]

15 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition. Boston, MA, USA. 2015. 3431–3440.
16 王灿辉, 张敏, 马少平. 自然语言处理在信息检索中的应用综述. 中文信息学报, 2007, 21(2): 35–45. [doi: 10.3969/j.issn.1003-0077.2007.02.006]
17 Ioffe S. Batch renormalization: Towards reducing minibatch dependence in batch-normalized models. Advances in Neural Information Processing Systems, 2017: 1942–1950.
18 Han D, Kim J, Kim J. Deep pyramidal residual networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 6307–6315.
19 Hoffer E, Ailon N. Deep metric learning using triplet network. Proceedings of International Workshop on Similarity-Based Pattern Recognition. Copenhagen, Denmark. 2015. 84–92.
20 Zhang CY, Bengio S, Singer Y. Are all layers created equal? arXiv: 1902.01996, 2019.
21 Huang G, Liu Z, Van Der Maaten L, *et al*. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 2261–2269.
22 Wang F, Jiang MQ, Qian C, *et al*. Residual attention network for image classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 6450–6458.
23 Zagoruyko S, Komodakis N. Wide residual networks. arXiv: 1605.07146, 2016.
24 Veit A, Wilber MJ, Belongie S. Residual networks behave like ensembles of relatively shallow networks. Advances in Neural Information Processing Systems. San Francisco, CA, USA. 2016.550–558.
25 Huang G, Sun Y, Liu Z, *et al*. Deep networks with stochastic depth. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 646–661.
26 Lin M, Chen Q, Yan SC. Network in network. arXiv: 1312.4400v3, 2013.
27 Krizhevsky A, Hinton G. Convolutional deep belief networks on CIFAR-10. Unpublished Manuscript, 2010. https://www.cs.toronto.edu/~kriz/conv-cifar10-aug2010.pdf.
28 Szegedy C, Vanhoucke V, Ioffe S, *et al*. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. 2818–2826.
29 Srivastava N, Hinton G, Krizhevsky A, *et al*. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014, 15: 1929–1958.
30 Xie S, Girshick R, Dollár P, *et al*. Aggregated residual transformations for deep neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 5987–5995.
31 Gastaldi X. Shake-shake regularization. arXiv: 1705.07485, 2017.
32 Zhang K, Sun M, Han TX, *et al*. Residual networks of residual networks: Multilevel residual networks. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(6): 1303–1314. [doi: 10.1109/TCSVT.2017.2654543]
33 Yamada Y, Iwamura M, Kise K. Deep pyramidal residual networks with separated stochastic depth. arXiv: 1612.01230, 2016.
34 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7132–7141.
35 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich, Germany. 2018. 3–19.
36 Tan MX, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv: 1905.11946, 2019.