DOI:10.19652/j.cnki.femt.2204508

基于深度神经网络和随机森林集成模型的ADS-B辐射源个体识别*

王锦卫[1,2] 杜奕航[2] 张 江[2] 乔晓强[2]
(1.南京信息工程大学电子与信息工程学院 南京 210044;2.国防科技大学第六十三研究所 南京 210007)

摘 要:针对辐射源个体识别中单一神经网络随着辐射源个体数量的增加,模型的识别准确率显著下降的问题,提出一种基于深度神经网络和随机森林集成模型的广播式自动相关监视(ADS-B)辐射源个体识别方法。该方法利用多种深度神经网络模型和随机森林对增强数据集进行训练,然后利用集成学习方法中的硬投票方法对各网络模型和随机森林识别得到的结果进行集成表决,使得识别结果更具有说服力,同时在在辐射源个体数量增加的情况下依旧保持较高的识别率。实验结果表明,在融合了DRSN、VGG、ResNet、GoogleNet、DenseNet 5类神经网络和随机森林后,相比于单一的神经网络,识别准确率能够提升了3%~20%,且在辐射源个体数量增加的情况下依然能保持较高的识别准确率。

关键词:辐射源个体识别;数据增强;神经网络;随机森林;集成学习
中图分类号:TP183;TN92 **文献标识码**:A **国家标准学科分类代码**:510.4010

Individual recognition of ADS-B emitter based on deep neural network and random forest ensemble model

Wang Jinwei[1,2] Du Yihang[2] Zhang Jiang[2] Qiao Xiaoqiang[2]
(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. The 63rd Research Institute, National University of Defense Technology, Nanjing 210007, China)

Abstract: Aiming at the problem that the recognition accuracy of single neural network model decreases significantly with the increase of the number of emitters in individual emitter recognition, this paper proposes an ADS-B emitter individual recognition method based on deep neural network and random forest ensemble model. This method uses a variety of deep neural network models and random forest to train the enhanced data set, and then uses votingclassifier to ensemble the results obtained by each network model and random forest recognition, which makes the recognition results more convincing. The experimental results show that, after intagrating the models of DRSN, VGG, ResNet, GoogleNet, DenseNet neural networks and random forest, the recognition accuracy of ensemble model can be improved by 3%~20% compared with the single neural network, and the recognition accuracy can still be maintained with the increase of the number of individual radiator.

Keywords: individual identification of radiation sources; data enhancement; neural network; random forest; ensemble learning

0 引 言

广播式自动相关监视系统(automatic dependent surveillance-broadcast system, ADS-B)是集成无线数据通信技术和卫星导航定位技术于一体的自动化监视系统,广泛应用于各种飞行设备。近年来随着空中交通的快速发展,飞行设备与日俱增,据调查数据显示,中国民航仅2022年上半年的旅客运输量就高达1.18亿次,航空货运输量高达307.7×10^4 t。准确识别ADS-B信号辐射源个体,实现飞行动态监视,可以有效避免空域安全事故的发生,进而为相关

收稿日期:2022-11-15
* **基金项目**:国家自然科学基金(61801496,61801497)、军委科技委基础加强计划领域基金(2019-JCJQ-JJ-221)项目资助

人员研判不明目标的属性、分析定位目标、还原空情态势提供有价值的判定依据，然而随着空域流量的激增，且信道环境日趋复杂，ADS-B辐射源个体识别的稳定性显著降低，该问题已成为该领域急需攻克的难题之一。

