压缩与激励网络的变体

Mahendran
NVmhkannan5@gmail.com

arXiv:2304.06502v2 [cs.CV] 3 Jul 2023

摘要

卷积神经网络能够学习空间特征，且这些特征在卷积核中高度交织。SE 模块打破了传统神经网络将全部结果传递给下一层的做法。相反，SE 模块通过压缩与激励（SE）模块，仅传递需要学习的重要特征。我们提出了 SE 模块的若干变体，以改进压缩与激励过程并提升性能。对层进行压缩或激励，使层权重能够平滑过渡。这些变体也保留了 SE 模块的特性。我们在残差网络上进行了实验，并将结果制成表格。

***K*eywords** 架构 ·深度学习

1 引言

自 AlexNet [1, 2] 获胜以来，深度神经网络架构不断演进。从 2012 年起，各种神经网络架构相继提出 [3, 4, 5, 6, 7, 8, 9, 10, 11, 12]。此后，人们普遍认为，更深的架构能够提升性能。近期的多项研究成果表明，宽度、尺度等其他网络特征也有助于提升性能 [4, 3, 6, 7]。根据 [9]，当前网络要取得更好的性能，并非必须增加深度。所提出的架构包含各种显著模块，这些模块是实现更好学习效果和提升性能的关键 [6, 7]。例如，ResNet 中用于解决权重退化问题的残差模块 [3] ，以及通过不同卷积核学习并降低计算成本的 Inception 模块 [2] ，为构建性能更优的网络铺平了道路。SE 网络中的压缩与激励模块 [13] 不同于现有模块，因为它只将重要信息传递到下一层。

基于卷积层的网络在检测、分类等许多视觉任务中表现更好 [1, 9]。在卷积层中，卷积核滤波器使该层能够从输入中学习特征。滤波器学习到的特征在相邻卷积核之间具有空间连接，并且彼此局部依赖。这就是为什么即使是 dropout 这样的正则化技术，在卷积神经网络上也无法发挥作用 [14]。另一方面，全连接（FC）层具有全局依赖性。在大多数神经网络架构中，网络初始阶段使用卷积层，随后接入全连接层。卷积层学习所有相互依赖的特征，然后将结果传递给全连接层，由其学习具有全局依赖性的特征。

在 SE 网络中 [13], , SE 模块包含全连接层，并添加在网络的卷积层之间。这种机制还会使该层的输入值经过压缩，并在最后激励结果。线性全连接层传递重要的全局信息，因为压缩会省去不重要的信息。将这些信息在激励后传递给卷积层，可以提升该层的表征能力，从而增强网络的表征能力。这就是压缩与激励网络机制的核心思想。压缩与激励网络通过压缩网络特征并将其恢复至原始形式，对网络中的特征进行重新校准。它利用卷积中的通道间依赖关系来增强网络的学习能力。

本文提出了四种 SE 模块，它们利用全局依赖特征，使网络获得更好的泛化能力。这些变体延迟了压缩或激励操作，并在 SE 模块之间额外添加了一个卷积层。延迟压缩或激励过程可以使该层传递更优的值，因为在延迟过程中操作会执行两次。在将结果传递给其他层时，所传递的信息更加集中。其他压缩与激励变体同样如此。此外，我们还提出了一种称为凸起模块的新模块。

