

# 基于滤波器分布拟合的神经网络剪枝算法

张佳钰<sup>1</sup>, 寇金桥<sup>2</sup>, 刘宁钟<sup>1</sup>

(1. 南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106;

2. 北京计算机技术及应用研究所 方舟重点实验室, 北京 100854)

**摘要:**随着人工智能技术的迅猛发展,深度神经网络在不断地加深与变宽,模型的计算量快速增加,神经网络模型的高存储和高功耗的需求也随之产生。网络剪枝是实现模型压缩和加速的一种有效方法。常见的剪枝方法遵循“较小规范-不重要”的标准来对滤波器进行修剪,认为范值较小的滤波器重要性较低,可以安全地修剪掉。针对删去重要性较小的滤波器容易导致滤波器范数分布不均衡的问题,文中提出了一种拟合原始滤波器范数分布的剪枝算法。该算法不仅可以筛选出拟合了原始范数分布的滤波器,还能删去冗余的滤波器。实验表明该算法在两个数据集上的模型压缩效果均优于对比实验。其中,在 CIFAR-10 数据集上压缩基于 ResNet110 的图像分类模型的效果明显,最终在减少了 62% 以上的 FLOPs 的情况下,相对准确率仅降低了 0.14%。

**关键词:**深度学习;模型压缩;网络剪枝;分布拟合;滤波器剪枝

中图分类号: TP391.4

文献标识码: A

文章编号: 1673-629X(2022)12-0136-06

doi: 10.3969/j.issn.1673-629X.2022.12.021

## Deep Convolutional Neural Networks Pruning Algorithm Based on Filter Pruning via Distribution Fitting

ZHANG Jia-yu<sup>1</sup>, KOU Jin-qiao<sup>2</sup>, LIU Ning-zhong<sup>1</sup>

(1. School of Computer Science and Technology, Nanjing University of Aeronautics and

Astronautics, Nanjing 211106, China;

2. Fangzhou Key Laboratory, Beijing Institute of Computer Technology and Application, Beijing 100854, China)

**Abstract:** With the rapid development of artificial intelligence technology, deep neural networks are constantly deepening and widening, and the computational amount of the model is increasing rapidly. Therefore, the demand of high storage and high power consumption of the neural network model is also generated. Network pruning is an effective way to achieve model compression and acceleration. The filters are usually pruned by following the "smaller norm-less important" criterion, where filters with smaller norm values are considered less important and can be safely pruned out. And the deletion of filters with smaller importance easily leads to the problem of unbalanced distribution of filter norms. In this regard, a pruning algorithm for fitting the original filter norm distribution is proposed, which not only retains the filters that can fit the distribution of filter weights of the original network, but also deletes the redundant filters. The experiments demonstrate that the proposed method outperforms the comparison experiments in terms of model compression on both datasets. Among them, the effect of compressing the ResNet110-based image classification model on the CIFAR-10 dataset is obvious, which ends up with a relative accuracy reduction of only 0.14% with over 62% reduction in FLOPs.

**Key words:** deep learning; model compression; network pruning; distribution fitting; filter pruning

## 0 引言

深度神经网络更深更宽的架构为计算机视觉任务带来了卓越的性能<sup>[1-2]</sup>。例如, Krizhevsky 等人<sup>[3]</sup>在 2012 年 ImageNet 挑战赛<sup>[4]</sup>中取得了突破性的成果,使用的网络包含 6 000 万个参数,有五个卷积层和三个

全连接层。在 Labeled Faces in the Wild (LFW) 数据集上的顶级人脸验证结果包含数亿个参数,使用了卷积层、局部连接层和全连接层的组合<sup>[5-6]</sup>。训练这样一个模型以获得合理的性能是非常耗时的。在只依靠全连接层的架构中,参数的数量可以增长到数十亿。

收稿日期: 2021-12-21

修回日期: 2022-04-22

基金项目: 国家自然科学基金(61375021)

作者简介: 张佳钰(1996-),女,硕士,CCF 会员(B5296G),通讯作者,研究方向为计算机视觉;寇金桥,硕士,研究员,研究方向为嵌入式系统架构设计、异构加速及软件开发;刘宁钟,博士,教授,研究方向为计算机视觉和模式识别。

