第 41 卷第 1 期　　遥　测　遥　控　　Vol. 41, No. 1
2020 年 1 月　　Journal of Telemetry, Tracking and Command　　January 2020

Email:ycyk704@163.com　TEL:010-68382327　010-68382557

应用知识蒸馏的深度神经网络波束形成算法

柏沫羽，刘　昊，陈浩川，张振华
（北京遥测技术研究所　北京 100076）

摘要：自适应波束形成技术广泛应用于雷达领域的旁瓣抗干扰中。当回波数据量增多时，传统的波束形成算法无法进行快速处理，而应用深度神经网络模型通过数据的预训练则可以快速地进行波束形成，因此根据波束形成原理设计深度神经网络，并利用知识蒸馏的方式对深度神经网络进行压缩，使压缩后的模型既有原始模型良好的泛化性能而且又有更快的计算速度。仿真结果表明，相比于传统的 LMS 算法，在实验环境下，未经模型压缩的深度神经网络自适应波束形成算法的计算速度提高了约 7 倍，基于模型压缩的深度神经网络自适应波束形成算法的计算速度提高了约 20 倍。

关键词：信号处理；深度神经网络；自适应波束形成；知识蒸馏

中图分类号：TN911.7　　**文献标识码：**A　　**文章编号：**CN11-1780(2020)01-0066-07

Beamforming algorithm for deep neural network using knowledge distillation

BAI Moyu,　LIU Hao,　CHEN Haochuan,　ZHANG Zhenhua
（Beijing Research Institute of Telemetry, Beijing 100076, China）

Abstract: Adaptive beamforming technology is widely used in sidelobe anti-interference in the radar field. When the amount of echo data increases, the traditional beamforming algorithm cannot perform fast processing, and the deep neural network model can quickly perform beamforming through data pre-training. Therefore, this paper designs a deep neural network according to the beamforming principle. The deep neural network is compressed by means of knowledge distillation, so that the compressed model has both good generalization performance and faster calculation speed. The simulation results show that compared with the traditional LMS algorithm, the computational speed of the adaptive beamforming algorithm for deep neural networks without model compression is improved by about 7 times and the computational speed of the adaptive beamforming algorithm based on model compression is improved by about 20 times in the experimental environment.

Key words: Signal processing; Deep neural network; Adaptive beamforming; Knowledge distillation

引　言

自适应波束形成技术是阵列信号处理的重要分支，近年来已成为新一代多功能自适应相控阵雷达的关键技术之一。自适应波束形成技术能够通过调整接收通道权系数来有效地实现干扰抑制等功能，在雷达、无线通信、声纳、地震勘测等系统中得到了广泛的应用。最小均方误差算法 LMS（Least Mean Square Algorithm）是自适应波束形成算法中一种被广泛应用的迭代算法。但是当所得到的回波数据量增多时，传统的波束形成算法无法进行及时处理。而深度神经网络模型可以在前期对大量的数据进行训练，之后利用训练好的模型就可以快速准确地进行波束形成，比传统的波束形成算法更加快速。不过，为了满足模型的准确性，深度神经网络模型需要大量的参数，这将占用过多的资源和训练时间，因此应用知识蒸馏的方式对模型进行压缩，建立“学生”网络，使模型在保持精确性的同时又具有更快的计算速度，使算法满足在大数据的情况下进行快速波束形成的需求，具有理论上和工程上的双重研究意义。

将深度神经网络与自适应波束形成技术相结合，具有提升自适应波束形成算法高效性的前景。2004 年 Suksmono 等人引入了多层感知机来替换传统的 LMS 算法的单层模型[1]，在收敛速度上有所提

