



吉林大学学报(工学版)
Journal of Jilin University(Engineering and Technology Edition)
ISSN 1671-5497,CN 22-1341/T

《吉林大学学报(工学版)》网络首发论文

题目: 简化型残差结构与快速深度残差网络
作者: 杨怀江, 王二帅, 隋永新, 闫丰, 周跃
DOI: 10.13229/j.cnki.jdxbgxb20210027
收稿日期: 2021-01-13
网络首发日期: 2021-05-17
引用格式: 杨怀江, 王二帅, 隋永新, 闫丰, 周跃. 简化型残差结构与快速深度残差网络. 吉林大学学报(工学版). <https://doi.org/10.13229/j.cnki.jdxbgxb20210027>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

简化型残差结构与快速深度残差网络

杨怀江^{1,2}, 王二帅^{1,3}, 隋永新^{1,2}, 闫丰^{1,2}, 周跃^{1,2}

(1. 中国科学院 长春光学精密机械与物理研究所, 长春 130033; 2. 长春国科精密光学技术有限公司, 长春 130033; 3. 中国科学院大学, 北京 100049;)

摘要：当前深度残差网络模型训练缓慢，为解决这一问题该文设计了一种新型的残差结构。与典型的残差结构相比，该结构仅含有一个 Batch Normalization 和 ReLU 模块，通过减少网络训练过程的计算量降低了耗时，提升了模型训练速度。在常用的 CIFAR10/100 图像分类数据库上进行了对比实验分析，以该方法构建的深度为 110 层的网络 CIFAR10 分类错误率为 5.29%，CIFAR100 分类错误率为 24.80%，典型的 110 层深度残差网络分类错误率分别为 5.75% 和 26.02%；在训练耗时方面，该方法平均周期耗为 133.47s，典型的残差网络平均周期耗为 208.26s，提升了 35.91%；结果表明该网络结构在保证分类性能的基础上极大地提升了训练速度，具有较好的实用价值。

关键词：图像处理；图像识别；图像分类；卷积神经网络；深度残差网络

中图分类号：TP183

文献标志码：A

DOI: 10.13229/j.cnki.jdxbgxb20210027

Simplified residual structure and fast deep residual networks

YANG Huai-jiang^{1,2}, WANG Er-shuai^{1,3}, SUI Yong-xin^{1,2}, YAN Feng^{1,2}, ZHOU Yue^{1,2}

(1. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China*; 2. *Changchun National Extreme Precision Optics Co., Ltd, Changchun 130033, China*; 3. *University of Chinese Academy of Sciences, Beijing 100049, China*)

Abstract: Training time of deep ResNets is painfully slow. To address this problem, this paper proposes a novel residual structure. Compared with the typical residual structure, the structure only contains a Batch Normalization and ReLU, which reduces training time and improves the training speed by reducing the amount of calculation in the network training process. The comparative experiments are carried out on the CIFAR10/100 image classification database. The classification error rate of 110 layers networks constructed by this method on CIFAR10 and CIFAR100 is 5.29% and 24.80%, respectively. The classification error rate of 110-ResNet is 5.75% and 26.02%, respectively. Training the network takes 133.47 (this method) and 208.26 (ResNet) seconds per epoch,

收稿日期：2021-01-13.

基金项目：国家重点研发计划项目（2018YFF01011900）.

作者简介：杨怀江（1966-），男，研究员，博士.研究方向：光学信息融合，深紫外光刻技术.E-mail:994301018@qq.com

increased by 35.91%. The results show that the network structure greatly improves the training speed while ensuring the classification performance, and has better practical value.

Key words: image process; image recognition; image classification; convolutional neural networks; deep residual networks

0 引言

自 2012 年以来,卷积神经网络方法在图像处理领域得到广泛应用,取得了远超其他方法的优异成绩,是当前图像分类、识别、检测领域的研究热点。2015 年由何凯明团队提出的深度残差网络^[1](deep residual networks,ResNets)是卷积神经网络中最为重要的一种模型,该模型取得了 ImageNet 和 COCO2015 图像竞赛中分类、识别和检测等多个项目的第一名,引起了广泛学者们对该模型的重视,随后人们开展了关于 ResNets 的大量研究工作。2016 年,何凯明团队提出了 pre-activation ResNets 模型^[2],对 ResNets 模型的残差结构中激活层和卷积层的顺序做出调整,进一步提高了 ResNets 模型的收敛速度和性能; Sasha 等人提出的 Resnet in Resnet^[3],将 ResNets 推广为多个残差分支结构; Sergey 等人提出了 Wide-ResNets^[4],通过大幅增加残差结构中卷积核数量,提升了网络的分类性能; Veit 等人深入分析了 ResNets,认为该模型表现为大量较浅网络的集合体^[5,6],其性能主要依赖深度为 10 至 34 层的部分网络;国内方面有大量研究人员将 ResNets 模型用于交通信号灯^[7,8]、鱼类图像^[9]、医学图像^[10,11]、人脸识别^[12-14]、手势识别^[15]、昆虫识别^[16,17]等多种特定图像分类任务中。这些关于 ResNets 的研究主要探讨了 ResNets 的原理,推广了 ResNets 模型在图像分类识别领域的应用。

