文章编号：2095-2163(2021)07-0196-06 中图分类号：TP391 文献标志码：A

基于异构卷积的轻量级图像分类网络

喻明毫，高建瓴，胡承刚
（贵州大学 大数据与信息工程学院，贵阳 550025）

摘 要：目前大多数大型神经网络都存在参数量大、计算难度高等问题，想要在移动端设备使用，则会受到计算资源的限制。虽然现有轻量级网络出现解决了一定的计算量的问题，但同时其网络中大量使用 1×1 点卷积，使得其成为了现在轻量级网络的计算瓶颈。针对点卷积造成的计算瓶颈的问题，首先提出使用 GhostModel 来代替其中一部分点卷积，然后结合异构卷积对残差结构进行改进并提出 ResHetModel_A、B 两个改进的模块，使用改进模块构成轻量级网络 HSNet。最后对注意力特征图进行分析，在网络加入注意力机制来提高网络表达。在 CAFIR10 和 CAFIR100 数据集上的分类实验证明网络的有效性。最后在 ImageNet 大型数据集上实验表明 HSNet 具有一定的泛化性。

关键词：轻量级网络；点卷积；异构卷积；残差结构；GhostModel

Lightweight image classification network based on heterogeneous convolution

YU Minghao, GAO Jianling, HU Chenggang
(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

【Abstract】 The problems of large parameter quantity and high computational difficulty exist with most large neural networks. If large neural networks want to apply to mobile devices, they are constrained by computing resources. Although existing lightweight networks solve some computational problems, at the same time, its network uses massive 1×1 point convolutions, which has become a computational bottleneck of the current lightweight network. In order to solve the problem of computing bottleneck caused by point convolution, first propose to use GhostModel to replace part of the point convolution, then combined with heterogeneous convolution to improve the residual structure, propose two improved modules ResHetModel_A and B, and use the improved modules to form a lightweight network HSNet. Finally, the attention feature map is analyzed, and attention mechanism is added to the network to improve network expression. The classification experiments on the CAFIR10 and CAFIR100 datasets prove the effectiveness of the network. Finally, experiments on the ImageNet large dataset show that HSNet has a certain generalization.

【Key words】 lightweight network; point convolution; heterogeneous convolution; residual structure; GhostModel

0 引 言

计算机视觉的发展推动人工智能不断进化，而作为计算机视觉强大进步源泉的深度学习，则在计算机视觉领域子任务，诸如图像分类、目标检测、图像分割等方面做出了重大贡献。与神经网络相结合的图像处理算法相较于传统的图像处理算法有巨大的精度优势。在大数据的时代，利用神经网络在数据中学习图像特征，继而进行分类、检测、分割等任务。目前，基于深度学习的图像分类网络层出不穷，大量优秀的网络不断问世，人们研究的重点是如何将图像分类精度提高，不断加深、加宽模型，虽然网络在精度上表现越发出众，但网络效率问题也随即产生。在实践中，为了将基于深度学习的图像分类技术应用于移动运算设备中，就需要考虑计算资源限度。分析可知，时下的大型分类网络出于其庞大的参数量和计算量等原因仍然难以落地移动端设备，在此基础上，研究人员就将网络轻量化作为另一个研究方向，并研发提出了体积小、速度快的模型用于图像分类。

2012 年提出的 AlexNet[1] 取得了 ImageNet 图像分类赛的冠军，此后优秀的模型不断涌现。AlexNet 不仅是其后续网络的雏形，同时还提出了 Multi-Path 的方式，通过使用 2 个 GPU 进行并行训练，网络也分为 2 个支路。受此启发，接下来就开始使用多路分支的方式来构造高效的分类网络。2014 年，Simonyan 等人[2] 提出的 VGG 结构是在此基础上构造，在 ImageNet 图像分类数据集 Top5 错误率中达到了 6.8%，但同年的冠军是 Szegedy 等人提出的 GoogLeNet[3]，其 Top5 错误率达到 6.67%，借鉴

作者简介：喻明毫(1997-)，男，硕士研究生，主要研究方向：信息与通信工程，深度学习；高建瓴(1969-)，女，硕士，副教授，主要研究方向：数据分析、数据库应用；胡承刚(1996-)，男，硕士研究生，主要研究方向：自然语言处理。
通讯作者：高建瓴 Email：454965711@qq.com

