

基于深度学习的无线信号认知技术研究

作者姓名 赵纪伟

指导教师姓名、职称 杨清海 教授

申请学位类别 工学硕士

学校代码 10701
分 类 号 TP391

学 号 1501120264
密 级 公开

西安电子科技大学

硕士学位论文

基于深度学习的无线信号认知技术研究

作者姓名：赵纪伟

一级学科：信息与通信工程

二级学科：通信与信息系统

学位类别：工学硕士

指导教师姓名、职称：杨清海 教授

学 院：通信工程学院

提交日期：2018 年 4 月

Thesis/Dissertation LaTex Template for Postgraduates of XIDIAN UNIVERSITY

A Thesis submitted to
XIDIAN UNIVERSITY
in partial fulfillment of the requirements
for the degree of Master

By

Jackokie Zhao

Supervisor: Qinghai Yang Title:Professor

April 2018

西安电子科技大学 学位论文独创性（或创新性）声明

秉承学校严谨的学风和优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。学位论文若有不实之处，本人承担一切法律责任。

本人签名：_____ 日 期：_____

西安电子科技大学 关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权属于西安电子科技大学。学校有权保留递交论文的复印件，允许查阅、借阅论文；学校可以公布论文的全部或部分内容，允许采用影印、缩印或其它复制手段保存论文。同时本人保证，结合学位论文研究成果完成的论文、发明专利等成果，署名单位为西安电子科技大学。

保密的学位论文在____年解密后适用本授权书。

本人签名：_____ 导师签名：_____

日 期：_____ 日 期：_____

摘要

复杂电磁环境下的无线信号认知是优化频谱利用效率、识别和最小化干扰的重要方法。传统的调制识别方法主要集中在能量检测以及专家特征和决策准则的使用。随着深度学习发展，利用原始采样信号学习信号特征并进行识别调制方式得以实现，并且在低信噪比条件下的准确率优于传统的基于专家特征提取的统计机器学习模型。本文对基于深度学习的调制识别算法、特征融合框架、影响模型性能的因素等进行了研究，主要贡献如下：

1. 本文提出了一种 CAE-CNN 调制识别的算法框架，并提出了相应的训练算法。我们分别利用监督方法和无监督方法将信号特征可视化，发现通过无监督的卷积自编码器可以复现输入信号，获得的数据低维表示具有一定的类别区分性；通过有监督的 CNN 获取的信号低维表示，对不同类别的调制信号具有较强的区分度，利用 t-SNE 算法降到二维的流型中，相同类别的信号独立成一个簇。我们融合卷积自编码器与卷积神经网络，提出了 CAE-CNN 算法框架和相应训练算法。仿真结果显示，我们提出的算法，在高信噪比条件下具有较高的识别准确率（95% 以上），并且鲁棒性较强；在低信噪比条件下具备比传统方法更高的识别率，在 SNR 为 -4dB 时准确率仍在 85% 左右。在算法效率上，我们提出的算法训练时间在可接受范围以内，同时算法分类识别时间也较低。
2. 提出了一种传统特征与深度特征融合的框架。将传统方法中常用的调制识别的特征与 CNN 获取的深度特征进行适应性的 *Batch – Normalization*，然后分别针对 Softmax、随机森林、深度神经网络等融合算法构建融合模型，并对不同融合算法的进行了仿真仿真。仿真结果显示，基于随机森林和基于 Softmax 的融合算法性能由于基准的卷积神经网络，并且基于 RF 的融合算法具有相对最优的分类性能，无论在低信噪比还是高信噪比都具有较强的鲁棒性。同时，受融合算法特性等的影响，不同融合框架下的误分信号各不相同。
3. 我们从网络底层研究网络超参数对调制识别性能的影响，并从欠拟合与过拟合以及偏差与方差的角度解释了仿真结果。CNN 在调制识别时的性能似乎不受网络深度的限制，在卷积层数目为 3 时即可达到性能的最值，此时再增大卷积层的深度，系统性能可能会发生下降。第一层卷积核数目较小时，网络的性能几乎随着第二层卷积核数目的增大而减小。而随着第一层卷积核增大，改变第二层卷积核数目，系统分类准确率变化很小，只有轻微的波动。整体而言，卷积核宽度 $W^{(1)}$ 和 $W^{(2)}$ 较大的网络性能较好，但是当卷积的宽度增大到 7 之后，分类性能较为稳定，仅仅发生微小的

波动。

关键词：调制识别， 深度学习， 网络超参数， 特征融合

ABSTRACT

Wireless signal recognition in complex electromagnetic environments is an important method for optimizing spectrum utilization efficiency, identifying and minimizing interference. Traditional modulation identification methods mainly focus on energy detection and the use of expert characteristics and decision criteria. With the development of deep learning, the original sampled signal is used to learn the signal features and identify and modulate the signal. The accuracy under low signal-to-noise ratio is better than the traditional statistical machine learning model based on expert feature extraction. This paper studies the modulation recognition algorithms based on deep learning, feature fusion framework, and factors that affect the performance of the model. The main contributions are as follows:

1. This paper proposes a CAE-CNN modulation recognition algorithm framework and proposes a corresponding training algorithm. We visualized the signal characteristics using supervised methods and unsupervised methods, respectively, and found that the unsupervised convolutional self-encoder can reproduce the input signals. The low-dimensional representation of the obtained data has a certain class distinction; it is obtained through supervised CNN. The low dimension of the signal indicates that there is a strong degree of discrimination between the different classes of modulated signals, which is reduced to a two-dimensional flow pattern using the t-SNE algorithm, and the signals of the same class form a cluster independently. We fuse convolutional self-encoders and convolutional neural networks and propose a CAE-CNN algorithm framework and corresponding training algorithms. The simulation results show that the proposed algorithm has higher recognition accuracy (95% or more) under high signal-to-noise ratio, and has stronger robustness; it has higher performance than traditional methods under low signal-to-noise ratio conditions. The recognition rate is still about 85% when the SNR is -4dB. In terms of algorithm efficiency, the training time of our proposed algorithm is within an acceptable range, and the classification and recognition time of the algorithm is also low.
2. A framework for the fusion of traditional features and depth features is proposed. The *Batch – Normalization* that adapts the features of modulation recognition commonly used in traditional methods and the depth features obtained by CNN, and then construct a fusion model for fusion algorithms such as Softmax, random forest, and deep neural network, and

the fusion algorithms for different fusion algorithms. Simulations were performed. The simulation results show that the performance of the fusion algorithm based on random forest and Softmax is due to the benchmark convolutional neural network, and the RF-based fusion algorithm has relatively optimal classification performance, which is comparable in both low SNR and high SNR. Strong robustness. At the same time, due to the influence of the characteristics of the fusion algorithm, the misclassification signals under different fusion frameworks are different.

3. We study the effect of network hyperparameters on modulation recognition performance from the network bottom and explain the simulation results from the perspective of under-fitting and over-fitting, as well as deviation and variance. The performance of CNN in modulation recognition does not seem to be limited by the depth of the network. The maximum value of performance can be achieved when the number of convolution layers is 3. At this time, the depth of the convolutional layer is further increased, and the system performance may decrease. When the number of first-level convolutional cores is small, the performance of the network decreases almost as the number of second-level convolutional cores increases. With the increase of the number of convolutional nuclei at the first level, the number of second-level convolutional nuclei changes, and the accuracy of system classification changes only slightly, with only slight fluctuations. Overall, the larger network performance of the convolution kernel widths $W^{(1)}$ and $W^{(2)}$ is better, but the classification performance is increased when the convolution width is increased to 7. It is relatively stable and only minor fluctuations occur.

Keywords: Modulation identification, deep Learning, feature fusion

插图索引

1.1	基于似然比判决理论的调制识别系统框图	4
1.2	基于统计机器学习的调制识别系统框图	5
2.1	神经元示意图	11
2.2	ReLU 函数	11
2.3	Sigmoid 函数	11
2.4	前馈神经网络	12
2.5	自编码器	17
3.1	调制信号生成框图	24
3.2	不同调制方式的高 SNR 样本的时域波形	26
3.3	不同调制方式的高 SNR 样本的频谱	26
3.4	自编码器	28
3.5	基于自编码器的信号复现	28
3.6	基于自编码器的调制信号 t-SNE 二维流型展示	29
3.7	基于 CNN 的调制信号 t-SNE 二维流型展示码器	30
3.8	CAE-CNN 网络框架	31
3.9	CAE-CNN 分类结果	33
3.10	与其他方法的比较	34
3.11	0dB 条件下的结果	35
3.12	不同算法训练时间	36
3.13	不同算法分类时间	36
4.1	特征融合框架图	43
4.2	基于 LR 框架的特征融合框架图	46
4.3	基于 DNN 框架的特征融合框架图	47
4.4	基于随机森林框架的特征融合框架图	48
4.5	不同融合框架下的分类性能对比	49
4.6	不同融合模型分类混淆矩阵对比	50
5.1	本章所用 CNN 网络框架	52
5.2	不同卷积核宽度下，卷积层数目对分类准确率的影响	54

5.3	分类准确率随卷积层数目和卷积核宽度的变化	55
5.4	分类准确率随卷积核数目的变化情况	57
5.5	第一卷积层中卷积核数目对分类准确率的影响	57
5.6	第二卷积层中卷积核数目对分类准确率的影响	58
5.7	不同 $H^{(1)} \times H^{(2)}$ 条件下卷积核宽度对分类性能的影响	60
5.8	不同卷积核高度 $H^{(1)} \times H^{(2)}$ 下，卷积核宽度对分类准确率的影响	61

表格索引

3.1	SMW200A 参数	24
3.2	SMW200A 参数	25
3.3	系统参数配置	32

符号对照表

符号	符号名称
\in	属于
w	权重
x	样本
y	标签
M	特征维数
N	样本数量
η	学习率
\mathcal{F}^{-1}	逆傅里叶变换
γ	弱分类器更新率

缩略语对照表

缩略语	英文全称	中文对照
SVM	Support Vector Machine	支持向量机
EM	expectation–maximization	最大期望
WTS	Weighted Tensor Subspace	加权张量子空间
PCA	Principal Component Analysis	主成分分析
IPCA	Incremental PCA	增量主成分分析
HOG	Histogram of Oriented Gradient	方向梯度直方图
2D-LDA	2D Fisher Linear Discriminant Analysis	二维 Fisher 线性判别分析
AVT	Attentional Visual Tracking	注意视觉跟踪
RF	Random Forest	随机森林
FFT	Fast Fourier Transformation	快速傅里叶变换
MOSSE	Minimum Output Sum of Squared Error filter	最小平方误差滤波器
CFT	Correlation Filter Tracker	相关滤波跟踪器
DFT	Discrete Fourier Transform	离散傅里叶变换
KCF	Kernelized Correlation Filter	核相关滤波器
CLE	Center Location Error	中心位置误差
OP	Overlap Precision	重叠精度
DP	Distance Precision	距离精度
ASMM	Atkinson–Shiffrin Memory Model	AtkinsonShiffrin 内存模型
MUSTer	MULTi-Store Tracker	多贮存跟踪器
KNN	K-Nearest Neighbor	K-最近邻
HOG	Histogram of Oriented Gradient	方向梯度直方图
ALM	Augmented Lagrange Method	增强拉格朗日方法
ADMM	Alternating Direction Method of Multipliers	交替方向乘子算法

目录

摘要.....	I
ABSTRACT	III
插图索引.....	V
插图索引.....	V
表格索引.....	VII
表格索引.....	VII
符号对照表.....	IX
缩略语对照表.....	XI
第一章 绪论	1
1.1 研究背景及研究意义	1
1.1.1 研究背景	1
1.1.2 研究意义	2
1.2 无线信号调制识别的发展和研究现状	3
1.2.1 基于似然比判决理论的方法	4
1.2.2 基于特征提取的统计机器学习方法	5
1.2.3 基于深度学习的无线调制识别方法	6
1.3 本文主要工作及内容安排	7
第二章 无线信号调制识别以及深度学习理论	9
2.1 引言	9
2.2 调制识别	9
2.2.1 信道对调制信号的影响	9
2.2.2 信道建模	9
2.3 神经网络概述	10
2.3.1 神经元概述	10
2.3.2 前馈神经网络	12
2.3.3 反向传播算法	13
2.4 卷积神经网络	14
2.4.1 卷积运算	15
2.4.2 卷积特性	15
2.5 卷积自编码器	17

2.6	神经网络优化算法	18
2.6.1	统计梯度下降优化算法	18
2.6.2	自适应矩估计优化算法	19
2.6.3	RMSProp 优化算法	20
2.6.4	选择正确的优化算法	21
2.7	本章小结	21
第三章	基于 CAE-CNN 的无线信号调制识别	23
3.1	引言	23
3.2	调制信号	23
3.2.1	调制信号生成	23
3.2.2	信号采集	25
3.3	调制信号的表示	25
3.3.1	数据集可视化	25
3.3.2	调制信号的无监督表示	27
3.3.3	调制信号的监督引导稀疏表示	29
3.4	基于 CAE-CNN 的无线信号调制识别	30
3.4.1	CAE-CNN 网络框架	30
3.4.2	CAE-CNN 算法	31
3.4.3	算法运行环境及参数	31
3.5	结果及分析	33
3.5.1	分类结果	33
3.5.2	算法效率	35
3.6	本章小结	37
第四章	基于传统特征与深度特征融合的无线调制方式识别技术研究	39
4.1	引言	39
4.2	传统特征	39
4.2.1	基本时频特征	39
4.2.2	高阶累积量	41
4.3	特征融合理论	42
4.4	传统特征与深度特征融合框架	43
4.4.1	特征归一化	43
4.4.2	基于 Softmax 回归的融合框架	44
4.4.3	基于深度学习的融合框架	45
4.4.4	基于集成树的融合框架	46

4.5 结果及分析	48
4.5.1 分类性能比较	48
4.5.2 分类混淆矩阵	49
4.6 本章小结	50
第五章 调制识别的深度框架研究	51
5.1 引言	51
5.2 系统模型	51
5.2.1 模型的偏差与方差	52
5.2.2 过拟合与欠拟合	53
5.3 网络超参数对调制识别的影响	53
5.3.1 网络层数对调制识别的影响	54
5.3.2 卷积核数目对调制识别的影响	56
5.3.3 卷积核大小对调制识别的影响	59
5.4 本章小结	62
第六章 总结与展望	63
6.1 研究结论	63
6.2 未来展望	64
参考文献	65
致谢	69
作者简介	71

第一章 绪论

1.1 研究背景及研究意义

1.1.1 研究背景

通信的主要目的是传递信息。无论是在传统的有线通信系统还是在无线通信系统中，由于信道传输特性的限制，基带信号不能直接通过信道传输，而是需要对基带信号进行调制以将其搬到一定的频段进行有效地传输^[1]。同时，对基带信号的调制，还可以使通信系统在具有更高的传输速率的同时，提高频谱利用率。

随着通信技术的不断发展，终端的数目呈爆发性增长，用户对信息传输速率和传输稳定性的要求不断提高。为了满足这样的需求，通信信号的调制形式由简单到复杂，调制方式由传统的模拟调制发展到主流的数字调制方式，信道从有线信道到有线与无线信道混合组网。调制是指将发送的信号加载到高频（基带调制除外）信号以适应信道传播环境的一种技术。调制方式自动识别，是指在未给定调制信号所携信息的情况下，利用学习的算法，通过接收信号电磁特性判定信号的调制方式。模拟调制针对的是模拟信号，传统的模拟调制按照调制方式的不同，可以分为幅度调制(AM)、频率调制(FM)和相位调制(PM)等；数字调制针对的是数字信号，按照调制方式的不同可以分为幅移键控(ASK)、频移键控(FSK)、相移键控(PSK)和正交幅度调制(QAM)等。

调制信号的自动识别，是一种优化频谱利用效率，识别和最小化干扰，以及有效推动认知无线网发展的重要手段，在协作通信和非协作通信中都得到了广泛应用^[2]。在军事领域，调制识别被广泛地用作通信侦察，电子战和威胁分析，是一种智能信号分析和处理的关键技术^{[3][4]}。在侦听接收机的设计中，获取无线接收信号的调制方式是侦听接收机的重要功能之一。有效的无线信号调制方式识别为解调器正确选择解调算法提供了参数依据，可以使我们最终得到准确的情报信息。调制识别技术也有助于为电子战选择最佳的干扰模式或干扰消除算法，以保证我方设备之间通信的稳定性；同时抑制和干扰对方的通信以达到电子战通信对抗的目的。

因此，通信信号调制自动识别技术具有非常广阔的应用前景，将成为未来的民用无线通信以及军事通信中的重要组成部分。

深度学习(Deep Learning, DL)，是一种机器学习(Machine Learning, ML)领域的分支。2006年Hinton等人提出了深度置信网络(Deep Belief Network, DBN)^[5]，并引入了分层预训练技术，自此深度学习得到了广泛的应用。在目标检测、语音识别、机器翻译等许多不同的人工智能任务中，深度学习已经极大地提升了这些任务上的

最佳表现水平^[6]。

2012 年的 ILSVRC2012 竞赛中，Krizhevsky 等人引入了 Dropout 等方法，训练了一个大型的深度卷积神经网络 AlexNet，使用深度学习模型击败了 Google 团队并取得了该竞赛的冠军，使得图片识别错误率下降了 14%^[7]。2014 年 Christian Szegedy 等人提出 GoogLeNet，打破了常规的卷积层串联的模式，将 $1 \times 1, 3 \times 3, 5 \times 5$ 的卷积层和 3×3 的池化层并联组合到一起，以较大优势取得 ILSVRC2014 冠军并使 top-5 的错误率降到了 6.67%^[8]。2015 年，Kaiming He 等人设计了一个多达 152 层的 ResNet 架构，使图片识别的错误率也降到了 3.6%^[9]。研究者在将深度学习应用到目标检测、面部识别或语言模型等传统任务的同时，也在将其扩展到许多不同的现代领域和任务中，比如 Osako 等人使用循环神经网络来给语音信号降噪^[10]，Gupta 等人使用堆栈自编码器（stacked autoencoders）来发现基因表达的聚类模式^[11]，Gatys 等人使用一种神经模型来生成不同风格的图像^[12] 等等。深度学习在很多领域都取得了较大的发展。

1.1.2 研究意义

在无线通信中，调制识别是正确实现无线通信解调和保证正常接收的重要组成部分。基带信号通常是频率非常低且分量很多的频谱，这些含有很多低频分量的信号不适合在信道中直接传输，这就需要我们对基带信号进行调制，以适应信道的特性并在无线信道中进行有效传输。信号在经过调制并通过信道到达接收端，与原始的信号具有很大的不同。所以，只有在接收端确定信号的调制方式，才能利用相应的解调方法解调获取原始信息。基于深度学习的调制识别的意义主要体现在以下几方面：

(1) 提高通信设备的稳定性与通用性，降低建设与运维成本。通信技术的迅速发展导致了各种通信系统的共存，而这些通信系统的调制方式可能各不相同。由于解调需要与调制进行匹配才能准确地接收发射信号，这些差异性的调制技术也导致接收机类型大幅增加。于是，人们希望能够通过发展新的无线认知技术，基于通用的接收机平台来接收并识别不同的调制信号，以减少接收机类型的快速增长。

(2) 军事领域的电子对抗、电磁干扰与反干扰的基础。当检测到对方无线电信号或电磁信号时，需要我们对检测到的信号或电磁信息，进行调制识别以及载波频率与波特率等参数的正确估计，方便接下来的情报获取或者进行电磁干扰。

(3) 有助于如何在频域、时域和空域等多维度上有效分配频谱资源，提高频谱利用率。在认知无线领域，无线频谱资源根据具体业务的不同，主要划分为民用的广播电视、无线通信、卫星通信及军用的雷达与军事通信等不同频段。为了避免相邻频段或频道的互相干扰，不同的通信业务的工作频段相互隔离，并且不同频段之间预留一定的间隔频带；上下频段的频谱资源分配以及频谱利用率极不平衡，造成了频谱资源

的极大浪费。由于无线频谱资源有限，随着频谱资源的消耗殆尽，现有的频谱分配与管理机制已经成为制约无线通信进一步发展的重要因素。

(4) 利用深度学习的方法提升传统调制识别算法的性能上限。在通信系统中使用 DL 有很多优点。首先，由于通信基础设施以及终端的数量多，且通信数据速率高，因此可以很容易获取 DL 训练所需要的大量数据。其次，DL 可以自主提取特征，避免了手动特征选择这一繁琐且具有挑战性的任务。另外，新的、更复杂的信号和通信应用的出现，给认知无线带来了更大的挑战；有时，我们需要对一些未知的信号进行认知，而传统的基于特征的方法在新的场景下很难适应信号特性的变化。因此，我们可以利用深度学习的特征自提取能力，来增强对于未知信号的认知能力。

(5) 促进深度学习在通信领域的发展。在调制识别领域中，利用专家特征和判决准则进行调制识别得到了广泛的应用，并且在特定情况下可以实现很高的识别准确率。深度学习在计算机视觉和自然语言处理等领域的应用，绝大多数是基于数据进行特征与模型的学习，而不是利用专家知识提取特征进而训练模型，这与很多传统的机器学习算法（比如 DNN、支持向量机、决策树等）在算法思想上有很大不同。而且，这些基于数据的深度学习算法在相应的领域取得了最好的性能。因此，这就需要我们重新审视一下，是否也应该对通信领域的传统算法在思想上进行改变，基于数据去理解信号，去学习信号本身的特征。