收稿日期：2019-12-17

升，但是依然采用的是迭代的方法，并没有充分利用神经网络的非线性拟合能力。2015 年张宝军等人研究了利用径向基神经网络进行波束形成的方法[2]，但是这种方法的神经网络训练过程较为复杂，需要进行额外的 k-means 聚类等操作，并且为了不使用更深层的神经网络而引入了过多的人工先验假设。2018 年冯晓宇等人提出了在低快拍情况下利用径向基神经网络进行波束形成的方法[3]，这种方法仅仅是在低快拍情况下的改进，并没有对以上径向基神经网络存在的问题进行解决。对于模型压缩而言，在 2013 年，Denil 等人提出了“在很多的深度神经网络中存在着显著的冗余，仅仅使用很少一部分（5%）权值就足以预测剩余的权值”的观点[4]。根据上述观点，2015 年 Hinton 等人提出了知识蒸馏的概念[5]，通过引入“教师-学生”网络使模型参数大为减少，模型速度得到提升。因此本文将深度神经网络模型应用于波束形成算法中，并对模型进行压缩优化，使波束形成算法相较于传统的算法有更快的速度，在大数据情况下具有更好的性能。

本论文根据波束形成原理设计深度神经网络模型，并对模型进行压缩，建立“教师-学生”网络，“教师”网络模型 TNNBF（Teacher Neural Network Beamforming）使用了 Leaky-ReLU 激活函数，解决了模型训练过程中的梯度消失以及神经元提前失活的问题。运用 Adam 优化器提高模型训练的全局收敛性，加快了算法的速度，并结合 Dropout 正则化方法提升过参数化网络的泛化性能，之后根据原始数据和泛化数据联合训练了“教师-学生”网络，提出了经过模型压缩后的深度神经网络波束形成算法 SNNBF（Student Neural Network Beamforming）。

1　知识蒸馏

现阶段，深度神经网络在信号处理、语音识别和计算机视觉等领域都取得了非常好的表现。复杂的模型固然具有更好的性能，但是高额的存储空间和计算资源消耗是其难以有效地应用在各硬件平台上的重要原因。为了解决这些问题，许多业界学者研究模型压缩方法来最大限度地减小模型对于计算空间和时间的消耗[6]。

在使用神经网络训练大规模数据集时，为了处理复杂的数据分布：一种做法是建立复杂的神经网络模型，例如含有上百层的残差网络，这种复杂的网络往往含有多达几百万个参数；另一种做法往往会混合多种模型，将几个大规模的神经网络在同一个数据集上训练好，然后综合多个模型，得到最终的分类结果。但是这种复杂模型，一是在新的场景下重新训练成本过高，二是由于模型过于庞大而难以大规模部署。所以，最基本的想法就是将大模型学习出来的知识作为先验，将先验知识传递到小规模的神经网络中，之后在实际应用中部署小规模的神经网络。

基于上述思想，为了最大程度地减小模型复杂度，减少模型存储需要的空间，同时也致力于加速模型的训练和推测，2015 年 Hinton 等人提出了知识蒸馏的概念。所谓蒸馏就是将复杂网络中的有用信息提取出来迁移到一个更小的网络上，这样学习出来的小网络可以具备和大的复杂网络相接近的性能效果，并且也大大地节省了计算资源。这个复杂的网络可以看成是一个教师，而小的网络则可以看成是一个学生。对于“教师”网络的蒸馏过程，可以认为是通过温度系数 T，将复杂网络结构中的概率分布蒸馏出来，并用该概率分布来指导精简网络进行训练。整个通过温度系数 T 的蒸馏过程由如下公式实现[5]：

$$T\_\mathrm{Prob}=\frac{\exp(\frac{z_i}{T})}{\sum_j \exp(\frac{z_j}{T})} \tag{1}$$

损失函数的 loss 值 L 为

$$\boldsymbol{L}=\alpha \boldsymbol{L}^{(\mathrm{soft})}+(1-\alpha)\boldsymbol{L}^{(\mathrm{hard})} \tag{2}$$