这些工作依赖于具有数百万甚至数十亿参数的深度网络,而具有非常高的计算能力的 GPU 的可用性对其成功起到了关键性的作用。然而,这也导致了过于昂贵的计算成本,并使模型在移动设备上的部署变得困难,甚至不可能。通常情况下,用一台 NVIDIA K40 机器在 ImageNet 数据集上训练整个模型需要两到三天时间。即使是具有高效率的最新架构,如残差连接<sup>[7]</sup>或 inception 模块<sup>[8]</sup>,也有数百万个参数需要数十亿次浮点运算(FLOPs)。在进行模型部署应用时,硬件不能承载模型巨大的参数量。因此,研究一个能够将模型的参数量和浮点运算(FLOPs)降低,却又能够保持模型的精度的算法,是非常有实际价值的。

针对以上问题,在模型剪枝中,针对滤波器的选择策略,该文设计了能拟合原始滤波器范数分布的算法,并在此算法的基础上,对滤波器的修剪策略进行了相关的改进和设计,使其更加适用于图像分类的场景,从而减缓了小型设备的算力与模型的巨大运算量需求无法匹配的问题,并成功提高了模型的推理速度。

## 1 相关工作

伴随着深度学习技术在目标检测、图像分类、点云等领域的应用普及,在边缘设备上应用深度模型解决实际问题引起了人们的广泛关注并逐渐成为可能。模型压缩<sup>[9-11]</sup>的方法主要分为以下几类:参数修剪和量化、低等级因子化、转移或压缩卷积滤波器和知识提炼。该文主要研究基于图像分类的模型剪枝算法,从而实现对模型的压缩与加速。网络剪枝<sup>[12-15]</sup>可以分为两类,即权重修剪<sup>[16-17]</sup>和滤波器修剪<sup>[18-19]</sup>。权重修剪直接删除滤波器的权重值,可能会导致一种非结构化的稀疏。这种不规则的结构使得它难以利用高效的基本线性代数子程序(BLAS)库<sup>[20]</sup>。与此相反,滤波器修剪直接丢弃了整个选定的滤波器,留下一个具有规则结构的模型。因此,它不需要稀疏卷积库的支持,并且可以与现有的用于密集矩阵乘法的高效 BLAS 库协同工作。因此,滤波修剪更适合用来加速和减少模型大小。

目前,文献<sup>[18,21-22]</sup>通过遵循“较小规范-不重要”的标准来进行滤波器修剪,认为范值较小的滤波器重要性较低,可以安全地修剪掉。而 He 等人<sup>[23]</sup>提出通过几何中值修剪滤波器(FPGM)的方法来压缩模型,不是根据滤波器的重要性来修剪,而是修剪该层中距离数据中心较近的滤波器,该文认为这些滤波器信息是重合的,可能会产生冗余信息,所以这些滤波器不会对网络产生影响。但是笔者认为滤波器数据中心的数据可能是该层较为重要的数据,仅对这一部分数据进行剪枝容易使得滤波器数据质量下降。在训练后,

滤波器的权重总是服从某一个分布,而修剪一部分数据中心附近的数据,会导致分布不均衡。集中修剪某一个点附近的数据后,其分布不再拟合原始的数据分布。

## 2 模型和方法

以前的工作通常使用“较小规范-不重要”的标准来修剪卷积神经网络中具有较小值的滤波器。该文指出,通过拟合滤波器的范数分布的方式,可以获得更好的效果。基于此,该文提出了一种基于拟合滤波器范数分布的剪枝算法来加速深度卷积神经网络(CNN)的推理过程。具体来说,所提出的 FPDF(Filter Pruning via Distribution Fitting)使得筛选的滤波器权重范数的分布能够拟合原始网络的滤波器权重范数的分布。与以前方法不同的是,FPDF 通过修剪有冗余的滤波器来压缩模型,而不是筛选掉相对权重值较小或者范数值较小的滤波器。

在两个图像分类基准上进行实验,实验结果验证了 FPDF 的有效性和优越性。值得注意的是,在 CIFAR-10 上,FPDF 在 ResNet20 上减少了 62% 以上的 FLOPs,相对准确率仅降低了 0.14%。此外,在 CIFAR-100 上,FPDF 在 ResNet56 上减少了 63% 以上的 FLOPs,相对准确率仅降低了 1.27%。