1.2 无线信号调制识别的发展和研究现状

早期调制识别的任务是由操作人员在仪器的帮助下完成的：主要是通过观察和分析接收信号的时域波形和频谱形状，判断信号的调制方式，然后选择相应的解调器进行解调。然而，随着无线通信技术尤其是数字通信技术的快速发展，信号调制方式变得越来越复杂，很难通过人工的方法来准确判别调制方式的类别。C.S.Waver 等人于 1969 年在斯坦福大学发表了第一篇关于通信信号调制识别的论文^[13]，此后，调制方式的自动识别引起了人们的广泛关注，各种技术出版物上出现了许多关于调制识别的文章，其中许多结果已经被应用到实际工作中。

对于无线信号调制识别的研究，现有算法大致可以分为：基于假设检验的最大似然法，基于特征提取的模式识别方法，以及基于深度学习的调制识别方法。大多数基于假设检验的最大似然类方法计算复杂度较高，对模型失配问题较为敏感，这大大限制了它们在实际通信环境中的应用。基于特征提取的模式识别方法，通常关注能量在不同频段上的分布，并使用专家特征和判别准则来识别和区分特定的调制方式。在特定的条件下，基于特征的方法可以实现接近理论最佳的识别性能，并且其具备较强的鲁棒性，因此得到更广泛的应用。鉴于深度学习在其他领域取得的成果，深度学习可

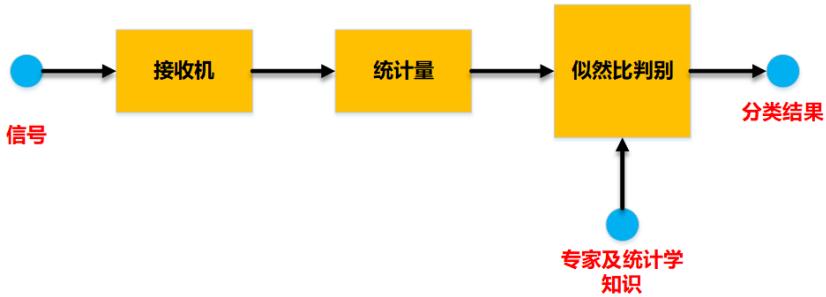


图 1.1 基于似然比判决理论的调制识别系统框图

以与硬件结合自适地进行学习提高传统算法的性能上限，并可以通过特定的正则化方法来降低过拟合提高模型的鲁棒性。因此，深度学习在通信领域近年来已经成为一个研究热点，在调制识别领域也有很多人投入到相应的研究中。

1.2.1 基于似然比判决理论的方法

基于似然比决策理论的方法也叫做最大似然假设检验法。其基本思想是使用概率论和假设检验理论来分析信号的统计特性，并依据损失函数最小化原则获取足够的统计信息，定义判决准则，对信号进行分类。基于似然比判决理论的调制识别系统框图如图1.1所示。

对于现有调制识别的发展，Kim 和 Polydoros^[14] 使用平均似然比检验来识别 BPSK 和 QPSK 信号。Hwang 等人通过似然方法解决了数字正交偏移调制信号的分类问题^[15]，Chung-Yu Huan 等人针对加性高斯白噪声中对 MPSK 调制进行分类的问题，提出了基于似然函数及其近似的算法^[16]。Schreyoegg 等人提出了一种分类 QAM 信号的星座图方法，主要通过分析幅度分布和相位直方图的 DFT 来对多种 QAM 信号进行区分^[17]。Wei 和 Mendel 将最大似然方法应用于数字正交调制的分类，并分析了最大似然分类器的渐近行为，当信噪比为 5dB 时，准确率接近 100%^[18]。Boiteau 等人提出了一种广义的似然比调制识别框架，作者首先对似然比函数中自然指数进行幂级数展开，然后对未知参数做期望平均处理，实质上应用了高阶统计量^[19]。

基于似然比决策理论的算法的优势是：在理论上保证贝叶斯最小误判准则，使分类结果最优；同时，它可以通过理论分析得到分类性能曲线。但是他同时具有很多局限性：首先，现有的似然比决策理论算法主要处理符号的同步采样序列，他们需要比模式识别方法更多的先验知识，这意味着我们需要预先知道信号的载波频率，符号速率和符号时序等先验信息；其次，基于未知参数似然比的分类，需要计算复杂的统计表达式，计算成本较高，很难对信号进行实时处理；第三，基于似然比决策的算法对模型失配和参数偏差较为敏感，即鲁棒性较差。基于似然比函数的分类，通常将参数值建模为高斯分布，并且需体现知道信噪比等参数。当实际信道噪声为非高斯噪声

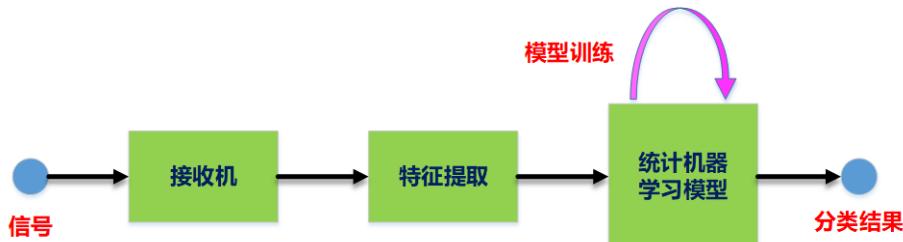


图 1.2 基于统计机器学习的调制识别系统框图

时，或者存在多径影响、多信号干扰以及 SNR 参数估计偏差较大时，分类的性能可能会急剧下降。

1.2.2 基于特征提取的统计机器学习方法

基于统计机器学习的调制识别系统大都具有相同的范式：首先从信号中提取先前选择的特征，然后利用训练好的机器学习模型进行调制识别。它主要包括特征提取子系统和机器学习子系统，其整体系统框架如图1.2所示。

特征提取子系统主要从未处理信号中提取所需的特征分量，如瞬时频率，瞬时相位和瞬时幅度等。模式识别子系统的主要功能是通过特征子系统提取的特征分量对模型进行训练；模型训练好以后，当需要判别的信号进入该子系统，我们的模型可以对不同的调制信号进行分类。

Liedtke 提出了一种用于未知调制方法分类的通用分类器，对数字调制信号识别方法首先进行公开讨论^[20]。该算法主要对幅移键控（ASK），具有小频偏的频移键控（FSK）和相移键控（PSK）等信号进行了识别，对于噪声、频率偏移、带宽不匹配以及邻域信号串扰等影响，具有较强的可靠性和鲁棒性。Azzouz 等人提出了一套用于识别不同类型数字调制方式的判别标准，它使用相位的非线性部分及其绝对值，归一化的瞬时振幅和频率等特征，在信噪比 $\text{SNR} = 10\text{dB}$ 时，调制类型分类正确率大于 90%^[21]。M.L.D. Wong 和 AK Nandi 等人提出了利用多层感知机进行调制识别的算法，该算法具有更好的泛化能力，并增加了一个新的基于统计累积量的特征集来识别十种不同的调制类型；即使在低信噪比条件下，该算法也具备较好的性能，在 $\text{SNR}=0\text{dB}$ 时，识别准确率可以达到 98%^[22]。汤卫东利用小波变化对调制信号进行了类内和类间识别，提高了低信噪比条件下的识别率^[?]。杨发权在其博士论文中，对高维数据块正交调制识别方法进行应用，并对 MLP 在调制识别中的应用进行了研究，实现了在低信噪比条件下基于星座图的调制识别^[23]。

高阶统计量是描述随机过程的高阶统计特性的数学工具，包括高阶矩和高阶累积量，以及高阶周期矩和循环累积量。Reichert 和 Juergen 首次提出使用高阶统计量来识别 2ASK, 2PSK, 4PSK, 2FSK, MSK 等信号，之后基于高阶统计量的调制信号

识别方法取得了很大进展^[24]。Swami 提出了一种基于四阶累积量的数字调制识别的算法，对 2ASK, 2PSK, 4PSK, MSK 和 2FSK 的信号进行识别，当信噪比为 7.3dB 以下时，识别率都能超过 97.9%^[25]。Gardner 和 Spooner 第一次将循环谱域的结构可分性应用到信号分类，之后 KIM 等人利用循环谱特性构建分类器来识别 SQPSK 和 2^k PSK 信号。

基于星座的调制识别将调制识别问题转化为形状识别问题。Mobasseri 首先提出了星座图作为调制识别的一种强特征，利用二项非均匀空间随机场恢复星座图，并利用机器学习算法作为分类器，在高噪声环境中具备较强的稳定性^[26]。接下来，Mobasseri 又证明模糊 C 均值聚类能够恢复未知星座，使用星座形状作为数字调制识别的强特征，通过 ML 规则对未知星座形状的信号进行分类，该算法适用于任意大小和维度的数字调制^[27]。

对于以上提到的各种分类特征，包括时频统计特征，高阶统计量，星座特征等，大部分是针对特定类型信号的，而不是对所有的信号都具备一定的辨识能力。另外，大多数现有的调制识别算法都假定信道是理想的高斯白噪声信道，并且只有少数算法研究了衰落信道。在实际应用中，无线信道的衰落现象不容忽视，多径效应使得传输信号间存在码间干扰等。在基于理想高斯白噪声信道环境的识别算法中，信噪比较高时识别性能较好；当信噪比较低时，算法估计的瞬时包络，相位和频率参数可能会存在较大的误差，使系统的识别性能急剧下降，并且稳定性差，不能满足实际应用的需求。

1.2.3 基于深度学习的无线调制识别方法

深度学习是机器学习中一种基于对数据进行表征学习的方法。深层神经网络是由一系列层组成网络，其中每一层通常是由已知的具有可调参数和非线性激活函数的线性单元组成，使得深度网络最后可以拟合高度非线性的函数。

深度学习在通信领域的应用，带来了新的机遇与变革。O’Shea 等人第一次将卷积神经网络（Convolutional Neural Network, CNN）引入调制识别，证明了深度学习可以应用在调制识别领域，并且在低信噪比条件下具有更加稳定的性能^[28]；此外，作者还提供了一个用于调制识别学习的基准数据集^[29]。Mendis 等人将深度置信网络（Deep Belief Networks, DBN）引入到调制识别，在多径信道下，信噪比在 0dB 以上时检测准确率达到 90% 以上，分类准确率达到 85% 以上^[30]。Afan 等人提出了一种基于非负约束自编码器的自动调制分类的方法^[31]。相较于传统的稀疏编码，该方法提高了稀疏性并使重构误差最小，在有限的信号长度和衰落信道条件下具有较高的准确率。Rajendran 等人提出了一个基于长短期记忆网络（Long Short-Term Memory, LSTM）的调制识别网络，在信号的信噪比从 0dB 到 20dB 的不同 SNR 条件下，平均分类准

确率为 90% 左右^[32]。Bin Tang 提出了一种基于生成对抗网络（Generative Adversarial Networks, GAN）增强的通信信号调制分类算法，弥补数据不足对网络训练的影响，作者提出增强数据的方法，使模型的分类精度可以提高 0.1%~6.0%^[33]。

以上的研究主要是从应用的层面对深度学习的既有方法进行不同领域的迁移应用，并没有提出一定的算法框架或者底层网络结构的改进，而且对于影响性能的因素，算法所能达到的上限，以及将深度学习与传统的调制识别方法结合方面都没有进行相应地研究。因此，需要我们对深度学习在调制识别中的应用在算法、框架、应用等层面进行更深入的研究，以提高系统的调制识别性能。

1.3 本文主要工作及内容安排

从现有的研究成果可以看出，虽然数字通信信号的调制识别算法有很多，但能够直接利用原始数据对调制信号进行识别的算法还比较少，大多都需要手动进行专家特征提取，进而训练机器学习模型进行调制识别。而在实际的非协作通信中，受环境的影响，我们学习的模型在不同的环境中分类性能差别很大，鲁棒性较差。现有的基于深度学习的调制识别算法，并没有对影响网络性能的因素进行分析，也没能有效利用过去几十年的研究成果。

本文提出了一种基于卷积自编码器和卷积神经网络框架（Convolutional Autoencoder - Convolutional Neural Network, CAE-CNN）的调制识别算法，并提出了相应框架下的网络训练方法；提出了一种将深度特征与传统特征进行融合的算法框架，并对特征融合的方式进行了研究；研究了网络的深度、网络底层模块结构等对调制识别性能的影响，分析了不同情况下造成网络识别性能变化的因素。

本文的具体内容安排如下：

第一章为绪论。首先对无线信号调制识别技术的研究背景进行了介绍，系统的阐述了现有的调制识别的算法，并对当前的研究现状进行了总结概括。

第二章首先介绍了调制信号的基本概念，然后分析了信道对于调制识别的影响；接下来，介绍了神经元、前馈神经网络、卷积神经网络、自编码器等深度学习的基本理论和基本网络架构；最后，介绍了最近较为流行的几种神经网络优化算法，并分析了各种算法的一些理论基础。

第三章提出了一种 CAE-CNN 调制识别的算法框架，并提出了相应的训练算法。我们分别利用监督方法和无监督方法将信号特征可视化，利用无监督的卷积自编码器可以复现输入信号，并获得原始信号的低维表示；通过有监督的 CNN 获取的信号低维表示，得到对于不同类别具有区分度的特征；利用 t-分布随机邻接嵌入（t-distributed stochastic neighbor embedding, t-SNE）算法，将信号的低维表示降到二维

的流型中，观察了信号的分布状况。我们融合卷积自编码器与卷积神经网络，提出了一种 CAE-CNN 算法框架和相应训练算法。

第四章提出了一种传统特征与深度特征融合的框架。首先概括了传统调制所用到的调制识别特征，然后阐述了体征融合的基本理论。接下来，将传统方法中常用的调制识别的特征与 CNN 获取的深度特征进行适应性的 *Batch – Normalization*，分别针对 Softmax、随机森林、深度神经网络等融合算法构建融合模型，并对不同的融合算法进行了仿真。

第五章从网络底层研究网络超参数对调制识别性能的影响，并从欠拟合与过拟合以及偏差与方差的角度来理解出现这些现象的原因。首先，构建了调制识别模型的偏差与方差模型，并阐述了过拟合与欠拟合的原因；然后，基于 CNN 网路框架研究了卷积核数目、大小以及卷基层深度对调制识别的影响。

第六章对本文的工作进行了总结，从深度框架、特征融合、网络超参数等三个点阐述了本文所做出的贡献。同时，对深度学习在调制识别领域的应用进行了展望，对以后打算开展的研究做了一定的概述。

第二章 无线信号调制识别以及深度学习理论

2.1 引言

在动态频谱接入 (Dynamic Spectrum Access, DSA) 中, 感知周围的发射终端, 以避免无线干扰并优化频谱分配, 是无线认知的重要组成部分。广播无线电、卫星信号、4G 无线信号、雷达用户以及附近其他潜在无线电干扰源等信号具有不同的调制形式和特征, 识别和区分这些信号是通信系统最基本的步骤, 因此, 必须针对信号的调制方式进行判别和解调。本章主要讨论信号调制的基本概念以及深度学习的基本理论。

2.2 调制识别

2.2.1 信道对调制信号的影响

无线信道模型是对无线信道的抽象描述, 它能很好地反映真实环境中的信号传输规律。无线通信数据信息主要以电磁波为载体通过无线信道传输。由于无线信道的环境复杂多变, 电波以不同的传输方式 (直射, 反射, 散射等) 到达接收点, 使得接收到的信号与发射的信号不同。因此, 只有准确预测无线信号的无线传播特性, 如路径损耗和相位延迟, 才能为无线网络提供合理的设计, 部署和管理策略。

信道效应具备不确定性, 在通信系统中是不可逆的。真实的通信系统在进行信号传输时会经历许多影响, 这给恢复和表示原始信号带来了很大难度。热噪声在接收器处产生相对平坦的高斯白噪声, 其形成信号的底噪。由于温度和半导体物理材料自身特性, 发射器和接收器的特性可能产生波动, 从而引起振荡器偏移导致符号时序偏移, 采样速率偏移, 载波频率偏移和相位差等。这些效应可能导致信道之间的时间移位, 缩放, 线性混合、旋转等效应, 给信息传输稳定性带来不利影响。最后, 根据在接收机处发射信号的到达模式, 信号经过实际信道可能会经历随机滤波, 产生幅度, 相位变化和多普勒频移。这就是通常所说的多径衰落或频率选择性衰落, 其主要发生在当信号的传播路线上出现建筑物、车辆等障碍物, 阻碍了信号的视距传播, 造成信号在空间中的反射, 发生时频特性的变化。

2.2.2 信道建模

无线信号的调制识别可以看作是一个 N 类的决策问题。其中, 输入是一个接收信号的复时间序列。也就是说, 我们以离散时间步长对无线电信号的同相和正交分量

进行采样，获得原始信号的 $2 \times N$ 的复数值向量。我们将接收信号用方程(2-1)表示：

$$r(t) = s(t) * c + n(t) \quad (2-1)$$

其中，调制信号 $s(t)$ 的生成，通过将连续信号或离散时间序列信号调制到具有变化的频率、相位、振幅、或多个变换的正弦波上得到。缩放因子 c 是指信号上的一些路径损耗或恒定增益项， $n(t)$ 是反映热噪声的加性高斯白噪声过程。在工程应用领域，这个简化的表达式在基于专家特征的决策统计方法中被广泛使用。

然而，实际的信道环境却比较复杂。发射信号 $s(t)$ 在传播过程中经历多个信道效应，最后在接收端被接收为 $r(t)$ 。这些信道效应包括：时间延迟，尺度缩放，相位旋转，频率偏移，加性热噪声，信道脉冲响应，以及所有的随机时变过程等。这些效应对信号的作用可以近似表示成方程 (2-2)：

$$r(t) = e^{j * n_{Lo}(t)} \int_{\tau=0}^{\tau_0} s(n_{Clk}(t - \tau)) h(\tau) + n_{Add}(t) \quad (2-2)$$

方程(2-2)考虑了许多对于模型来说很重要的现实世界的影响：通过残留载波随机游走过程调制 $n_{Lo}(t)$ ，通过残留时钟振荡器随机游走重采样 n_{Clk} ，与时变的旋转非恒定幅度脉冲响应 $h(t - \infty)$ 卷积，以及加性噪声 $n_{Add}(t)$ （可能不是白噪声）。每个都可能导致未知的时变误差。考虑到现实世界中存在的无线信道的影响时，会使我们的接收信号表示复杂化。

考虑到传播信道的复杂性，对专家特征提取并进行分类决策建模是很难的。这通常会迫使我们简化假设，构建易于处理的如方程(2-1)所描述的基本模型；然而，基本模型很难刻画复杂的信道特征，这样就造成了算法性能的上限较低，鲁棒性较差。在本文中，我们主要关注包括所有上述影响的模拟传播环境中的实测数据，利用数据反映信道本身的特征，而不是从理论上的进行信道建模指纹特征提取等。

2.3 神经网络概述

2.3.1 神经元概述

神经网络 (Neural Networks, NN)，是一种模仿人类大脑神经行为，进行分布式并行信息处理的算法数学模型，它起源于 Rosenblatt 等人在 1957 年提出的感知器模型^[34]。以监督学习为例，假设我们有训练样本集 $(x^{(i)}, y^{(i)})$ ，那么神经网络算法能够提供一种复杂且非线性的假设模型 $h_{W,b}(x)$ ，它具有参数 W, b ，可以以此参数来拟合我们的数据。图2.1即是“神经元”的图示：

此“神经元”是一个以 x_1, x_2, \dots, x_n 及截距 b 为输入值的运算单元，其输出为：

$$h_{W,b}(x) = f(W^T x) = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (2-3)$$

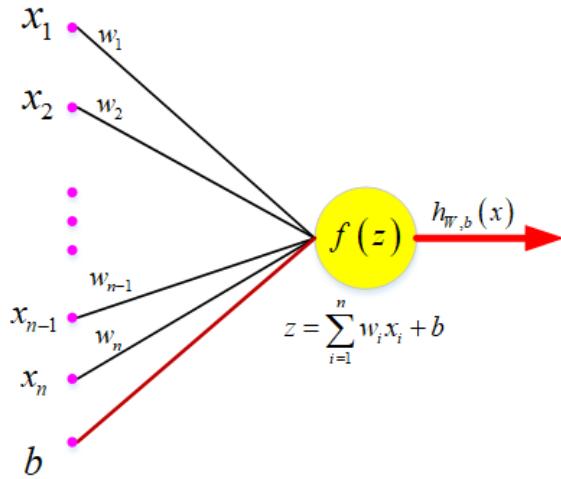


图 2.1 神经元示意图

其中函数 $f(z)$ 被称为“激活函数”，常用的激活函数有 ReLU、Sigmoid、tanh 等。在本论文中，激活函数 $f(z)$ 选取可以降低“梯度消失”与“梯度爆炸”效应的 ReLU 函数^{[35][7][36]}：

$$f(z) = \begin{cases} 1 & z \leq 0 \\ z & z > 0 \end{cases} \quad (2-4)$$

可以看出，ReLU 激活函数在输入值小于零时输出为 0，在大于 0 时为一条经过原点的直线，其相较于 Sigmoid 激活函数而言，具有如下优点：单侧抑制，具备较宽的兴奋边界，有利于梯度的传播并对网络引入稀疏性，缓和了梯度弥散问题，加快了网络的训练速度。

