

文章编号: 2095-4980(2021)01-0054-06

## 基于轻量化深度学习网络的调制信号识别模型

张思成<sup>1</sup>, 林云<sup>1</sup>, 康健<sup>2</sup>, 涂涯<sup>1</sup>

(1.哈尔滨工程大学 信息与通信工程学院, 黑龙江 哈尔滨 150001; 2.北京宇航系统工程研究所, 北京 100076)

**摘要:** 电磁态势分析是信息化战争中至关重要的工作, 如何利用深度学习技术有效实现调制信号识别是其中一项关键技术。首先将调制信号转化为带有颜色信息的星座图形式, 并用深度学习方法, 选用 VGG16 和 AlexNet 两个卷积神经网络完成调制识别任务。结果显示, 当信噪比大于等于 0 dB 时, 可以达到 99% 以上的识别准确率。由于军用设备对于计算性能和存储性能把控较为严格, 因此采用全零矩阵平均百分比的方法对深度学习模型进行压缩。结果显示, 在不损失识别准确率的前提下, 信噪比为 0 dB 时, 对于模型参数量, AlexNet 可以压缩 3 466 倍, VGG16 可以压缩 20 156 倍; 对于浮点运算量, AlexNet 可以压缩 2 314 倍, VGG16 可以压缩 13 475 倍。表明本研究方法对调制信号识别具可行性以及高效性。

**关键词:** 调制信号识别; 深度学习; 卷积神经网络; 模型轻量化

**中图分类号:** TN911.3

**文献标志码:** A

**doi:** 10.11805/TKYDA2019293

## Modulation signal recognition model based on lightweight Deep Learning network

ZHANG Sicheng<sup>1</sup>, LIN Yun<sup>1</sup>, KANG Jian<sup>2</sup>, TU Ya<sup>1</sup>

(1.College of Information and Communication Engineering, Harbin Engineering University, Harbin Heilongjiang 150001, China;  
2.Beijing Institute of Astronautical Systems Engineering, Beijing 100076, China)

**Abstract:** Electromagnetic situational analysis is a crucial task in information warfare, and modulation signal recognition by using Deep Learning(DL) is one of the key technologies. In this paper, the modulation signals are firstly transformed into the form of constellation diagrams with color information, and two Convolutional Neural Networks(CNNs), VGG16 and AlexNet, are selected to complete the modulation signal recognition task by using DL. The results show that a recognition accuracy higher than 99% can be achieved when the Signal-to-Noise Ratio(SNR) of noise is greater than or equal to 0 dB. Since the computational performance and storage performance of military devices are more stringent in controlling, the Average Percentage of Zeroes(APoZ) method is adopted to compress the DL model. The results show that with 0 dB SNR, AlexNet can be compressed by 3 466 times and VGG16 can be compressed by 20 156 times for model parametric quantities, and by 2 314 times and 13 475 times for floating-point operations, respectively, without losing recognition accuracy. In summary, the proposed method is both feasible and efficient in modulation signal recognition.

**Keywords:** modulation signal recognition; Deep Learning; Convolutional Neural Networks; model lightweighting

信息化时代, 电磁空间成为信息化条件下的战略高地, 夺取和保持电磁空间优势, 是打赢信息化战争的必要条件。调制识别技术对于正确、精准描述当前战场状态, 掌握敌我双方部队的表现和属性等, 起着十分重要的作用<sup>[1]</sup>。对于调制识别, 已有大量的研究与发现, 目前的研究主要可以分为 2 大类: 基于似然比方法和基于特征识别方法。相比于基于似然比方法, 特征识别的方法是当前研究的主要方向。目前, 已有方法包括基于瞬时特征、基于小波变换、基于高阶累积量、基于循环平稳分析和基于星座图<sup>[2]</sup>。近年来, 深度学习(DL)方法逐渐用于各个