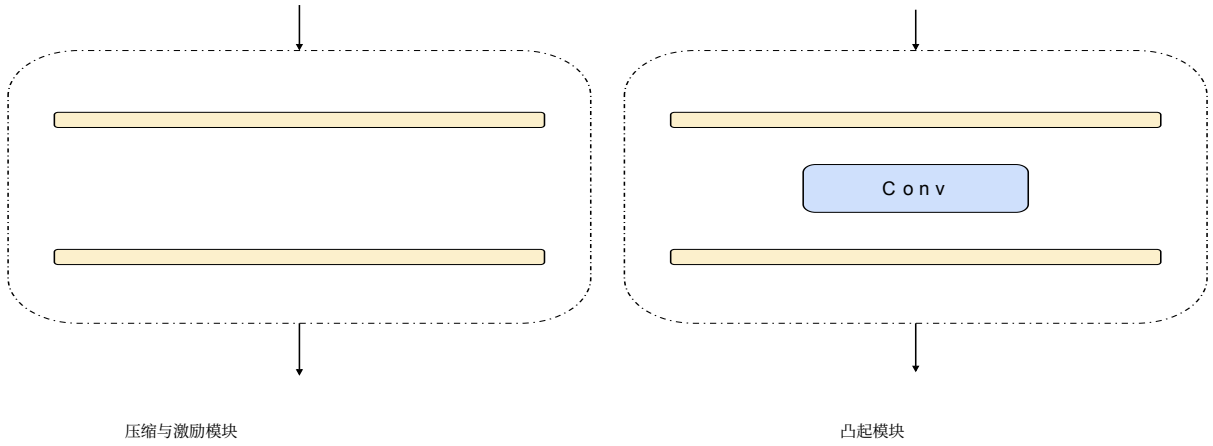

图1：本图展示了压缩与激励模块与所提出模块之间的比较。凸起模块的顶部和底部包含全连接层，与压缩与激励模块类似。顶部的稠密层/全连接层汇集输入，并将其传递给激励方法，使压缩后的值恢复原始形状，以供下一层使用。

凸起模块从压缩操作开始，其中该层通过全连接（FC）层从前置层收集空间信息。所提出的方法在压缩与激励模块中包含全连接层操作。FC 层学习到的全局依赖特征在 SE 模块中未得到利用。我们在线性层之后执行全连接操作，使该层能够从全局表征中学习。这些信息对于网络更好地学习和泛化至关重要。丰富且整合后的信息被传递到同等大小的 FC 层。该层使网络能够利用这些关键信息，然后再将其恢复为原来的形状。该层之后连接的是 SE 模块激励部分的线性层。这个凸起模块不仅传递关键信息，还能使网络从这些信息中学习。

我们还在图1中详细展示了凸起模块。之所以称其为凸起模块，主要是因为在 SE操作之间新增了一层，它在模块中起到了减速带的作用。图中将该模块与 SE 模块进行了比较。如图所示，凸起模块在压缩层和激励层之间设置了单个卷积神经网络层。在普通的 SE 模块中，网络的激励阶段紧接在压缩层之后。而在所提出的网络模块中，在压缩过程中加入了一个大小为 3x3 的卷积层，以提升网络性能。

2 三层卷积神经网络

我们在CIFAR-10 上的简单三层卷积神经网络中，尝试了在压缩和激励之间进行学习的所提出的方法。我们尝试了三层卷积神经网络上所提出模块的各种变体。我们使用批量大小为64的设置进行实验，并将学习率以 0.7 的衰减率进行衰减，初始值为 1.0。我们使用 Adadelta优化器，并采用阶梯式学习率。实验进行 50 个训练轮次。

所提出的模块被添加到三层卷积神经网络中并进行性能测试。在默认的3层网络中，如果需要测试某个模块，我们会用该模块替换第2个卷积层，并在 CIFAR-10上进行实验。例如，三层卷积神经网络中的凸起模块如图2所示。结果汇总如表1所示。从结果可以明显看出，所提出的压缩和激励网络的其他变体能够以更好的方式提升性能。

该方法还有其他变体。现有的替代方案是减小卷积核尺寸 [15, 13]。还有一种称为门控机制（高速网络） [16] 的方法。在某些情况下，使用预训练模型时，会冻结某些层，因为在数据集上进行微调时，输入不会经过这些层。模型微调期间不需要这些层在训练过程中存储的知识。这与门控机制有些类似。有些模块也会...