收稿日期：2021-04-10

AlexNet 的思想提出 Inception 模块。He 等人提出的 ResNet[4] 使用分支的形式构造残差模块，并使用残差网络构成 ResNet。虽然大型网络在精度上不断提高，单为了在有限的计算资源下达到更好的检测效果，设计一种轻量化的网络相比大型网络能够在相同检测效果下消耗更低的资源。2016 年，Iandola 等人[5] 提出了 SqueezeNet，其模型大小只有0.5 M，主要网络由 FireModel 构造，其中大量使用了 1×1 卷积。Mobilenet[6] 中将深度卷积分成 2 步，提出深度可分离卷积，深度可分离卷积关注卷积方式，操作就是将普通深度卷积转换为逐通道卷积与点卷积。前者卷积核数与通道数相同，后者使用点卷积来混合通道信息，将每一个输入特征图信息在输出有所体现。此后有多人又提出了 ShuffleNet 系列[7-8] 使用通道混洗操作，其思想是将卷积分组，加强每个通道之间的信息交互，通道混洗的操作使各个输入通道信息在输出通道有所体现。这样可以减少通道数量的同时不损失通道信息。上述大多轻量化设计中大量地使用 1×1 卷积来压缩参数量，使得网络 *FLOPs* 暴增，1×1 显然成为轻量型网络设计的计算瓶颈，于是在基于如何去除卷积同时不增加参数量和 *FLOPs* 的研究上，Vahid 等人[9] 结合快速傅里叶变换（FFT）中的蝶形运算与卷积操作提出 Butterfly Transform 来无限逼近卷积并应用于卷积神经网络中，来降低计算复杂度。Li 等人[10] 又提出 MicroNet 思想是分解矩阵，具体操作是将卷积核矩阵分解为 2 组自适应卷积。Han 等人[11] 在研究 CNN 提取的特征图中发现大量特征图存在冗余的情况，于是使用 Ghost 幻影图来代替冗余的特征图。

本文结合异构卷积与 GhostModel 构造一种轻量级的分类网络，此分类网络中不会大量使用 1×1 卷积，同时，网络风格类似于 ResNet，但网络模型远小于 ResNet，将此轻量级网络命名为 HSNet。

1　相关工作

1.1　异构卷积

异构卷积是由 Singh 等人[12] 提出的一种不同于传统的卷积方式，图像分类中的滤波器含有大量 3×3 的卷积核，异构卷积主要是使用 1×1 卷积核，3×3 卷积核进行排列，以此减少参数。此方式使用了通用逻辑门的思想，将复杂操作简单化。原始异构卷积如图 1 所示。

图 1 中，P 是超参数，通过 P 来控制 1 × 1 卷积的数量，通过在 3×3 卷积之间插入 1×1 可以达到减少参数的效果，本文基于 GhostModel 对异构卷积进行改进，改进后的异构滤波器如图 2 所示。在图 2 中，将其中的 1 × 1 卷积使用如图 3 所示的 GhostModel 替换。