收稿日期: 2019-08-14; 修回日期: 2019-12-01

作者简介: 张思成(1996-), 男, 在读硕士研究生, 研究方向为深度学习在通信领域的应用。email:2015080325@hrbeu.edu.cn

研究领域,已成为调制识别研究的一大热点。文献[3]将大而密的信号时间序列输入到卷积神经网络(CNN)进行盲学习,实现了良好的分类效果。文献[4]将信号转换为具有网络拓扑的图像并输入到 CNN 和支持向量机(Support Vector Machine, SVM)中,测试表明 CNN 分类优于 SVM。文献[5]将信号转换为星座图、灰度图像和增强灰度图像,并使用 CNN 进行训练和测试。实验结果表明,该方法可以有效地对 8 种调制信号进行分类。但深度学习模型通常具有很大的规模,对设备性能的要求也很高,这极大地限制了其在军用设备上的应用。为使其用于资源受限的设备上,研究者们一直研究如何对模型规模进行压缩。文献[6]中,研究人员将 CNN 参数的数据位从 32 位缩短到 10 位,并保持了分类能力。文献[7]将 Dropout 方法应用于 CNN。此方法通过随机失活网络中的连接来减小模型的规模。此外,还有一些方法从模型的体系结构入手,设计更紧凑、更高效的模型组件,如 GoogLeNet 中的 Inception 和 SqueezeNet 中的 Fire 模块,但这种方法需要对 DL 理论知识有较为扎实的基础<sup>[8-9]</sup>。

基于以上方法和理论,本文主要研究深度学习解决调制信号识别问题,并在此基础上使用全零矩阵平均百分比(APoZ)方法对模型进行压缩。

## 1 模型

### 1.1 数据集与信号形式转换

使用 Matlab 仿真生成实验所需的数据集,选取的调制信号包括 4 进制振幅键控(Amplitude Shift Keying, ASK)、二进制相移键控(Binary Phase Shift Keying, BPSK)、正交相移键控(Quadrature-PSK, QPSK)、偏移四相相移键控(Offset-QPSK, OQPSK)、8 进制相移键控(8PSK)、16 进制正交振幅调制(Quadrature Amplitude Modulation, QAM)和 32 进制 QAM 共 8 种类型。噪声环境为:信噪比为 $-6 \sim 6$  dB,步进为 2 dB 的加性高斯白噪声。为得到适合于深度学习模型的数据形式,将信号转化为带有颜色信息的星座图形式。第一步绘制信号的星座图,然后根据星座图中每一个点的归一化点密度对该点进行上色<sup>[10]</sup>。

图 1 为求一个点的归一化点密度示意图。第  $i$  个信号采样点的归一化密度  $\rho(i)$  的数学表达式为:

$$\rho(i) = \frac{\sum_{j=1}^N f[h(i)-h(j) < r \& \& |v(i)-v(j)| < r]}{N} \quad (1)$$

式中:  $h(i)$  为获取第  $i$  个信号采样点的横坐标值;  $v(i)$  为获取第  $i$  个信号采样点的纵坐标值;  $r$  为计算归一化点密度正方形区域边长的一半;  $h(j)$  为获取第  $j$  个信号采样点的横坐标值;  $v(j)$  为获取第  $j$  个信号采样点的纵坐标值;  $N$  为点的总数。 $f(x)$  函数满足:

$$f(x) = \begin{cases} 1, & x = \text{True} \\ 0, & x = \text{False} \end{cases} \quad (2)$$

计算出每一个点的归一化密度之后,根据图 2 的色条为该点上色,即可形成具有 RGB 信息的图像。

为每一个数据样本生成 10 000 个采样点,并使用上述方法对数据进行处理。为满足实验要求,在每一个选定的信噪比下生成 8 种调制信号的训练集、交叉验证集和测试集,其中训练集中每种信号 9 000 张,交叉验证集中每种信号 1 000 张,测试集中每种信号 1 000 张。图 3 为信噪比为 6 dB 下的处理效果。

### 1.2 深度学习模型

本文选择 AlexNet 和 VGG16 两个深度学习网络完成调制信号分类任务。2 个模型均具有较强的分类性能,但是在模型规模上却有着较大差异,这也将有助于在实验对比过程中分析并发现一些问题。

1) AlexNet 网络:在 2012 年被 Alex 提出,并赢得了 2012 届 ILSVRC 大赛的冠军。top5 预测的错误率为 16.4%,性能远超第一名。AlexNet 网络具有 5 层卷积层和 3 层全连接层<sup>[11]</sup>。