然而当前训练 ResNets 模型仍面临着耗时长的问题,ResNets 训练耗时过长消耗了大量的计算资源和时间成本,限制了 ResNets 在图像分类、图像识别领域的应用。例如在 ImageNet 数据库采用多块高性能的 GPU 同时训练深度 152 层的 ResNets 也要花费数周时间。目前减少训练用时的方法主要是微调法和 stochastic depth 法,微调法即以其他大型数据

库预先训练的公开模型为基础,利用任务数据进一步训练,对模型参数做微调。当任务数据与公开模型数据库差异较大时,其分类效果往往不太理想。Stochastic depth^[18]是黄高等人提出的,在模型训练过程中随机跨过若干残差结构,通过控制每个残差结构的保存概率变相缩减 ResNets 深度,以此达到提升网络训练速度的目的;在测试时采用所有残差块,相当于构成一组具有不同深度的 ResNets 网络,保证了模型的性能。

本文在 pre-activation ResNets 的基础上细致地探讨了残差结构中激活层的功能与必要性,提出了仅含有一层 Batch Normalization 和 ReLU 单元的简化型残差结构,并以此构建深度残差网络,在 CIFAR 数据库上测试了这种新型残差结构的性能,结果表明该网络结构在保证分类性能与传统 ResNets 相当的同时,其训练耗时大幅下降。

1 ResNets 模型

经典的 ResNets 系列模型是何凯明等人提出的,它由多个残差结构堆叠而成,其中残差结构可以用公式(1)表示:

$$Y = X_i + F(X_i, W_i) \quad (1)$$

式中 Y 和 X_i 分别是第 $i(i=1,2,3,\dots)$ 残差块的输出和输入, F 和 W_i 分别是第 i 残差结构的残差函数和相关参数。

1.1 残差结构

残差结构主要有基本型和瓶颈型两种,基本型包含两个依次相连的激活层—卷积层结构,瓶颈型包含三个依次相连的激活层—卷积层结构,其中激

活层包括批规范化(Batch Normalization, BN)部分和线性整流单元(Rectified linear units, ReLU)部分, 如图(1)所示。

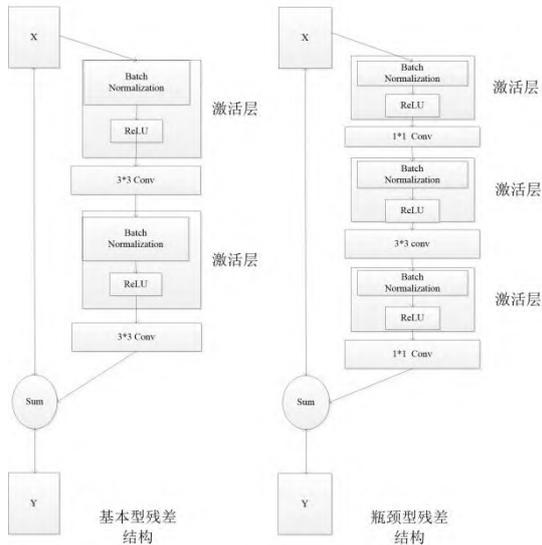


图 1 基本型与瓶颈型残差结构

Fig. 1 Basic and bottleneck residual structure

与基本型残差结构相比, 瓶颈型残差结构包含更多的卷积层和激活层, 以此构建的深度残差网络耗时更多, 因此本文仅考虑基本型残差结构。基本型残差函数可以用式(2)表示:

$$F(X_i, W_i) = W_{i,2}(f_2(W_{i,1}f_1(X_i))) \quad (2)$$

式中 W_i 和 f_i 代表残差结构中的卷积层参数和激活函数。卷积层通过对特征图层各像素点邻域内的若干特征值做加权计算, 在训练过程中按误差反向传播, 计算各参数的偏导数, 据此更新参数, 最终达到特征提取的目的。以常用的 3×3 卷积核为例, 对应的卷积计算和参数更新如式(3)所示:

$$y_{c,x,y} = \sum_{chn=1}^N \sum_{i=-1}^1 \sum_{j=-1}^1 w_{c,chn,i,j} x_{chn,x+i,y+j} + b_c$$

$$\Delta w_{c,chn,i,j} = \sum_{x,y} \Delta y_{c,x,y} x_{chn,x+i,y+j}$$

$$\Delta b_c = \sum_{x,y} \Delta y_{c,x,y}$$

$$\Delta x_{chn,x,y} = \sum_{c=1}^N \sum_{i=-1}^1 \sum_{j=-1}^1 \Delta y_{c,x-i,y-j} w_{c,chn,i,j}$$