### 2.1 基于拟合滤波器范数分布的剪枝算法

FPGM<sup>[23]</sup>指出处于中位数附近的滤波器包含了冗余的信息,这些滤波器是可以被替代的,并提出了基于几何中位数的剪枝方法。如果将中位数附近的滤波器减去,会造成滤波器分布上的一种空缺。位于中位数附近的滤波器对网络的贡献较大,当剪枝率较大的时候,FPGM 会将重要的滤波器误删,对计算结果造成较大的波动。图 1 分别展示了原滤波器范数分布(图 1a),按照文中算法删除滤波器后的范数分布(图 1b),按照 SFP 删除滤波器后的范数分布(图 1c)。由图 1a 和图 1b 的分布可以看出,文中算法使得滤波器删除前后的数据分布趋于一致。SFP 删去的数据都是滤波器范数排序靠后,处于底部的数据,其分布已经和图 1a 的分布区分开了。删除尾部范值较小的数据之后,分布曲线的尾部明显消失,并且其规范值的偏离是显著的。

### 2.2 算法流程

算法的总体流程如图 2 所示。

针对每一层网络的所有滤波器,按照滤波器的范数数值进行排序,然后按照设置好的间隔选择要被删去的滤波器。 $M$ 、 $N$  为间隔参数,表示每  $M$  个数据取前  $N$  个, $N/M$  也表示压缩率。间隔参数将滤波器分为  $M$  个组,每一个组中的数据都是相似的,对滤波器范数分布的贡献趋于相同,所以仅删去其中的前  $N$  个数据,

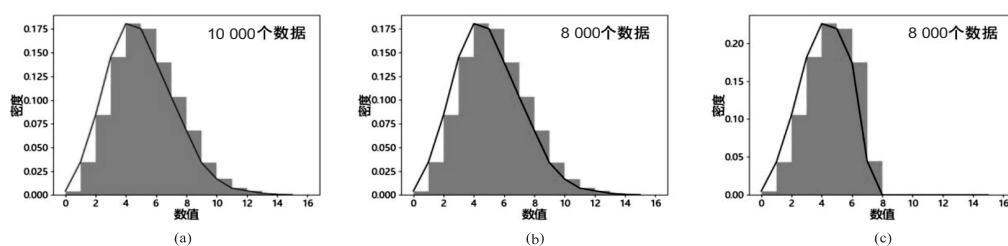


图 1 删去部分数据前后的数据分布

能够使其保持原有的滤波器范数分布,即仅使用后  $M - N$  个数据来代表组内  $M$  个数据。当压缩率为 20% 时,数据筛选方法是间隔四个数据取一个 ( $1/5 = 20\%$ )。当删除率为 50% 时,间隔为 2,表示每两个数据取出一个数据。

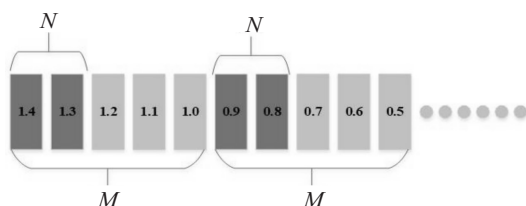
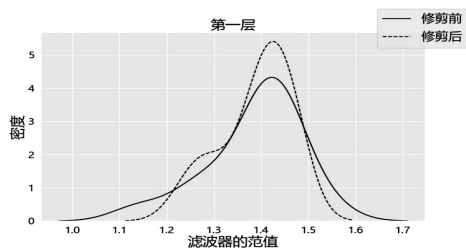
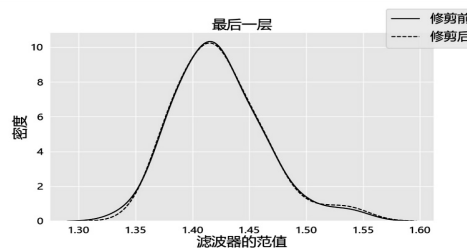