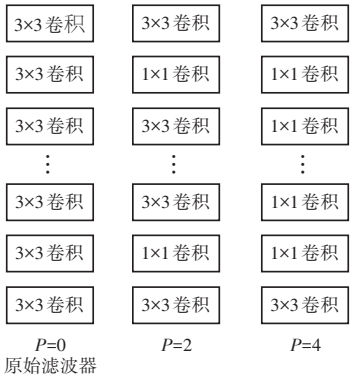


图 1　异构卷积

Fig. 1　Heterogeneous convolution

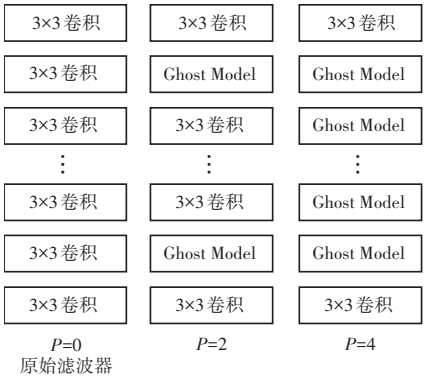


图 2　改进后的异构卷积

Fig. 2　Improved heterogeneous convolution

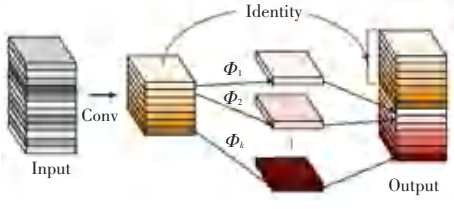


图 3　GhostModel

Fig. 3　GhostModel

1.2　GhostModel

传统卷积方式产生的特征图含有大量冗余，表现为同一滤波器产生的特征图非常相似，相似特征图之间可以通过一系列线性变换，Han 等人[11] 提出使用 GhostModel 来代替一部分卷积，GhostModel 思想是先使用少量的卷积生成原始特征图，然后使用这些特征图生成“幻影图”来代替原来冗余的特征，此方式可以减少大量参数以及 *FOLPs*。 图 3 中表示了 GhostModel。

2 网络结构设计

2.1 节中主要根据 Xception[13] 网络对第一层卷积 StemBlock 进行改进。在上一节中,已经对异构卷积进行了改进,在 2.2 节中会使用改进的异构卷积与 Ghost Model 构造一个模块,并以此模块搭建网络。对此拟展开分析论述如下。

2.1 StemBlock

Xception 网络中使用 Inception 结构来构成整个网络,其中网络前两层是由 2 个 3×3 卷积构成,用其初步提取特征,受 Szegedy 等人研究成果[14] 的启发,本文构造了一种轻量级的 StemBlock,2 种不同的 StemBlock 如图 4 所示。图 4 左侧图中,SConv 表示深度可分离卷积,对比原始 StemBlock,如图 4 右侧图所示,本文的结构可以获得不同的特征表达,并减少参数量。

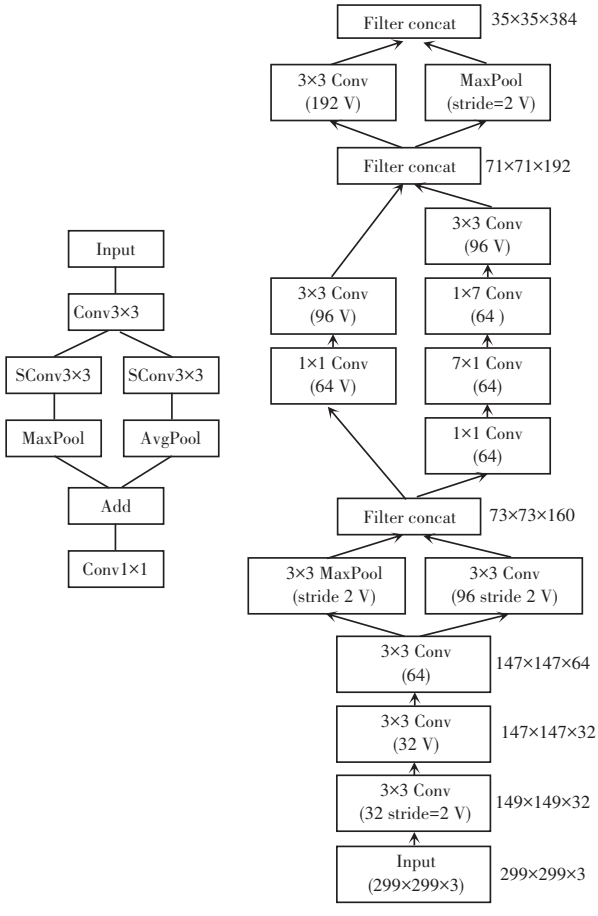


图 4 2 种不同的 StemBlock

Fig. 4 Two different StemBlock

2.2 异构卷积模块

特征图注意力机制和多路径表示对视觉识别非常重要,特征图注意力机制一般有通道注意力机制以及空间注意力机制,都是通过池化生成一个权重系数向量,再与原特征图相乘得到注意力图,本文在网络中使用的通道注意力 ECA 模块如图 5 所示。