算法的具体过程可以简单概述为：

①首先用较大的 T 来训练模型，这时候复杂的神经网络能产生更均匀分布的软目标。

②之后小规模的神经网络用相同的 T 值来学习由大规模神经网络产生的软目标，接近这个软目标从而学习到数据的结构分布特征。

③最后在实际应用中，将 T 值恢复到 1，对数据进行测试。

从算法的具体过程中可以得到，数据本身是其结构信息和数值的一种混合物，结构关联信息通过概率分布被蒸馏分离出来。T 值很大时，相当于用很高的温度将关键的分布信息从原有的数据中分离出来，之后在同样的温度下用新模型融合蒸馏出来的数据分布，最后恢复温度，让两者充分融合起来。知识蒸馏这种模型压缩方法本质上相当于对数据进行了增强，加入了类别之间关联性的先验信息。将大规模网络学习到的这种关系包装到数据中，用这种更强的数据来训练小规模的模型，充分考虑到了类间的距离和类内的方差信息，从而提升了小规模模型的性能，达到了蒸馏的效果。与直接使用预训练模型的结构和权重相比，这是一种相对更高级的知识迁移方式。此外，Hinton 提出的知识蒸馏方法是针对分类问题的，本文将知识蒸馏的思路应用于回归问题的深度神经网络模型中，使设计完成的深度神经网络波束形成算法具有更好的性能，即算法有更快的计算速度并且其占用更少的计算资源。

2 应用知识蒸馏的 NNBF 算法

2.1 问题描述

深度神经网络[7]是一种能够构建复杂非线性关系的模型，在通过一定数量的样本训练之后，它也可以推断未知数据之间的未知关系，拥有较强的泛化性能。波束形成技术是一种通过回波信息和约束关系来合成波束的一种技术，传统的波束形成算法运算量大，运算时间长，占用资源多，在接收到大量回波数据时无法快速地进行实时处理。因此利用深度神经网络对传统的波束形成技术进行改进，之后再对深度神经网络进行压缩，去除模型中的冗余，可以使波束形成的时间缩短，还可以根据所得的回波数据不断更新网络模型，使训练出来的神经网络可以更好地应对各种情况，具有良好的稳健性。

根据波束形成的原理，抽象深度神经网络模型。深度神经网络的模型输入为回波信号 $\boldsymbol{X}$, n 为接收信源的数目，batch 为训练样本的数量。此外，本文将 SNNBF 算法和传统的波束形成算法——LMS 算法进行比较对照。不失一般性，以一维线阵为例，天线阵元数为 16，模型的输出为形成的波束信号 $\boldsymbol{Y}$，自适应波束形成权重因子为 $\boldsymbol{Z}$，$\boldsymbol{W}$ 为模型的权重，其中 $\boldsymbol{X}\in \mathrm{R}^{\mathrm{batch}\times n}$，$\boldsymbol{Y}\in \mathrm{R}^{\mathrm{batch}\times 1}$，$\boldsymbol{W}\in \mathrm{R}^{n\times 1}$。

2.2 “教师”网络模型

根据波束形成的基本原理，建立深度神经网络模型。首先将相同的期望信号方向、干扰信号方向的数据进行分组，每一组训练样本先采用 LMS 算法获得期望权重因子 $\boldsymbol{Z}$ 向量，然后将 $\boldsymbol{Z}$ 作为新的训练样本目标，训练框架可用下述公式进行表示：

$$\hat{\boldsymbol{Y}} = \boldsymbol{Z}\boldsymbol{X} \tag{3}$$

$$\boldsymbol{h}_n = \mathrm{dropout}\left[\sigma(\boldsymbol{w}\boldsymbol{h}_{n-1} + \boldsymbol{b}_{n-1})\right] \tag{4}$$

$$\hat{\boldsymbol{Z}} = \sigma(\boldsymbol{h}_{n-1}\boldsymbol{w}_n + \boldsymbol{b}_n) \tag{5}$$

其中，公式（3）表示 LMS 算法的原理，公式（4）和公式（5）表示神经网络的原理，$\hat{\boldsymbol{Z}}$ 表示神经网络最后求得的阵元权重。为了使神经网络有足够强的非线性拟合能力以及更好的泛化性，“教师”神经网络由七层全连接隐藏层组成，每一层的参数矩阵分别为 $[16*512]$,$[512*512]$,$[512*384]$, $[384*256]$, $[256*128]$,$[128*128]$,$[128*64]$；最后一层线性变换层参数为 $[64*16]$。隐藏层的每一层都使用 Leaky-ReLU 作为非线性激活函数，并使用 Dropout 方法提高泛化性能。