1.2 残差结构中的激活单元

残差结构中激活层由 BN 和 ReLU 组成, 下面分别讨论这两种结构元素的功能。

1.2.1 ReLU

关于神经系统科学的研究^[19]指出皮质神经元的激活机制能够用整流器近似, 这一想法在卷积神经网络中得以应用。AlexNet^[20]首次在卷积神经网络中应用了 ReLU 激活函数, 具体表现为式(4), 其效果如图(2)所示。

$$f(x) = \max(0, x) \quad (4)$$

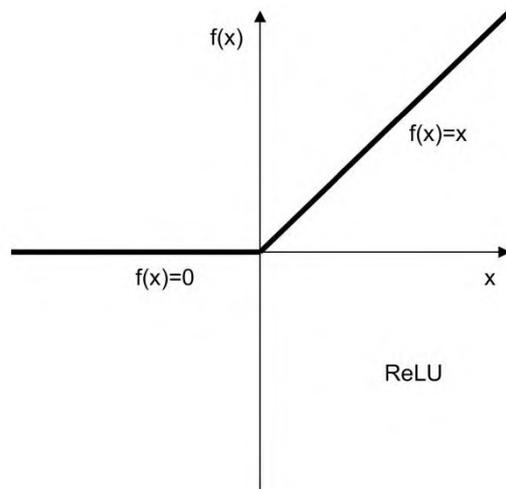


图 2 ReLU 示意图

Fig.2 Schematic diagram of ReLU

特征数据经过 ReLU 单元, 产生负值的部分被强制

输出为 0，正值部分保持不变。采用 ReLU 比采用 sigmoid 和 tanh 等激活方法的网络模型收敛速度快很多。ReLU 由于其实现简单，收敛快的特点被 NiN^[21]、AllCNN^[22]、Inception^[23]、Resnet 等诸多网络模型采用。在经典的残差结构当中，每个激活部分都包含一个 ReLU 层，其目的是提升卷积神经网络的非线性能力。然而，如文献^[2]所提到，在残差结构尾部附加 ReLU，会导致残差函数仅能产生非负偏置，影响残差网络性能，严重时会使网络无法收敛。该文指出 ReLU 置于卷积层之前能有效提高残差结构提取非线性特征的能力，从而促进残差网络的收敛，改善其分类性能。

1.2.2 BN

卷积神经网络中 BN 结构^[24]是 Sergey 等人在 2015 年提出的，该结构将卷积层的输入正规化为近似高斯分布，从而使得训练更为容易，在参数初始化、学习率设置方面也更为宽松，提高了网络性能。随后 BN 在 VGG^[25]、Inception、ResNet 等模型中得到广泛应用。

记 X_i 和 Y_i 为 BN 操作对应的输入和输出，则 BN 可以用式(5)表示：

$$\hat{X}_i = \frac{X_i - E(X_i)}{\sqrt{\text{Var}(X_i)}} \quad (5)$$

$$Y_i = \text{BN}_{\gamma, \beta}(X_i) = \gamma \hat{X}_i + \beta$$

式中 E , Var , γ , β 分别代表期望、方差、尺度缩放和偏置参数，在参数初始化时 γ , β 被设置为 1 和 0。网络训练期间，BN 对需要计算输入图层的期望和方差，并据此调节 γ , β ，最终达到稳定模型的功能。进一步的研究表明^[26]，BN 方法成功的根本原因在于 BN 是网络模型参数优化问题下的参数重新确定过程，该过程使得参数解空间更为平滑和稳定。

当前的卷积神经网络模型中 BN 与卷积层一般是组合使用，然而在正向特征传递过程的每个批次，BN 都要重新计算特征图层输入的期望和方差，在

反向梯度传输过程的每个批次，BN 也要计算各参数的偏导，带来了巨大的额外计算量需求，使得模型训练的单个周期耗时明显提升。

2 简化型残差结构

2.1 两种简化型残差结构

ReLU 和 BN 的主要作用是提升网络的非线性能力和网络的稳定性。然而第一激活层的输入是残差网络中上一个残差结构的复合输出，它受此前的残差网络影响，已经具有较强的非线性能力，分布也比较稳定。本文认为第一激活层对于一个残差结构分支来说几乎没有增强其线性能力的作用，尤其是对于深度较深的网络；另一方面第一激活层中引入了 BN，带来了额外的计算量，必然降低训练速度。第二激活层的输入是残差结构中第一卷积层的输出，其分布会受到残差结构中第一卷积层参数更新的直接影响，因此有必要采用 BN 和 ReLU 提高其稳定性和非线性能力。