图2.2和图2.3分别是 ReLU 及 Sigmoid 的函数图像：

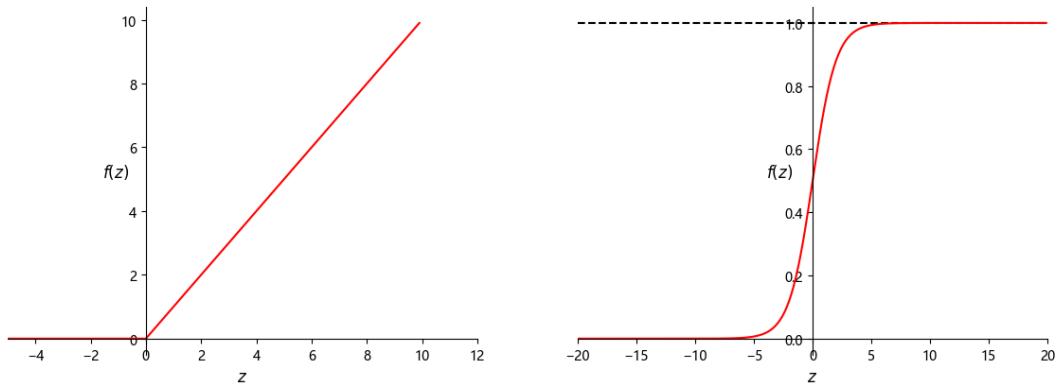


图 2.2 ReLU 函数

图 2.3 Sigmoid 函数

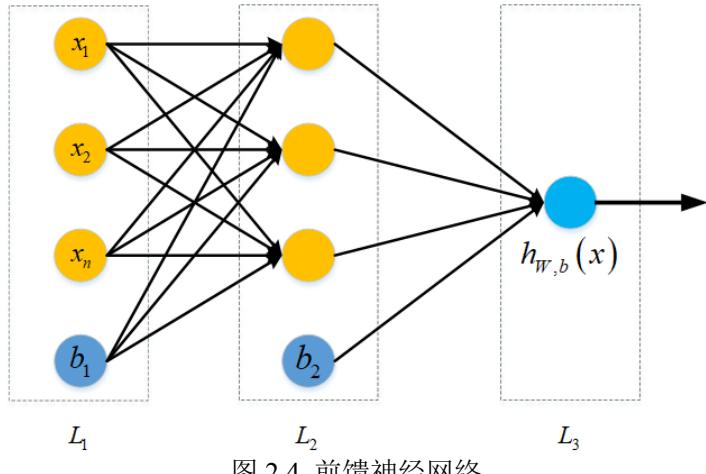


图 2.4 前馈神经网络

2.3.2 前馈神经网络

前馈神经网络，是通过联结多个单一的“神经元”，使得一个“神经元”的输出作为另一个“神经元”输入的网络系统。图2.4展示了一个简单的神经网络示意图。

假设用圆圈来表示神经网络的输入，偏置节点（截距项）就是标上“ b ”的圆圈。在神经网络中，最左边的一层作为输入层，输出层则是最右边的一层（本例采取只有一个节点的输出层）。中间层称为隐藏层，是由所有节点组成的。顾名思义，隐藏层就意味着它们的输出值无法在训练样本集中出现。并且，若不计算偏置单元，在该神经网络例子中，一共有 $n(n = 3)$ 个输入单元， $n(n = 3)$ 个隐藏单元，以及一个输出单元。

我们用 l_* 来表示网络的层数，本例中，我们将第 i 层记为 l_i ，于是 l_0 表示输入层， l_L 表示输出层。利用 $a_i^{(l)}$ 表示第 l 层第 i 单元的激活值（输出值）。当 $l = 1$ 时， $a_i^{(1)} = x_i$ ，即将输入值的第 i 个特征作为第 i 个输入值。对于给定参数集合 W, b ，利用所构建的神经网络的输出作为 $h_{W,b}(x)$ 的结果。

第 l 层第 i 单元输入加权和（含偏置单元）用 $z_i^{(l)}$ 表示。上面的计算步骤叫作前向传播。如果用 $a^{(1)} = x$ 表示输入层的激活值，那么给定第 l 层的激活值 $a^{(l)}$ 后，通过方程(2-5)的计算，可得第 $l + 1$ 层的激活值 $a^{(l+1)}$ 。

$$z^{(l+1)} = W^{(l)}a^{(l)} + b^{(l)} = \sum_{j=l}^n W_{ij}^{(l)}x_j + b_i^{(l)} \quad (2-5)$$

$$a^{(l+1)} = f(z^{(l+1)}) \quad (2-6)$$

线性代数对矩阵-向量的处理有一定的计算优势。将参数转为矩阵形式，然后利用这种优势精确并快速求解较复杂的神经网络。

上述讨论的是一种简单的、单隐层神经网络结构，接下来将构建另一种结构（神经元间的联接模式）更为复杂多样的神经网络，即包含多个隐层的深度神经网络。最

常见的一个例子是 n_l 层的神经网络，第 1 层是输入层，第 n_l 层是输出层，中间的每个层 l 与层 $l+1$ 紧密相联。这假如我们需要计算在该种模式下神经网络的输出结果，根据前文描述的原理和网络层之间的等式关系，一步一步前向传播，从而得到第 L_2 层第 L_{n_l} 层的激活值。通过这种方式，我们需要的神经网络就构建成功。同时，这种神经网络称为“前馈神经网络”，因为神经元之间没有形成闭环或者回路的连接。

2.3.3 反向传播算法

反向传播算法（Backpropagation, BP）是一种与最优化方法（如梯度下降法）结合使用的，用来训练人工神经网络的常见方法，最早由 Rumelhart、Hinton 等人正式提出^[37]。

假设我们有一个固定的样本集 $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ ，它包含 m 个样本了，可以利用批量梯度下降算法求解该网络。具体来说，对于单个样本 (x, y) ，我们假设其损失函数 $J(W, b; x, y)$ 为方差代价函数，则有该样本的损失函数 $J(W, b; x, y)$ ：

$$J(W, b; x, y) = \frac{1}{2} \|h_{W,b}(x) - y\|^2. \quad (2-7)$$

如果训练集包含 m 个样本，那么所有样本的整体损失函数 $J(W, b)$ 为：

$$\begin{aligned} J(W, b) &= \left[\frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}, y^{(i)}) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \end{aligned} \quad (2-8)$$

上述定义中，第一项是均方差项，第二项是权重衰减项（即正则化项），它的存在可以使过拟合以及权重绝对值均减小， λ （即权重衰减系数）是这两项的平衡因子。上述所介绍的代价函数在分类和回归问题中经常用到。在二分类问题中，如果我们使用 Sigmoid 函数作为激活函数，由于 Sigmoid 激活函数的值域为 $[0, 1]$ ，那么可以使用 $y = 0$ 或 $y = 1$ ，来代表两种类别的标签；如果我们使用双曲正切激活函数 \tanh ，那么可以选用 -1 和 $+1$ 作为类别标签。

反向传播算法的思路如下：假设样例 (x, y) ，我们先通过“前向传播”的运算方法计算出网络中所有的激活值，其中包括 $h_{W,b}(x)$ 的输出值。之后，针对第 l 层的每一个节点 i ，我们计算出其“残差” $\delta_i^{(l)}$ ，此残差表明了该节点对最终输出值的误差产生了多少影响。对于最终的输出节点，我们可以直接算出网络产生的激活值与实际值之间的差距，我们将这个差距定义为 $\delta_i^{(n_l)}$ （第 n_l 层表示输出层）。对于隐藏单元我们如何处理呢？我们将基于节点残差的加权平均值计算 $\delta_i^{(l)}$ ，这些节点以 $a_i^{(l)}$ 作为输入。以下几个步骤清楚地表述了反向传播算法：

进行前馈传导计算，利用前向传导公式，得到 L_2, L_3, \dots 直到输出层 L_{n_l} 的激活值。对输出层（第 n_l 层），计算：

$$\delta^{(n_l)} = -(y - a^{(n_l)}) \bullet f'(z^{(n_l)}) \quad (2-9)$$

对于 $l = n_l - 1, n_l - 2, n_l - 3, \dots, 2$ 的各层，计算：

$$\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) \bullet f'(z^{(l)}) \quad (2-10)$$

计算最终需要的偏导数值：

$$\nabla_{W^{(l)}} J(W, b; x, y) = \delta^{(l+1)} (a^{(l)})^T, \quad (2-11)$$

$$\nabla_{b^{(l)}} J(W, b; x, y) = \delta^{(l+1)}. \quad (2-12)$$

实现中应注意：在以上的第 2 步和第 3 步中，我们需要为每一个 i 值计算其 $f'(z_i^{(l)})$ 。前提是 $f(z)$ 是 sigmoid 函数，同时 $a_i^{(l)}$ 已经通过前向传导运算中得到。最后，我们就梯度下降算法的使用进行较全面的总结。在下面的伪代码中， $\Delta W^{(l)}$ 是一个与矩阵 $W^{(l)}$ 维度相同的矩阵， $\Delta b^{(l)}$ 是一个与 $b^{(l)}$ 维度相同的向量。注意这里“ $\Delta W^{(l)}$ ”是一个矩阵，而不是“ Δ 与 $W^{(l)}$ 相乘”。批量梯度下降法中的一次迭代可以通过以下步骤完成：

对于所有 l ，令 $\Delta W^{(l)} := 0, \Delta b^{(l)} := 0$ （设置为全零矩阵或全零向量）对于 $i (i \in \{1, 2, \dots, m\})$ ，使用反向传播算法计算 $\nabla_{W^{(l)}} J(W, b; x, y)$ 和 $\nabla_{b^{(l)}} J(W, b; x, y)$ 。计算 $\Delta W^{(l)} := \Delta W^{(l)} + \nabla_{W^{(l)}} J(W, b; x, y)$ 。计算 $\Delta b^{(l)} := \Delta b^{(l)} + \nabla_{b^{(l)}} J(W, b; x, y)$ 。更新权重参数：

$$W^{(l)} = W^{(l)} - \alpha \left[\left(\frac{1}{m} \Delta W^{(l)} \right) + \lambda W^{(l)} \right] \quad (2-13)$$

$$b^{(l)} = b^{(l)} - \alpha \left[\frac{1}{m} \Delta b^{(l)} \right] \quad (2-14)$$

接下来，我们可以对梯度下降法算法不断迭代，减小损失函数 $J(W, b)$ 的值，直到达到一定阈值或者终止条件，进而求解神经网络。

2.4 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN），是一种专门用来处理具有类似网格结构数据的神经网络，最早由 LeCun 在 1989 年提出^[38]。卷积神经网络至少在网络的某一层中使用卷积运算来替代一般的矩阵乘法运算，在时间序列数据（可以认为是在时间轴上有规律地采样形成的一维网格）和图像数据（可以看作是二维的像素网格）等领域取得了较好的成果。

2.4.1 卷积运算

卷积是一种对两个实变函数的线性运算，通常用星号表示：

$$s(t) = (x * w)(t) = \int x(\tau)w(t - \tau)d\tau \quad (2-15)$$

在卷积网络的术语中，卷积的第一个参数（在这个例子中，函数 x ）通常叫做输入，第二个参数（函数 w ）叫核函数。输出 $(x * w)$ 有时被称作特征映射。如果我们假设 x 和 w 都定义在整数时刻 t 上，则卷积的离散形式：

$$s(t) = (x * w)(t) = \sum_{\tau=-\infty}^{\infty} x(\tau)w(t\tau). \quad (2-16)$$

在机器学习应用中，输入通常是由样本的多维参数值组成的高维数组，而核通常是由优化算法对模型参数进行学习得到的多维数组。通常我们称之为做张量。因为在输入与核中的每一个元素都必须明确地分开存储，所以通常情况下假设除了那些存储数值的有限点集以外的其他值都置为零。这样，我们在实际操作中要想得到卷积中的无限求和，就可以通过求和有限个数组元素来实现。在处理图像数据时，我们经常一次在多个维度上进行卷积运算。如果把一张二维的图像 I 作为输入，同时使用一个二维的核 K ，则有：

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n). \quad (2-17)$$

卷积是可交换的，我们可以等价地写作：

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n). \quad (2-18)$$

由于我们将核进行了翻转，所以出现了卷积运算的可交换性。从 m 增大的角度来看，输入的索引在增大，但是核的索引在减小。我们将核翻转的唯一目的是实现可交换性。

2.4.2 卷积特性

(1) 卷积是一种无限强的先验

先验被认为是强或者弱取决于先验中概率密度的集中程度。弱先验具有较高的熵值，例如方差很大的高斯分布。这样的先验允许数据对于参数的改变具有或多或少的自由性。强先验具有较低的熵值，例如方差很小的高斯分布。这样的先验在决定参数最终取值时起着更加积极的作用。一个无限强的先验需要对一些参数的概率置零并且完全禁止对这些参数赋值，无论数据对于这些参数的值给出了多大的支持。

我们可以把卷积网络类比成全连接网络，但对于这个全连接网络的权重有一个无限强的先验。这个无限强的先验是说一个隐藏单元的权重必须和它邻居的权重相同，但可以在空间上移动。这个先验也要求除了那些处在隐藏单元的小的空间连续的接受域内的权重以外，其余的权重都为零。

总之，我们可以把卷积的使用当作是对网络中一层的参数引入了一个无限强的先验概率分布。这个先验说明了该层应该学得的函数只包含局部连接关系并且对平移具有等变性。

(2) 卷积性质

卷积运算通过三个重要的思想来帮助改进机器学习系统：稀疏交互、参数共享、等变表示。传统的神经网络使用矩阵乘法来建立输入与输出的连接关系。对于卷积，参数共享的特殊形式使得神经网络层具有对平移等变的性质。如果一个函数满足输入改变，输出也以同样的方式改变这一性质，我们就说它是等变的。特别地，如果函数 $f(x)$ 与 $g(x)$ 满足 $f(g(x)) = g(f(x))$ ，我们就说 $f(x)$ 对于变换 g 具有等变性。对于卷积来说，如果令 g 是输入的任意平移函数，那么卷积函数对于 g 具有等变性。当处理时间序列数据时，这意味着通过卷积可以得到一个由输入中出现不同特征的时刻所组成的时间轴。如果我们把输入中的一个事件向后延时，在输出中仍然会有完全相同的表示，只是时间延后了。而这也正好对应到我们无线信号中的时移不变性。因此，我们可以很好的将??应用到调制信号识别中。

(3) 卷积特征学习

一般而言，卷积神经网络的训练中最耗时的是特征学习。通常输出层的计算代价相对不高，因为在通过若干层卷积池化之后，输入到该层的特征数量较小。当使用梯度下降执行有监督训练时，每一次进行梯度计算需要完整地运行整个网络的前向传播和反向传播，以使残差从输出层反馈到输入层。

利用那些无监督方式训练得到的特征是一种常用的减少卷积网络训练成本的方法。

通常，可以不通过监督训练而得到卷积核的基本策略主要有三类。第一类是简单地随机初始化它们；第二类是人工干预，比如设置卷积核的特定的方向和尺度来进行边缘检测；最后，可以使用无监督的方法来学习卷积核。

使用无监督的标准来学习特征，我们可以认为是将 CNN 分成卷积层与分类层两部分，网络允许我们将卷积层的特征参数确定与随后的分类层的参数学习相分离。这样，便可以只需提取一次全部训练集的特征，作为随后分类层的新的训练集。假设最后一层类似逻辑回归（Logistic Regression, LR）或者支持向量机，那么分类层的学习通常是一种凸优化问题。

随机卷积核经常在卷积神经网络中表现得很好。由 Saxe 的工作可知，如果我们

对由卷积和随后的池化组成的层赋予随机权重，网络就会具备平移不变性和频率选择性^[39]。他们认为这提供了一种低时间成本的方法来选择卷积网络的结构：首先通过仅训练最后一层来评估几个卷积网络结构的性能，然后选择最好的结构并使用相对耗时的方法来训练整个网络。

这种选择卷积神经网络结构的方法，在作者看来时间成本相对较低：首先，通过单纯地训练最后一层的方式来评估几个卷积网络结构的性能；然后，比较这些网络结构并选取最优的网络，进行整个网络的训练。

2.5 卷积自编码器

自动编码器（autoencoder, AE）是一种无监督的学习算法，由 Rumelhart 在 1986 年第一次提出^[37]。其优化目标是通过一些较低的中间隐层维度，使用均方误差（Mean Squared Error, MSE）等损失函数，最小化网络输出的重构误差。它的目标是一种尽可能还原原始输入信号，主要用于数据的降维或者特征的提取。为了还原输入信号，编码后的数据必须保留原始数据的主要特征。如图2.5所示，自编码器主要由两部分组成：一个数据维度变换的编码器，以及一个生成重构信号的解码器。它内部还有一个隐藏层，可以产生原始数据编码后的低维表示，本质上都是对输入信号做某种非线性变换。

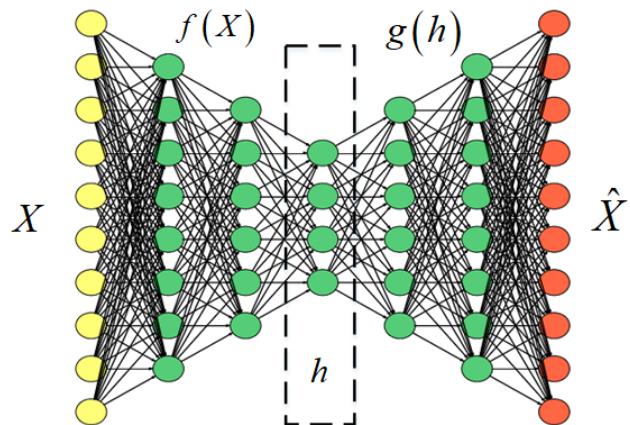


图 2.5 自编码器

图2.5中，将原始的训练数据集样本作为自编码器的输入，从输入层 X 到隐层 h 再到输出层 \hat{X} 的网络就是自编码器模型。编码器将输入信号 X 变换成编码信号 h ，

而解码器将编码信号 h 转换成输出信号 \hat{X} 。即：

$$\begin{aligned} h &= f(X) \\ \hat{X} &= g(h) = g(f(X)) \end{aligned} \quad (2-19)$$

通常，自编码器利用反向传播算法，将误差进行反向传播，并使用随机梯度下降算法等，以找到接近等式(2-20)中的最佳网络参数 $\hat{\theta}$ 。

$$\hat{\theta} = \arg \min_{\theta} \sum (X - \hat{X})^2 \quad (2-20)$$

卷积自编码器（Convolutional Auto-Encode, CAE）是自编码器中的一种，主要是在编码器中加入了卷积层，使得网络可以学习到数据的卷积相关特征。编码器主要是由卷积层和池化层交叉组成的神经网络。卷积的作用相当于一个滤波器，而池化则是提取不变特征。CNN 和 CAE 之间最主要的区别在于前者是进行有监督的卷积核学习，并且将提取的特征进行组合从而用来分类。事实上，CNN 通常被称为是一种监督学习；而 CAE 则通常被用来训练从输入数据中提取特征，从而重构输入数据。

CAE 中的卷积层具有时移不变性以及受约束的参数搜索空间（相对于全连接层），因此非常适合于无线电时间序列信号表示。我们使用 dropout 并在输入层加入噪声 [7] 对网络进行正则化，来增强模型的泛化能力。

2.6 神经网络优化算法

在深度学习领域中，有很多现有的参数优化算法，Ruder 对常用的一阶和二阶、自适应与固定速率等学习算法进行了理论的阐述和比较^[40]。在大多数机器学习问题中，我们主要关注任务的性能度量 P ，它是定义在测试集上对于模型性能的一个衡量标准。但是，有时我们的性能 P 并不能达到，或者问题本身无解，这时，我们只是间接地优化 P 。我们希望通过降低代价函数 $J(\theta)$ 来提高 P 。这一点与纯粹的优化不同，纯优化最小化目标 $J(\theta)$ 本身。训练深度模型的优化算法通常也会包括一些针对机器学习目标函数的特定结构进行的优化。即寻找神经网络上的一组参数 θ ，它能显著地降低代价函数 $J(\theta)$ ，该代价函数通常包括整个训练集上的性能评估和额外的正则化项。

2.6.1 统计梯度下降优化算法

统计梯度下降（Stochastic Gradient Descent, SGD）及其变种是机器学习尤其是深度学习中最常用的一种优化算法^[41]。它通过抽取数据的 m 个小批量（独立同分布的）样本，然后计算它们梯度均值，获得梯度的无偏估计。算法2.1展示了 SGD 算法的迭代更新的方法。