目前辐射源个体识别技术飞速发展，涌现了一大批较好的辐射源个体识别新方法。陈小惠等[1]利用改进后的极限学习机(extreme learning machine，ELM)来训练辐射源信号的复杂度特征，能够较为准确地识别信号类别，然而随着辐射源个体数量的增加识别效果不佳。随着卷积神经网络(convolutional neural network，CNN)[2]的兴起和成熟，通常将CNN网络与星座图特征[3]相结合，利用神经网络对星座图进行特征提取，在识别效果上得到了较大的提升。文献[4]使用深度残差网络[5](residual network，ResNet)对I、Q路数据进行综合特征提取，以此提高数据的利用率，得到了较高的准确率。唐震等[6]在深度残差网络模型中引入注意力机制[7]和软阈值函数[8]，改进后的模型(deep residual shrinkage networks，DRSN)具有较高的识别准确率和抗噪声性能。文献[9]提出了一种基于矢量图的SEI方法，利用神经网络提取矢量图中的特征进行识别，取得了较好的识别效果。文献[10]直接输入IQ数据进行端到端的个体识别，虽然有很好的识别效果但是由于所提取特征的不确定性，模型的泛化性能较差。文献[11]是提出一种基于模糊函数等高线与栈式降噪自编码器的识别方法，在保留信号能量特征的基础上能够达到一定的识别率，增加了神经网络识别分类在信号识别领域的可解释性。许敏等[12]提出一种基于密集连接结构与注意力机制的特定辐射源识别算法，在注意力机制的作用下对4类辐射源的识别准确率高达95.12%。使用以上方法虽然在一定程度上提高了模型的识别效果，但是在实际应用中其识别性能均随着辐射源个体数量的增加显著下降。因此提高模型的稳定鲁棒性，使之能够在不同的辐射源个体数量情况下依旧保持较高的准确率就变得至关重要。而集成学习(ensemble learning)能综合各个模型的优点，使得模型的鲁棒性更强，已被用于文本和图像识别领域，并取得较好的识别效果。例如，文献[13]提出了一种轻量化的增量式集成学习算法，实现了关于类别数量的线性计算复杂度，在降低训练成本的同时在一定信噪比范围内依然能保持相对较高的准确率。文献[14]提出了一种基于蚁群特征选择并行分类集成学习的孪生辐射源个体识别方法，由二分类扩展到多分类，提高了模型的适用性，获得了较好的识别效果。尽管以上方法都能够较好的识别辐射源个体，但是该方法提取特征过于复杂，提取的特征参数的阈值很容易随着数据或者环境的变化而改变，导致识别效果下降，并且需要手动调节。结合集成学习和深度学习的优势，本文提出了一种基于深度神经网络和随机森林的集成模型ADS-B辐射源个体识别方法，新构建的集成模型VRDGD-RF相较于单一的神经网络，识别准确率有了较为明显的提升，且在辐射源个体数量增加的情况下依旧能够保持较高的识别率。

1 ADS-B信号模型及数据预处理

1.1 ADS-B信号模型

本文数据采集基于1090ES数据链的ADS-B信号，工作在1 090 MHz的ADS-B信号由报头和数据位两部分组成，8 μs的报头共计有4个脉冲，脉冲的宽度是0.5 μs，4个脉冲的位置固定，分别间隔1.0、3.0、4.5 μs。数据位分别由5位的下行链路格式字段(DF)、3位的附加标识符段(CA)、24位的唯一标识符段(ICAO)、56位扩展的消息段(ME)以及24位的奇偶校验字段(PI)构成，如图1所示。

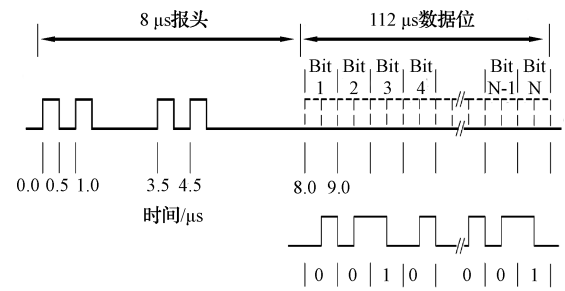


图1 ADS-B信号波形格式

数据位采用的是脉冲位置调制(pulse position modulation，PPM)，由于数据速率为1 Mbps，因此每条消息存在时间为120 μs。为了排除地址模式对识别结果的影响以及保护数据隐私，数据已根据信号的帧结构去除了唯一标识符(ICAO)对应的信号片段。

1.2 数据预处理

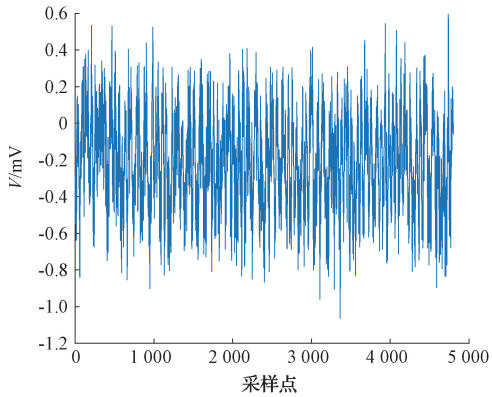
首先为了降低数据幅度对特征的影响，对样本数据进行归一化处理，每个信号样本维度为2×4 800的I/Q数据，得到两架飞机ADS-B信号的I、Q数据图如图2所示。