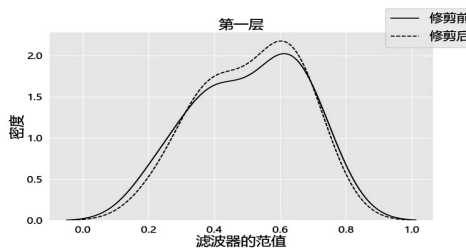
图 2 FPDF 选择需要删除的滤波器的策略图示



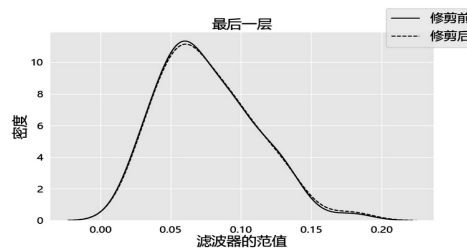
(a) ResNet56 网络第一层滤波器分布



(b) ResNet56 网络最后一层滤波器分布



(c) ResNet110 网络第一层滤波器分布



(d) ResNet110 网络最后一层滤波器分布

图 3 ResNet56 和 ResNet110 在 CIFAR-10 数据集上网络第一层和最后一层的滤波器分布  
(实线表示原始滤波器分布,虚线表示通过 FPDF 选取的滤波器分布)

### 3 实验结果与分析

#### 3.1 数据集描述

CIFAR-10 (Canadian Institute for Advanced Research - 10 classes) 是一个用于识别普适目标的数据集。共包含 10 个类别的 RGB 彩色图片:飞机、汽车、鸟类、猫、鹿、狗、蛙类、马、船和卡车。数据集中一共有 50 000 张训练图片和 10 000 张测试图片,图片的尺寸为  $32 \times 32$ 。图片所包含的都是现实世界中真实的物体,含有较大的噪声,并且物体的比例与特征都不尽相同。

图 3 展示了应用 FPDF 的滤波器选择算法前后的滤波器范数分布。ResNet56 和 ResNet110 的第一层仅有 16 个滤波器,最后一层有 64 个滤波器。图 3a 和图 3c 为网络第一层的数据,图 3b 和图 3d 为网络最后一层的数据。可以看出,删去 20% 的滤波器之后,其分布随着滤波器数量的增加可以拟合删除前的分布。

与 SFP<sup>[22]</sup> 不同,该文采用的是硬剪枝,在选择拟合原模型分布的滤波器之后对模型进行一次性的裁剪。在剪枝之后,对得到的小模型进行微调,微调可以使模型收敛。具体步骤为对小模型进行训练,直到模型收敛。

CIFAR-100 (Canadian Institute for Advanced Research - 100 classes) 是一个微小图像数据集的子集。该数据集包含 100 个类别的图片,每个类别有 600 张图片,其中 500 张用于训练,另外 100 张用于测试。图片尺寸为  $32 \times 32$ 。该数据集中包含两个维度的类别:100 个粗粒度类别和 200 个细粒度类别,每一张图片都有一个粗粒度类别和一个细粒度类别。

#### 3.2 实验细节

实验基于 ResNet20、ResNet32、ResNet56、ResNet110 网络,在 CIFAR-10 和 CIFAR-100 两个数据集上对 FPDF 进行验证。对比实验的方法包括

SFP<sup>[22]</sup>、FPGM<sup>[23]</sup>、PEEC<sup>[18]</sup>、CP<sup>[24]</sup> 等加速算法。CIFAR 数据集的输入为  $32 \times 32$  的图像,而原始的 ResNet 第一层卷积核为  $7 \times 7$ ,这样的卷积结构对于 CIFAR 数据集的性能表现较差,所以将 ResNet 第一层网络的卷积核修改为  $3 \times 3$ ,并去掉之后的最大池化层。该文对这个修改后的网络进行剪枝以适用于基于 CIFAR 的图像分类场景。

### 3.3 训练设置

文中实验参数设置与 FPGM 文中的参数保持一致。在 CIFAR-10 的实验中,使用的预训练模型为 FPGM 代码仓库中提供的模型,因此不需要重新训练基线模型。在 CIFAR-100 的实验中,按照本节中的设置训练基线模型。文中实验在 Ubuntu16.04 操作系统下进行,使用 TITAN XP 进行训练,使用 Pytorch1.4.0 框架实现。学习率(learning rate)初始值设置为 0.02,分别在第 60 轮、120 轮以及 160 轮的时候分别衰减为 0.000 2、0.000 02 和 0.000 002;动量值为 0.9;使用 SGD 表示优化器;批大小(batch size)设置为 128。