对于模型中的激活函数而言，本文采用 Leaky-ReLU 激活函数作为隐藏层的输出[8]。这种激活函数在神经元抑制区域依然拥有非零的梯度值，使得隐藏层的神经元在训练过程中不会大量死亡，可以让更多的神经元得到充分训练。对于模型中的优化算法而言，由于深度神经网络是一个非凸优化问题，拥有很多的局部极值点以及鞍点，普通的梯度下降算法很容易让模型陷入局部极值，所以应该采用带动量的一阶优化算法，使算法能够跳出局部极值以及鞍点，得到更优质的解。SGD 算法是一种

固定学习率的经典算法[9]，而 Momentum 方法是一种通过添加动量[10]、提高收敛速度的算法，Adagrad 算法让不同的参数拥有不同的学习率[11]，并且通过引入梯度的平方和来作为衰减项，而在训练过程中自动降低学习率。AdaDelta 算法[12]则对 Adagrad 算法进行改进，让模型在训练后期也能够有较为合适的学习率。Adam 方法就是根据上述思想而提出的[13]，对于每个参数，其不仅仅有自己的学习率，还有自己的 Momentum 量，这样，在训练的过程中，每个参数的更新都更加具有独立性[14]。它的自适应学习率调节功能可以使神经网络训练梯度在下降初期更加迅速，在后期更加稳健，并且不会提前停止；对于收敛性而言，Adam 优化算法的动量部分能够使模型收敛到相较于普通梯度下降算法更优的局部最优解上，提高了模型的性能。本文使用 Adam 算法作为网络模型的优化函数。

在训练过程中发现设计的深度神经网络相较于训练样本而言是过参数化的，很容易过拟合。为了降低深度神经网络过拟合的风险，本文采用了 Dropout 方法来进行深度神经网络的正则化。Dropout 算法是一种神经网络的正则化方法[15]，其功能是防止神经网络的过拟合。基于上述各个流程的操作，深度神经网络模型的总体原理如图 1 所示。

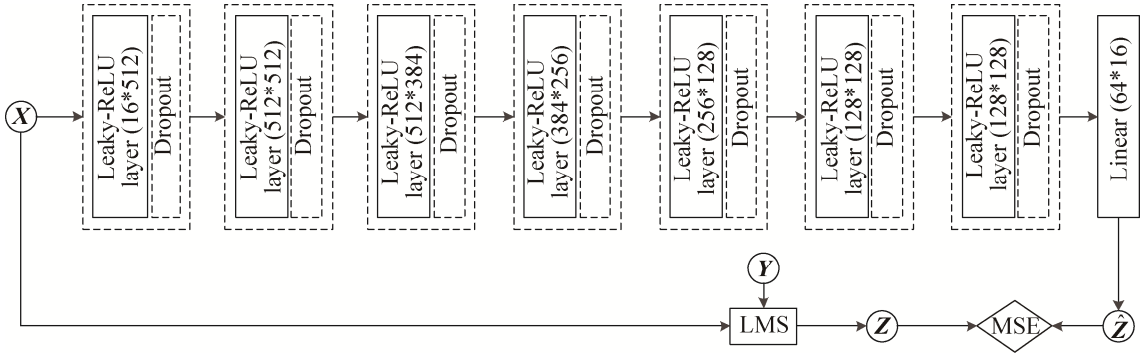

图 1　“教师”网络模型总体原理

Fig. 1　Overall schematic of the teacher model

2.3　“学生”网络模型

“学生”网络的模型结构为简化版本的“教师”网络，采用四层隐藏层的神经网络，每一层隐藏层采用更少的参数矩阵，并且去除 Dropout 模块。其具体结构如下：每一层隐藏层的参数矩阵为：$[16*384],[384*256],[256*128],[128*128]$；最后一层线性变换层参数为 $[128*16]$，隐藏层的非线性激活函数为 Leaky-ReLU 函数。