算法 2.1 随机梯度下降 (SGD) 在第 k 个训练迭代的更新

Require: 学习速率 ϵ_k **Require:** 初始参数 θ **while** 停止准则未满足 **do**从训练集中采包含 m 个样本 $\{x^{(1)}, \dots, x^{(m)}\}$ 的小批量，其中 $x^{(i)}$ 对应目标为 $y^{(i)}$ 计算梯度估计： $\hat{g} \leftarrow +\frac{1}{m} \nabla_{\theta} \sum_i L(f(x^{(i)}; \theta), y^{(i)})$ 应用更新： $\theta \leftarrow \theta - \epsilon \hat{g}$ **end while**

学习速率在 SGD 算法中是一个至关重要的参数。之前，我们介绍的 SGD 使用固定的学习速率。在实际应用中，有时随着学习的深入需要降低学习速率，因此我们将第 k 步迭代的学习速率记作 ϵ_k 。

这是因为 SGD 中梯度估计引入的噪声源 (m 个训练样本的随机采样) 并不会在极小点处消失。相比之下，当我们使用批量梯度下降到达极小点时，整个代价函数的真实梯度会变得很小，之后为 **0**，因此批量梯度下降可以使用固定的学习速率。保证 SGD 收敛的一个充分条件是

$$\begin{aligned} \sum_{k=1}^{\infty} \epsilon_k &= \infty \\ \sum_{k=1}^{\infty} \epsilon_k^2 &< \infty \end{aligned} \tag{2-21}$$

在实际应用中，一般会设置阈值 τ 。学习速率将持续衰减，直到第 τ 次迭代。学习速率衰减的形式如公式(2-22)所示。

$$\epsilon_k = (1 - \alpha)\epsilon_0 + \alpha\epsilon_{\tau} \tag{2-22}$$

其中 $\alpha = \frac{k}{\tau}$ ，表示衰减因子。一般在 τ 步迭代之后使 ϵ 保持常数，使优化算法以较小的恒定速率学习模型参数。

2.6.2 自适应矩估计优化算法

自适应矩估计 (Adaptive Moment Estimation, Adam)，是一种自适应学习率的算法^[42]。Adam 基于低阶矩的自适应估计，是一种基于一阶梯度的随机目标函数优化算法。该方法实施起来很简单，计算效率高，存储需求小，善于处理稀疏梯度，并且非常适合于数据和参数较大时的参数优化，网络超参数通常不需要进行手动调整。虽然有时学习速率需要手动修改默认值，但它对超参数的选择具有较强的鲁棒性。Adam 的整体流程如算法2.2所示。在算法2.2中，动量直接并入了梯度一阶矩 (指数加权) 的估计。将动量加入 RMSProp 最直观的方法是将动量应用于缩放后的梯度。其次，Adam 包括偏置修正，修正从原点初始化的一阶矩 (动量项) 和 (非中心的) 二阶矩的估计。

算法 2.2 Adam 算法

Require: 步长 ϵ (建议默认为: 0.001)**Require:** 矩估计的指数衰减速率, ρ_1 和 ρ_2 在区间 [0, 1) 内。 (建议默认为: 分别为 0.9 和 0.999)**Require:** 用于数值稳定的小常数 δ (建议默认为: 10^{-8})**Require:** 初始参数 θ 初始化一阶和二阶矩变量 $s = 0, r = 0$ 初始化时间步 $t = 0$ **while** 没有达到停止准则 **do**从训练集中采包含 m 个样本 $\{x^{(1)}, \dots, x^{(m)}\}$ 的小批量, 对应目标为 $y^{(i)}$ 。计算梯度: $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x^{(i)}; \theta), y^{(i)})$ $t \leftarrow t + 1$ 更新有偏一阶矩估计: $s \leftarrow \rho_1 s + (1 - \rho_1)g$ 更新有偏二阶矩估计: $r \leftarrow \rho_2 r + (1 - \rho_2)g \odot g$ 修正一阶矩的偏差: $\hat{s} \leftarrow \frac{s}{1 - \rho_1^t}$ 修正二阶矩的偏差: $\hat{r} \leftarrow \frac{r}{1 - \rho_2^t}$ 计算更新: $\Delta\theta = -\epsilon \frac{\hat{s}}{\sqrt{\hat{r}} + \delta}$ (逐元素应用操作)应用更新: $\theta \leftarrow \theta + \Delta\theta$ **end while**

2.6.3 RMSProp 优化算法

RMSprop 是一个广泛应用的却未被发表的自适应学习率的算法, 该算法由 Geoff Hinton 在其 Coursera 课堂的课程 6e 中提出^[43]。RMSProp 将学习率分解成一个平方梯度的指数衰减的平均, 利用衰减系数来控制历史信息的获取多少, 以使其能够在找到凸碗状结构后快速收敛, 适合处理非平稳目标。

RMSProp 算法的标准形式如2.3所示。

算法 2.3 RMSProp 算法

Require: 全局学习速率 ϵ , 衰减速率 ρ **Require:** 初始参数 θ **Require:** 小常数 δ , 通常设为 10^{-6} (用于被小数除时的数值稳定)初始化累积变量 $r = 0$ **while** 没有达到停止准则 **do**从训练集中采包含 m 个样本 $\{x^{(1)}, \dots, x^{(m)}\}$ 的小批量, 对应目标为 $y^{(i)}$ 。计算梯度: $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x^{(i)}; \theta), y^{(i)})$ 累积平方梯度: $r \leftarrow \rho r + (1 - \rho)g \odot g$ 计算参数更新: $\Delta\theta = -\frac{\epsilon}{\sqrt{\delta+r}} \odot g$ ($\frac{1}{\sqrt{\delta+r}}$ 逐元素应用)应用更新: $\theta \leftarrow \theta + \Delta\theta$ **end while**

2.6.4 选择正确的优化算法

在本节中，我们讨论了一系列算法，通过自适应每个模型参数的学习速率以解决优化深度学习中的难题。然而，对于哪一种优化算法具有最优的性能，目前仍没有确定。Schaul 等人对现有的一些神经网络优化算法在几个学习任务上进行了比较^[44]。结果显示，一般情况下，自适应学习速率的算法（如以 RMSProp 和 AdaDelta 等）具备较强的鲁棒性；在各个任务上，每种优化算法所取得的性能相近，没有某个确定的算法能绝对地优于其他算法。

目前，使用率较高的优化算法主要有 SGD、包含动量项的 SGD、基于动量的 RMSProp、AdaDelta 和 Adam 等。此时，选择哪一个算法则主要取决于使用者的喜好以及对不同优化算法的熟悉程度（以便调节超参数）。本文中，根据不同的条件，综合使用以上几种优化算法，以期能够达到性能的最优。

2.7 本章小结

本章对于论文中所用到的相关理论知识进行了较为系统的阐述，主要介绍了信号调制及深度学习的理论基础。我们首先介绍了信道对于调制识别的影响，并基于数据利用网络自学习信道的影响，避免了直接对于信道建模的误差影响；接下来，介绍了神经元、前馈神经网络、卷积神经网络、自编码器等深度学习的基本理论和基本网络架构；最后，介绍了最近较为流行的几种神经网络优化算法，并分析了各种算法的理论基础。

第三章 基于 CAE-CNN 的无线信号调制识别

3.1 引言

无线通信领域的研究人员已经开始将深度神经网络应用于认知无线电，并取得了一定的成果^{[45][46][47]}。Oshea 最近证明了利用原始数据进行有监督无线调制识别的可行性，作者利用原始信号经希尔伯特变换后得到的 I 与 Q 路信号作为训练样本，调制方式作为标签，训练 CNN 分类器。结果显示，其分类性能超越了传统的基于专家特征的决策树、SVM 等分类模型。然而，作者仅仅是用了传统的 CNN 框架，并没有对分类性能以及网络框架进行进一步的研究。

本章针对无线信号调制识别问题，提出了一种基于 CAE 与 CNN 融合的无线信号调制识别的算法框架，并将此框架下的识别准确率及鲁棒性等与传统的基于特征的识别方法进行比较分析。

3.2 调制信号

无线通信信号实际上是经过调制信号与信道综合作用生成的。尽管在机器学习中我们一般会建议使用真实数据，但是在无线电通信领域中，由于标记数据匮乏，而且在现实中采集真实可靠的数据，是一种相当费时费力的工作，并且也很难得到监管部门的许可，因此在我们的应用场景中，使用仿真数据进行实验较为普遍。本文所用基础数据，经过纯粹的软件仿真，利用基础硬件进行信号生成与采集，相较于经过纯粹的软件仿真所得数据，具备较高的可靠性。

3.2.1 调制信号生成

我们利用 SMW200A 信号生成器，构建通信系统框架，获取近似真实的仿真信号。我们以与真实系统完全相同的方式确定性地引入调制，脉冲整形，携带数据以及与现实通信系统相同的发射参数。我们将真实的语音和文本数据集调制到信号上，这样，接收信号不仅是一系列的确知信号，并且包含了信息，使我们的调制信号更接近真实环境中的信号。

信号源我们使用信号生成器生成随机信号，SMW200A 信号生成器自带信道模型，这包括许多所需的信道，例如高斯信道，频率选择性信道、多径信道、瑞利信道等。由于高斯信道统计特性较为确定，下文当中使用的数据，我们是利用高斯信道生成的数据。但是，基于多径信道仿真的数据我们已经获取，后续将会利用更接近真实

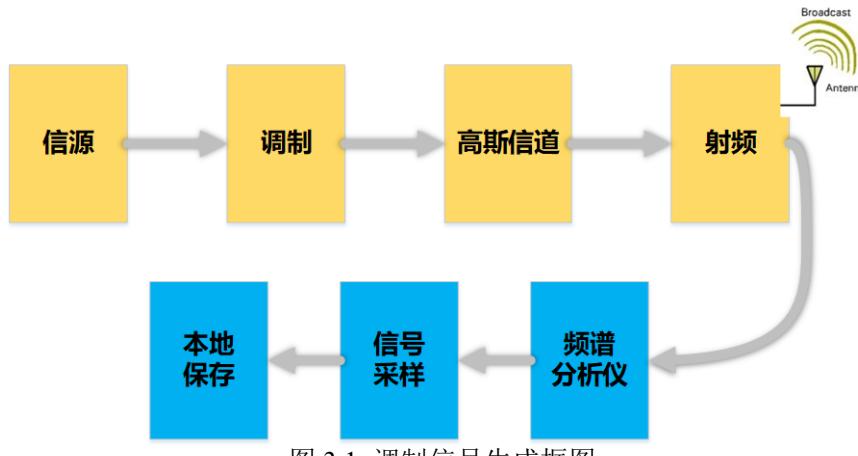


图 3.1 调制信号生成框图

的数据继续进行相应的研究。

整个调制信号生成的系统框图如图3.1所示。

我们生成的数据集主要由 7 种调制方式组成：其中包括 6 类数字调制和 1 类模拟调制。这些调制方式都被广泛应用于我们周围的无线通信系统中，对于调制识别而言具备一定的代表性。这些调制类别为 AM-SSB, BPSK, CPFSK, GFSK, PAM4, QAM64, QPSK。

SMW200A 信号发生器的需要调节的参数很多，在此我们列出了高斯信道条件下主要设置的部分系统参数，如表3.1所示。

表 3.1 SMW200A 参数

Baseband	symbol Rate	5M Sym/s
	Trigger In	Auto
	Clock	Internal
	Data Source	RRBS
	SNR	-20dB ~18dB
	Modulation	AM-SSB, BPSK, CPFSK, GFSK, PAM4, QAM64, QPSK
	Filter	Gauss, Multi-Path, Rayleigh roll off factor=0.3
AWGN A	System Bandwidth	10MHz
	Mode	Additive Noise
Others	IQ Mod	open
	RF A	open
	Carrier	1GHz
	Level	(-20 ~10db) = -10dBm (-9 ~5db) = -5dBm (-4 ~18db) = 0dBm

3.2.2 信号采集

在射频接收端，我们使用 FSW50 频谱分析仪接收经过多径信道后的样本，并查看其星座图和功率谱密度确定采样的准确性，最后将 IQ 两路采样数据保存到硬盘中。并利用频谱分析仪对采样数据保存到本地，用于后续的调制识别研究。FSW50 频谱分析仪的系统参数如表3.2所示：

表 3.2 SMW200A 参数

IQ Mod	Sample Rate	40 MHz
	Number of samples	40000000
	Duration of signal	1 s
	Data format complex	float32
	Scaling factor	1 V
	Center frequency of I/Q capture	1 GHz
	Modulation	Adaptive to Signal Source
	Data Format	.tar

我们的符号率为 5M/s，采样率大约为 40M/s，则每一个样本的持续时间约为 25μs，每个样本大约包含 125 个符号，它们含有经过信道带来的随机时间偏移，缩放，旋转，相位和噪声等影响。

数据以大约每个符号 8 次采样的速率进行采样，数据源的平均发送功率为 0dB。

因此，我们使用 1000 个采样点的滑动窗口，在采样信号序列上滑动，每次移动 1000 个采样点，来获取训练样本。最后我们将数据样本，以 32 位浮点数的复数形式保存成 mat 文件，总数据集大约 30GB。

3.3 调制信号的表示

不同的调制信号具有不同的时频特征。本节，我们将原始数据可视化，了解不同信号的时频特征；同时，我们利用 CAE 以及 CNN 获取信号的无监督表示，并展示不同信号在 CAE 特征空间中的分布状况，从而进一步了解不同网络框架对调制信号进行特征提取的不同状况。

3.3.1 数据集可视化

对于每一种调制方式，我们随机抽出一个样本，并对其时域(图3.1)和频域(图3.2)进行展示。我们可以发现，不同调制方式之间具备许多相似性，同时也具备一定差异性。有些信号我们是可以通过肉眼进行模糊判别；但是，受脉冲形变，失真和其他信道影响，有些信号即便是人类专家也很难从视觉上分辨属于何种调制类别。

如图(图3.2)所示，在时域中，我们可以看到 BPSK 等信号具备较明显的特征，而

AM-SSB 等信号在视觉上让人感觉像是噪声，很难直接判断出来。

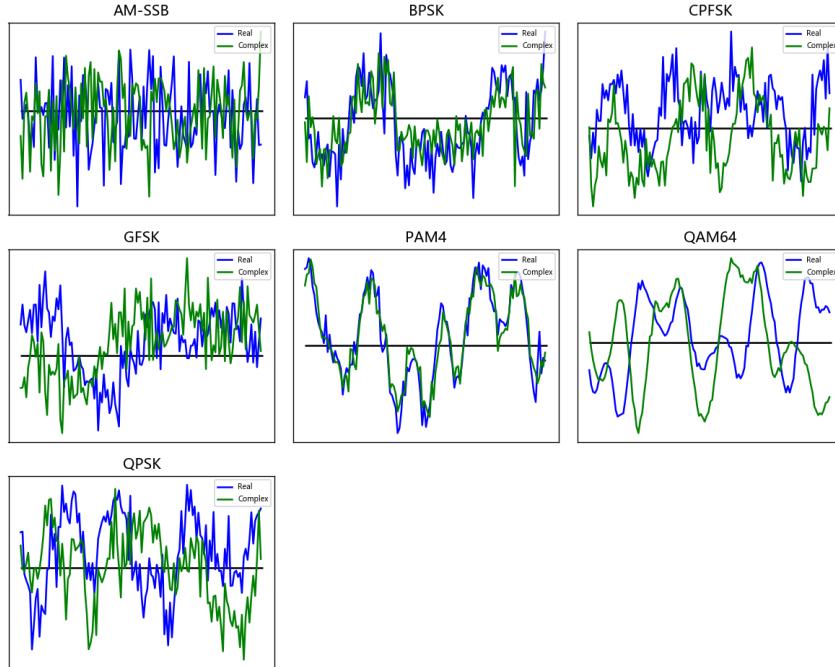


图 3.2 不同调制方式的高 SNR 样本的时域波形

如图3.3所示，在频域中，每一个信号都具备一个带宽限制的功率包络，其形状为调制识别提供了一定的信息，但是对于人类专家来说，从视觉来区分不同的调制信号是一个困难且繁琐的判定方法。

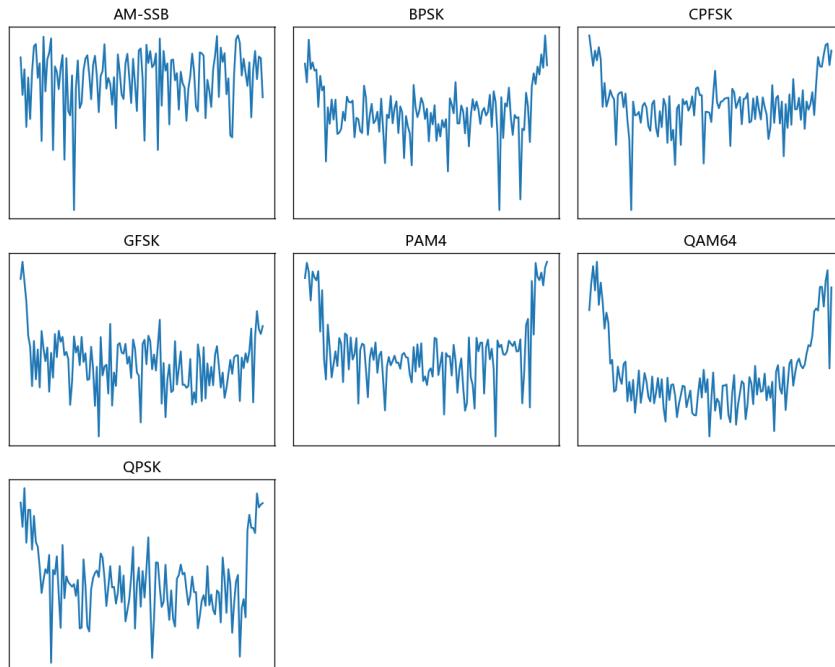


图 3.3 不同调制方式的高 SNR 样本的频谱

3.3.2 调制信号的无监督表示

无监督的稀疏表示是指在没有使用类标签的情况下，利用无监督的方法来学习数据集的稀疏表示。这可以通过使用基于数据依赖性的降维技术来完成，例如主成分分析（Principal components analysis, PCA）或独立分量分析（Independent components analysis, 缩写：ICA）。但是，这些方法只能对数据进行线性降维，如果说处于低维流型的数据与原始空间本身不具备线性关系，那么这种情况下就不适合使用线性降维方法。

CAE 非常适合于减小参数空间，获取的卷积特征具有时移不变性。在 CAE 的训练过程中，我们尽量减少信号重构的 MSE，但由于我们的主要目标是获得原始信号的聚类稀疏表示，因此我们对重构误差作出简化假设：限制重构误差最小的情况下尽量降低隐藏层的维度。然而，由于很难确定重构误差的最小值，所以，我们只能人为的指定隐层的维度来确定我们的稀疏表示维度，在维度确定的情况下调整参数使重构误差尽量小。本文中，我们利用卷积自编码器对输入的信号进行重构，学习一组原始信号的非线性稀疏表示。

自动编码器是一种无监督的学习算法，其中神经网络的优化目标是通过一些更有约束的中间维度，使用 MSE 等损失函数，最小化输出处的重构误差。通常，自编码器利用反向传播算法，将误差进行反向传播，并使用 SGD 算法等，以找到接近等式(3-1)中的最佳网络参数。

$$\hat{\theta} = \arg \min_{\theta} (\sum (X - f(X, \theta))^2) \quad (3-1)$$

通过约束网络的中间层维度，从而可以通过提取用于聚类的中间稀疏编码，来获得原始数据的非线性降维。在这种情况下，使用相似的调制信号，可以由相似的卷积核和特征图来表示，因此，他们分布在该压缩空间的相近区域中。自编码器中的卷积层具有时移不变性以及受约束的参数搜索空间（相对于全连接层），因此非常适合于无线电时间序列信号表示。我们使用 dropout 等方法，并在输入层加入噪声对网络进行正则化，来增强模型的泛化能力。图3.4显示了卷积自动编码器的体系结构。

我们使用 RMSProp 和 Adam 梯度下降求解器进行优化，两者都获得了相似的结果，下文中我们默认使用具备自适应学习速率的 Adam 优化器进行训练。

图3.5显示了两个输入维度为 2×128 的训练样本，中间隐层维度为 1×30 ，以及输出维度为 2×128 的训练结果展示图。

通过图3.5，我们可以发现，卷积自编码器可以很好的复现原始信号；我们提取的特征可以很好地重构原始数据，即可以很好地表征原始数据。这说明我们的卷积自编码器可以通过无监督的方式学习信号的低维嵌入表示。