### 3.4 剪枝设置

将 FPDF 应用到四个网络中,在 CIFAR-10 上的实验设置 40% 和 50% 的压缩率,表示将模型的残差块的滤波器剪去 40% 和 50%。在 CIFAR-100 的实验上设置了 50% 和 60% 的压缩率,表示将模型的残差块的滤波器修剪去 50% 和 60%。在滤波器修剪的过程中,为了与 SFP、FPGM 的设置保持一致,为网络的每一层设置相同的剪枝率。将大模型修剪成小模型后会对模型进行微调,促使模型收敛。

### 3.5 评价标准

准确率:对于一个给定的数据集,分类器将其分类正确的样本数与总样本数的比例。计算公式为:

$$\text{Accuracy} = \text{TP} / (\text{TP} + \text{FP})$$

其中,TP 表示所有被预测为正类的正类实例个数,FP 表示所有被预测为负类的正类实例个数。下文中的准确率 Accuracy 使用简写 accu. 代替。

FLOPs; Floating Point Operations,每秒浮点运算次数,通常用来衡量网络模型的计算复杂度,也可以用来间接地衡量网络模型的速度。

### 3.6 结果分析

实验结果如表 1 和表 2 所示。表格第一列均表示不同的网络模型,ResNet20、ResNet32、ResNet56、ResNet110,第二列均表示剪枝的方法,第三列均表示该网络下训练的大模型的准确率,第四列均表示模型压缩之后的准确率,第五列均表示模型压缩后下降的准确率,第六列均表示模型压缩后的 FLOPs,第七列均表示模型压缩后减少的 FLOPs。在 CIFAR-10 的实验结果中 FPDF 在剪枝率为 50% 的时候,文中方法的准确率都优于对比方法。在基于 ResNet110 的剪枝上,修剪去 62.8% 的 FLOPs,可以将准确率恢复至 93.54%。在基于 CIFAR-100 的实验结果中,仅有基于 ResNet20 的剪枝效果略低于 SFP,但是在基于层数较多的网络上的剪枝效果均优于 SFP 和 FPGM,并且 FPDF 在 ResNet110 上压缩了 72.3% 的 FLOPs 的准确率比压缩了 62.8% 的 FLOPs 的准确率高了 0.2%。

表 1 在 CIFAR-10 数据集上剪枝的结果比较

模型	方法	基线 accu. / %	加速 accu. / %	accu. 下降 / %	FLOPs	减少 FLOPs / %
ResNet20	SFP(50%)	92.20	91.11	1.09	1.43E7	64.5
	FPGM(50%)	92.20	91.45	0.75	1.43E7	64.5
	FPDF(50%)	92.20	91.48	0.71	1.43E7	64.5
ResNet32	SFP(50%)	92.63	90.65	1.97	2.49E7	63.7
	FPGM(50%)	92.63	90.76	1.86	2.49E7	63.7
	FPDF(50%)	92.63	90.78	1.84	2.49E7	63.7
ResNet56	PFEC	93.04	91.31	1.75	9.09E7	27.6
	CP	92.80	90.90	1.90	-	50.0
	SFP(40%)	93.59	92.26	1.33	5.94E7	52.6
	FPDF(40%)	93.59	92.43	1.15	5.94E7	52.6
	SFP(50%)	93.59	92.24	1.35	4.62E7	63.1
	FPGM(50%)	93.59	92.04	1.54	4.62E7	63.1
	FPDF(50%)	93.59	92.26	1.32	4.62E7	63.1
ResNet110	SFP(50%)	93.68	93.50	0.18	9.40E7	62.8
	FPGM(50%)	93.68	93.46	0.22	9.40E7	62.8
	FPDF(50%)	93.68	93.54	0.14	9.40E7	62.8