之后对“教师-学生”网络进行联合训练，为了让“学生”网络在原有的有限数据集中学习到“教师”网络同样的泛化性能，数据集的构造是非常重要的环节。本文采用和分类问题的知识蒸馏相同的思路，进行数据构造。数据集由两个部分组合而成：原始数据样本和“教师”泛化样本。原始数据样本由传统的 LMS 算法计算得到，也就是训练“教师”网络所用到的样本；“教师”泛化样本的结构和原始数据样本相同，其中回波信号部分 $\boldsymbol{X}'$ 的生成公式为

$$\boldsymbol{X}' = \boldsymbol{X} + \boldsymbol{N}\ , \qquad \boldsymbol{N} \sim \mathrm{N}(0,\boldsymbol{\alpha}) \tag{6}$$

其中，$\boldsymbol{N}$ 为服从于 0 为均值、$\boldsymbol{\alpha}$ 为方差的高斯分布，$\boldsymbol{\alpha}$ 可看作回归问题知识蒸馏中的温度参数，$\boldsymbol{\alpha}$ 越大说明模型越倾向于训练原始训练样本周围更大范围的泛化样本；$\boldsymbol{\alpha}$ 趋近 0 时泛化样本退化为原始的训练样本。将 $\boldsymbol{X}'$ 输入“教师”网络得到泛化样本的权重向量 $\boldsymbol{Z}'$。将泛化样本和原始样本组成一个训练批次样本，即为“教师-学生”网络模型的训练样本。

之后对模型中的损失函数进行设计，损失函数设计为“教师”网络和“学生”网络的均方误差值，然后将“教师”网络中的所有参数都固定，不进行梯度更新；并将数据同时输入“教师”网络和“学生”网络，并使用 Adam 优化算法进行模型优化。按照上述流程“教师-学生”网络的整体训练架构如图 2 所示。

训练结束之后，“学生”网络中的隐藏层和最后的线性变换层中的参数可以提取出来作为蒸馏之后的模型，理论上蒸馏后的“学生”网络模型能够拥有和原来的“教师”网络模型同等的泛化误差，并且大大降低了计算开销。因此基于模型压缩后的“学生”深度神经网络波束形成算法相比于未经压缩的深度神经网络波束形成算法有更好的性能。