本文提出的删除激活层来降低残差结构计算量的两种简化残差结构，如图(3)所示，其中左图是删除第一激活层所得，右图是删除第二激活层所得。与传统的残差结构相比，这两种残差结构减少了残差分支中激活层的数量。这种改动一方面减少近一半数量的激活层数据迁移，节省了训练过程中的内存开支；另一方面减少了残差分支中 BN 层对输入期望和方差的计算量，使得模型训练耗时得以降低。

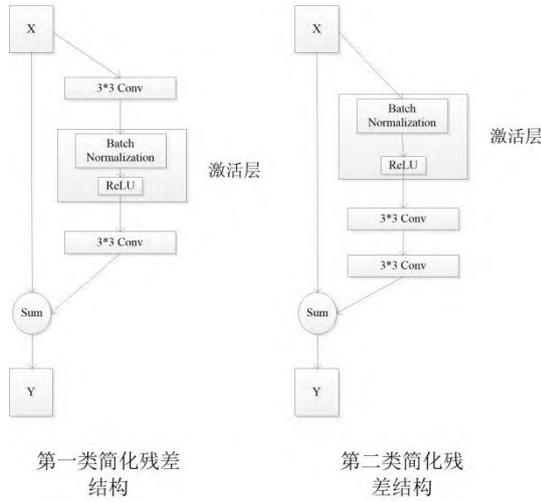


图 3 两种简化型残差结构

Fig. 3 Two kinds of simplified residual structure

2.2 原理分析

残差结构具有式(2)的形式，基本型残差结构与简化残差结构的区别仅在于激活函数。基本型的激活函数是 BN 与 ReLU 的耦合；简化残差结构中被删减的激活层可以用同等变换代替，其它部分与基本型完全一致。

记残差结构输入为 X ，第一卷积层输入为 X_1 ，第二卷积层输入为 X_2 ，输出为 Y ，对应的卷积操作和卷积参数为 $Conv(X, w, b)$ ，结合式(1)~(5)，有基本型残差结构函数如式(6)所示，第一类简化残差结构和第二类简化残差结构如式(7)和式(8)所示。

$$X_1 = \max(0, BN(X))$$

$$X_2 = \max(0, BN(Conv(X_1, w_1, b_1)))$$

$$Y_{\text{基本型}} = X + Conv(X_2, w_2, b_2)$$

$$X_1 = X$$

$$(6) \quad X_2 = \max(0, BN(Conv(X_1, w_1, b_1)))$$

$$Y_{\text{第一类}} = X + Conv(X_2, w_2, b_2)$$

$$(7)$$

$$X_1 = \max(0, BN(X))$$

$$X_2 = Conv(X_1, w_1, b_1) \quad (8)$$

$$Y_{\text{第二类}} = X + Conv(X_2, w_2, b_2)$$

随着训练周期的增加，模型越来越稳定，残差结构输出的特征值大多数为正值，此时基本型残差结构中第一激活层对输入的影响可以近似为仿射变换，第一激活层和第一卷积层的耦合作用可以等效为一个卷积层，如式(9)所示：

$$X_1 = \max(0, BN(X)) \approx \gamma X + \beta$$

$$Conv(X_1, w_1, b_1) = Conv(X, \gamma w_1, \gamma b_1 + \beta) \quad (9)$$

结合式(6)、(9)，得到基本型残差结构如式(10)所示：

$$X_2 = \max(0, BN(Conv(X, \gamma_1 w_1, \gamma_1 b_1 + \beta_1)))$$

$$Y_{\text{基本型}} = X + Conv(X_2, w_2, b_2) \quad (10)$$

对比式(7)和式(10)，发现两式具有相似的形式，两者差异仅为卷积层参数，当 γ, β 分别为 1 和 0 时两式完全相等，即第一类简化残差结构能够取代基本型残差结构。