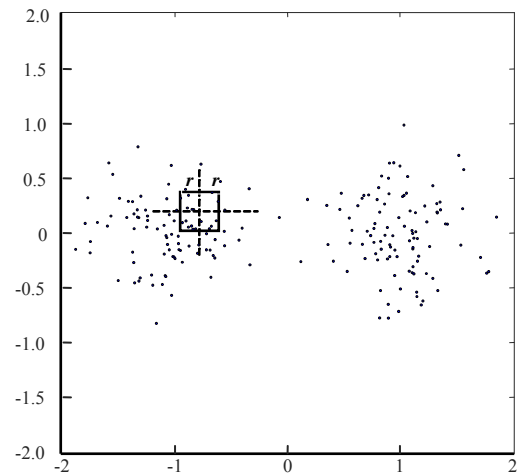


Fig.1 Normalized point density  
图 1 归一化点密度示意图

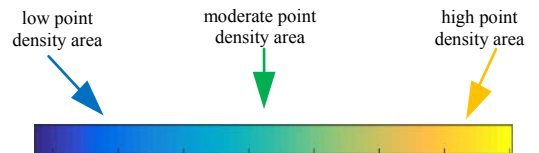


Fig.2 Color bar  
图 2 颜色条

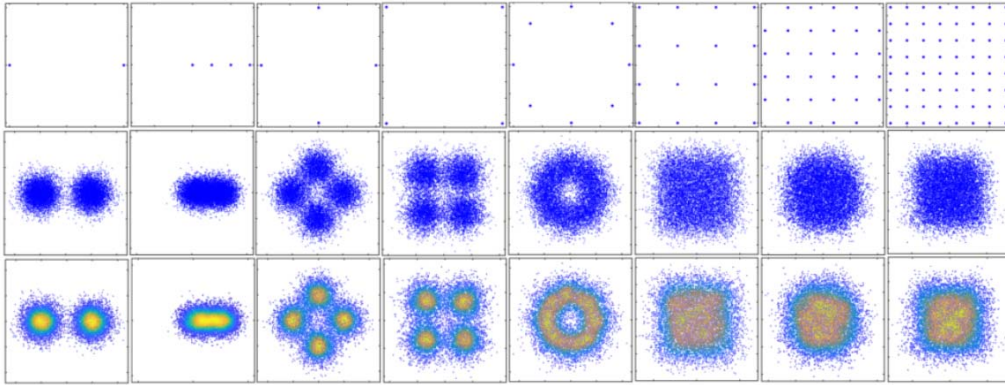


Fig.3 Effect of data processing at 6 dB SNR  
图 3 信噪比 6 dB 下的数据处理效果

2) VGG16: VGG 网络由牛津大学于 2014 年提出, 该模型在简单性和实用性方面尤为突出, VGG 在图像分类和目标检测任务方面具有非常好的性能。VGG 有许多衍生版本, 其中最受欢迎的是 VGG16。VGG16 具有 13 个卷积层和 3 个全连接层<sup>[12]</sup>。

## 2 深度学习模型压缩方法

使用基于交叉验证集的 APoZ 指标来评估神经网络中每一个卷积核和神经元的重要性。如果一个数据输入到卷积神经网络, 在计算过程中某一个卷积核或神经元的输出为 0, 且 0 不会经过乘法运算得到非零数据, 而偏置又是非常小的数据。此时可以认为这个卷积核或神经元对于该数据的分类起到的作用很弱。进一步, 如果将全部交叉验证集的数据依次输入到神经网络模型中, 某一个卷积核或神经元的输出大部分情况都为 0, 可以认为这个卷积核或神经元在这项分类任务中的作用很弱, 并且可以被剔除。基于交叉验证集, 某一个卷积核或神经元输出为 0 的概率即为全零矩阵平均百分比<sup>[13]</sup>。第  $c$  层中第  $i$  个卷积核或神经元的全零矩阵平均百分比  $APoZ_c^{(i)}$  为:

$$APoZ_c^{(i)} = APoZ(O_c^{(i)}) = \frac{\sum_k \sum_j f(O_{c,j}^{(i)}(k) = 0)}{H \times W} \quad (3)$$

式中:  $O_c^{(i)}$  为第  $c$  层中第  $i$  个卷积核或神经元的输出中的第  $j$  维数值;  $O_{c,j}^{(i)}(k)$  为第  $k$  个数据在第  $c$  层中第  $i$  个卷积核或神经元的输出中第  $j$  维数据;  $f(x)$  函数满足式(2);  $W$  为第  $c$  层中第  $i$  个卷积核或神经元的输出的维度;  $H$  为交叉验证集中数据样本的数量。