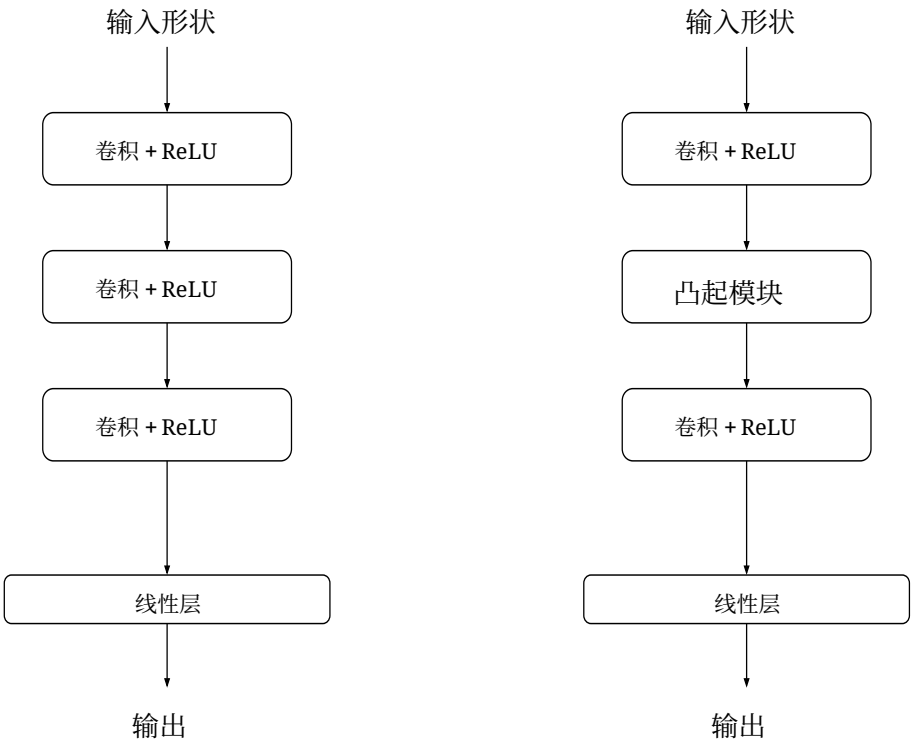

图2：三层卷积神经网络。左侧：三层卷积神经网络的默认结构。右侧：采用所提出的凸起模块、配置相同的自定义网络结构。

方法	Loss	准确率
默认	0.8091	72.52
挤压与激励模块	0.8068	72.58
慢速挤压与激励	0.8093	72.20
挤压与慢速激励	0.8131	72.63
慢速挤压与慢速激励	0.7987	72.97
凸起模块	0.8044	72.65

表1：这些数值表示在采用默认结构、现有 SE 模块及所提出的变体的三层卷积神经网络上进行的实验结果。

类似于 nn. Inception 模块中使用的这种方法、SE 模块、门控机制和冻结层（预训练模型）。

3 文献综述

深度学习网络：在 AlexNet [1] 赢得 ILSRVC 2012 竞赛后，人们开始对开发深层网络产生兴趣。成功之后，出现了许多沿着相同路径构建更深层网络的架构。其中有一些著名架构 [2, 4, 9, 3, 7, 6, 13] 充分利用了深度学习的优势。VGG [12]网络用一系列 3x3 卷积替代 AlexNet 中较大的滤波器，并赢得了 ILSRVC 2014 竞赛。残差网络使网络能够在不断加深的同时有效学习。残差模块具有跳跃连接，可传递经过特定层学习到的特征，从而解决梯度消失问题。ResNet 的变体之一是 ResNext，它采用分割-变换-合并策略重复相同的模块，性能优于 ResNet。ResNext 模块增加了宽度，并将其置于残差模块内部，因此可以非常有效。Inception 网络通过在 Inception 模块中引入不同核尺寸的卷积，使网络能够在局部形成稀疏结构。Xception 网络通过使用深度可分离卷积替代 Inception 模块，加快了训练速度。先进网络中的模块通过引入新技术不断得到改进。