归一化之后，通过对I、Q数据的观察发现，I、Q路之间的关系存在着某些共性和差异性，且I、Q路的数据均为样本的特征数据，为提高数据的利用率，采用左右拼接的方式将I、Q路信号值整合在一起形成I/Q拼接图，得到两类不同ADS-B信号的I/Q数据拼接如图3所示。

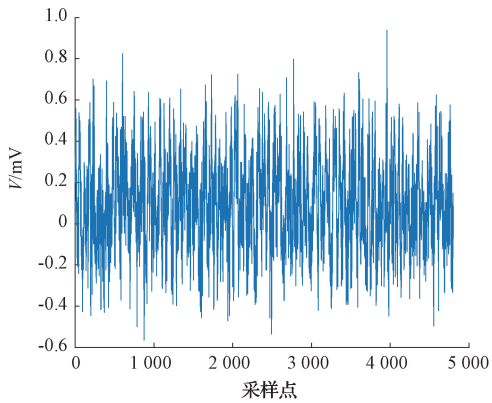
为了防止出现过拟合现象，应当在不产生失真的情况下对已存在的数据样本进行扩增，本文使用的样本数据扩增方法有亮度增强、旋转(90°和180°)、翻转(90°和180°)、灰暗和对比度增强相结合的方法。在使用以上方法进行扩增后，辐射源个体样本数据可以扩增到原来的12倍。以第1架飞机的ADS-B信号拼接后的I/Q样本数据为例，扩增后的数据结果如图4所示。

2 ADS-B信号识别模型

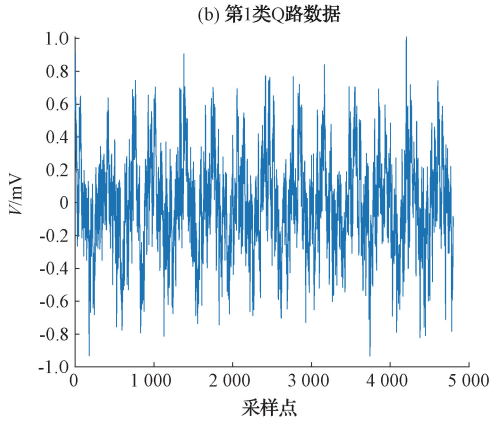
综合利用常见的卷积神经网络和随机森林对第1节所述ADS-B辐射源信号进行识别。整体识别流程如图5所示。集成模型识别流程主要分为4部分。