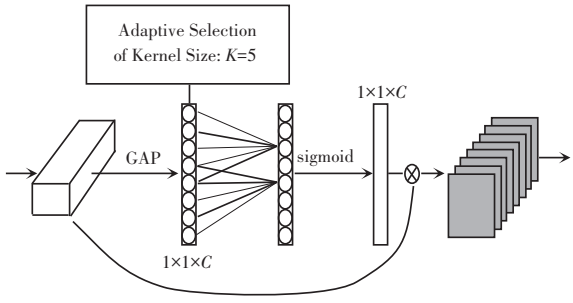
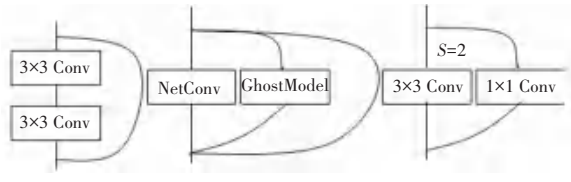


图 5 ECA-Layer

Fig. 5 ECA-Layer

分析可知,神经网络会出现随着网络加深,训练集准确率下降的现象,何凯明等人指出网络加深会出现梯度消失的情况,并提出了残差网络 ResNet,其中包含了一个直连路线,网络输出等于输入加上卷积后的输出。本文在残差网络的基础上结合异构卷积构造一个轻量级模块 ResNetModel,模块由异构卷积和 GhostModel 组成。

原始残差网络如图 6(a)所示,本文改进的残差网络如图 6(b)、图 6(c)所示。原始残差网络通过直连抵消网络层数过深导致的梯度消失现象,本文结合异构卷积和幻影图操作改进残差网络,网络主要由图 6(b)、图 6(c)的模块构成,在每个阶段的初始阶段,使用 ResNetModel_B 残差模块降采样,接着使用 ResNetModel_A 重复,加强特征表达,在模块 ResNetModel_A 中,还会使用注意力机制来提升视觉表达。网络整体结构见表 1。表 1 中, *FLOPs* 为 231.87 M。



(a) ResBlock (b) ResNetModel_A (c) ResNetModel_B

图 6 原始残差网络与改进的残差网络

Fig. 6 Original residual network and improved residual network

表 1 网络整体结构

Tab. 1 Overall network structure

Satge	In size -Out size	ECA
Stem	224-112	0
1	112-56	1
2	56-28	1
3	28-14	1
4	14-7	0
FC	1 280×1×1	-

3　实验结果分析

实验环境操作系统为 Ubuntu18.04，使用 GPU 训练，深度学习框架为 Pytorch。首先为了测试注意力对网络的影响，进行了消融实现，设置不同阶段添加 ECA 模块进行实验。实验数据集使用 CAFIR10 数据集，在此数据集中总共有 10 类目标。实验结果见表 2。表 2 中，符号“√”表示使用 ECA。

表 2　ECA 对模型影响

Tab. 2　ECA effect on the model（symbol

with ECA	1	2	3	4	精度/%
HSNet					87.54
HSNet_W	√				88.23
HSNet_W		√			88.55
HSNet_W			√		**90.20**
HSNet_W				√	89.06

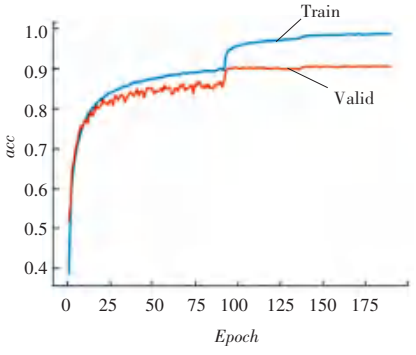
从表 2 的实验结果来看，使用 ECA 模块确实会使模型精度提高，但是使用的方式需要经过实验验证，本文在 1、2、3 阶段使用 ECA，4 阶段不使用时效果最好。表 1 中模型结构是其最优的形式，使用表 1 中的网络与其他分类网络在 CAFIR10 数据集以及 CAFIR100 数据集上进行实验，验证模型的有效性。结果见表 3。