SqueezeNet 使用瓶颈方法来构建小型网络。它由 Fire 模块组成，该模块先将特征压缩到一定数量的平面，再使用大小为 1 的卷积核滤波器进行扩展，随后使用大小为 3 的滤波器。该模型体积更小，并能有效地将信息传递到下一层。压缩与激励网络提出了 s。其他先压缩再激励的变体采用了类似的方法[a]。其核心思想是在模型压缩和激励时传递重要特征，从而恢复网络的原始状态。已有研究提出了类似的架构，包括网络中的网络，它在网络中包含微型网络。这个微型网络由遍布整个网络的 1x1 滤波器组成。正如 SqueezeNet 论文所述，该模块能够识别通道中的非线性依赖关系，同时具有模型动态性。通过传递重要信息，这提升了网络的表征能力。 [作者]提出了高速网络，它使用门控机制来调节信息流。双路径网络采用在内部连接路径的方式。该网络兼具 ResNet 和 DenseNet 的能力，还通过层间的内部连接学习新特征，从而提升网络性能。密集连接卷积网络也通过层间的重构连接继承了类似能力，从而改善网络的表征属性。

在网络中，压缩方法内部会临时构建模块；但在某些情况下，压缩方法的输出所传递的重要信息并未得到利用，而是直接恢复为原始形状。现有方法只有在恢复形状后才会训练这些重要信息。针对特定目的，已有方法还提出了其他变体 [5, 6, 7, 8, 9, 10, 11, 17]。 [18] 提出了针对 AlexNet [1] 和 GoogLeNet [2] 的变体，以提升韩文字符识别的性能。

4 SE 的变体

因此，需要在网络的压缩与激励模块中加入变体。压缩与激励能够增强网络的表征能力。提出这些变体是为了避免网络权重发生突然变化。SE 模块声称只向后续层传递重要信息。所提出的模块尝试缓慢地将形状缩减至相应的通道，以保留一部分不重要的值。保留这些不重要特征的目的是避免各层过拟合，并帮助它们更好地学习。 [通过实验对此进行详细说明]
如图所示，我们提出了四种压缩与激励模块。图中展示了 SE 模块与这四种模块的对比。在提出的四种模块中，有三种在压缩和激励过程中有所不同。其余一种模块则加入了全连接层以提升性能。下面将详细介绍这几种模块。

慢速挤压与激励 - 在该模块中，压缩分为两个步骤进行。首先将尺寸缩小一半，然后在下一步中将输入剩余的一半尺寸缩小。压缩步骤中给定的缩减值会直接用于后续的激励步骤。

慢速挤压与慢速激励 - 该模块同时改进了压缩和激励过程。根据给定的缩减值，先完成一半的压缩，再对剩余的一半进行压缩。同样，在激励阶段，先对缩减后输出的一半进行激励，然后根据半激励形状的输出恢复为原始形状。