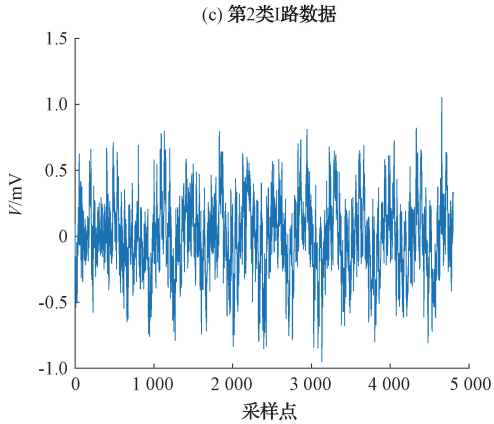
(a) 第1类I路数据



(b) 第1类Q路数据

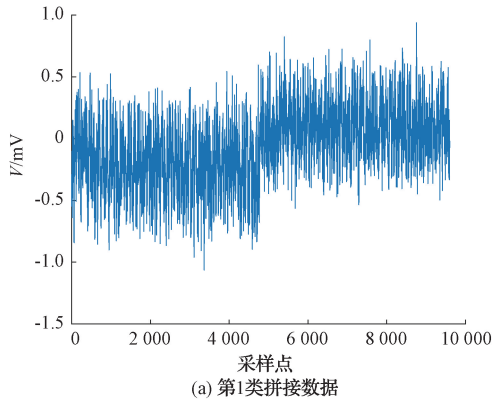


(c) 第2类I路数据

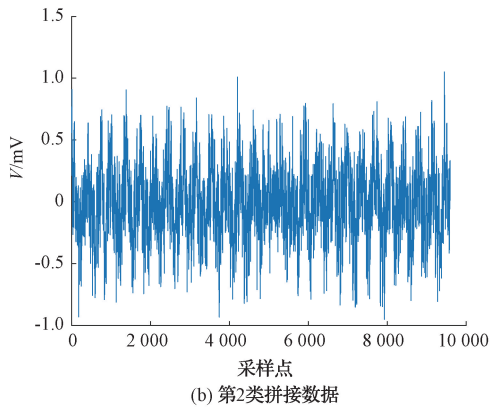


(d) 第2类Q路数据

图 2　两架飞机 ADS-B 的 IQ 信号数据



(a) 第1类拼接数据



(b) 第2类拼接数据

图 3　两类 ADS-B 信号的 I/Q 拼接图

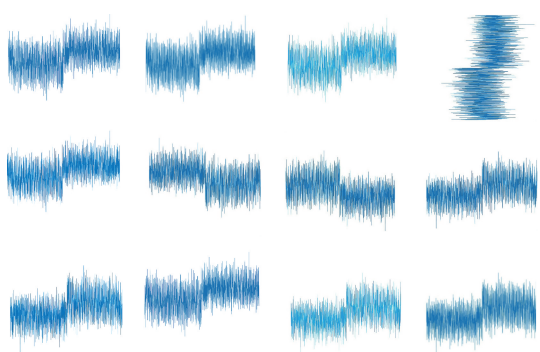


图 4　扩增后的数据集

1)数据预处理。主要通过将 ADS-B 信号的 I、Q 路数据以时间为横轴，归一化后的数值作为纵轴，将两路数据以左右拼接的形式拼接在一起，然后使用数据增强的方式将数据集扩增为原来的 12 倍。

2)数据集划分。将增强后的数据集按照 8∶1∶1划分为训练集、验证集和测试集。

3)集成模型构建以及数据集测试。即利用集成算法将神经网络和随机森林进行集成，构建集成模型进行训练。

4)模型性能评估和检测。即利用集成模型训练得到的参数对集成模型的性能进行评估。

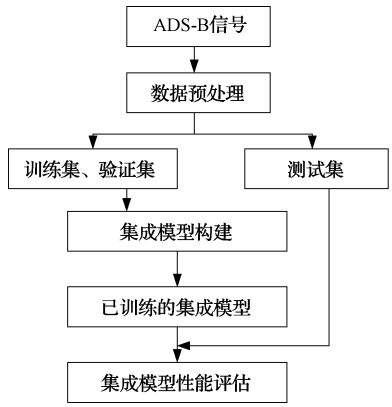


图 5　集成模型识别流程

2.1　集成模型构建

本文提出的集成模型 VRDGN-RF 的框架如图 6 所示，整体框架由 5 种神经网络和随机森林构成。本文在设计集成模型框架时尝试用多种不同的深度神经网络搭配来构建集成模型，综合考虑模型收敛情况和识别率性能，最终确定由 VGG19[15]、DRSN18[6]、ResNet34[5]、GoogleNet[16]、DenseNet121[17]来构建集成模型 VRDGD-RF，神经网络名称后面的数字代表该网络对应的层数。图 6 中 bootsraping 表示将样本多次重复抽样来建立新的样本，类似于重采样，可以提高模型的鲁棒性；block1、block2、block3、block4、block5 分别表示上述 5 种神经网络；block6 则由 5 个相同的随机森林(RF)构成。