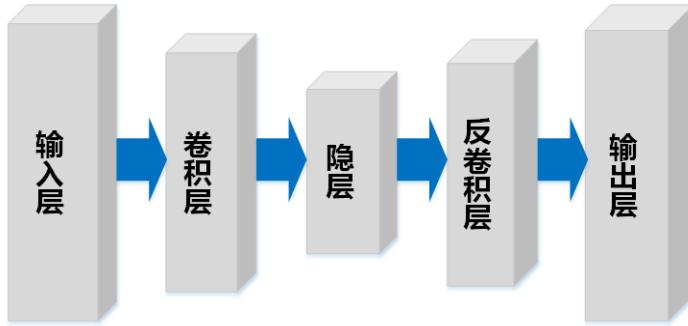


图 3.4 自编码器

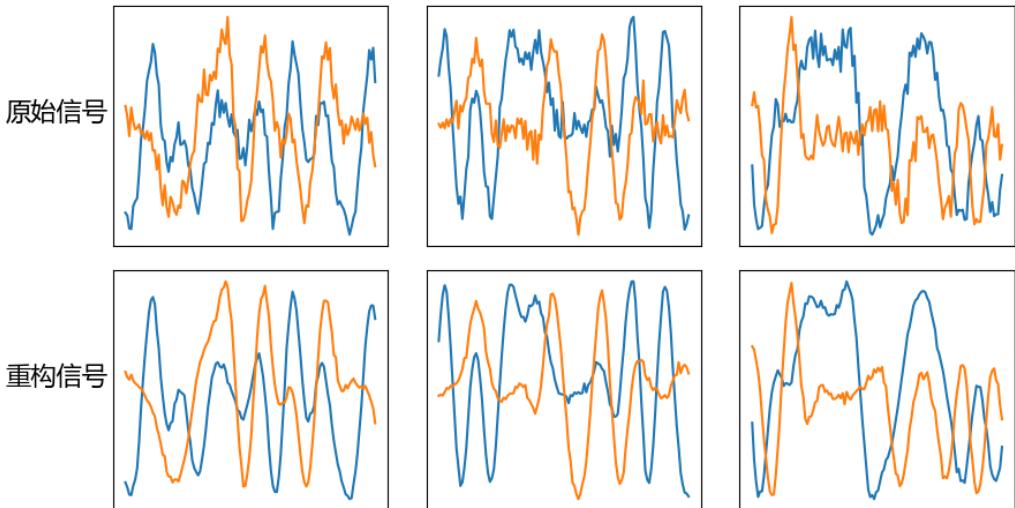


图 3.5 基于自编码器的信号复现

为了可视化我们学到的卷及特征，并对这些特征的类可分性进行直观展示，我们将数据的低维嵌入特征，利用 t-SNE 算法映射到二维流型，并在平面坐标系中展示。在低维嵌入空间中分布在相近的区域的样本，分布在二维流型中相近的区域。因此，我们可以通过观察不同类别的数据样本在 t-SNE 可视化之后的二维流型上的分布，来反映样本无监督表示的类可分性。

我们训练卷积自编码器从每一类样本中随机采样 100 个样本，将其通过训练的卷积自编码器获得其低维无监督表示，并利用 t-SNE 映射到二维流型，最终的效果如图3.6。

在这种情况下，我们看到几个类如 QAM64, AM-SSB, PAM4 和 BPSK 已经形成了独立的、大部分可分的簇，可以利用 DBSCAN 等聚类方法聚类为单独的类；而其他类则会出现类别混淆，并且难以通过聚类方法分离类别簇。尽管我们的无监督表示类可分性效果不太好，但考虑到这些特征从来没有被训练用来区分不同类别的样本，我们就已经获得了数据一定程度的类可分性，这已经算是一个可以接受的结果了。

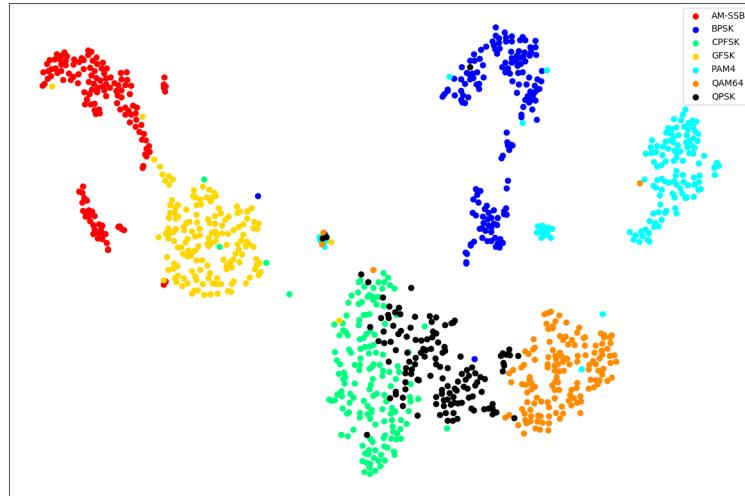


图 3.6 基于自编码器的调制信号 t-SNE 二维流型展示

3.3.3 调制信号的监督引导稀疏表示

当我们具有一部分监督数据的时候，我们也可以使用监督训练时学习到的判别特征生成一个稀疏表示空间。OShea 在他的工作中，利用有标记样本以监督方式训练卷积神经网络，可以达到很好的分类效果。CNN 主要是由卷基层与 DNN 层构成；由于 CNN 本身可以对测试样本进行分类，这就相当于在进行 Softmax 层的分类之前，我们已经获取了原始数据具有类别区分度的特征。因此，我们可以利用监督的方式，获取原始数据的监督引导特征。

在训练好分类网络以后，我们移除最后的 softmax 层，保留剩余的这一部分网络。这样，在样本经过训练好的网络，最后隐层输出的特征即为原始数据的稀疏表示。我们利用监督方式训练网络，并获取监督引导特征空间，获取数据监督引导的稀疏表示。

我们从每一类样本中随机采样 100 个样本，通过图 X 的网络将其映射到监督引导特征空间，并利用 t-SNE 映射到二维流型，最终的效果如图 X：

在这种情况下，我们几乎可以把每个调制类别的样本在二维流型中利用聚类算法分开。当然，其中也有一部分的数据是混淆的，比如类别 X 中也有部分样本散落到类别 Y 中。

可能是因为在获取监督引导特征空间时，我们的目标是正确区分不同的调制类别，所以我们获取的监督引导特征对于不同类别的样本是有一定的区分度的，即不同调制类别的样本分布在特征空间的不同区域，这就表现为在 t-SNE 之后不同类别样本分布在二维流型的不同区域。当然，随着训练网络时样本类别的增加，我们获取的具有类区分度的特征将会得到更好的泛化。

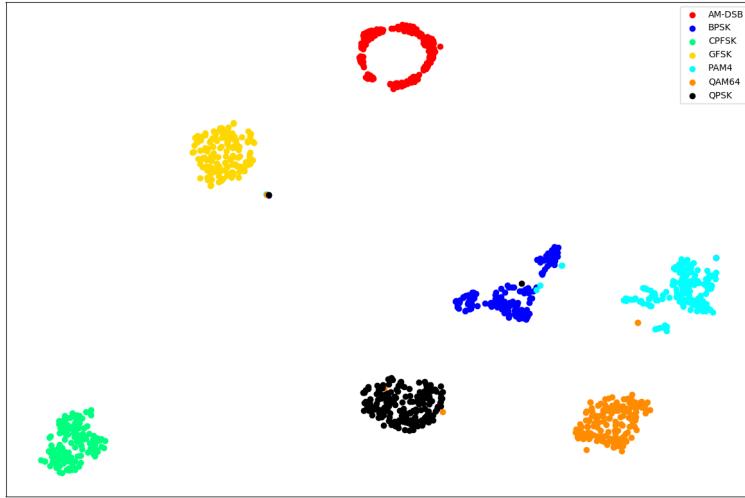


图 3.7 基于 CNN 的调制信号 t-SNE 二维流型展示码器

3.4 基于 CAE-CNN 的无线信号调制识别

在上一节中，我们分别使用监督引导和无监督引导的方法获取数据的低维表示，其本质上就是基于降维的非线性特征提取过程。在本节中，通过融合 CAE 与 CNN，我们联合重构误差与分类误差，提出了一种新的调制识别网络框架和训练算法。

3.4.1 CAE-CNN 网络框架

在上一节中，我们发现数据样本在低维嵌入空间中的表示具有一定的类可分性。因此，我们可以将 CNN 分为特征提取与分类两个步骤。对于 CAE 与 CNN 的融合，我们的本质是希望能够分类的同时保证特征提取尽量多地包含数据的原始信息。为此，我们通过在 CNN 的交叉熵损失中加入 CAE 的重构误差损失作为我们的整体损失；通过改进 CAE 与 CNN 的训练算法，降低重构误差与分类误差，达到更好的分类效果。图3.8展示了我们网络的结构。

在 CAE-CNN 中的卷基层中，我们使用了 Dropout，用以降低模型的过拟合影响；并在卷积权重增加了权重值 W 的 2 范数 $\|w\|_2^2$ 作为惩罚项，使权重尽量小；同时，借鉴稀疏自编码器（Sparse Autoencoder, SAE）的思想，我们在第一个密集连接层加入权重的 $F1$ 范数范数惩罚来鼓励解的稀疏性^[48]。

那么我们有 CAE 的损失函数：

$$J_{CAE}(X, \hat{X}; \theta_{CAE}) = \frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \hat{x}^{(i)}\|_2^2 + \frac{\lambda}{2} \|w\|_2^2 + \eta \|J_f(x)\|_F^2 \quad (3-2)$$

其中， $J_f(x)$ 是隐层输出值对于权重的雅克比矩阵， $\|J_f(x)\|_F^2$ 表示该雅克比矩阵 F 范数的平方。

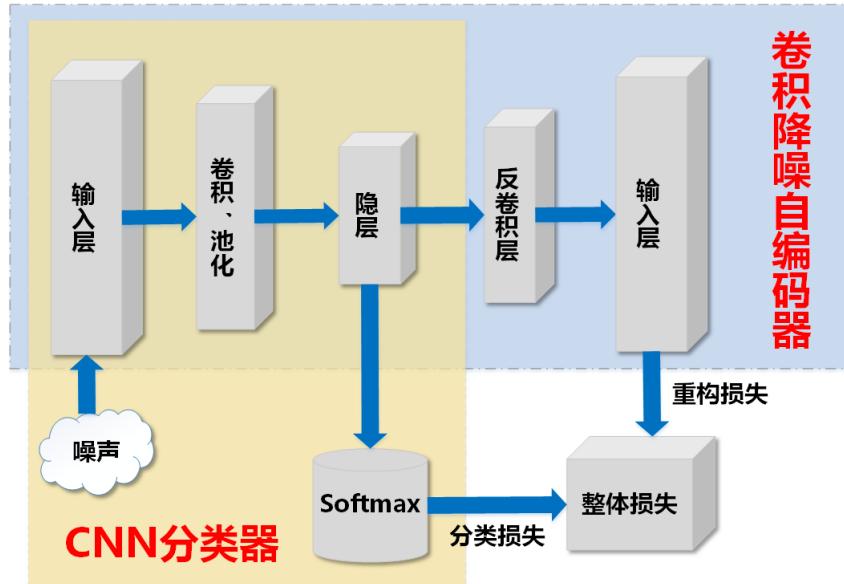


图 3.8 CAE-CNN 网络框架

由于 CNN 的损失函数我们使用交叉熵损失函数，同时将卷基层权值的二范数作为损失的一部分，那么我们有 CNN 的损失函数：

$$J_{CNN}(X, Y; \theta_{CNN}) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \| h_{W,b}(x^{(i)}) - y^{(i)} \|^2 \right) \right] + \frac{\sigma}{2} \| w \|_2^2 \quad (3-3)$$

系统的总体损失函数为 CAE 的损失跟 CNN 损失的加权和，如方程(3-4):

$$J(X, \hat{X}, Y; \theta_{CNN}, \theta_{CAE}) = \alpha J_{CNN}(X, Y; \theta_{CNN}) + \beta J_{CAE}(X, \hat{X}; \theta_{CAE}) \quad (3-4)$$

3.4.2 CAE-CNN 算法

在提出的 CAE-CNN 框架中，我们有分类损失与重构损失，因此在模型训练时，需要我们同时考虑到两种损失，并进行反向传播得到优化后的网络参数。对应于我们的 CAE-CNN 框架，我们提出了3.1训练算法。

3.4.3 算法运行环境及参数

使用分类交叉熵损失函数和 Adam[15] 求解器进行训练 (在我们的数据集上略胜过 RMSProp[12])。我们在 tensorflow[16] 计算框架上运行网络的训练和预测，使用 NVIDIA Cuda[8] 组件，在 Nvidia GTX1080ti 显卡上加速运算。接下来的仿真我们都使用这一套软硬件组合进行仿真。我们的机器系统配置如表3.3所示。

算法 3.1 CAE-CNN 训练算法

Input: $\alpha, \beta, \lambda, \delta, \eta, K$
Output: $J(X, \hat{X}, Y; \theta_{CNN}, \theta_{CAE}), J_{CNN}(X, Y; \theta_{CNN}), J_{CAE}(X, \hat{X}; \theta_{CAE})$

Require: 初始化 $\alpha, \beta, \lambda, \delta, \eta$
Require: 初始化 $K = 1000, \Delta K = 20, k = 0, \Delta J = 0.001, \Delta S = 1000$
Require: 训练数据集及测试数据集构建
Require: 构建 CAE 框架及损失函数 $J_{CNN}(X, Y; \theta_{CNN})$, CAE 框架及损失函数 $J_{CAE}(X, \hat{X}; \theta_{CAE})$, 以及整体损失函数 $J(X, \hat{X}, Y; \theta_{CNN}, \theta_{CAE})$, 最终构建完整的计算图 Graph

```

1: while K>0 do
2:   for  $k = 0; k < K; k = k + \Delta K$  do
3:     计算  $J_{CAE}(X, \hat{X}; \theta_{CAE})$ , 并训练 CAE 网络
4:   end for
5:   for  $k = 0; k < 1000 - K; k = k + \delta K$  do
6:     计算  $J_{CNN}(X, Y; \theta_{CNN})$ , 并训练 CNN 网络
7:   end for
8:    $K = K - 20$ 
9: end while
10:  $S^* = 0, J_{CNN}^* = 1e5$ 
11: for  $S = 1; \Delta S \leq 5000 || S > 10^5; S ++$  do
12:   计算  $J_{CNN}$ , 并训练 CNN
13:   if  $J_{CNN} < J_{CNN}^*$  then
14:      $J_{CNN}^* = J_{CNN}$ 
15:      $S^* = S$ 
16:   else
17:      $\Delta S = S - S^*$ 
18:   end if
19:   if  $S \% 10 == 0$  then
20:     计算  $J_{CAE}$ , 并训练 CAE
21:   end if
22: end for

```

表 3.3 系统参数配置

参数	配置
System	Centos7.3
CPU	Intel® Xeon(R) CPU E5-2683 v3
GPU	GeForce GTX 1080 Ti × 2
Memory	32GB
Storage	128G SSD & 2TB HardDisk
External Storage	Toshiba HDTH310E 2TB
Python	python3.6
Library	Tensorflow1.4

3.5 结果及分析

我们使用 Adam 优化器，CAE-CNN 模型训练了大约需要 50 分钟。我们设置 *BatchSize* 大小为 1024 的样本，每一次训练集大约需要 15 秒。在 epoch 大于 200 之后，我们确实观察到一些过度拟合，因此在算法3.1中我们设置最大的循环次数小于 300 次 epoch。

3.5.1 分类结果

我们调制方式总共有 7 种，每一类调制信号在噪声信噪比为 $-20dB \sim 18dB$ 之间每隔 $2dB$ 进行数据采样，每个信噪比下采样 0.1 秒，每个样本共包含 1000 个采样值，大约有 4000 个样本。我们将 80% 的样本作为训练集，10% 的样本作为验证集集，10% 的样本作为测试集。即训练样本大约有。这些样本均匀分布在从-20dB 到 +18dB 的 SNR 中，并被标记以便我们可以评估特定子集上的性能。

经过 CAE-CNN 的网络训练，我们在测试数据集上的所有信噪比之间的分类准确率大致达到了 72.3%。这表示的是我们分类器在-20dB 到 18dB 之间样本的一个平均准确率，由于信噪比在-2dB 到-6dB 之间分类器性能迅速下降至 30% 左右，并在较低信噪比条件下维持一个很低的分类水平。因此，从这方面而言，这个准确率已经很高了。最终，我们 CAE-CNN 分类结果如图3.9

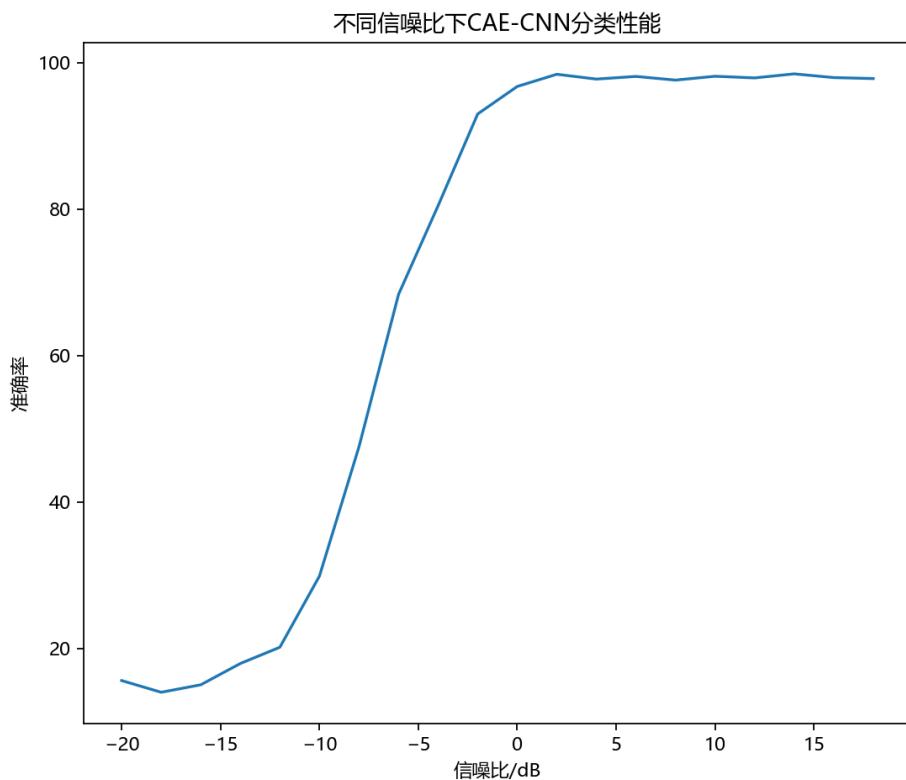


图 3.9 CAE-CNN 分类结果

从图3.9中我们可以看出，信噪比在0dB以上时分类准确率都能达到96%以上。在-4dB时，仍然有85%以上的准确率，这相对于传统的基于特征的分类器而言，具有很大的提升。

同时，我们与前人提出的一些算法进行比较，本文中我们选用了基于特征的支持向量机（Support Vector Machine, SVM）、决策树（Decision Tree, DTREE）、随机森林（RandomForest, RF）、K-近邻（K-NearestNeighbor, KNN）、深度神经网络（Deep Neural Network, DNN）。分类器所用特征在第四章我们将系统阐述。最终，分类结果如图3.10所示。

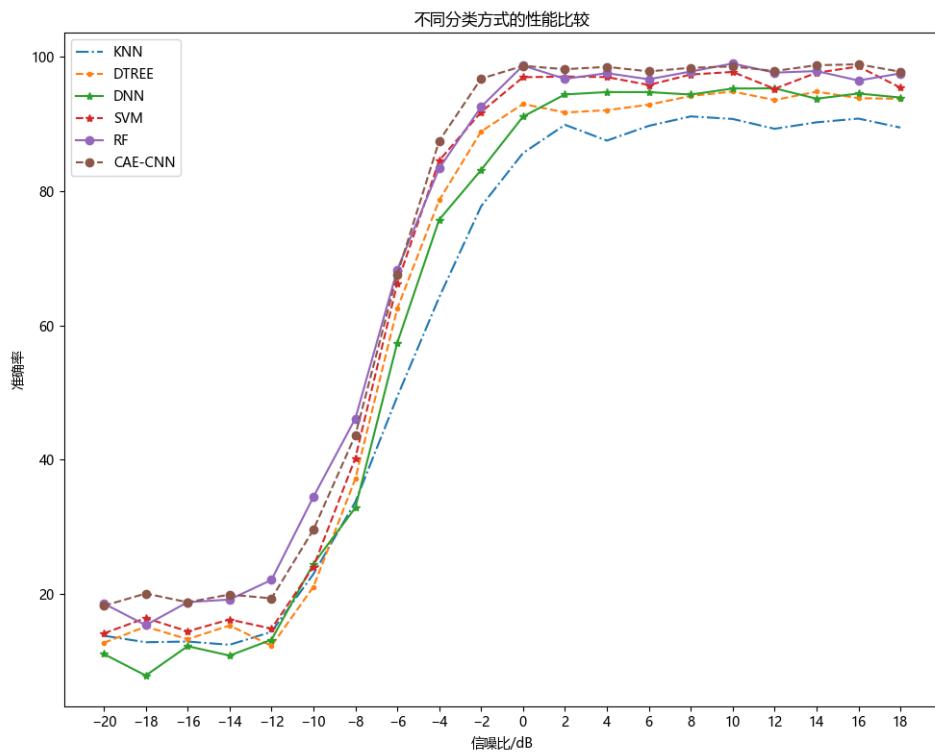


图 3.10 与其他方法的比较

从图3.10中我们可以看出，KNN 大部分情况下分类性能较差；在高信噪比（大于0dB）条件下，SVM、CAE-CNN 和 RF 分类性能相差不大，但是 CAE-CNN 性能相对更好一些，而且曲线比较平滑，分类性能较为稳定；而在低信噪比条件下，CAE-CNN 同时也具备较好的表现，大部分情况下优于其他的分类器。由此可见，本文提出的调制识别算法具有较好的识别性能。