表 3　CAFIR10 实验结果

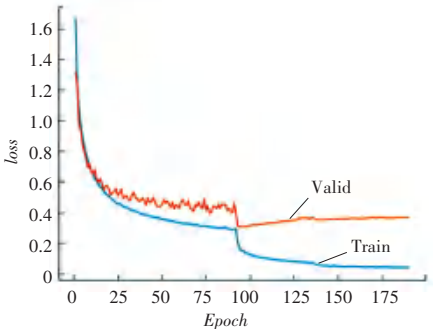
Tab. 3　CAFIR10 experimental results

网络	*FLOPs*	精度/%
ResNet50	4.09 G	84.16
Ghost ResNet50	3.92 G	91.10
VGG16	15.48 G	93.50
Ghost-VGG16	12.38 G	92.00
GhostNet	149.23 M	88.30
HSNet	231.87 M	90.20

图 7 是网络 HSNet 在 CAFIR10 数据集中训练集和验证集上的精度曲线和损失曲线。在验证集 HSNet 网络精度可以达到 90.2%。训练集和验证集上的混淆矩阵如图 8 所示。



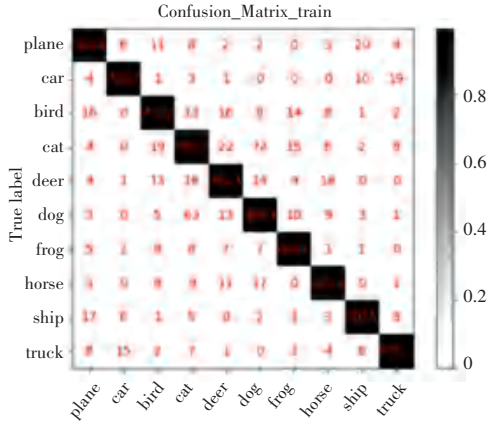
（a）精度曲线



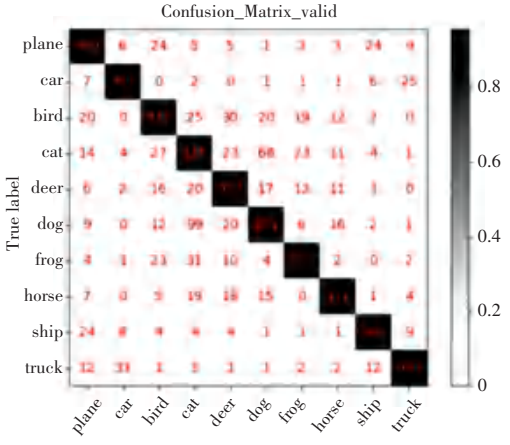
（b）损失曲线

图 7　CAFIR10 数据集训练结果

Fig. 7　CAFIR10 dataset training results



（a）训练集



（b）验证集

图 8　训练集和验证集混淆矩阵

Fig. 8　Training set and validation set confusion matrix

表 3 中给出了 ResNet、VGG 以及使用 Ghost Model 替换之后模型的精度对比。由结果来看，本文网络在 *FLOPs* 远低于 ResNet、VGG 及其 Ghost 替换模型的情况下没有很大的精度损失，在与同为轻量级网络的 GhostNet 对比下，本文实验精度更好。

ResNet 在 CAFIR100 和 CAFIR10 数据集上的实验中，网络使用了不同的通道数，其中 CAFIR10 数据集上 ResNet56 的第一个卷积层输出通道为 16，

CAFIR100 数据集实验中 ResNet50 第一个卷积层输出通道为 64,所以在 *FLOPs* 上有差距。初始通道数为 16 时,ResNet50 的参数量为 1.53 M。由表 3 和表 4 的结果来看,减少通道数可以减少参数以及 *FOLPs*,但是实验效果不理想。

本文还在 CAFIR100 数据集上进行分类实验,CAFIR100 数据集一共有 100 类,每个类包含 600 个图像,每类由 500 个训练图以及 100 张测试图组成。实验评价标准使用 Top1 错误率和 Top5 错误率。使用本文 HSNet 对比了轻量级网络以及非轻量级网络,实验结果参见表 4。