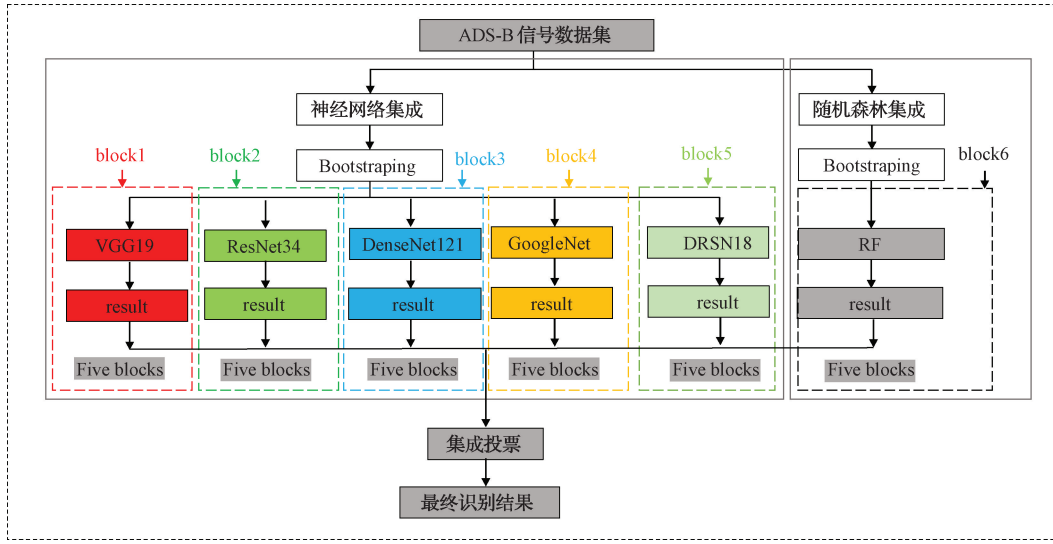


图 6　VRDGD-RF 集成模型框架图

通过以上的集成模型框架可以看出，首先将数据集使用 bootstrap 算法将数据进行重复抽样，形成新的数据样本，这样可以有效提高模型的鲁棒性，然后将分别送入神经网络模块和随机森林模块进行识别分类，每个 block 包含 5 个子神经网络，每个子神经网络都会输出一个预测结果，集成模型 VRDGD-RF 在这一阶段会获得 30(6×5)个预测结果，再将这些预测结果使用多数投票法进行投票，得到集成模型的最终识别结果。

2.2　神经网络和随机森林模块

集成学习主要是将多个子学习器进行组合优化本文的子学习器模块主要分为卷积神经网络和随机森林两个模块。其中神经网络模块中的 VGG、ResNet、DenseNet、GoogleNet 已较为熟知在此不做赘述，主要介绍 DRSN18 模块，其网络结构如图 7 所示。其中输入图片规格为 120×120，通道数为 3。Conv 为卷积层，BN 为归一化层，ReLU 为激活函数，K 为卷积核，S 为步长，Avgpool 为平均池化层，FC 为全连接层，RS-Block 为残差收缩模块，RS-block 主要是在 ResNet 残差模块的基础上添加了注

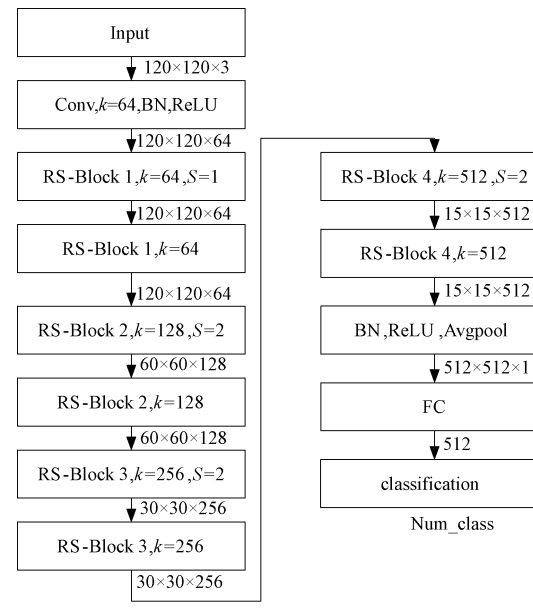


图 7　DRSN 网络结构

意力机制算法和软阈值函数，可以有效提高模型的抗噪声性能和鲁棒性。

另一模块为随机森林模块，主要作用是能够随机采样和随机特征提取，可以通过大量的基础树模型找到最稳定可靠的结果，以此提高模型的鲁棒性和泛化性能。

2.3 基于集成学习模型结构优化的方法

对集成学习模型的网络结构进行优化主要是对整个集成网络子学习器的组合优化，即使用集成进化算法在模型融合过程中选择最优的子学习器组合。本文将神经网络和随机森林作为子学习器，对6种子学习器的输出结果进行多数投票得到最终分类结果，其系统结构如图8所示。