剔除的卷积核和神经元对模型的最终计算结果有一些影响, 剔除后, 模型会产生一定微弱畸形, 需对模型进行重新训练, 将剔除的卷积核和神经元的作用分担到其他保留的卷积核和神经元中<sup>[14]</sup>。由于模型剪枝操作不可逆, 较好的方法是每次只剔除数量较少的卷积核和神经元, 不断迭代。这个过程相对耗时, 但可以得到一个比较好的压缩效果。图 4 为模型轻量化流程的示意图。

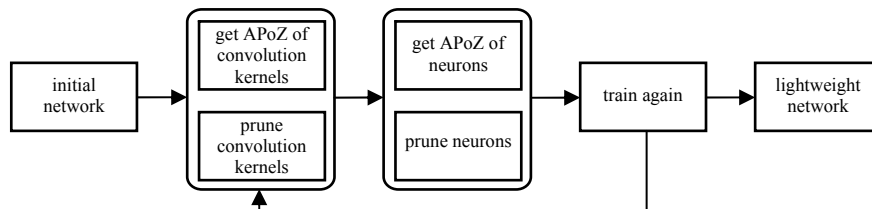


Fig.4 Schematic diagram of the model lightweighting process  
图 4 模型轻量化流程的示意图

## 3 实验

### 3.1 原始网络分析

实验训练并测试 2 个卷积神经网络, 从识别准确率、召回率和精准度 3 方面分析模型的性能, 从模型的参数量和浮点运算量(FLOPs)分析模型的规模。

3.1.1 准确率

两模型在不同信噪比下的分类准确率曲线见图 5。经过对比分析可以发现：a) 两模型在信噪比为-6 dB 的噪声下，识别准确率均高于 80%，并随信噪比的增加而增加。当信噪比增加至 0 dB 时，两模型的准确率均达到 99%以上。b) 在不同信噪比下，VGG16 的准确率均略高于 AlexNet。

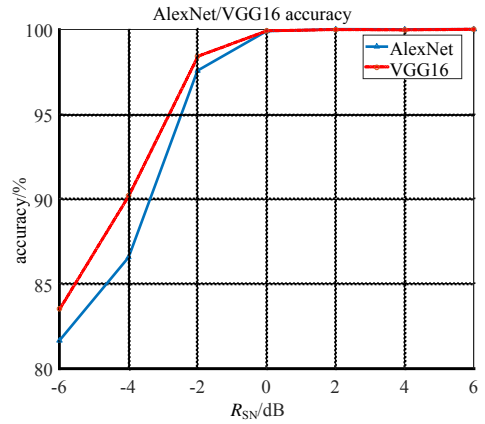


Fig.5 Classification accuracy curves of AlexNet and VGG16  
图 5 AlexNet 和 VGG16 的分类准确率曲线

3.1.2 召回率和精准度

对模型的召回率和精准度进行分析，其中召回率表征模型对某一类数据的分类性能，精准度表征模型某一类输出的可信度。对于均匀分布的测试集，当模型的准确率接近 100%的程度时，模型的召回率和精准度也会接近 100%。实验选择信噪比为-6 dB 下的模型进行分析，两模型在信噪比为-6 dB 的噪声下，召回率和精准度见图 6 和图 7。可以发现：a) 2 个模型具有基本相同的召回率和精准度；b) 2 个模型对于 QPSK, OQPSK 和 8PSK 三类信号的分类性能较差，很容易产生误判，且判别输出为这 3 类信号时可信度也较低；c) 对于其他 5 种信号，两模型均能进行很好的分类且可信度较高。

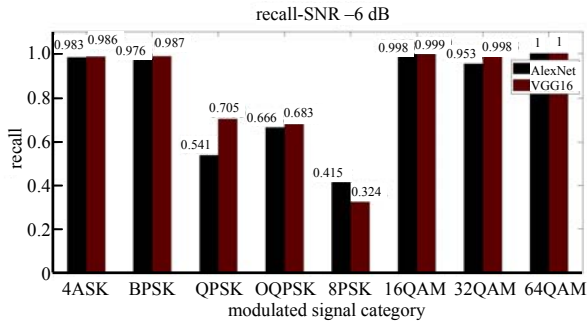


Fig.6 Recall of AlexNet and VGG16 under -6 dB SNR  
图 6 AlexNet 和 VGG16 在-6 dB 信噪比下的召回率