我们计算得到0dB的测试样本的分类混淆矩阵，并将其可视化，如图3.11。

可以看到 SNR 为0dB时，在混淆矩阵中我们有一个较为纯净的对角线，仅有少量的 QPSK 样本误分为 BPSK 和 CPFSK。这可能与 QPSK 本身的相位和幅值同时变化有关，加之噪声带来一定的影响，使得分类器产生了个体样本的误判。

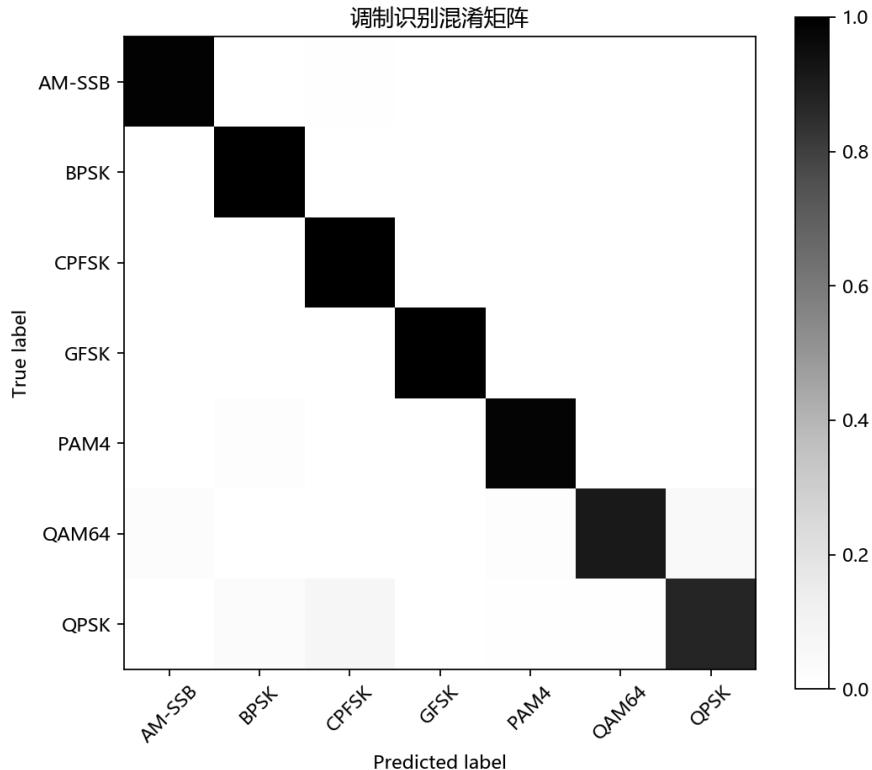


图 3.11 0dB 条件下的结果

3.5.2 算法效率

由于调制识别的应用场景大都要求我们需要能够对信号进行实时响应，因此算法的运行效率将会是影响识别性能的重要因素。在本节中，我们将通过比较不同算法的训练时间以及分类时间，查看算法的运行效率。由于不同算法的训练时间和样本分类时间在数量级上可能不同，所以接下来进行时间分析时，我们的时间坐标轴都是以对数坐标为基准的。

影响算法训练时间因素很多，包括能够实现并行、是否进行加速、算法复杂度、硬件设备、数据样本数目等等。本文中所有测试算法均在同一平台下运行。深度学习的一个普遍缺点是需要很大的计算量，然而在本文中，我们的网络相对较小，数据集的量也较少，在 DNN 和 CAE-CNN 的训练使用 Tensorflow 框架，并启用了 GPU 并行加速；KNN、SVM、RF、DTREE 我们使用 scikit-learn 库中的 API。最终的训练时间如图3.12。

从图3.12中我们可以看到，KNN 的训练时最快的，这是因为算法本身几乎不需要进行运算。CAE-CNN、DNN、SVM 耗时都相对较多，这是因为这几种算法本来就需要很大的计算量，而且 SVM 没法进行 GPU 的加速训练，最终导致了 SVM 的训练耗时最多。其次，DNN 中由于全连接层的存在，相较于 CAE-CNN 模型需要更多的训练量，因此训练时间也较长。而 RF 我们选用了 100 棵树进行训练，每一棵树的深度

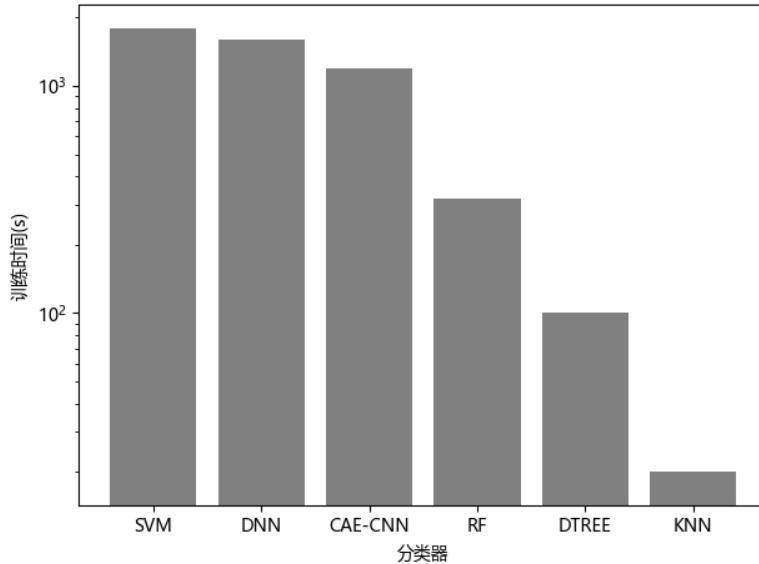


图 3.12 不同算法训练时间

限制在 5 以内，因此，其训练时间相较于 DTREE 要多一些；但是考虑到 RF 可以实现并行，在我们的平台中训练时间相较于 DTREE 增加的并不多。可以看出，CAE-CNN 分类器训练时间虽然较长，但是相比于 SVM 以及 DNN 算法仍然要少一些，并且会带来一定的性能提升。

接下来，我们比较一下算法分类时的运行时间。其中 CAE-CNN 由于样本是以一个 batch 共 1024 个样本输入进网络的，我们对每个样本的分类时间进行了平均。最终的分类时间如图3.13所示。

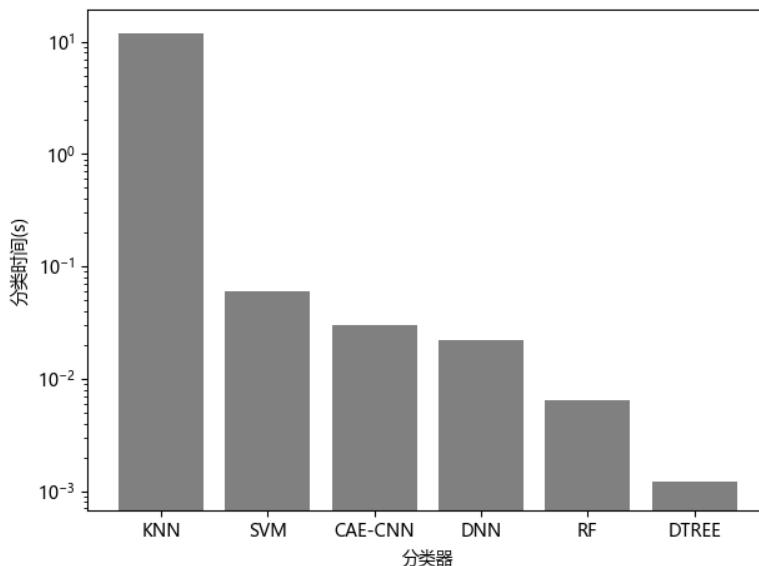


图 3.13 不同算法分类时间

DTREE 具有最低的分类时间，而 KNN 所需分类时间最多。这主要是因为 KNN

在分类时，是将所有的训练样本进行比较，而且分类时间与样本数目多少成正相关，所以具有很高的分类时间；决策树本身因为模型比较简单，只需要在树的少量分支上进行一定的运算，因此可以对数据进行快速分类。SVM、DNN、CAE-CNN 算法在进行分类时，需要进行一定的数据运算，因此分类时间也较长。虽然 CAE-CNN 相较于其他算法在性能上没有明显的优势，但是考虑到小样本的分类本身启动 GPU 就耗费一定的时间，而且其分类时间并没有很大，这也算一个可以接受的水平。

3.6 本章小结

本章首先研究了通过无监督和有监督的方式所提取的数据特征之间的差异，发现通过无监督的方法对数据样本进行特征提取可以重构原始数据样本，并获得原始样本具有一定类可分性的特征，但是不同类别之间的数据会产生一定的混淆。而通过有监督的方式获取的数据特征，对于不同调制方式而言，具有较强的类可分性，利用 t-SNE 算法将其映射到二维流型中，发现同一类别的数据样本分布在低维流型中相近的区域，不同类别的样本分布在低维流型中的不同区域。利用以上结论，我们提出了一种 CAE-CNN 调制识别的算法框架，并提出了一种相应的训练算法。结果显示，我们提出的 CAE-CNN 算法，在高信噪比条件下具有较高的识别准确率，并且鲁棒性较强；在低信噪比条件下也具备更高的识别率。在算法效率上，我们提出的算法训练时间在可接受范围以内，同时算法运行时间较短，算法运行复杂度较低。

第四章 基于传统特征与深度特征融合的无线调制方式识别技术研究

4.1 引言

在机器学习领域，经常使用融合的方法解决分类或者回归等问题，按不同层次划分，主要有数据融合、特征融合、模型融合。特征融合主要是基于信息融合思想，通过融合不同特征对于问题的不同表征形式，从不同层次、不同特征空间、不同时间尺度、差异性全局特征和局部特征等不同层面，对特征进行重新组合，来提高模型的性能和鲁棒性^{[49][?]}。

在上一章中，我们可以得到，CAE 对调制信号所提取的特征可以重构原始信号并具有一定的类可分性；CNN 提取的特征具备较强的类可分性。因此，我们可以利用 CAE、CNN 等网络，提取数据样本具备一定类可分性的深度特征。

在本章中，我们将 CAE-CNN 算法提取的卷积特征与传统特征进行融合，研究深度特征与传统特征结合对分类准确率的影响；同时，我们也对不同的特征融合框架进行探索，研究不同的融合方式对于特征融合效果的影响。

4.2 传统特征

基于统计机器学习的调制识别方法，主要思想是特征提取以及模型训练；基于 D-S 证据理论和判决理论的方法主要依赖特征的选择和判决门限的设定。因此，对于传统的调制识别方法，特征的选择与提取是影响系统性能的重要条件，决定了算法所能达到的上限。不同调制方式我们所提取的特征主要包括：时域特征、频域特征、高阶统计量等。本章针对这几类特征集中的部分特征，利用原始数据进行提取，作为特征融合的传统特征集。

4.2.1 基本时频特征

我们假设原始信号 $x(t)$ 解析表示为：

$$s(t) = x(t) + i * y(t) \quad (4-1)$$

其中， $y(t)$ 为原始实信号的希尔伯特变换（Hilbert Transform）。

无线信号的调制信息主要以信号的瞬时幅度、瞬时频率和瞬时相位等特征进行表征，合理地选择和构造信号的的统计量是获得信号调制方式的重要方法^[50]。

(1) 零中心归一化瞬时幅度谱密度的最大值

γ_{max} 表征了信号瞬时幅度变化的统计特性，反映了无线信号的调制包络的变化情况。我们可以通过设置一定的判决门限，来区分恒定包络和非恒定包络的不同调制信号：

$$\gamma_{max} = \max |DFT(A_{cn}(t))|^2 / N_s \quad (4-2)$$

其中， N_s 为取样点数， $A_{cn}(t) = A(t)/m_a - 1$ 为零中心归一化瞬时幅度， $m_a = \frac{1}{N_s} \sum_{i=-1}^{N_s} A(t)$ 为瞬时幅度 $A(t)$ 的均值，用均值 m_a 对瞬时幅度 $A(t)$ 进行归一化的主要目的是消除信道的噪声、多径等干扰的影响。

(2) 零中心归一化非弱信号瞬时幅度标准差 δ_{da}

$$\delta_{da} = \sqrt{\frac{1}{C} \left[\sum_{A_n(i) > a_t} A_n^2(i) \right] - \left| \frac{1}{C} \sum_{A_n(i) > a_t} A_n(i) \right|} \quad (4-3)$$

其中， C 是所有 N 个样本中属于非弱信号的数目，瞬时幅度 $A_n(t)$ 大于幅度判决门限电平 a_t 的信号为非弱信号。 δ_{da} 表征了一个符号周期内的样本点中样本的幅值变化信息，可以利用该统计量来区分一个符号周期内归一化中心瞬时幅度的调制方式，以及归一化中心瞬时幅度不为零的调制方式。

(3) 零中心归一化瞬时幅度绝对值的标准差 δ_{aa}

$$\delta_{aa} = \sqrt{\frac{1}{N} \left[\sum_{i=1}^N A_{cn}^2(i) \right] - \left[\frac{1}{N} \sum_{i=1}^N |A_{cn}(i)|^2 \right]} \quad (4-4)$$

δ_{aa} 表征了原始信号的幅值信息，可用于区分非归一化的绝对幅值信息，以及归一化的绝对幅值信息的不同调制类别。

(4) 零中心归一化瞬时幅度的四阶紧致性 μ_{42}^a

$$\mu_{42}^a = E [A_{cn}^4(t)] / E [A_{cn}^2(t)] \quad (4-5)$$

μ_{42}^a 是用来表示瞬时幅度分布的密集性的一个统计量，可用于区分瞬时幅度高密集分布信号以及瞬时幅度分布比较分散的信号。

(5) 零中心归一化瞬时频率均值的平方与方差之比 R_f

$$R_f = u_f^2 / d_f \quad (4-6)$$

在公式4-6中， u_f 和 d_f 分别表示信号零中心归一化瞬时频率的均值和方差。这两个参数可以用来判断信号所包含的频率信息。

(6) 零中心非弱信号段归一化瞬时频率绝对值的标准差 δ_{af}

$$\delta_{af} = \sqrt{\frac{1}{c} \left[\sum_{a(i)>a_i} f_N^2(i) \right] - \frac{1}{c} \left[\sum_{a(i)>a_i} |f_N(i)|^2 \right]} \quad (4-7)$$

其中 δ_{af} 表示信号的绝对频率信息，通常用来区分归一化中心瞬时频率绝对值为常数的调制类别，以及包含绝对、直接频率信息的调制类别。

4.2.2 高阶累积量

由于码元序列的高阶统计量能够反映星座图的分布特征，并具有抗噪声等优点，所以适合于区分幅度相位调制方式。由于高阶统计量特征的特点，通常基于高阶统计量的调制识别算法具备较强的鲁棒性，因此得到了广泛的应用^{[51][52][53][?][54]}。

对于零均值的高斯白噪声信号，三阶及以上的高阶累积量为零。依据这一特性，获取接收信号 $r(t)$ 的高阶累积量就等价于获取发送信号 $s(t)$ 的高阶累积量。从而，高阶累积量对于零均值高斯白噪声具有一定的抑制作用，达到一定程度上的干扰抑制效果。不同调制类别的信号，其高阶累积量的值不同，而且不受零均值高斯白噪声的影响；因此，高阶累积量可以作为区分不同调制方式的特征值，用于调试识别算法的学习。

若平稳随机过程 $x(t)$ 具有零均值，那么其 k 阶矩可以定义为：

$$M_{kx} = \tau_1, \tau_2, \dots, \tau_{k-1} = E[x(t), x(t + \tau_1), \dots, x(t + \tau_{k-1})] \quad (4-8)$$

若考虑延时 $\tau_1 = \tau_2 = \dots = \tau_{k-1} = 0$ ，则 $x(t)$ 的 p 阶混合矩：

$$M_{pq} = E \{ [x(t)]^{p-q} [x^*(t)]^q \} \quad (4-9)$$

其中， $x^*(t)$ 为 $x(t)$ 的共轭， $E*$ 为求期望。

根据 Swami^[25] 的论文中的高阶矩与高阶累量的定义，我们可以得到零均值平稳随机过程 $x(t)$ 的各阶累积量的定义。

(1) 二阶累积量

$$\begin{aligned} C_{20} &= Cum(X, X) = M_{20} = E[X(t)X(t)] \\ C_{21} &= Cum(X, X^*) = M_{21} = E[X(t)X^*(t)] \end{aligned} \quad (4-10)$$

(2) 四阶累积量

$$\begin{aligned} C_{40} &= Cum(X, X, X, X) = M_{40} - 3M_{20} \\ C_{41} &= Cum(X, X, X, X^*) = M_{41} - 3M_{20}M_{21} \\ C_{42} &= Cum(X, X, X^*, X^*) = M_{42} - M_{20}^2 - M_{21}^2 \end{aligned} \quad (4-11)$$

(3) 六阶累积量

$$\begin{aligned} C_{60} &= Cum(X, X, X, X, X, X) = M_{60} - 15M_{40}M_{20} + 30M_{20}^3 \\ C_{63} &= Cum(X, X, X, X^*, X^*, X^*) = M_{63} - 9M_{42}M_{21} + 9|M_{20}|^2 M_{21} + 12M_{21}^3 \end{aligned} \quad (4-12)$$

(4) 四阶累积量与二阶累积量之比

$$R_{mn} = \left| \frac{C_{42}}{C_{21}} \right| \quad (4-13)$$

4.3 特征融合理论

传统的模式分类方法，主要通过特征提取，利用支持向量机、人工神经网络等算法进行分类器的训练。但是，传统的模式分类方法通常基于人工设计的特征，经过特征提取算法得到原始数据某些方面的特征，由于人们对于数据本身认知的片面性，很难利用特征完整的表征数据本身的分布。这样，纯粹的基于人工特征提取训练的分类器就很难完全准确地对数据样本进行分类或者回归。

解决此问题的一种思路就是使用特征融合算法，利用多维特征构成的特征空间，来获取数据本身包含的性质，弥补某些固有特征对于数据特征反映不完整的缺陷。特征融合方法从特征层面进行处理，它既可以利用已有的算法提取数据特征，相较于重新设计特征和特征提取的算法，具有较低的成本。在本节中，我们主要研究基于深度学习的特征融合理论。

基于深度学习的特征融合算法，主要是将特征融合的思想引入到神经网络模型中，并在模型中选择特定隐层融合其他特征，利用多源特征来训练网络参数。

特征融合理论主要基于两个基本假设：多特征融合通常比单特征具备更好的分类性能；融合的特征之间相关性尽量小。前者是进行特征融合的出发点和前提，只有融合后的特征具备这样的特性，我们才具备融合的必要性。后者是基于流形理论的假设，当特征之间的关联性较高时，融合之后的特征空间在高维流形中趋于扁平化，可能会产生特征冗余，而实际上可以用更低的维度来表征原始数据。因此，我们应该从多角度选择低相关性的特征进行融合，使数据能够在融合特征形成的高维流形中处于一个有利于我们任务（分类、回归等）的分布。

假设原始的数据样本集为 X ，那么对于任意 $x \in X$ ，我们假设用于提取特征的隐层单元集合为 H_{hidden} ，其维度为 N ， $h_i(x)$ 表示网络中第 i 个单元的输出值，则隐层神经单元构成的深度特征集为 $H = h_i(x) | h_i \in H_{hidden}$ 。通过深度网络，我们可以获取原始特征空间到深度特征空间的映射：

$$h(x) = \bigcup_{0 < i < N} w_i * h_i(x) \quad (4-14)$$

当数据通过网络映射到深度特征空间的输出值，便是我们在新空间中的特征向量，即数据的深度特征。

同时，我们以本章中第2小节的特征提取算法，对数据的传统特征进行提取，每一个特征用 $f_i(x)$ 表示，作为不同维度的特征，假设我们的传统特征集为 F ，对于任意样本 $x \in X$ ，我们根据第2小节中的特征提取算法有：

$$f(x) = \bigcup_{0 < i < |Z|} \mu_i * f_i(x) \quad (4-15)$$

这样，通过融合传统特征 F 与深度特征 H 得到的融合特征集为 Z ：

$$Z = F \cup H = \{z | (z \in H) \vee (z \in F)\} \quad (4-16)$$

最终，我们通过不同的融合算法对多源特征进行融合。特征融合的基本框架如图4.1所示：

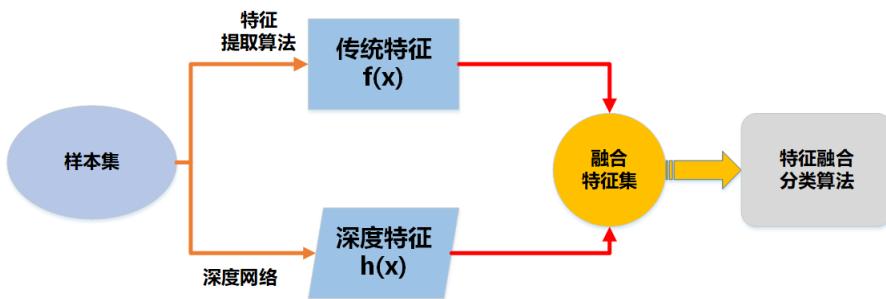


图 4.1 特征融合框架图

其中，融合算法有很多种，比如LR、支持向量机等。在下一节中，我们将进行更深入地探讨。

4.4 传统特征与深度特征融合框架

基于统计模式的调制识别方法又称为基于特征的调制识别方法。此类算法族通常将无线信号的调制识别看作一个基于统计的模式识别问题。通常，整个调制识别系统由两部分组成：基于统计的特征提取子系统和基于统计机器学习的模式识别子系统。特征提取子系统的目的是从原始数据样本中提取已知的、能表征信号调制类型的特征，可以认为是从信号所在的原始空间，到人工提取特征所构成的特征空间的一个映射。