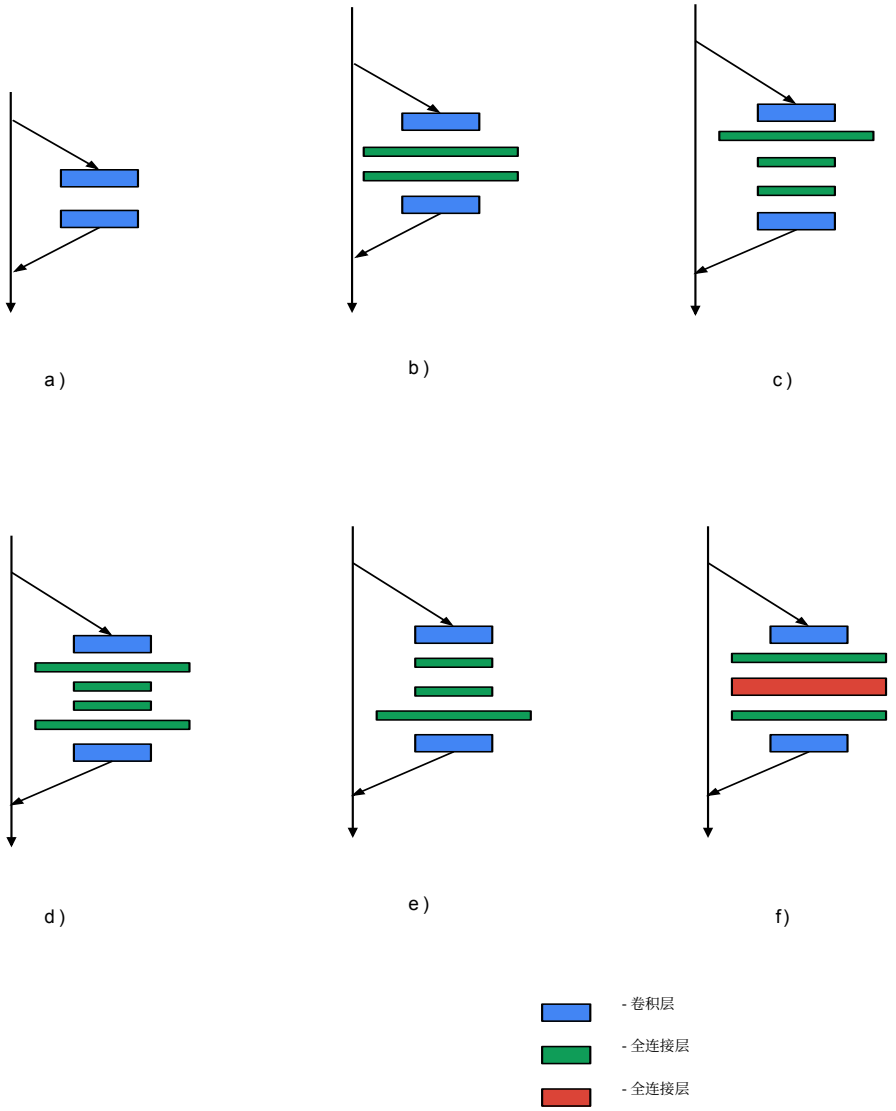

图 3：可视化展示了所提出的压缩与激励模块变体 [13]，以及应用于残差模块的现有模块。a) 残差网络模块，包含两个卷积层，用于基本块。b) SE [13] 的压缩与激励模块，包含用于压缩和激励的稠密层。c) 慢速挤压与激励——该模块通过应用两个用于压缩操作的稠密层，延迟压缩过程。d) 慢速挤压与慢速激励——该变体的压缩与激励模块分别包含两层全连接层。e) 挤压与慢速激励——该版本包含慢速激励过程，因此使用两个用于激励的稠密层。f) 这是压缩与激励的新版本，包含一个工作中的全连接层，该层接收压缩层的输入，并将输出发送至激励层，以恢复原始形状。

挤压与慢速激励 - 在此模块中，按照给定的缩减形状完成压缩，而激励过程则略微放缓。将（1x1xch）的压缩结果激励至 reduction//2 的值。随后，该输出再经过原始 reduction 的激励，以保持形状不变，从而传递至下一个堆叠层。

压缩-学习-激励 / 凸起- 与之前的方法不同，这种压缩与激励模型在中间包含一个全连接层。由于全连接层充当默认 SE 模块的凸起，因此它也被称为凸起模块。正如研究人员此前在 [sepaper]，中所解释的，全连接层能够学习全局表征，并将其传递给后续层。在此压缩-学习-激励模块中，学习到的值会直接用于压缩状态，随后再激励至正常状态。