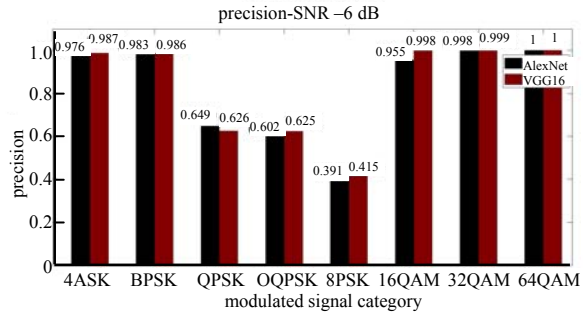


Fig.7 Precision of AlexNet and VGG16 under -6 dB SNR  
图 7 AlexNet 和 VGG16 在-6 dB 信噪比下的精准度

3.1.3 模型规模

采用模型的参数量和浮点运算量 2 个指标分析模型的规模，其中模型的参数量表征模型对于设备存储性能的要求，浮点运算量表征模型对于设备计算性能的要求。两模型的参数量和浮点运算量见表 1。可以发现，VGG16 网络模型在模型参数量和浮点运算量两方面均为 AlexNet 网络的 2.3 倍左右，而 2 个模型的性能却很相近，一定程度上也能证明模型的冗余。

表 1 AlexNet 和 VGG16 的参数量和浮点运算量

network	parameter	FLOPs
AlexNet	58,314,142	349,865,586
VGG16	134,239,320	805,730,100

3.2 轻量化网络的测试与分析

采用基于交叉验证集的 APoZ 指标来评估网络模型中的卷积核和神经元，每次剪掉模型中全零矩阵平均百分比在前 20%的卷积核和神经元，经过重训练后，测试发现模型准确率下降 1%，则不再继续；否则，继续剪切。

3.2.1 召回率和精准度

本实验在控制识别准确度不变的情况下对模型进行剪枝，因此两模型的识别准确率曲线与图 5 相同，实验直接分析召回率和准确度两项指标。同样选择信噪比为-6 dB 下的模型进行分析。两模型在-6 dB 下的召回率和准确度见图 8 和图 9。

对比图 6~图 7、图 8~图 9 可以发现，两模型经过剪枝操作后，对于 QPSK, OQPSK 和 8PSK 三种信号的召回率和准确度之间有着此消彼长的现象，其他信号依然保持很好的分类性能。这说明数据集的可分性决定模型对该数据集分类的整体性能，而各项指标之间能够保持一种平衡。

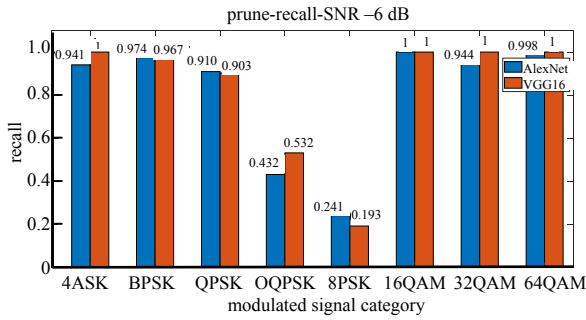


Fig.8 Recall of lightweight AlexNet/VGG16 under -6 dB SNR  
图 8 轻量化模型在-6 dB 信噪比下的召回率

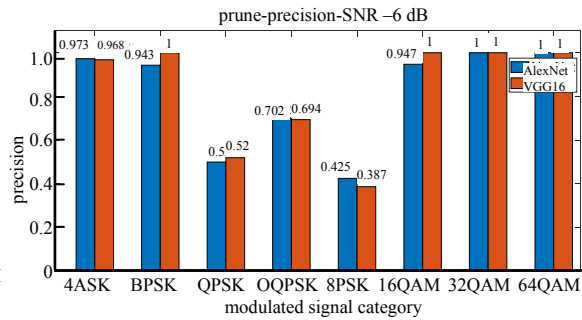


Fig.9 Precision of lightweight AlexNet/VGG16 under -6 dB SNR  
图 9 轻量化模型在-6 dB 信噪比下的精准度