表 4 CAFIR100 数据集实验结果

Tab. 4 Experimental results of CAFIR100 datasets

		参数/M	*FLOPs*	Top1 错误率/%	Top5 错误率/%
Lightweight	GhostNet	4.00	149.23 M	36.22	12.33
网络	MobileNet	3.30	2.34 G	34.02	10.56
	MobileNetV2	2.36	2.42 G	31.92	9.02
	SqueezeNet	0.78	2.69 G	30.59	8.36
	ShuffleNet	1.00	2.22 G	29.94	**8.35**
	ShuffleNetV2	1.30	2.23 G	30.49	8.49
	HSNet(本文模型)	3.00	231.87 M	**29.25**	8.41
un-Lightweight	VGG16	34.00	15.48 G	27.07	8.84
网络	ResNet50	23.70	63.99 G	22.61	6.04
	ResNeXt50	14.80	37.34 G	22.23	6.00
	ResNeXt101	42.90	70.36 G	**22.22**	**5.99**

在 CAFIR100 数据集上的训练损失曲线如图 9 所示,本文提出的 HSNet 与现有的轻量级网络在参数量的对比上处于中间水平,但是由 *FLOPs* 标准来评价本文轻量级网络优于现有的轻量级网络,精度同样具有优势,具有一定的实际应用价值。相比大型网络参数量具有很明显的优势,在参数量以及 *FLOPs* 相差巨大的情况下,实验效果并没有损失多少。

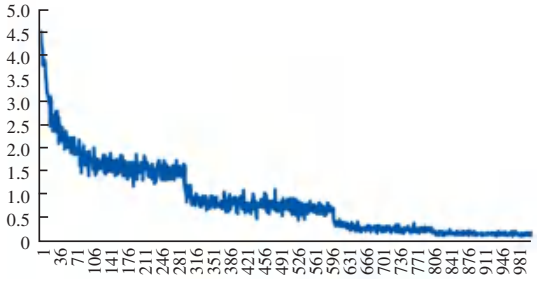


图 9 CAFIR100 训练损失曲线

Fig. 9 CAFIR100 training loss curve

为了验证 HSNet 在大型数据集上的稳定性以及泛化性,最后使用 HSNet 在 ImageNet 大型数据集上进行分类实验,实验结果见表 5。实验选取表 4 中的几个轻量化网络和非轻量化网络,采取 Top1 的精度和 Top5 的精度。试验结果表明网络在大型数据集上具有一定的稳定性。

表 5 ImageNet 实验结果

Tab. 5 ImageNet experimental results %

网络	Top1*Acc*	Top5*Acc*
GhostNet 0.5×	66.2	86.6
GhostNet 1.0×	73.9	91.4
ShuffleNetV2 1.0×	67.8	87.7
ShuffleNetV1 0.5× (g = 8)	58.8	81.0
ShuffleNetV2 0.5×	61.1	82.6
ResNet50	76.1	92.8
Vgg11	70.3	91.5
HSNet	69.7	88.9

4 结束语

本文受 Inception 网络结构的启发,构造了网络初始的 StemBlock 层,使用卷积池化的方式获取不同的特征表达,在网络的主干部分,结合改进后的异构卷积和 GhostModel 对原残差网络进行改进,提出 ResHetModel_A、B 两种新型的残差结构,使用这两种残差结构叠加,构成了本文提出的 HSNet,在 CAFIR10 和 CAFIR100 数据集上的实验证明了本文模型的有效性,在大型 ImageNet 数据集上说明轻量型网络 HSNet 具有一定的稳定性与泛化性。

参考文献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with Deep Convolutional Neural Networks[J]. Neural Information Processing Systems, 2012, 141:1097-1105.

[2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[3] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going Deeper with Convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015:1-9.

[4] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Identity mappings in deep residual networks[M]//LEIBE B, MATAS J, SEBE N, et al. Computer Vision-ECCV 2016. ECCV 2016. Lecture Notes in Computer Science. Cham:Springer, 2016, 9908: 630-645.

[5] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and<0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.

[6] HOWARDA G, ZHU Menglong, CHEN Bo, et al. MobileNets: Efficient Convolutional Neural Networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[7] ZHANG Xiangyu, ZHOU Xinyu, LIN Mengxiao, et al. ShuffleNet: An extremely efficient Convolutional Neural Network for mobile devices[C]//CVPR. Salt Lake City, UT:IEEE, 2018: 1-9.