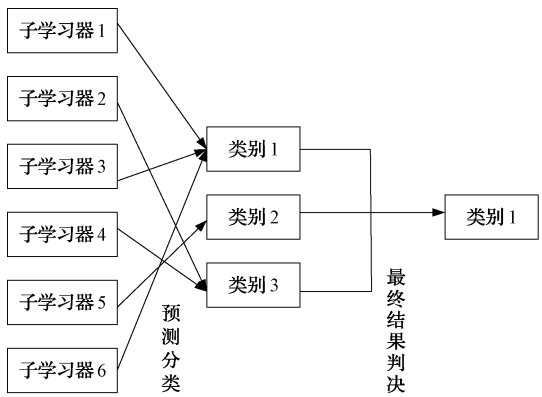


图8 硬投票集成原理

采用相对多数投票法(plurality voting)，其数学模型为：

$$H(x)=C_{\arg j}\max\sum_{i=1}^{T}h_i^j(x) \tag{1}$$

式中：$h_i^j(x)$表示输入的样本x在类别C_j上的输出，选取票数最多的识别结果作为最终的识别结果，对于出现相同票数的情形，则在其中随机选取一个作为最终识别结果。

3 实验结果与分析

本文使用的ADS-B数据集来自100架某民航客机的ADS-B辐射源信号，在训练集和测试集中每种类别的样本数量大致相同，每架飞机大约120个ADS-B数据样本，每个样本为数据维度2×4 800的I/Q数据。训练集、验证集和测试集样本的比例设置为8∶1∶1，且每种样本的数据在训练集、验证集和测试集中都大致相等。深度神经网络集成模型的训练和测试则使用Pytorch框架，运行系统为Windows11，GPU为NVIDIA RTX A6000，初始学习率设置为0.001，且每20个epoch进行一次更新，损失函数采用的是交叉熵损失函数(Cross Entropy Loss)，优化器选用Adam。

3.1 损失收敛分析

集成模型的训练损失和验证损失随迭代次数变化如图9所示，从图9可以看出，集成模型的训练损失在0.03处收敛，而集成模型的验证集损失则在0.19处收敛。由此可以看出集成模型在测试集上的表现较好，集成模型的收敛性能良好。

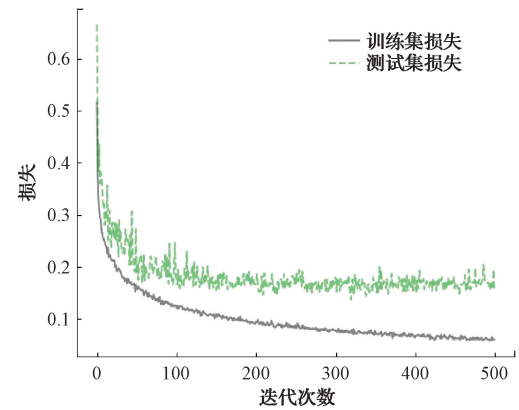


图9 集成模型的损失曲线

3.2 识别率对比及鲁棒性对比分析

为了验证所提模型的有效性，仿真了集成模型VRDGD-RF和5种不同神经网络模型对30类ADS-B信号进行分类识别性能，其结果如图10所示。从图10可以看出，集成模型的识别效果最佳，准确率高达94.52%。其次为DRSN18和ResNet34，两者准确率相差无几，最高准确率分别为91.21%和90.56%，稍差一点的VGG19则保持在80%，而DenseNet121的性能较集成模型弱将近18%，只有77%左右的准确率，效果最差的网络为GoogleNet，准确率只有不到70%。说明集成模型的性能相对于单一深度神经网络有较大提升。

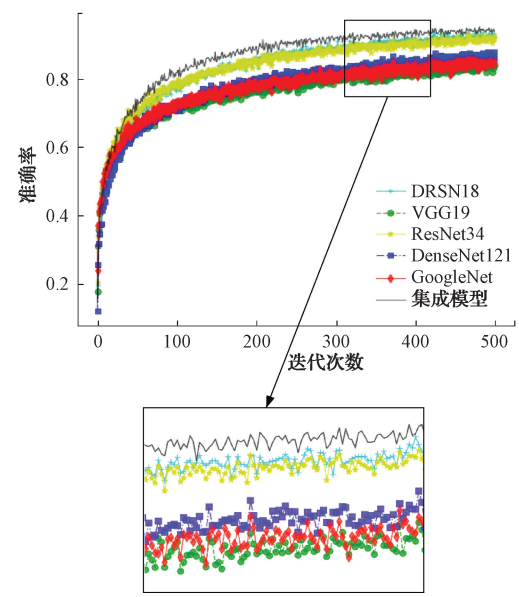


图10 5种不同网络模型及集成后模型的准确率

为进一步验证识别本文模型的鲁棒性，在不同信噪比(signal-to-noiseration，SNR)条件下进行对比实验。首先

将原始数据进行IQ拼接后，然后加入高斯白噪声，ADS-B信号数据集中噪声信噪比范围为－9～9 dB，间隔为3 dB。该方法下的30类的ADS-B信号，共21 000(100×30×7)个样本，同样按8∶1∶1的比例划分为训练集、验证集和测试集。将本文算法与改进型AlexNet、添加注意力机制后的ResNet34以及DRSN18网络进行对比实验，得到各模型在不同信噪比上的识别准确率曲线如图11所示。