3.2.2 轻量化模型规模

分析模型经过剪枝后，模型的参数量和浮点运算量的压缩倍数。两模型对于这 2 项指标在不同信噪比下的压缩倍数见图 10 和图 11。可以发现：a) 两模型的压缩倍数均随着信噪比的提升而升高。b) 信噪比为 -6 ~ 0 dB 时，两模型的压缩倍数均保持在几十倍；当信噪比大于 0 dB 时，两模型的压缩倍数急剧上升，AlexNet 的压缩倍数可达到数千倍，而 VGG16 的压缩倍数可达到数万倍。由此可知，数据的可分性越强，所需的模型规模越小。

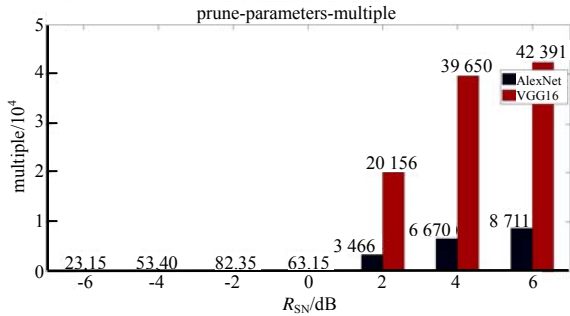


Fig.10 Compression times of the quantities of parameters  
图 10 模型参数量的压缩倍数

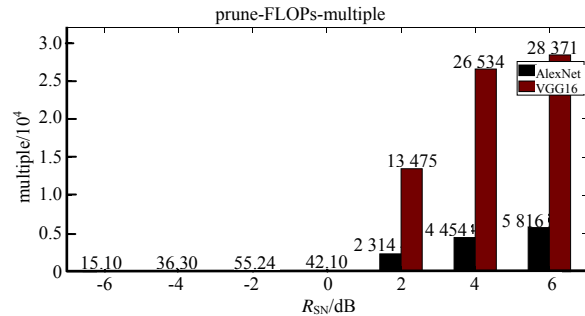


Fig.11 Compression times of the quantities of FLOPs  
图 11 浮点运算量的压缩倍数

3.3 探究性实验

为进一步分析和理解压缩方法，本文进行了一个探究性实验。在探究性实验中，设计了一个与剪枝后具有相同规模的网络，随机初始化其参数，对其进行训练和测试。剪枝后的网络和自设计同架构网络的测试结果的混淆矩阵见图 12 和图 13。经过对比可以发现，自设计的同架构网络无法达到经过剪枝后得到的模型的分类效果。对此结果，经过查阅资料，本文做出假设性分析：网络需要更多的参数逼近全局最优解或更优解。虽然最优解处某一参数最终为零，但如果该参数初始化设定为零，则模型无法越过局部最优解与全局最优解或更优解之间的障碍。网络在开始训练时需要更大的模型容量，使模型能够与数据形成很好的拟合，拟合后，通过本实验用的剪枝方法可以基于数据集对模型进行去除冗余。但如果模型在开始训练时就进行剪裁，导致模型容量不足，对数据进行拟合时则达不到同样的效果<sup>[15]</sup>。

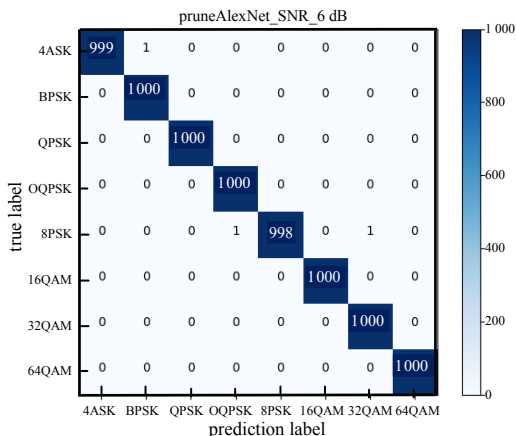


Fig.12 Effect of the model after pruning  
图 12 剪枝后模型的效果

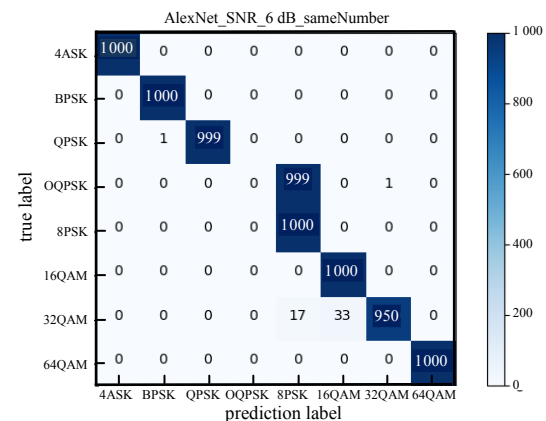


Fig.13 Effect of self-designed network with the same architecture  
图 13 具有相同规模的自设计模型的效果