4.1 训练细节

我们在 ResNet-18 上实现了所提出的模块。残差网络能够在特定实例中进行学习，并将学习结果传播至附近的层。我们认为，仅凭残差网络就能更好地展示实验结果。实验使用 CIFAR-10、CIFAR-100 和 SVHN 数据集。由于计算能力有限，我们采用 ResNet-18 作为标准架构进行测试。批量大小设为 32，并按照数据集的均值和标准差进行归一化。我们使用三个数据集，其中 CIFAR-100 的数据变换方式与 CIFAR-10 不同。输入数据采用随机调整大小裁剪，图像尺寸为 224。裁剪后再随机进行水平翻转。CIFAR-100 数据集的训练数据也采用归一化处理。测试 CIFAR-100 时，将数据集调整为 256 的尺寸，然后进行 224 的中心裁剪，最后进行归一化。对于 SVHN 数据集，训练集和测试集均采用与 CIFAR-10 类似的处理方式。

5 实验

本节将展示 SE 变体（包括凸起模块）在 ResNet-18 网络上的实现。所提出的 SE 模块变体性能大幅优于其他模型。本节还介绍了各模型的实现和架构细节。

6 实现

残差网络以能够保留已学习的权重而闻名，因为它在网络加深时可以避免权重退化问题。所提出的残差模块由卷积层构成，会在经过若干层后传递已学习的参数。这些参数与特定深度相关，因此更加有效，因为在某些层之后传递的特征只会用于该层。

我们在残差模块中实现了所提出的变体：慢速压缩与激励、慢速挤压与慢速激励、挤压与慢速激励，以及凸起模块。由于残差模块可能有助于避免权重退化，因此确保残差模块传递关键信息十分重要。图 3 展示了在残差模块之上添加凸起模块等变体的实现。我们在 ResNet-18 上测试了所提出的方法，并采用了所有最新的附加技术。我们使用批归一化技术，按批次对卷积层的输出进行归一化。批归一化得到广泛应用，因为它优于其他归一化技术。我们对该 ResNet-18 模块在三个数据集上的结果进行了测试，并将结果列于表中。

方法 Top-1 Top-5 默认 73.130 97.620
SE 模块 75.580 97.210 慢速压缩与激励
74.370 96.910 压缩与慢速激励 75.630
97.160 慢速挤压与慢速激励 76.020 97.700
凸起模块 76.020 97.810

Tabl图 2：在采用所提出的 SE 模块变体的 ResNet-18 上对 CIFAR-10 数据集进行测试的结果 e.

方法	Top-1	Top-5
默认	72.870	92.380
SE 模块	74.460	93.190
慢速挤压与激励	73.340	92.960
压缩与慢速激励	73.500	93.910
慢速挤压与慢速激励	73.250	92.860
凸起模块	73.960	93.200

表 3：在 ResNet-18 上使用所提出的 SE 模块变体，对 CIFAR-100 数据集进行测试的结果。

6.1 CIFAR - 100 结果

表3 展示了采用上述配置时的结果。这些数值证明，所提出的模块确实对基础网络产生了影响。在 Top-5准确率方面，其结果与 SE 模块相当，有时甚至优于 SE 模块。

6.2 SVHN结果

方法	Loss	准确率
默认	91.211	98.851
SE 模块	91.795	98.855
慢速挤压与激励	92.048	98.959
挤压与慢速激励	92.244	98.871
慢速挤压与慢速激励	92.187	98.782
凸起模块	91.576	98.882

Ta表 4：在采用所提出的 SE 模块变体的 ResNet-18 上，对 SVHN 数据集进行测试的结果。

从结果来看，对于 SVHN 数据集，所提出的模块使模型性能提升了最多 1.01 个百分点。这种提升是因为每个残差层都确保仅将图像中的关键特征传递给网络。

7 结论

我们提出了残差网络上压缩与激励模块的变体。这些残差结构上的变体使网络能够传递重要信息。所提出的 SE 变体可像 SE 模块一样应用于现有网络。我们介绍了压缩与激励的变体，它们与 MobileNetV2 、MobileNetV3 以及 InceptionV2、InceptionV3 类似。我们认为，压缩与激励及其变体有望成为网络传统输入流中的基石，因为它们只传递重要信息。