结合式(8)、(9)得到第二类简化残差结构如式(11)所示：

$$X_2 = Conv(X, \gamma_1 w_1, \gamma_1 b_1 + \beta_1)$$

$$Y_{\text{第二类}} = X + Conv(X_2, w_2, b_2) \quad (11)$$

即经过一段时间的训练，模型较为稳定时第二类简

化残差结构其残差分支近似为两个连续的卷积层。连续两个小尺寸卷积核的作用可以等效为一个尺寸较大的卷积核作用，两个连续的 3×3 卷积核等效为一个 5×5 的卷积核的参数替换方法如式(12)所示：

$$\begin{aligned}
 X_{2,h,w} &= \sum_{i_1,j_1=-1}^1 w_{1,i_1,j_1} X_{1,h+i_1,w+j_1} + b_1 \\
 Y &= \sum_{i_2,j_2=-1}^1 w_{2,i_2,j_2} X_{2,h+i_2,w+j_2} + b_2 \\
 b_3 &= \sum_{i_2,j_2=-1}^1 w_{2,i_2,j_2} b_1 + b_2 \\
 w_{3,i_3,j_3} &= \sum_{i_1+i_2=i_3,j_1+j_2=j_3} w_{1,i_1,j_1} w_{2,i_2,j_2} \\
 Y &= \sum_{i_3,j_3=-2}^2 w_{3,i_3,j_3} X_{1,h+i_3,w+j_3} + b_3
 \end{aligned} \tag{12}$$

结合式(11)、(12)，第二类简化残差结构可表示为式(13)：

$$Y_{\text{第二类}} = X + \text{Conv}(X, w_3, b_3) \tag{13}$$

表明此时第二类简化残差结构退化成残差分支仅包括一个卷积核尺寸为 5×5 的卷积层，如图(4)所示：

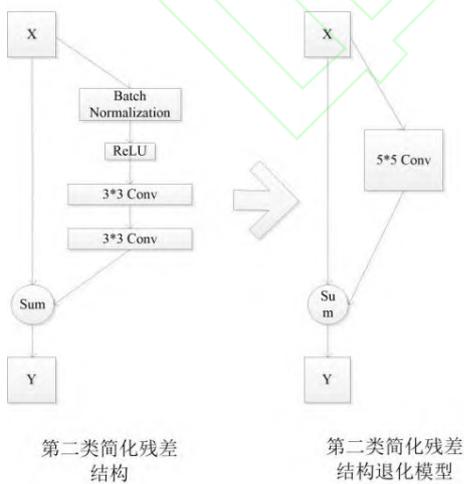


图 4 第二类简化型残差结构退化模型

Fig. 4 The degradation model of the second

simplified residual structure

由式(3)可知单一的卷积层近似为仿射变换，其输出是对应位置输入的加权和，当偏移量参数 b 较小时，卷积操作可以看作是线性变换。即第二类简化残差结构因缺乏卷积层间的激活单元，其非线性能力受到影响，在模型较为稳定时第二类简化残差结构几乎可以看作线性变换，无法通过训练进一步提升提取非线性特征的能力，从而影响模型的性能。

综上所述，本文认为利用第一类简化残差结构训练深度残差网络，能够在保持网络分类性能的前提下，减少训练过程中的计算量，降低内存开销，提高网络训练速度；第二类简化残差结构会造成网络分类性能下降，在其他方面与第一类简化残差结构一致。

3 实验分析

本文利用 CIFAR10/100 数据集分别训练了几组分类模型，详细对比了基本型 ResNet 与基于简化残差结构的快速深度残差网络准确率，训练周期耗时等性能。

3.1 CIFAR 数据集

CIFAR10/100 图像数据集由尺寸为 32×32 的彩色图像组成。其中，CIFAR10 包括飞机、汽车、鸟、猫、鹿、狗、青蛙、马、轮船、卡车等 10 个常见类别，每个类别含有 5000 张训练图像和 1000 张测试图像，最终构成了共包括 50000 张图像的训练样本库和 10000 张图像的测试样本库；CIFAR100 包括更为精细的 100 个图像类别，每个类别含有 500 张训练图像和 100 张测试图像。经过严格的人工筛选和标记，每张图像都带有唯一正确的标签。图像数据增强是卷积神经网络模型降低过拟合现象最常用的手段，本文在图像数据增强方面采用常规的图像裁剪与随机水平翻转；在图像数据预处理方面，本文利用训练图像的各通道像素均值和方差归一化整个数据库。

3.2 快速深度残差网络与实验环境

前文提到简化型残差结构与基本型残差结构相比，其训练计算量较少，使得训练用时得以缩减，因此本文将利用简化型残差结构设计的深度残差网络称为快速深度残差网络。该模型由低级特征提取层、中间层、和分类器三部分组成。其中低级特征提取层即为一个卷积层，中间层分为三个阶层，每个阶层由相同数量的残差块组成；分类器结构为 BN—ReLU—全局平均池化层—全连接层。记中间层每个阶层的残差块数量为 n ，因残差块含有两个卷积层，低级特征提取层含有一层卷积层，分类器中的全连接层也算作一层卷积层，则网络模型的总深度(仅计算卷积深度)为 $D=6n+2$ 。CIFAR 数据集图像尺寸较小，采用的卷积核尺寸均为 3×3 。低级特征提取层的卷积核数量与第一阶层中卷积核数量都为 16；第二阶层卷积核数量为 32，第三阶层卷积核数量为 64。模型训练的批尺寸为 128，权重衰减为 10^{-4} ，参数更新方法为 Nesterov momentum，momentum 权重设置为 0.9，训练周期数设置为 160，学习率前 80 个周期为 0.1，81 周期至 120 周期为 0.01，121 周期至 160 周期为 0.001。以上参数设置与 ResNet 完全一致，本文在 CIFAR10 上训练的深度为 110 层的快速深度残差网络模型示意图如图(5)所示。