表 2 在 CIFAR-100 数据集上剪枝的结果比较

模型	方法	基线 accu./%	加速 accu./%	accu. 下降/%	FLOPs	减少 FLOPs/%
ResNet20	SFP(50%)	68.73	65.11	3.62	1.43E7	64.5
	FPGM(50%)		65.17	3.56	1.43E7	64.5
	FPDF(50%)		65.25	3.48	1.43E7	64.5
	SFP(60%)		63.65	5.08	1.05E7	73.9
	FPGM(60%)		63.63	5.10	1.05E7	73.9
	FPDF(60%)		63.40	5.33	1.05E7	73.9
ResNet32	SFP(50%)	70.87	67.98	2.89	2.49E7	63.7
	FPGM(50%)		67.96	2.91	2.49E7	63.7
	FPDF(50%)		68.00	2.87	2.49E7	63.7
	SFP(60%)		66.72	4.15	1.84E7	73.1
	FPGM(60%)		66.17	4.70	1.84E7	73.1
	FPDF(60%)		66.88	3.99	1.84E7	73.1
ResNet56	SFP(50%)	71.82	70.36	1.45	4.62E7	63.1
	FPGM(50%)		70.10	1.71	4.62E7	63.1
	FPDF(50%)		70.54	1.27	4.62E7	63.1
	SFP(60%)		68.45	3.36	3.43E7	72.6
	FPGM(60%)		68.90	2.91	3.43E7	72.6
	FPDF(60%)		68.96	2.85	3.43E7	72.6
ResNet110	SFP(50%)	74.46	72.02	2.43	9.40E7	62.8
	FPGM(50%)		72.04	2.41	9.40E7	62.8
	FPDF(50%)		72.05	2.40	9.40E7	62.8
	SFP(60%)		71.86	2.59	7.00E7	72.3
	FPGM(60%)		71.68	2.77	7.00E7	72.3
	FPDF(60%)		72.25	2.20	7.00E7	72.3

#### 4 结束语

剪枝算法通常是依据滤波器的重要性进行滤波器的删除,经典的剪枝方法仅依照滤波器权重值或范数值的大小对滤波器的重要性进行评估,往往导致所选取的滤波器的范数分布发生变化,造成一定的精度损失。因此,该文提出了一种拟合滤波器范数分布的方法,用来解决范数分布不均衡问题。并详细介绍了该方法的算法流程,展示了其与现有算法的剪枝效果区别。通过在公开数据集(CIFAR-10, CIFAR-100)上进行实验,展示了该方法能够更好地维持原范数分布,并证明了该方法相比于现有方法,在图像分类任务上有更高的精度,且具有一定的鲁棒性,可以泛化到多个数据集上。

综上所述,提出的滤波器选择策略改进了现有的模型剪枝算法,使剪枝后的小模型维持较高的精度,有助于实际模型的产品落地。

#### 参考文献:

- [1] DONG X, YANG Y. Searching for a robust neural architecture in four GPU hours[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Long Beach; IEEE, 2019: 1761-1770.
- [2] LUO Y, ZHENG L, GUAN T, et al. Taking a closer look at domain shift: category-level adversaries for semantics consistent domain adaptation[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Long Beach; IEEE, 2019: 2507-2516.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of advances in neural information processing systems 25: 26th annual conference on neural information processing systems 2012. Lake Tahoe: [s. n.], 2012: 1106-1114.
- [4] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [5] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: closing the gap to human-level performance in face verification[C]//Proceedings of 2014 IEEE conference on computer vision and pattern recognition. Columbus: IEEE, 2014: 1701-1708.
- [6] LU Y, KUMAR A, ZHAI S, et al. Fully-adaptive feature sharing in multi-task networks with applications in person at-