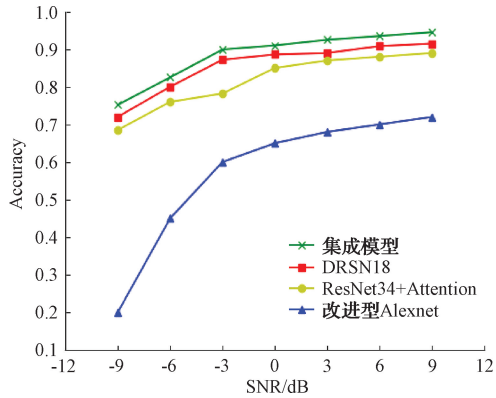


图11 不同模型在不同信噪比下的识别准确率

从图11可以看出，不同模型的识别准确率随着信噪比的增大而不断增加。本文提出的集成模型VRDGD-RF在不同的信噪比下的准确率均优于其他模型，证明了该模型的有效性和鲁棒性。由此可见将多种分类器使用集成算法结合，可以有效增加模型的分类识别率及鲁棒性。

3.3 识别率随辐射源个体数量影响分析

为了验证所提模型识别性能随辐射源个体数量的变化情况，仿真了不同数量辐射源个体情况下，集成模型VRDGD-RF的识别性能，其结果如图12所示。由图12可见，随着辐射源个体数量的增加，虽然集成模型的准确率呈下降趋势，但在辐射源个数增长至50时，集成模型的准确率仍能达到80%以上。由此可见，集成模型VRDGD-RF的泛化能力良好，在辐射源个体数量增加的情况下依旧能保持较高的识别率，且识别率的下降较为平缓。

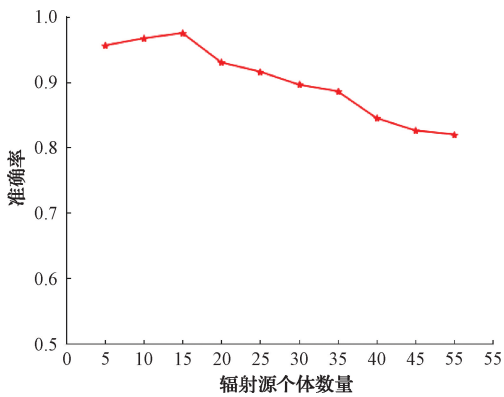


图12 随辐射源种类增加的准确率

3.4 综合指标对比分析

为了评估集成模型VRDGD-RF的综合性能，通过对准确率(Acc)、精确率P、召回率R和$F1$值多个评估指标对集成模型的性能进行衡量，同时与单一神经系统进行比较，各指标的计算公式如下：

$$Acc=\frac{TP+TN}{TP+FP+TN+FN}\times 100\% \tag{2}$$

$$P=\frac{TP}{TP+FP}\times 100\% \tag{3}$$

$$R=\frac{TP}{TP+FN}\times 100\% \tag{4}$$

$$F1=\frac{2PR}{P+R}\times 100\% \tag{5}$$

式中：TP、FP、TN、FN分别为真正类、假正类、真负类和假负类的样本数量。其中TP表示样本标签和预测结果都预测为正的样本；TN表示样本标签和预测结果都预测为负类的样本；FN表示样本标签为正而预测结果的样本；FP表示样本标签为负类而预测结果为正类的样本。

表1所示为采用集成模型上和单一模型对30架飞机的辐射源个体进行识别的性能参数。

表1 不同网络的性能对比 (%)

网络模型	准确率	精确率	召回率	$F1$值
DRSN18	91.21	93.53	90.21	91.84
VGG	76.43	78.11	81.31	79.68
ResNet34	90.56	90.74	89.74	90.24
GoogleNet	84.62	85.77	81.24	83.44
DenseNet121	89.77	90.42	88.93	89.67
集成模型	94.52	95.86	95.41	95.63