[8] MA Ningning, ZHANG Xiangyu, ZHENG Haitao, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design[M]//FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision-ECCV 2018. ECCV 2018. Lecture Notes in Computer Science. Cham:Springer, 2018, 11218:122-138.

[9] VAHID K A, PRABHU A, FARHADI A, et al. Butterfly transform: An efficient fft based neural architecture design[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 12021-12030.

[10] LI Yunsheng, CHEN Yinpeng, DAI Xiyang, et al. MicroNet: Towards Image Recognition with Extremely Low FLOPs[J]. arXiv preprint arXiv:2011.12289, 2020.

[11] HAN Kai, WANG Yunhe, TIAN Qi, et al. GhostNet: More Features from Cheap Operations[J]. arXiv preprint arXiv: 1911.11907, 2020.

[12] SINGH P, VERMA V K, RAI P, et al. HetConv: Heterogeneous Kernel-Based Convolutions for Deep CNNs [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA:IEEE, 2019:4830-4839.

[13] CHOLLET F. Xception: Deep learning with Depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, Hawai:IEEE, 2017, 1: 1800-1807.

[14] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inceptionv4, Inception-ResNet and the impact of residual connections on learning[C]//AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. California USA:AAAI, 2017: 4278-4284.

(上接第 195 页)

[45] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional architecture for fast feature embedding[C]// ACM International Conference on Multimedia. Orlando, Florida: ACM, 2014: 675-678.

[46] SPIVAK M D. A comprehensive introduction to differential geometry[M]. Houston, Texas: Publish or Perish, inc, 1970.

[47] BERTINETTOL, HENRIQUES J F, VALMADRE J, et al. Learning feed-forward one-shot learners[C]//Advances in Neural Information Processing Systems. Barcelona: MIT Press, 2016:523-531.

[48] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[C]//Advances in Neural Information Processing Systems. Barcelona: MIT Press, 2016: 3630-3638.

[49] SUKHBAATAR S, WESTON J, FERGUS R, et al. End-to-end memory networks[C]//Advances in Neural Information Processing Systems. Montréal, Canada: Google, 2015:2440-2448.

[50] WESTON J, CHOPRA S, BORDES A. Memory networkss[J]. arXiv preprint arXiv:1410.3916, 2014.

[51] MILLER A, FISCH A, DODGE J, et al. Key-value memory networks for directly reading documents[C]// Conference on Empirical Methods in Natural Language Processing. Austin, Texas, USA: ACL, 2016:1400-1409.

[52] RAMALHO T, GARNELO M. Adaptive posterior learning: Few-shot learning with a surprise-based memory module[C]// International Conference on Learning Representations. Louisiana, United States: ICLR, 2019:1-14.

[53] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning[C]// Advances in Neural Information Processing Systems. Long Beach: Microsoft, 2017:4077-4087.

[54] BERTINETTO L, HENRIQUES J F, VALMADRE J, et al. Learning feed-forward one-shot learners[C]// Advances in Neural Information Processing Systems. Barcelona: MIT Press, 2016:523-531.

[55] MUNKHDALAI T, YUAN X, MEHRI S, et al. Rapid adaptation with conditionally shifted neurons[C]//International Conference on Machine Learning. Stockholm, Sweden: Intuit, 2018: 3661-3670.

[56] EDWARDS H, STORKEY A. Towards a neural statistician[C]// International Conference on Learning Representations. Toulon, France: Bengio, 2017:1-14.

[57] REED S, CHEN Y, PAINE T, et al. Few-shot autoregressive density estimation: Towards learning to learn distributions[C]// International Conference on Learning Representations. Vancouver, BC, Canada: Google, 2018:1-11.

[58] GORDON J, BRONSKILL J, BAUER M, et al. Meta-learning probabilistic inference for prediction[C]//International Conference on Learning Representations. Louisiana, USA: DeepMind, 2019:1-22.

[59] ZHANG Rruixiang, CHE Tong, GHAHRAMANI Z, et al. MetaGAN: An adversarial approach to few-shot learning[C]// Advances in Neural Information Processing Systems. Montreal, Canada: NIPS, 2018:2371-2380.