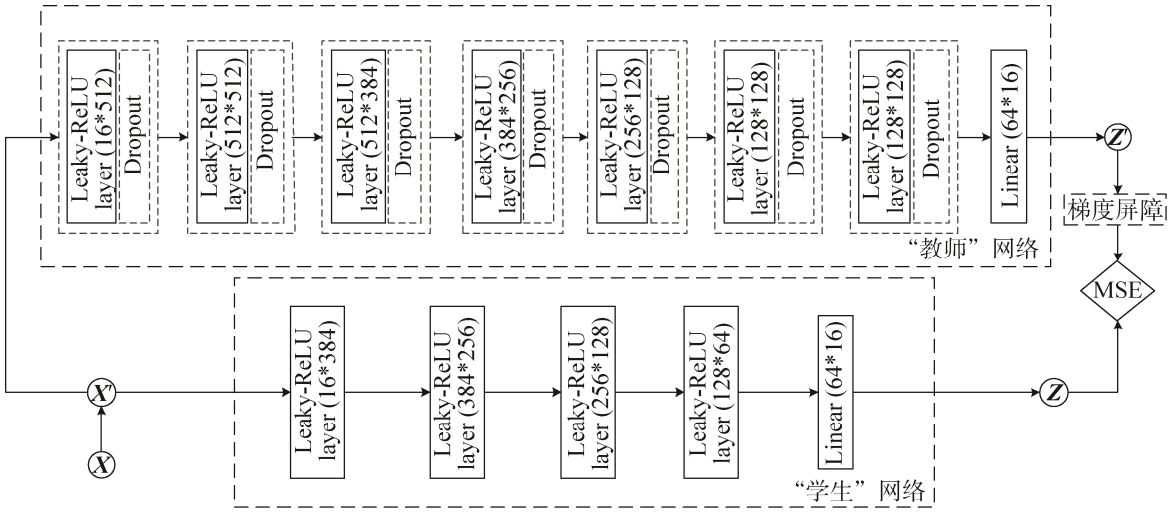

图 2　模型的总体原理

Fig. 2　Overall schematic of the model

3　仿真与分析

3.1　仿真条件

以一维线阵为例进行仿真。其中天线阵元数为 16，阵元间距为半波长，信噪比为 10dB，干噪比为 30dB，取 6 个不同目标方向和干扰方向的 6000 个训练样本和 60 个测试样本，其来波方向分别为 0°、10°、20°、30°、40°、50°方向，对应干扰方向分别为–50°、–40°、–30°、–20°、–10°、0°方向。此外，在下述仿真中，验证 SNNBF 算法的可行性以及意义时，均采用传统的 LMS 算法作为基准。下述所有仿真图均基于此条件进行仿真分析。

图 3 期望信号方向为 0°，干扰信号方向为–50°，分别采用 LMS 算法、TNNBF 算法和 SNNBF 算法得到的天线方向图，可以看到 LMS 算法、TNNBF 算法和 SNNBF 算法都可以在期望信号方向进行很好的波束形成，并且在干扰信号方向都可以进行很好的抑制，因此通过图 3 可知 SNNBF 算法有良好的波束形成性能。

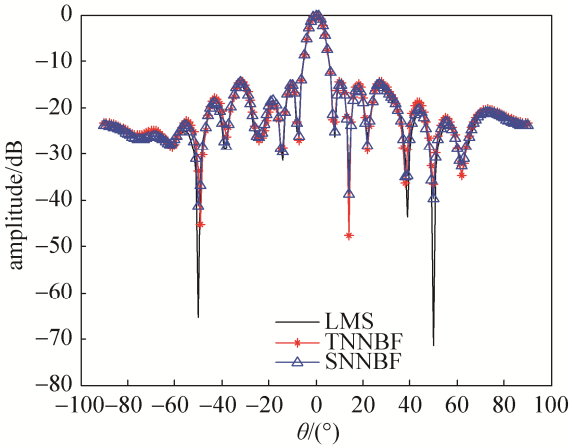

图 3　LMS 算法、TNNBF 算法和 SNNBF 算法天线方向图

Fig. 3 Antenna pattern of LMS algorithm, TNNBF algorithm and SNNBF algorithm

3.2　SNNBF 算法和 TNNBF 算法性能对比

“教师”网络是过参数化的，理论上“学生”网络能够以更少的参数规模达到类似于“教师”网络的泛化性能。选择 60 组不同信号源和干扰源的样本进行测试，统计最终合成的信号的均方误差，对两种算法在不同迭代步长的情况下损失值的大小进行实验，图 4 为 SNNBF 算法和 TNNBF 算法的性能对比图。从图中可以看出，“学生”网络在四分之一“教师”网络的参数规模下提供了和“教师”网络类似的波束形成性能，经过试验，“学生”网络在测试集上的均方误差为 1.429371785，“教师”网络在测试集的均方误差为 1.291752884，均方误差差距在 10%以内。

为了进一步验证知识蒸馏在深度神经网络波束形成问题上的意义，本文重新训练了一个参数规模和“学生”网络一样的小网络，测试结果如图 5 所示。可以看出，直接训练的小网络由于没有“教师”网络提供的泛化训练样本，所以在个别测试样本中的误差明显高于“学生”网络，测试数据集中小网络的平均均方误差是“学生”网络的 1.57 倍。因此“教师-学生”网络训练模式在波束形成的模型压缩问题上是有效的。