4.4.1 特征归一化

为了减小特征值数量级对分类性能的影响，降低梯度爆炸和梯度弥散对模型的影响，我们将深度特征和传统特征进行批归一化（Batch Normalization，BN）处理。

假设网络的输入样本为 x , 对于特定的 $mini-batch : \beta = \{x_{1,2,\dots,m}\}$ 数据流向融合的分类器之前, 对其进行归一化:

$$\hat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\delta_\beta^2 - \epsilon}} \quad (4-17)$$

$x_i^{(k)}$ 表示融合算法的第 i 个输入单元, $\mu_\beta \leftarrow \frac{1}{m} \sum_{i=1}^m x_i$ 表示该单元 $mini-batch$ 样本输入的均值, $\delta_\beta^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_\beta)^2$ 表示该单元 $mini-batch$ 样本输入的标准差, ϵ 是为了防止除数为 0 而增加的偏移值, 此处我们取 0.001。

对于 DNN 融合算法, 我们直接将 BN 作为一个完整的网络层。由于 BN 本身改变了数据的分布, 可能会影响到网络原有特征的表达能力了。为了恢复网络的表征能力, 在对数据归一化后, 又进行了还原操作 $y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i)$, 其中 γ, β 也是通过网络的训练学习到的参数。最后, 利用 ADAM 优化算法直接学习 BN 的参数和 DNN 的参数, 获得特征与样本标记的映射关系。

而对于 Softmax 等融合算法, 由于很难进行 γ, β 变量的学习, 所以, 我们首先将数据通过网络, 并获得所有样本的深度特征; 然后, 基于全局的角度对包括传统特征在内的所有特征, 进行如方程(4-17)所示的全局归一化 (Normalization)。

4.4.2 基于 Softmax 回归的融合框架

由于我们调制识别并不是一个二分类问题, 传统的 SVM 以及逻辑回归 (Logistic Regression, LR) 并不适应我们的应用场景。Softmax 回归 (Softmax Regression, SR) 模型, 是 LR 在多分类问题上的推广。在 SR 中, 我们解决的是多分类问题 (相应与 LR 解决的二分类问题), 类标 y 可以取 k 个不同的值 (而不是 2 个)。因此, 对于训练集 $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, 我们有 $y^{(i)} \in \{1, 2, \dots, k\}$ 。

对于给定的测试输入 x , 我们想用假设函数针对每一个类别 j 估算出概率值 $p(y = j|x)$ 。也就是说, 我们想估计 x 的每一种分类结果出现的概率。因此, 我们的假设函数将要输出一个 k 维的向量 (向量元素的和为 1) 来表示这 k 个估计的概率值。具体地说, 我们的假设函数 $h_\theta(x)$ 形式如方程(4-18):

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|x^{(i)}; \theta) \\ p(y^{(i)} = 2|x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k|x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (4-18)$$

其中, $\theta_1, \theta_2, \dots, \theta_k \in \Re^{n+1}$ 是模型的参数。 $\frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}}$ 这一项对概率分布进行归

一化，使得所有概率之和为 1。我们有 *Softmax* 回归中将 x 分类为类别 j 的概率为：

$$p(y^{(i)} = j | x^{(i)}; \theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \quad (4-19)$$

那么 *softmax* 回归算法的损失函数 $J(\theta)$ 由方程(4-20)可得：

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] \quad (4-20)$$

其中, $1\{\cdot\}$ 是指示函数 (Indicator Function), 其取值规则为: $1\{\text{False}\} = 0, 1\{\text{True}\} = 1$ 。

我们通过添加一个权重衰减项 $\frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2$ 来修改损失函数, 这个衰减项会惩罚过大的参数值, 现在我们的损失函数 $J(\theta)$ 变为方程(4-21):

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2 \quad (4-21)$$

增加权重衰减项 ($\lambda > 0$) 之后, 损失函数就变成了严格的凸函数, 这样就可以保证方程具有唯一解。对 $J(\theta)$ 的最小化问题, 目前还没有理论上的闭式解, 最常用的方法是使用迭代的优化算法 (例如梯度下降法) 进行求解。对 $J(\theta)$ 进行求导, 我们得到梯度方程(4-22)。

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m [x^{(i)}(1\{y^{(i)} = j\} - p(y^{(i)} = j | x^{(i)}; \theta))] + \lambda \theta_j \quad (4-22)$$

通过最小化 $J(\theta)$, 可以得到 *softmax* 回归模型。最终, 基于 SF 的特征融合框架如图4.2所示:

4.4.3 基于深度学习的融合框架

随着深度学习的发展, 已经有部分学者对基于深度学习的特征融合方法进行了研究^{[55][?]}。Simonyan 等人首先提出了一种基于双流架构的 CNN 模型, 一定程度上解决了视频中的动作识别问题^[56]; Feichtenhofer 等人在此基础上改进了特征的融合算法, 提出了一种基于空间特征和时间特征融合的识别算法^[57]。

Nielson 提到如果神经网络达到一定条件, 可以用于模拟任意函数^{??}。DNN 是对输入的数据样本, 通过隐层对数据进行非线性变换, 利用交叉熵或者均方误差等损失函数对隐层参数进行训练, 来拟合数据到任务的映射关系, 并输出我们希望的结果。

基于深度学习框架的融合方法, 在融合点之前分别进行特征学习。由本章第 3 小节定义, 通过深度网络学习到的特征集为 H , 通过传统算法提取的特征集为 F 。

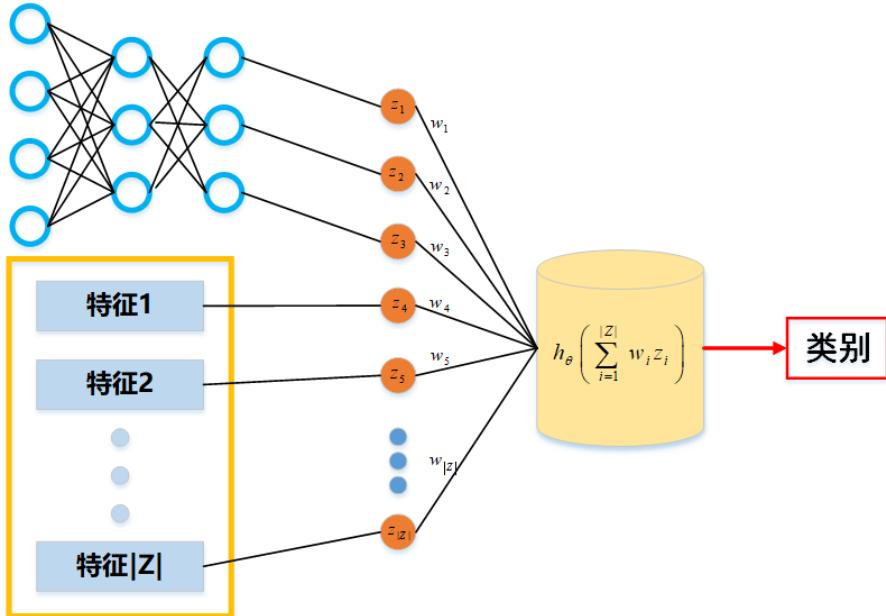


图 4.2 基于 LR 框架的特征融合框架图

为降低因为特征尺度关系对后期网络学习造成的影响，我们在特征融合之前在框架中加了一个特征预处理层，来对特征进行归一化处理。此时，我们获得的归一化特征集变为：

$$\hat{Z} = \{f_{norm}(z) | z \in H \cup F\} \quad (4-23)$$

融合层将独立学习的归一化特征进行融合，输入到神经网络中。我们通过对包括 LSTM、DNN、CNN 等框架进行验证之后，最终使用的是卷积神经网络，并在其中加入池化层，使用 dropout 降低模型的过拟合风险。在之后的深度网络层，由于样本数据有限，我们仅仅用了 2 层的全连接层，但是从分类结果来看，其效果还是达到预期。

整个基于深度学习的特征融合框架如图4.3所示：

4.4.4 基于集成树的融合框架

集成学习（Ensemble Learning），主要通过构建并结合多个基学习器来完成学习任务。按照算法的思想划分，集成学习主要分为两类：Bagging 算法族和 Boosting 算法族^[58]。

Bagging^[59] 是一类并行式集成学习的算法，主要基于采样数据集训练基学习器，并对这些基学习器进行组合（比如分类问题投票法、回归问题平均法）作为最终结果。Bagging 主要是通过多个基学习器的集成来降低模型的方差。

Boosting^[60] 的思想是，首先训练简单的基学习器，然后根据基学习器的表现调整训练样本分布，使得基学习器误分或者错误预测的样本得到更多的关注。接下来，基于调整分布后的训练样本来训练下一个基学习器，直至基学习器数目达到事先指定

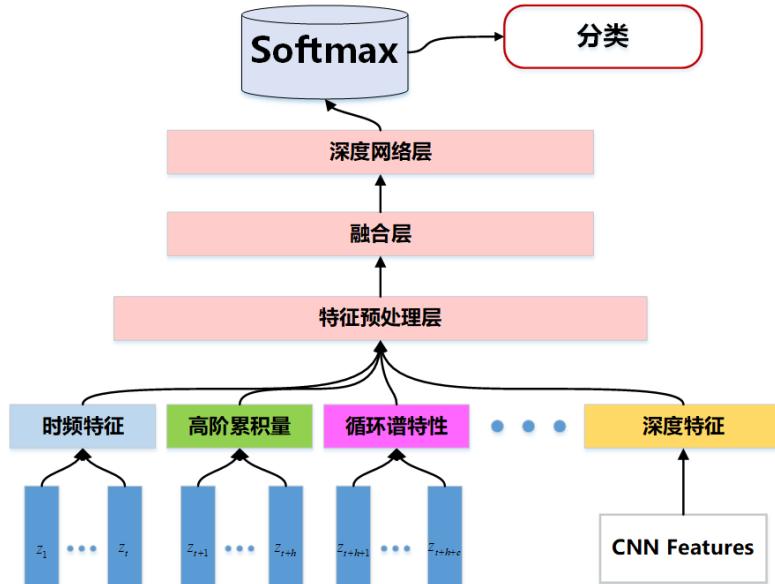


图 4.3 基于 DNN 框架的特征融合框架图

的值 T 。最终，将学得的 T 个基学习器进行加权组合。Boosting 主要通过过个学习器的加性组合降低偏差，其主要的代表算法是 AdaBoost^[61]。

随机森林 (Random Forest, RF)^[62] 是一个基于 Bagging 思想的，通过随机方式建立的，包含多棵决策树的集成分类器。首先，它通过自助重采样 (Bootstrap Resampling) 方法，从原始训练样本集 S 中有放回地随机抽取 N_t 个样本生成新的训练样本集 S_t ，重复自助重采样生成 T 个子训练集 S_1, S_2, \dots, S_T ；然后，利用 T 个自动生成的训练子集，分别训练相应数据集下的子学习器，组成随机森林。最后，利用自学习期的结果进行融合（分类问题可以用投票，回归问题可以用回归）。对于我们的调制分类任务而言，随机森林简单、容易实现、计算开销小，正符合我们对集成树融合效果的验证性要求。

对于调制分类问题，我们将 CART 树 (Classification And Regression Tree, CART)^[63] 作为基分类器，并引入样本采样和特征采样，增加样本扰动和特征扰动对集成学习算法的影响。基分类器的集成方法主要有平均法、投票法、staging 学习法等方法，平均法主要是针对回归问题，而 staging 的方法是利用基分类器的输出作为新的样本学习一个新的分类器，会增加模型的复杂度。因此，在调制分类问题中，我们使用投票法来作为基分类器的结合方法。

假设我们总共学习 K 个基学习器，则每个并行训练得到的基学习器为 T_i ，那么利用投票法所得的分类结果如方程(4-24)所示：

$$C(x) = \arg \max_c \sum_{i=1}^K I(T_i(x), c) \quad (4-24)$$

其中, $I(T_i(x), c) \in \{0, 1\}$, 若基分类树 T_i 将样本 x 预测为类别 c , 则 $I(T_i(x), c) = 1$, 否则 $I(T_i(x), c) = 0$ 。

RF 不仅通过样本扰动(通过对原始训练集进行采样)实现随机性, 而且通过对属性进行随机采样来增加特征的扰动, 通过多棵数进行投票来预测分类结果。这就使得最终的集成分类器的泛化性能, 可通过增加个体学习器之间的差异程度而得到进一步的提升。

图4.5展示了我们基于随机森林算法的特征融合框架图:

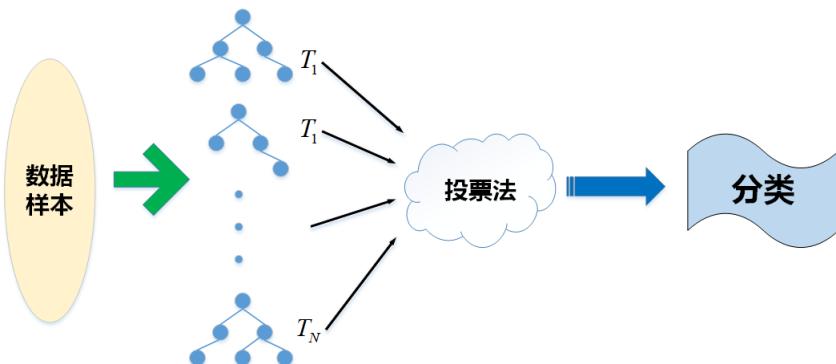


图 4.4 基于随机森林框架的特征融合框架图

4.5 结果及分析

首先, 我们提取原始数据的深度特征和传统特征, 并对其基于4.4.1小节中的基准进行特征的标准化; 然后, 分别基于 Softmax、RF、DNN 三种融合框架, 学习相应的融合模型, 获取特征融合之后的分类器; 最后, 我们利用我们的测试数据集来对特征融合模型的性能进行验证。

4.5.1 分类性能比较

为了进行比较, 我们将论文 [47] 中的 CNN2 作为基准, 与融合框架进行比较。对于测试数据集, 我们得到的不同信噪比条件下的分类准确率如图4.5所示。

通过图4.5, 我们可以发现: 在高信噪比条件下, 除了基于 DNN 融合框架的性能较差, 其他三种框架的差别并不太大, 但是基于 RF 融合框架的模型性能更好一些, 且分类性能较稳定; 在低信噪比条件下, 基于 RF 和 CNN 融合框架的算法性能相近, 优于其它两种算法。

RF 本身是基于集成的思想, 模型本身基于降低方差的思想, 通过随机的样本采样和特征采样来降低泛化误差, 对特征的波动及噪声影响泛化能力更强, 因此性能可能较好。从图4.5中我们也可以发现, RF 相对于其他的融合模型无论是在低信噪比还是在高信噪比性能都具有一定优势。而对于 DNN 的融合模型, 我们发现其性能较差,

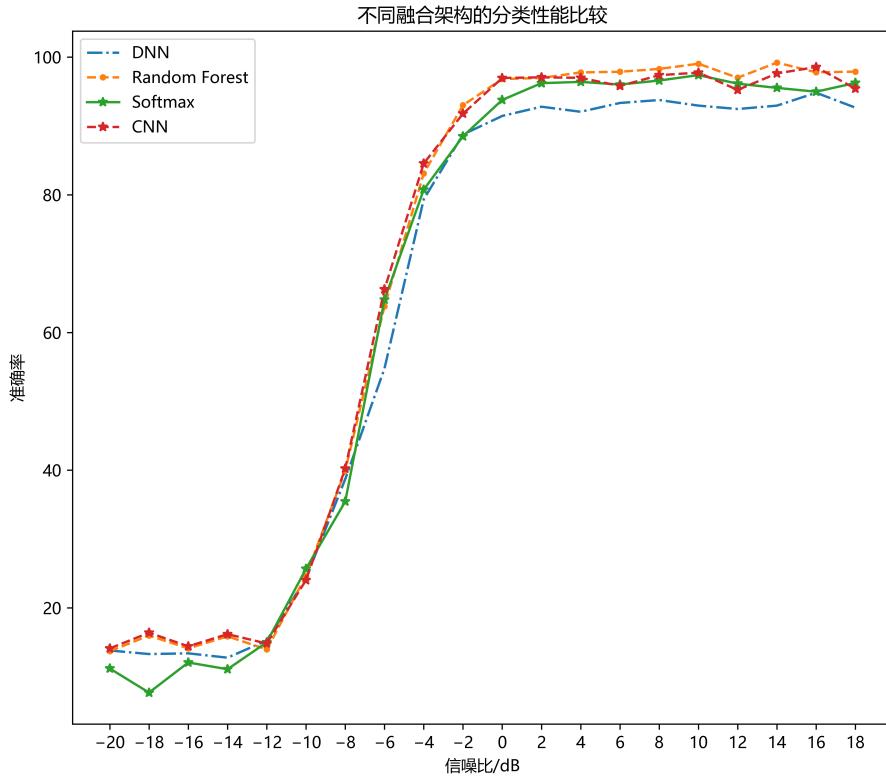


图 4.5 不同融合框架下的分类性能对比

这可能是因为 DNN 本身拟合能力较强，而且随着网络深度增加，训练难度也会增加，泛化能力也较差，因此很难得到较好的性能。

4.5.2 分类混淆矩阵

我们将测试数据 0dB 的分类结果，构建混淆矩阵，得到结果如图4.6。

通过图4.6我们可以发现三种分类器都能得到较为纯净的对角线，但是 DNN 融合模型相对而言差一些，这与上一小节中的结果所得的分类结果相符，主要问题在 QPSK 误分为其他类别。而基于 RF 和 Softmax 的模型，混淆矩阵对角线非常纯净。基于 RF 的模型只有极少量的 QAM 样本误分为 QPSK 和 AM-SSB，基于 Softmax 的模型只有少量 PAM4 样本误分为 AM-SSB。

从 0dB 分类结果来看，三种融合框架下的模型发生误分的信号种类并不相同，这可能是因为模型的特性不同所致；同时，特征标准化处理的方式不同，可能也对误分的结果产生一定的影响。由于误差非常小，我们很难从视觉上很明显地观察出误分结果。