由表1可见，集成模型的各项性能指标较任一深度神经网络模型的性能指标均有所提升。其中集成模型VRDGD-RF的准确率较DRSN18、VGG、ResNet34、GoogleNet、DenseNet121分别提高了3.31%、18.09%、5.96%、15.86%和4.75%；其$F1$值较其他5种模型分别提高了3.79%、15.95%、5.39%、12.19%和5.96%，其中精确度和召回率都比单一的神经网络的效果更佳，证实在经过集成后的识别效果在准确率、精确率、召回率和$F1$值都较单一的神经网络有了更为明显的提升。

4 结 论

本文基于集成学习的思想，提出一种基于集成模型的辐射源个体识别方法，在充分利用ADS-B信号的I、Q路数据的情况下对原有的ADS-B信号进行数据扩增，然后利用多种深度神经网络模型和随机森林对增强数据集进行训练，然后采用硬投票的集成学习方法对各网络模型和随机森林识别得到的预测结果进行集成表决，得到最终识别结果。实验结果表明，本文所建立的集成模型VRDGD-RF相较单一的深度学习神经网络无论在识别准确率精确

率、召回率和 $F1$ 值都有明显的提升，同时在辐射源个体数量增加的情况下集成模型依旧能够保持在较高的识别率，且下降趋势较为平缓。本文的使用数据目前只停留在数据的时域图上，因此后续可以考虑到信号的其他变换形式例如频域、多域融合等方面进行一些可行性的探讨，将会是未来的研究内容。

参 考 文 献

[1] 陈小惠，彭杰，薛毓楠. 基于复杂度的通信辐射源目标识别方法[J]. 国外电子测量技术，2021，40(5)：22-26.

[2] TAN M X，LE Q V. EfficientNet：Rethinking model scaling for convolutional neural networks[C]. International Conference on Machine Learning. PMLR，2019：6105-6114.

[3] PENG S L，JIANG H Y，WANG H X，et al. Modulation classification based on signal constellation diagrams and deep learning.[J]. IEEE Transactions on Neural Networks and Learning Systems，2019，30(3)：718-727.

[4] 翁琳天然，彭进霖，何元，等. 基于深度残差网络的ADS-B信号辐射源个体识别[J]. 航空兵器，2021，28(4)：24-29.

[5] HE K M，ZHANG X Y，REN S Q，et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2016：770-778.

[6] 唐震，乔晓强，张涛，等. 基于深度残差收缩网络的辐射源个体识别方法[J]. 电子测量技术，2022，45(9)：168-174.

[7] HU J，SHEN L，ALBANIE S，et al. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：7132-7141.

[8] ZHAO M H，ZHONG S S，FU X Y，et al. Deep residual shrinkage networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics，2020，16(7)：4681-4690.

[9] 潘一苇，杨司韩，彭华，等. 基于矢量图的特定辐射源识别方法[J]. 电子与信息学报，2020，42(4)：941-949.

[10] LIN Y，TU Y，DOU Z. An improved neural network pruning technology for automatic modulation classification in edge devices[J]. IEEE Transactions on Vehicular Technology，2020，69(5)：5703-5706.

[11] 普运伟，郭江，刘涛涛，等. 基于模糊函数等高线与栈式降噪自编码器的雷达辐射源信号识别[J]. 仪器仪表学报，2021，42(1)：207-216.

[12] 许敏，李博涵，王凯，等. 特定辐射源个体识别算法研究[J]. 电子测量与仪器学报，2021，35(10)：116-123.

[13] 丁嘉辉. 集成学习在辐射源识别中的应用研究[D]. 西安：西安电子科技大学，2021.

[14] 徐雨芯，顾楚梅，曹建军，等. 基于蚁群特征选择并行分类集成学习的孪生辐射源个体识别[J]. 兵工学报，2022，43(12)：3132-3141.

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

[16] SZEGEDY C，VANHOUCKE V，IOFFE S，et al. Rethinking the inception architecture for computer vision[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2016：2818-2826.

[17] HUANG G，LIU Z，WEINBERGER K Q. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：4700-4708.

作 者 简 介

王锦卫，硕士研究生，主要研究方向为零样本学习、辐射源个体识别和调制信号识别。
E-mail：1694977599@qq.com