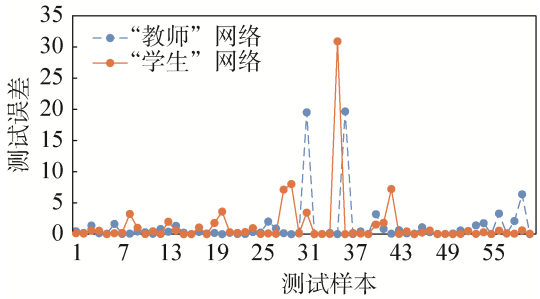

图 4　SNNBF 算法和 TNNBF 算法性能对比

Fig. 4　Performance comparison diagram between SNNBF algorithm and TNNBF algorithm

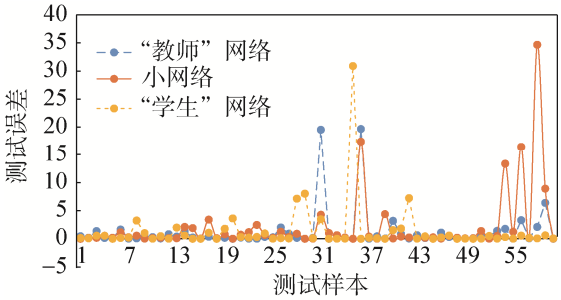

图 5　相同规模的“学生”模型和小模型性能对比

Fig. 5　Performance comparison chart of the student model and the small model of the same scale

本文也测试了多种“学生”网络规模下泛化误差的变化情况。使用的“学生”网络规模分别为：“学生”网络：[16∗384],[384∗256],[256∗128],[128∗128],[128∗16]；“学生”网络 2：[16∗256] [256∗128] [128∗128] [128∗64] [64∗16]；“学生”网络 3：[16∗256] [256∗128] [128∗128] [128∗64] [64∗16]。从图 6 和图 7 中可以看出，随着“学生”网络参数规模的降低，测试误差也相应上升，但是结果优于同样参数规模的直接训练的神经网络。因此，在实际工程环境下可以根据需求调整“学生”网络的规模，在速度和效果上进行较好的权衡。

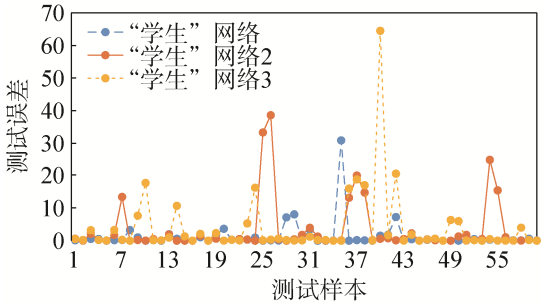

图 6　不同规模“学生”网络模型性能

Fig. 6　Performance of the student network model of different sizes

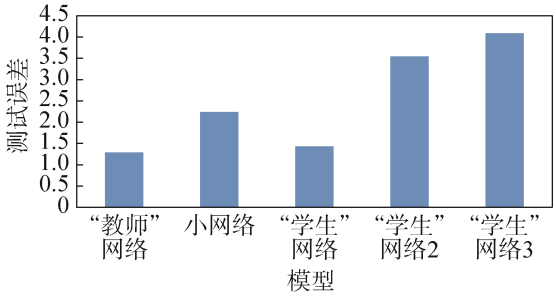

图 7　不同规模网络模型算法性能

Fig. 7　Performance model of network model algorithms of different scales

使用知识蒸馏压缩方法可以在波束形成精度几乎无损失的情况下，大幅度降低计算代价，加快运算速度，并且拥有远高于直接训练的小网络的精度。同时，根据实际使用场景对于精度的需求，使用知识蒸馏框架可以方便地从一个大网络中蒸馏出不同精度的小网络，让模型在精度和运算效率之间做出权衡。

4　总结

自适应波束形成技术是一种良好的空域抗干扰技术，其广泛应用于航天导航、飞行器测控、地面通信和新体制雷达等领域。本文以 LMS 算法为根本，利用深度学习模型对 LMS 算法进行改进，并利用知识蒸馏的方式对模型进行压缩，使算法在大数据的情况下能够快速稳健地进行波束形成。