## 4 结论

本文将调制信号转化为具有颜色信息的星座图形式,采用深度学习方法,基于此数据集对卷积神经网络进行训练和测试。选择 VGG16 和 AlexNet 两个卷积神经网络完成分类任务,当信噪比大于等于 0 dB 时,两模型的识别准确度均能达到 99%以上。为进一步减小模型规模,使其适用于资源受限的小型嵌入式设备,本文使用全零矩阵平均百分比的指标对模型中卷积核和神经元进行评估,剪切掉其中的冗余。实验中,当信噪比大于 0 dB 时, AlexNet 网络的压缩倍数可达到数千倍,而 VGG16 的压缩倍数可达到数万倍。

### 参考文献:

- [1] 冯德俊,朱江,李方伟. 战场电磁态势感知关键技术研究[J]. 数字通信, 2013,40(5):20-23. (FENG Dejun,ZHU Jiang,LI Fangwei. Key technology research of electromagnetic situation awareness in battlefield[J]. Digital Communications, 2013,40(5):20-23.)
- [2] 潘俊. 通信信号调制方式识别方法探析[J]. 数字通信世界, 2019,169(1):96. (PAN Jun. Research on identification method of communication signal modulation[J]. Digital Communication World, 2019,169(1):96.)
- [3] O'SHEA T J,CORGAN J,CLANCY T C. Convolutional radio modulation recognition networks[C]// Engineering Applications of Neural Networks. Aberdeen,United Kingdom:[s.n.], 2016:213-226.
- [4] PENG S,JIANG H,WANG H,et al. Modulation classification using convolutional neural network based deep learning model[C]// Wireless & Optical Communication Conference. Newark,NJ,USA:IEEE, 2017.
- [5] PENG Shengliang,JIANG Hanyu,WANG Huaxia,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.
- [6] COURBARIAUX M,BENGIO Y,DAVID J P. Training deep neural networks with low precision multiplications[EB/OL]. (2015-09-23)[2019-06-24]. <https://arxiv.org/abs/1412.7024>.
- [7] SRIVASTAVA N,HINTON G,KRIZHEVSKY A,et al. Dropout:a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014,15(1):1929-1958.
- [8] SZEGEDY C,IOFFE S,VANHOUCKE V. Inception-v4,Inception-ResNet and the impact of residual connections on learning[C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco,California,USA:[s.n.], 2016:4278-4284.
- [9] IANDOLA F N,HAN S,MOSKEWICZ M W,et al. SqueezeNet:AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[EB/OL]. [2019-01-02]. <https://arxiv.org/pdf/1602.07360.pdf>.
- [10] LIN Y,TU Y,DOU Z,et al. The application of deep learning in communication signal modulation recognition[C]// 2017 IEEE International Conference on Communications in China(ICC). Qingdao,Shandong,China:IEEE, 2017:1-5.
- [11] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. Lake Tahoe,Nevada,USA:NIPS, 2012:1097-1105.
- [12] SIMONYAN K,ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2019-01-02]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [13] HU H,PENG R,TAI Y W,et al. Network trimming:a data-driven neuron pruning approach towards efficient deep architectures[J]. arXiv preprint, arXiv:1607.03250, 2016.
- [14] 李聪颖. 基于卷积神经网络的模型压缩研究及应用[D]. 郑州:河南大学, 2018. (LI Congying. Research and application of model compression based on convolutional neural network[D]. Zhengzhou,China:Henan University, 2018.)
- [15] 张瀛. 基于动态剪枝算法的神经网络预测模型[J]. 复旦学报(自然科学版), 2015,54(3):313-319. (ZHANG Ying. Neural network prediction model based on dynamic pruning algorithm[J]. Fudan Journal(Natural Science Edition), 2015,54(3):313-319.)