- tribute classification[C]//Proceedings of 2017 IEEE conference on computer vision and pattern recognition. Honolulu; IEEE,2017;1131–1140.
- [7] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas; IEEE,2016;770–778.
- [8] SZEGEDY C,LIU W,JIA Y,et al. Going deeper with convolutions[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Boston;IEEE,2015;1–9.
- [9] 田冠中. 面向深度卷积神经网络的模型压缩关键技术研究[D]. 杭州:浙江大学,2021.
- [10] 龙 鑫. 深度卷积网络模型的压缩和加速技术研究[D]. 长沙:国防科技大学,2020.
- [11] 林绍辉. 深度神经网络压缩与加速关键技术研究[D]. 厦门:厦门大学,2019.
- [12] 姜春晖. 深度神经网络剪枝方法研究[D]. 合肥:中国科学技术大学,2019.
- [13] 罗建豪. 面向深度卷积神经网络的模型剪枝算法研究[D]. 南京:南京大学,2020.
- [14] 吴卫贤,赵 鸣,黄晓丹. 基于量化和模型剪枝的卷积神经网络压缩方法[J]. 软件导刊,2021,20(10):78–83.
- [15] 严阳春,郭荣佐,杨锦霞. 基于权重关联性的卷积神经网络模型剪枝方法[J]. 小型微型计算机系统,2021,42(7):1500–1504.
- [16] CARREIRA-PERPIÑÁN M Á, IDELBAYEV Y. Learning-compression algorithms for neural net pruning[C]//Proceedings of 2018 IEEE conference on computer vision and pattern recognition. Salt Lake City;IEEE,2018;8532–8541.
- [17] HAN S,POOL J,TRAN J,et al. Learning both weights and connections for efficient neural network[C]//Proceedings of advances in neural information processing systems 28; annual conference on neural information processing systems 2015. Montreal;[s. n.],2015;1135–1143.
- [18] LI H,KADAV A,DURDANOVIC I,et al. Pruning filters for efficient ConvNets[C]//Proceedings of 5th international conference on learning representations. Toulon;[s. n.],2017.
- [19] YU R,LI A,CHEN C,et al. NISP: pruning networks using neuron importance score propagation[C]//Proceedings of 2018 IEEE conference on computer vision and pattern recognition. Salt Lake City;IEEE,2018;9194–9203.
- [20] CHEN Y,WEN X,ZHANG Y,et al. CCPrune: collaborative channel pruning for learning compact convolutional networks[J]. Neurocomputing,2021,451;35–45.
- [21] YE J,LU X,LIN Z,et al. Rethinking the smaller-norm-less-informative assumption in channel pruning of convolution layers[C]//Proceedings of 6th international conference on learning representations. Vancouver;ICIR,2018.
- [22] HE Y,KANG G,DONG X,et al. Soft filter pruning for accelerating deep convolutional neural networks[C]//Proceedings of the twenty-seventh international joint conference on artificial intelligence. Stockholm;[s. n.],2018;2234–2240.
- [23] HE Y,LIU P,WANG Z,et al. Filter pruning via geometric median for deep convolutional neural networks acceleration[C]//Proceedings of IEEE conference on computer vision and pattern recognition. Long Beach; IEEE,2019;4340–4349.
- [24] HE Y,ZHANG X,SUN J. Channel pruning for accelerating very deep neural networks[C]//Proceedings of IEEE international conference on computer vision. Venice;IEEE,2017;1398–1406.
- +++++
- (上接第 80 页)
- [13] MATHY L,IANNONE L. LISP-DHT: towards a DHT to map identifiers onto locators[C]//Proceedings of the 2008 ACM CoNEXT conference. New York; Association for Computing Machinery,2008;1–6.
- [14] JAKAB L,CABELLOS-APARICIO A,CORAS F,et al. LISP-TREE: a DNS hierarchy to support the lisp mapping system[J]. IEEE Journal on Selected Areas in Communications,2010,28(8):1332–1343.
- [15] DANNEWITZ C,KUTSCHER D,OHLMAN B,et al. Network of information (netinf) – an information-centric networking architecture[J]. Computer Communications,2013,36(7):721–735.
- [16] D'AMBROSIO M,DANNEWITZ C,KARL H,et al. MDHT: a hierarchical name resolution service for information-centric networks[C]//Proceedings of the ACM SIGCOMM workshop on information-centric networking. New York; Association for Computing Machinery,2011;7–12.
- [17] SHARMA A,TIE X,UPPAL H,et al. A global name service for a highly mobile internetwork[J]. ACM SIGCOMM Computer Communication Review,2014,44(4):247–258.
- [18] VU T,BAID A,ZHANG Y,et al. Dmap: a shared hosting scheme for dynamic identifier to locator mappings in the global internet[C]//2012 IEEE 32nd international conference on distributed computing systems. Los Alamitos;IEEE,2012;698–707.
- [19] 尧 星. 新型互联网层级式映射系统的设计与实现[D]. 北京:北京交通大学,2018.
- [20] 陈嘉祺. 智慧协同网络分布式映射系统强一致性研究[D]. 北京:北京交通大学,2017.