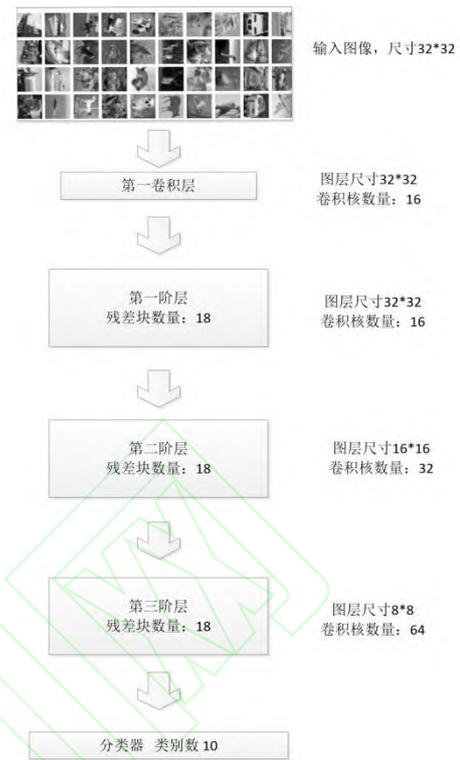


图 5 110 层快速深度残差网络模型示意图

Fig. 5 Schematic diagram of fast residual network with 110 layers

实验采用的操作系统为 windows 7，CPU 为 intel core i3-8350k，4.0GHz，16G 内存，GPU 为 NVIDIA GTX 1070，8G 显存，卷积神经网络开发框架采用基于 matlab 的 matconvnet-1.0-beta25。

3.3 两种简化型残差结构实验结果

为考察两种简化型残差结构的实际性能，本文分别利用基本型残差结构、第一类简化型残差结构、第二类简化型残差结构组建了 $n=3$ 的残差结果的残差网络，即网络深度为 $D=20$ 层。

本文对三种网络模型分别训练了 5 次以减小网络参数初始化对网络分类性能的影响，然后考察每种网络的平均错误率和平均周期用时，其结果如表(1)所示；并从每种网络模型中选取具有最小测试错误率的一个，绘制了如图(6)所示的测试分类错

误率曲线。

表 1 三种残差结构的分类错误率和周期耗时对比

Table 1 Comparison of classification error rate and training time per epoch of three residual structures

	分类错误率/%	周期耗时/s
基本型	7.62±0.15	30.07
第一类残差结构	7.68±0.20	22.08
第二类残差结构	10.04±0.18	22.10

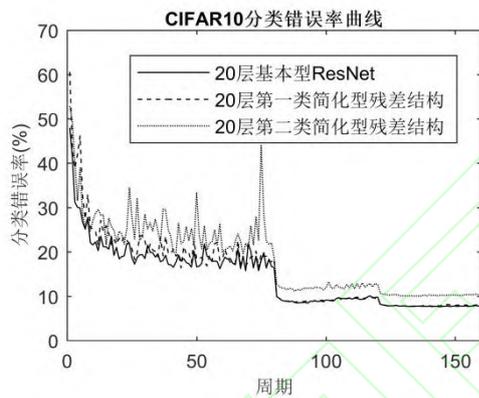


图 6 三种残差结构在 CIFAR10 上的测试分类错误率曲线

Fig. 6 Test classification error cure of the three residual structures on CIFAR10

从表(1)可以看出，第一类简化型残差结构在错误率方面为 7.68%与典型残差结构 7.62%相当；周期耗时为 22.08s 比典型残差结构的 30.07s 缩减了 26.68%，大幅提高了训练速度。第二类简化型残差结构周期耗时为 22.10s 与第一类简化型残差结构相当，但其错误率为 10.04%，表现较差。从图(6)可以看出，随着训练周期的增加，三种网络模型分类错误率持续波动下降并逐渐趋于稳定，因学习率的缩小使得网络参数微调步长变短，分类错误率在第 81 周期和第 121 周期出现了明显下降。第一类简化型残差结构与基本型残差结构曲线基本重合，在前 80