参考文献

[1] Krizhevsky A.; Sutskever I. 和 Hinton G.E. 使用深度卷积神经网络进行 ImageNet 分类。《*ACM* 通讯》，60:84–90，2017年。[2] Szegedy C.; Liu W.; Jia Y.; Sermanet P.; Reed S.; Anguelov D.; Erhan D.; Vanhoucke V. 和 Rabinovich A. 通过卷积不断深入。发表于 *IEEE*计算机视觉与模式识别会议论文集，第 1–9 页，2015年。[3] He K.; Zhang X.; Ren S. 和 Sun J. 用于图像识别的深度残差学习。发表于 *IEEE*计算机视觉与模式识别会议论文集，第 770–778 页，2016年。[4] Tan M. 和 Le Q. EfficientNetV2：更小的模型与更快的训练。发表于国际机器学习会议，第 10096–10106 页，2021年。[5] Liu Z.; Mao H.; Wu C.Y.; Feichtenhofer C.; Darrell T. 和 Xie S. 面向 2020 年代的卷积神经网络。发表于 *IEEE/CVF*计算机视觉与模式识别会议，第 11976–11986 页，2022年。[6] Xie S.; Girshick R.; Dollár P.; Tu Z. 和 He K. 用于深度神经网络的聚合残差变换。发表于 *IEEE*计算机视觉与模式识别会议论文集，第 1492–1500 页，2017年。

[7] Zagoruyko S. 和 Komodakis N. 宽残差网络。*arXiv* 预印本 *arXiv:1605.07146*，2016年。[8] Chollet F. Xception：采用深度可分离卷积的深度学习。发表于 *IEEE*计算机视觉与模式识别会议论文集，第 1251–1258 页，2017年。[9] Howard A.G.; Zhu M.; Chen B.; Kalenichenko D.; Wang W.; Weyand T.; Andreetto M. 和 Adam H. MobileNets：面向移动视觉应用的高效卷积神经网络。*arXiv* 预印本 *arXiv:1704.04861*，2017年。[10] Sandler M.; Howard A.; Zhu M.; Zhmoginov A. 和 Chen L.C. MobileNetV2：倒置残差与线性瓶颈。发表于 *IEEE*计算机视觉与模式识别会议论文集，第 4510–4520 页，2018年。[11] Howard A.; Sandler M.; Chu G.; Chen L.C.; Chen B.; Tan M.; Wang W.; Zhu Y.; Pang R.; Vasudevan V. 和 Le Q.V. 探索 MobileNetV3。发表于 *IEEE/CVF*国际计算机视觉会议，第 1314–1324 页，2019年。[12] Simonyan K. 和 Zisserman A. 用于大规模图像识别的超深卷积网络。*arXiv* 预印本 *arXiv:1409.1556*，2014年。[13] Hu J.; Shen L. 和 Sun G. 压缩与激励网络。发表于 *IEEE*计算机视觉与模式识别会议论文集，第 7132–7141 页，2018年。[14] Ghiasi G.; Lin T.Y. 和 Le Q.V. DropBlock：一种用于卷积网络的正则化方法。神经信息处理系统进展，2018年。[15] Lin M.; Chen Q. 和 Yan S. 网络中的网络。*arXiv* 预印本 *arXiv:1312.4400*，2013年。[16] Srivastava R.K.; Greff K. 和 Schmidhuber J. 训练超深网络。神经信息处理系统进展，2015年。[17] Chen H.Y. 和 Su C.Y. 一种增强型混合型MobileNet。发表于 *2018*年第九届国际意识科学与技术会议（*iCAST*），第 308–312 页，2018年。[18] Lee S.G.; Sung Y.; Kim Y.G. 和 Cha E.Y. 改进韩文字符识别性能的 AlexNet 和 GoogLeNet 变体。信息处理系统期刊，14:205–217，2018年。