SNNBF 算法采用深度学习的相关技术，设计了自适应波束形成的权重推断网络，利用知识蒸馏的原理建立了“教师-学生”神经网络模型。其中，“教师”网络使用 Adam 优化器增强深度神经网络训练的全局收敛性，然后用 Leaky-ReLU 激活函数解决深度神经网络的梯度消失问题，并利用 Dropout 方法抑制波束形成深度神经网络的过拟合现象，使自适应波束形成的权重推断网络在准确性和泛化性上均有较好的性能。这种神经网络模型存在冗余，因此利用“知识蒸馏”的方式对模型进行压缩，生成“学生”网络，这一网络既包含“教师”网络的精确性，又具有更快的计算速度，在同样的计算资源下，TNNBF 算法将 LMS 算法收敛速度提高了约 7 倍，SNNBF 算法将 LMS 算法收敛速度提高了约 20 倍，并且在未来随着训练数据的增加，权重推断网络的泛化性能以及准确性能够继续提高，具有较大的理论和工程的应用价值。

参考文献

[1] BAYU S A, HIROSE A. Intelligent beamforming by using a complex-valued neural network[J]. Journal of Intelligent and Fuzzy Systems, 2004, 15(3-4): 139–147.

[2] 张宝军, 卢梦怡, 陈治清, 等. 基于径向基函数神经网络的波束形成算法[J]. 西安邮电大学学报, 2015, 20(6): 33–36.
ZHANG Baojun, LU Mengyi, CHEN Zhiqing, et al. Beamforming algorithm based on RBF neural network[J]. Journal of Xi'an University of Posts and Telecommunications, 2015, 20(6): 33–36.

[3] 冯晓宇, 谢军伟, 张晶, 等. 低快拍下模糊径向基神经网络波束形成算法[J]. 火力与指挥控制, 2018, 43(4): 132–135,140.
FENG Xiaoyu, XIE Junwei, ZHANG Jing, et al. Beamforming algorithm based on fuzzy RBF neural network in the situation of limited snapshots[J]. Fire Control & Command Control, 2018, 43(4): 132–135,140.

[4] MISHA D, BABAK S, LAURENT D, et al. Predicting parameters in deep learning. Advances in Neural Information Processing Systems[C]. 2013: 2148–2156.

[5] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. Computer Science, 2015, 14(7): 38–39.

[6] 叶远征. 基于卷积神经网络的目标检测算法研究与应用[D]. 绵阳: 西南科技大学, 2019.
YE Yuanzheng. Research and application of target detection algorithm based on convolutional neural network[D]. Mianyang: Southwest University of Science and Technology, 2019.

[7] HINTON G, GEOFFREY E, SIMON O, YEE WHYE T. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006: 1527–1554.

[8] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models[C]//In Proc. ICML, 2013, 30(1): 3.

[9] PARAS. Stochastic gradient descent[J]. Optimization, 2014.

[10] PHANSALKAR V V, SASTRY P S. Analysis of the back-propagation algorithm with momentum[J]. IEEE Transactions on Neural Networks, 1994, 5(3): 505–506.

[11] WILSON A C, ROELOFS R, STERN M, et al. The marginal value of adaptive gradient methods in machine learning[J]. 2017.

[12] ZEILER M D. ADADELTA: an adaptive learning rate method[J]. Computer Science, 2012.

[13] KINGMA D, BA J. Adam: a method for stochastic optimization[J]. Computer Science, 2014.

[14] 史浩强. 陀螺仪若干典型故障智能诊断与预测技术[D]. 西安: 西安理工大学, 2019.
SHI Haoqiang. Intelligent diagnosis and prediction technology for some typical faults of gyroscopes[D]. Xi'an:Xi'an University of Technology, 2019.

[15] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The journal of machine learning research, 2014, 15(1): 1929–1958.

[作者简介]

柏沫羽 1993 年生，在读硕士，研究方向为雷达信号处理。

刘 昊 1976 年生，博士，研究员，研究方向为相控阵天线与微波技术。

陈浩川 1979 年生，研究员，研究方向为雷达总体设计。

张振华 1977 年生，研究员，研究方向为雷达系统与信号处理。