周期因初始学习率较大，分类错误率曲线有小范围起伏，后 80 个周期起伏很小，稳定在 7.6%附近；第二类简化型残差结构稳定性较差，分类错误率在前 80 个训练周期有很大起伏，121 周期之后稳定在 10%附近，明显高于前两者。

以上实验结果与前文分析一致，表明第一类简化型残差结构能够替代基本型残差结构，在保持网络分类性能的前提下大幅缩减网络训练耗时。

3.4 快速深度残差网络实验结果

为进一步验证第一类简化型残差结构组建的快速深度残差网络随深度加深的表现，在 CIFAR10 和 CIFAR100 数据集上设计了卷积深度为 D=56 层和 D=110 层的快速深度残差网络，即每个阶层残差块数量 n=9 和 n=18。因训练用时随网络深度大幅增加，为节省时间考虑这些网络模型仅训练一次。在 CIFAR10 和 CIFAR100 数据集上分别训练 160 个周期后，本文考察了两类模型分类错误率和周期耗时，如表(2)所示；并绘制了 110 层快速深度残差网络和 ResNet 的测试分类错误率曲线，如图(7)、(8)。

表 2 基本型深度残差网络与快速深度残差网络对比

Table 2 Comparison of basic ResNet and fast deep residual networks(FastResNet)

	CIFAR10 错误率/%	CIFAR100 错误率/%	周期耗时/s
56 层 ResNet	6.31	27.93	84.22
56 层 FastResNet	5.81	26.91	58.84
110 层 ResNet	5.75	26.02	208.26
110 层 FastResNet	5.29	24.80	133.47

结果表明训练相同深度的快速深度残差网络和

基本型深度残差网络，快速深度残差网络耗时大幅降低。其中 56 层的快速深度残差网络周期耗时为 58.84s，基本型深度残差网络周期耗时 84.22s，相对减少了 30.04%，深度为 110 层的快速残差网络周期耗时 133.47s，基本型周期耗时 208.26s，相对减少 35.91%。在分类错误率方面，快速深度残差网络也取得比基本型深度残差网络略好的效果，110 层快速深度残差网络分类错误率为 5.29%，基本型分类错误率为 5.75%，相对提升了 8%。

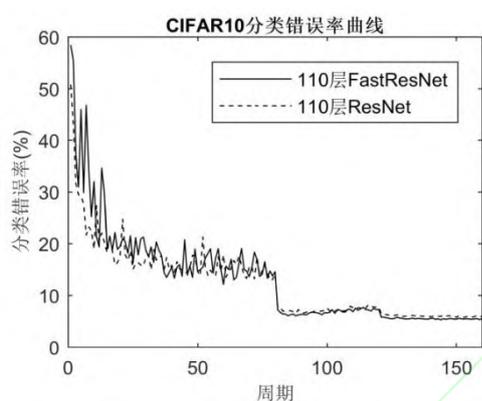


Fig.7 The classification error curve of 110-ResNet and 110-FastResNet on CIFAR10

图 7 110 层 ResNet 与 FastResNet 在 CIFAR10 上的分类错误率曲线

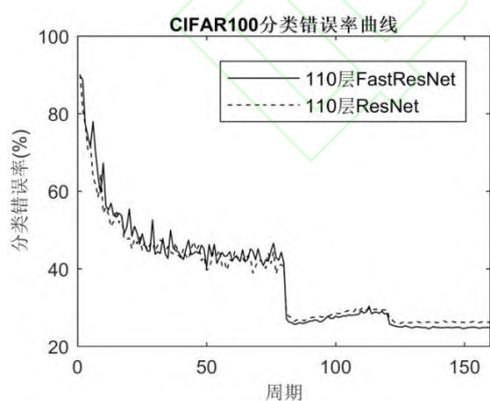


Fig.8 The classification error curve of 110-ResNet and 110-FastResNet on CIFAR100

图 8 110 层 ResNet 与 FastResNet 在 CIFAR100 上的分类错误率曲线

在 CIFAR10 数据集上，本文还训练了深度达到 668 层的快速残差网络和基本型深度残差网络，其中快速深度残差网络分类错误率为 5.09%，周期耗时为 2020.92s，基本型深度残差分类错误率为 6.03%，网络周期耗时为 3953.02s；网络训练速度相对提升了 48.88%。结果表明，网络深度越深快速深度残差网络训练耗时提升的幅度越明显，其分类错误率也越低。此外，668 层基本型深度残差网络分类错误率高于 110 层的 5.75%，分类性能出现退化，即过拟合现象严重。本文认为超深的基本型深度残差模型包含过多的激活层，归一化参数繁多，导致出现较为严重的过拟合现象；快速深度残差网络减少了激活层的数量，一定程度上避免了这一问题。

