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.

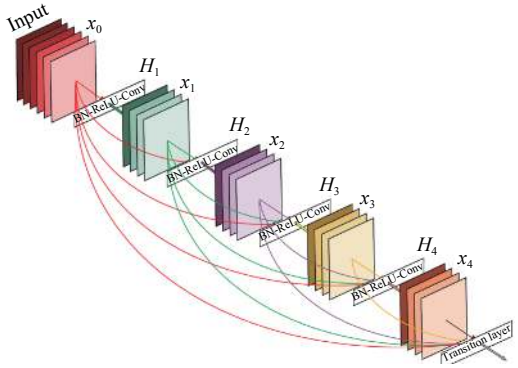

图 6 密集型网络结构图[21]

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.

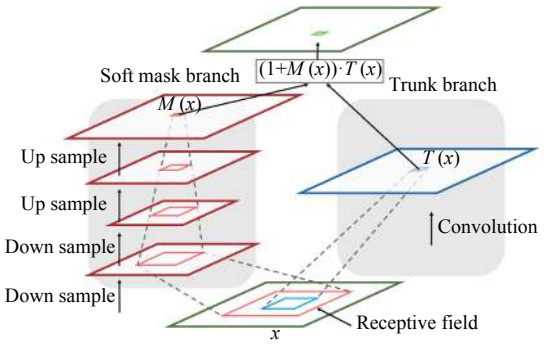

图 7 注意力残差单元结构图[22]

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

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.