结束语

本文提出了移除基本型残差结构中部分激活层的简化型残差结构，以此构建了快速深度残差网络，并从理论分析和实验论证这两方面讨论了简化型残差结构的合理性。实验表明，删除残差结构第一激活层不会降低网络的分类性能，删除第二激活层会使得网络模型退化，导致分类错误率上升；训练快速深度残差网络比基本型深度残差网络周期耗时少，这一优势随着网络深度的加深更为突出；其分类性能也比基本型残差网络要好，网络深度越深快速深度残差网络分类错误率越低，一致性好。本文提出的快速深度残差网络方法在保证分类性能的前提下能够大幅降低残差网络训练时间成本，具有一定的实际价值。进一步的研究将在快速深度残差网络的应用中展开。

参考文献：

- [1] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [2] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]// European conference on computer vision. Cham: Springer, 2016: 630-645.

- [3] TARG S, ALMEIDA D, LYMAN K. Resnet in resnet: Generalizing residual architectures [EB/OL].[2016-03-25].<https://arxiv.org/pdf/1603.08029>.
- [4] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[EB/OL].[2017-06-14].
<https://arxiv.org/pdf/1605.07146>.
- [5] Veit A , Wilber M , Belongie S . Residual Networks Behave Like Ensembles of Relatively Shallow Networks[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. Newyork: Curran Associates Inc, 2016:550-558.
- [6] Wu Z , Shen C , Hengel A V D . Wider or Deeper: Revisiting the ResNet Model for Visual Recognition[EB/OL].[2016-11-30].<https://arxiv.org/pdf/1611.10080.pdf>
- [7] 王若瑜. 基于 Resnet-50 的智能驾驶红绿灯分类研究[J]. 电子测试, 2019, 408(Z1):145-147.
- [8] 宋倩, 黄昶, 余慧瑶. 基于 TensorFlow 的交通标志形状识别[J]. 信息通信, 2017, 000(012):286-288.
- [9] 王文成, 蒋慧, 乔倩, 等. 基于 ResNet50 网络的十种鱼类图像分类识别研究[J]. 农村经济与科技, 2019, 030(019):60-62.
- [10] 常川. 基于 ResNet 深度网络的人类蛋白质图谱图像分类方法研究[J]. 医学信息学, 2019, 40(007):45-49.
- [11] 王 恒, 李 霞, 刘晓芳, 等. 基于 ResNet50 网络的乳腺癌病理图像分类研究[J]. 中国计量大学学报, 2019, 30(1):72-77.
- [12] 张枫, 田联房, 杜启亮. 基于残差网络与中心损失的人脸识别[J]. 计算机工程与设计, 2019, 40(06):1689-1695.
- [13] 陈甜甜. 基于深度学习的说话人识别研究[D]. 北京邮电大学, 2018.
- [14] 朱超平, 杨艺. 基于 YOLO2 和 ResNet 算法的监控视频中的人脸检测与识别[J]. 重庆理工大学学报(自然科学), 2018, 32(08):170-175.
- [15] 熊才华, 巩言丽, 廉华, 等. 基于 ResNet-50 改进的 Faster R-CNN 手势识别算法[J]. 计算机时代, 2019, 000(009):1-4.
- [16] [1]刘子毅. 基于图谱特征分析的农业虫害检测方法研究[D]. 浙江大学, 2017.
- [17] 张银松, 赵银娣, 袁慕策. 基于改进 Faster-RCNN 模型的粘虫板图像昆虫识别与计数[J]. 中国农业大学学报, 2019, 24(05):115-122.
- [18] HUANG G, SUN Y, LIU Z, et al. Deep networks with stochastic depth[C] //European conference on computer vision. Cham: Springer, 2016: 646-661.
- [19] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]// Proceedings of the thirteenth international conference on artificial intelligence and statistics. 2010. 249-56.
- [20] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Proceedings of the Advances in neural information processing systems. 2012: 1097-1105.
- [21] LIN M, CHEN Q, YAN S. Network in Network[EB/OL]. [2014-03-04].
<https://arxiv.org/pdf/1312.4400>.
- [22] SPRINGENBERG J T, DOSOVITSKIY A, BROX T, et al. Striving for simplicity: The all convolutional net[EB/OL]. [2015-04-13].
<https://arxiv.org/pdf/1412.6806>.
- [23] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [24] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. [EB/OL]. [2015-03-02].
<https://arxiv.org/pdf/1502.03167>.
- [25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2015-04-10].
<https://arxiv.org/pdf/1409.1556>.
- [26] SANTURKAR S, TSIPRAS D, ILYAS A, et al. How does batch normalization help optimization?[C]// Proceedings of the Advances in Neural Information Processing Systems. 2018:2483-93