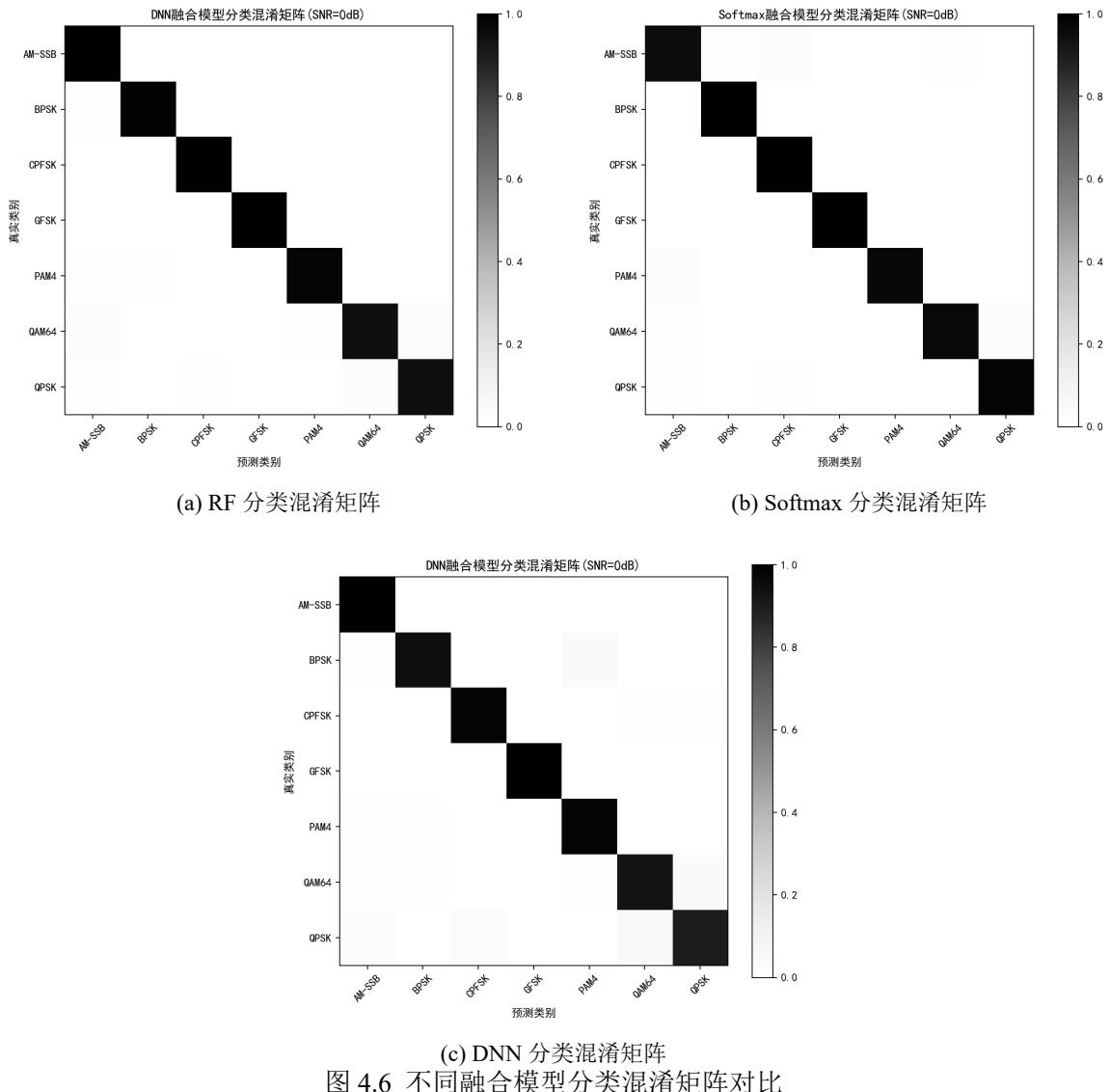


图 4.6 不同融合模型分类混淆矩阵对比

4.6 本章小结

基于特征层面的融合模型，能够综合利用多源特征，实现多特征的优势互补，可以获得更鲁棒、更准确的识别结果。本章我们首先展示了融合模型中所用到的传统时频特征和高阶累积量等特征，并阐述了基本的特征融合理论，提出了三种特征融合的算法模型，并提出了相应的算法框架下的特征标准化方法。算法仿真结果显示，我们提出的特征融合算法中，基于 RF 和基于 Softmax 的性能优于 Oshea 提出的 CNN 模型，并且基于 RF 的融合算法具有相对最优的分类性能，无论在低信噪比还是高信噪比都具有较强的鲁棒性。同时，受融合算法特性等的影响，不同融合框架下的误分信号各不相同。

第五章 调制识别的深度框架研究

5.1 引言

尽管机器学习的目的是提供一种通用的问题解决算法，但目前性能最优的网络模型大部分仍然是对应于特定应用的。目前存在几种较完善的网络，比如包括多层感知机、RNN、CNN 及其许多变体以等，他们被应用在各个领域，取得了很好效果。然而，由于模型所面对的环境不同，数据不同，数据分布也不同；因此，其他领域的很多模型对于我们的调制识别任务并不一定完全适应，需要我们进行相应的探索。

在第三章中，我们基于网络模型的层面，提出了一种不同类型深度模型融合的深度网络框架，并将其应用到无线信号调制识别任务中，而且取得了一定的性能提升。在本章中，我们将以 LeCun 的经典五层网络结构为基准，从网络底层研究网络超参数对调制识别性能的影响，并从欠拟合与过拟合的角度来理解出现这些现象的原因。

5.2 系统模型

在大部分现有的深度神经网络中，卷积层是它们共有的基本网络单元。每个卷积层由若干个卷积核组成，每个卷积核通常非常小（ 1×1 到 5×5 是图像处理中的常见尺寸）。在传统的 DSP 应用中，卷积核通常设计得非常宽，而非设计成很深的多层结构。计算机视觉领域中，神经网络的一个明显趋势是建立更深的网络来学习更复杂的函数变换和层次特征关系。

标准卷积层 \square 的传递函数在方程(5-1)中给出：

$$f(I_i) = f(b_j + \sum_i (k_{ij} * I_i)) \quad (5-1)$$

其中， $f(I_i)$ 是第 i 个卷积核的输出特征图 (Feature Map)， b 和 k 表示卷积核的偏置和卷积核的权重参数， I_i 代表卷积核的输入值， $*$ 表示卷积运算，并且 $f()$ 表示诸如整流线性单元 (ReLU) 或 Sigmoid 单元等激活函数。

本章中，我们所有的网络结构都是基于图5.1的框架，基准的 CNN 在 Softmax 分类层之前有两个卷积层和一个全连接层，Softmax 层之后输出样本分类结果。

整个网络中，我们所变化的只是网络的层数、卷积核的高度和宽度、卷积核的数目等这几个超参数。我们所有的模型训练都将使用 Adam 优化器，因为它提供了梯度归一化和动量的方法，降低了像学习率这样的超参数对模型训练结果的影响。每个卷积层和隐藏层都使用整流线性单元 (ReLU) 作为激活函数，并使用 $dropout = 0.5$ 来降低模型的过拟合，而预测时，我们使所有的单元处于激活状态。

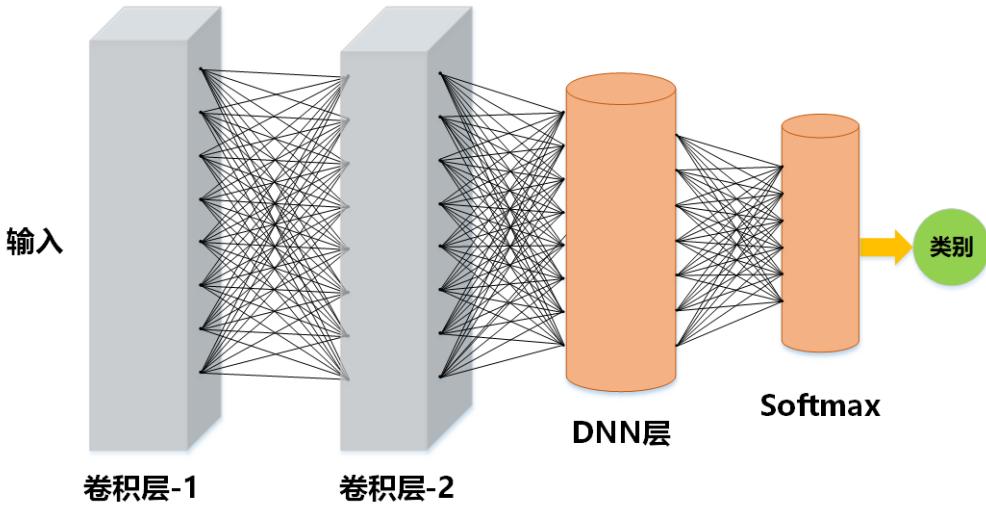


图 5.1 本章所用 CNN 网络框架

5.2.1 模型的偏差与方差

偏差恒量机器学习算法对于输入样本的期望输出与真实值的偏离程度，本质上表征了机器学习算法本身的拟合能力。方差表示算法性能随训练集的变化而产生的波动，本质上表征了数据扰动所对算法性能的影响。噪声则表示当前任务下，我们的训练数据与数据的真实值之间的差距，即任何学习算法所能达到的期望泛化误差的 F 下界，衡量了学习任务本身的难度。

我们假设调制识别的训练样本集为 X , Y 为训练样本的真实类别。对于训练集中的每一个样本 $x_i \in X$, 假设其真实类别为 \hat{y}_i , y_i 为在训练集中的类别。在训练集 X 上学得模型 $\phi(\dots)$, $\phi(x_i; X)$ 为输入样本为 x_i 时的输出, 即预测类别:

$$y_i = \phi(x_i), x_i \in X \quad (5-2)$$

我们假设网络预测的期望为:

$$\bar{\phi}(x) = E_X [\phi(x; X)] \quad (5-3)$$

如果我们使用相同数目的不同训练样本集 \hat{X} 进行训练, 那么我们有方差:

$$\delta_{\hat{X}}^2(x) = E_{\hat{X}} \left[(\phi(x, \hat{X}) - \bar{\phi}(x))^2 \right] \quad (5-4)$$

此时, 模型的噪声为 $\epsilon^2 = E_{\hat{X}} [(\hat{y} - y)^2]$, 表示训练数据与真实数据的误差情况。模型的期望输出与真实标记的差别为偏差, 即:

$$bias^2(x) = (\bar{\phi}(x) - \hat{y})^2 \quad (5-5)$$

我们假设噪声期望为 0, 即 $E_{\hat{X}}[\hat{y} - y] = 0$, 由 \blacksquare 我们有泛化误差为偏差、方差与噪声之和:

$$E(\phi; X) = bias^2(x) + \delta_{\hat{X}}^2(x) + \epsilon^2 \quad (5-6)$$

为了简化模型的分析，我们假设所有的训练样本都是无偏的，即对于任意训练样本 $(x_i, y_i) \in X$ ，我们有 $\hat{y}_i \equiv y_i$ 。此时我们有泛化误差 $E(\phi; X)$ ：

$$E(\phi; X) = bias^2(x) + \delta_X^2(x) \quad (5-7)$$

因此，我们可以从降低偏差和方差的角度出发，来提升模型的性能。

5.2.2 过拟合与欠拟合

在训练机器学习模型时，我们以降低训练误差为目标，获得模型的参数。然而，我们不仅希望学习算法在训练集上表现良好，更希望其在非训练的数据集上同样具备良好的性能。泛化正是衡量机器学习算法在未知数据上表现能力良好的一种定性标准。泛化误差可以定义为模型对于新输入未知样本的误差期望。欠拟合是指模型不能在训练集上获得足够低的误差，而过拟合是指测试误差大于训练误差。

模型的容量是指模型的拟合能力。容量较低的模型其拟合能力较弱，很难反映训练集的数据分布。容量高的模型，由于其拟合能力过强，可能会将噪声数据也学习到模型中，从而发生过拟合。我们可以通过调整模型的容量，控制模型是否倾向于过拟合或者欠拟合。

通过小节5.2.1中的结果可知，模型的泛化性能是由学习算法本身的拟合能力、数据的充分性以及学习任务本身的难度共同决定的。

对于给定的学习任务 T ，为了取得较好的泛化性能，我们需要使模型具备较小的偏差，即模型可以充分拟合训练数据，并且还要使模型的方差处于较小值，即使得数据扰动对模型产生的影响尽量小。

然而，实际中偏差与方差却具备相反的倾向性。在我们对算法进行训练时，如果训练不足，则机器学习模型的拟合能力不够强，训练数据的扰动不足以使模型产生显著变化，此时偏差主导了泛化误差；而随着训练程度的加深，模型的拟合能力逐渐增强，训练数据产生的扰动逐渐能被模型所学习，此时方差逐渐主导了泛化误差；在训练程度达到临界条件以后，如验证误差超过了训练误差一定阈值，模型的拟合能力已经非常强，训练数据发生的轻微扰动都会导致模型发生显著的变化，此时，模型可能学习到训练数据某些样本自身的、非全局性的特性，这就会产生模型的过拟合。

5.3 网络超参数对调制识别的影响

网络的超参数，如学习速率，每层卷积核的数量，卷积核的大小以及卷积层的数目等，会影响网络规模，特征提取的数目等，如果不具备一定的领域先验知识很难优化。

在本节中，我们将研究网络卷积核以及卷积层等相关超参数对分类性能的影响，这可以为我们后续的研究提供一定的网络整体架构的参考，节省了后续研究的时间成本，并有助于进一步探究新的网络底层架构，以提升调制识别的性能。

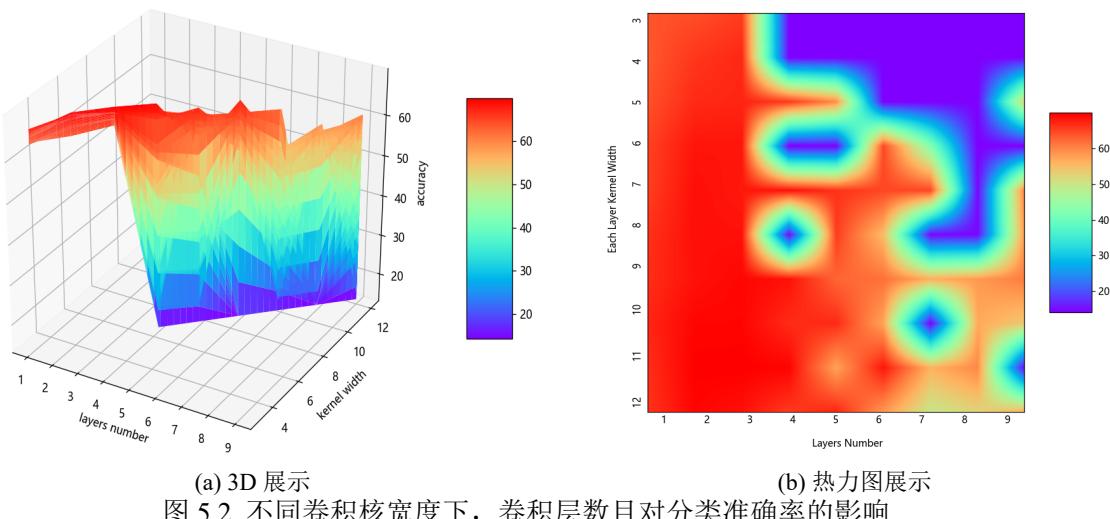
5.3.1 网络层数对调制识别的影响

在我们的调制识别场景中应用的是宽度较小的卷积核，由于小卷积核无法获取全局特征，所以我们通过增加网络层数，多层卷积核的局部感知野逐渐叠加后，较深层的卷积核的局部感知野也会逐渐扩大，进而学得全局特征。因此，卷积层的深度，在一定程度上决定了网络的拟合能力。

在本小节我们将探索网络层数对调制识别的影响。由于我们仅仅是探索卷积层深度对分类性能的影响，而且网络的训练本身是一个非常耗时的过程。因此，为了降低时间成本并简化分析提供一个定性的结果，我们在本小节中的仿真所有卷积核数目都为 32，并将卷积核的大小从 1×3 到 1×12 变化。

网络层次深度对训练数据遍历次数的要求差别很大，为了解决这个问题，我们将所有的网络 epoch 设置为一个较大值（本文中 $epoch = 500$ ），以满足那些对于复杂网络对于较高次数 epoch 训练的需求。同时，我们使用了预停止策略，假设训练过程中，训练损失在 N_{pre_stop} （本文中我们设置为 $N_{pre_stop} = 10$ ）个 epoch 内不降低，即在这段训练时间范围内，模型不能得到很好的优化，此时我们则会停止训练。这样便可以在尽量保证准确率的同时提高训练效率。

根据我们的预设条件，系统的仿真结果如图5.2所示。



通过图5.2我们可以发现，分类性能具有一个相对整体性的趋势：当卷积层大于 3 时，分类准确率几乎成下降趋势。通过图5.2a我们可以发现，系统性能在卷积层为

3, 卷积核的宽度为 6 以上时, 系统性能能够达到一个较高的水准; 同时我们可以发现卷积核宽度较大时, 卷积核层数在大于 3 时分类准确率几乎呈一个断崖式下跌。

在图5.2b中我们可以清楚地发现, 整幅图像在卷积层数目较大、卷积核数目较小时, 系统的性能较差。卷积层数目小于 4 时, 系统性能随卷积核宽度变化不大, 增大卷积核宽度并不能很大程度上提升系统分类性能; 卷积层较大时, 增加卷积层的宽度反而会降低系统性能。

接下来, 我们以卷积层数目作为变量, 观察不同卷积核宽度条件下的分类性能, 结果如图5.3。

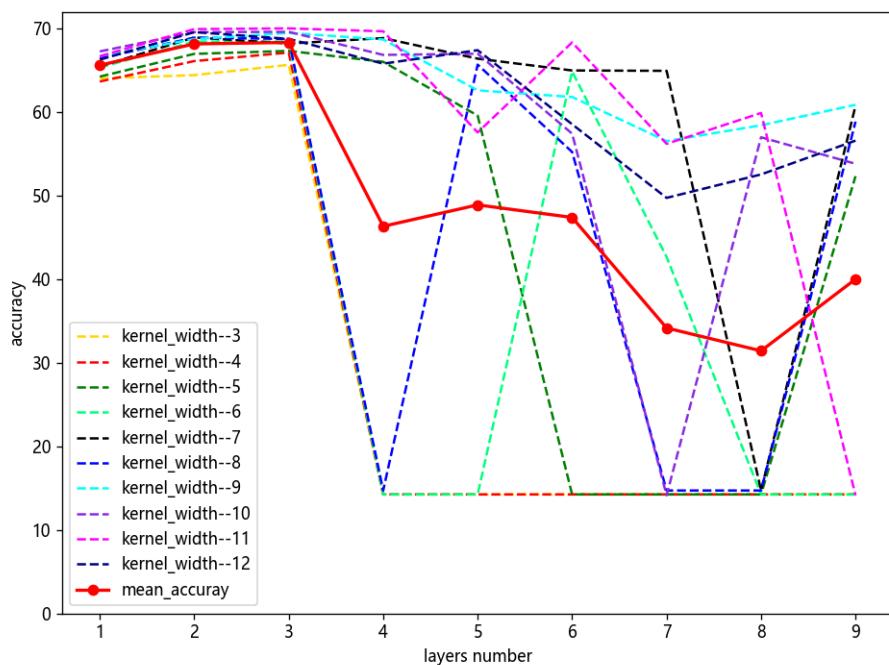


图 5.3 分类准确率随卷积层数目和卷积核宽度的变化

从图5.4可以发现, 整体而言, 不同卷积核宽度的平均分类准确率, 随着卷积层数目增加, 首先先轻微地上升, 然后出现较大幅度的下降。

增加卷积层的深度, 在卷积层数目 $N_{layers} \leq 3$ 时, 可以微小地改善分类性能。当卷积层深度 $N_{layers} \geq 4$ 时, 改变卷积层的数量在分类准确率几乎都呈下降趋势; 但是, 不同卷积核宽度的网络性能下降程度不同, 卷积核宽度较大的网络越, 随着卷积层深度的增大性能下降程度较低。在卷积层深度 $N_{layers} \leq 3$ 时, 我们发现分类性能几乎随着卷积核宽度的增加而增加, 直到卷积核宽度达到 11 时, 达到分类性能的上限; 同时, 我们发现分类性能在此时随着卷积层数增加而增加。

在卷积层数较低时, 分类性能有所增加, 这可能是因为模型的拟合能力不足以完全表征信号的特性。在卷积层数 $N_{layers} \geq 4$ 时, 网络性能大幅下降; 因为卷积核宽度较大时, 分类性能仍然处于一个较高的水准, 而且分类性能会随着卷积层数目的增加

而发生剧烈的突变，所以我们可以确定此时导致分类性能下降的主要原因是网络的训练难度增加；其次，网络性能在卷积层数目为3时达到了性能的最高点，之后性能一直呈下降趋势，我们认为这是网络发生过拟合的一个节点。这表明对于我们的数据而言，可能是由于调制数据通常只改变正弦曲线的幅度，频率或相位，因此对于调制方式而言数据的复杂度并不高，没有必要学习更高层次的深度特征。

5.3.2 卷积核数目对调制识别的影响

CNN是通过增加网络的深度来增强拟合能力的，而网络对于数据特征的提取，主要是通过不同的卷积核进行参数学习得到的，卷积核的数目在一定程度上表征了网络所能学习特征的数目，即网络对数据理解的维度高低。

上一小节我们探究的是卷积层数目对调制识别的影响，本节我们将对网络结构进行更细化的研究，探索卷积核数目对调制识别的影响。

在第三章中，我们发现第一层卷积核为 1×7 ，第二层卷积核为 2×6 时，分类器具备较好的性能。因此，在本小节中，我们所有的仿真都基于这样的卷积核结构进行，变化的仅仅是卷积核的数目。由于我们训练神经网络时间成本较高，很难对每一个卷积核数目参数进行遍历，所以我们选择一些有代表性的参数进行试验，主要目的是定性的分析卷积核数目的变化对于分类性能的影响。我们定义卷积核数目集合 $\kappa = \{16, 32, 48, 64, 96, 128, 196, 256\}$ ，第一个卷积层与第二个卷积层的卷积核数目，分别从集合 κ 中按照卷积核数目由小到大选取，卷积核变化范围为16~256。这样两个卷积层的核数目组合共有 $|\kappa|^2 = 64$ 种，即网络训练时总的遍历次数为 $|\kappa|^2$ 次，复杂度为 $O(n^2)$ 。同时使用自适应学习速率的Adam优化器进行训练，同样设置较大的epoch并应用预停止策略，来探索卷积核数目对分类性能的影响。

通过对不同卷积核数目组合的网络进行训练，我们最终得到分类器性能随卷积核数目变化的情况，图5.4显示了两个卷积层中卷积核数量从16增加到256时，样本分类准确率的变化情况。

通过图5.4，我们可以发现，在第一层卷积核数目为16，即第一层卷积核数目较小时，网络的性能几乎随着第二层卷积核数目的增大而减小。而随着第一层卷积核增大，改变第二层卷积核数目，系统分类准确率变化很小，只有轻微的波动。由于我们使用的是较小的卷积核，因此单卷积层的卷积核的局部感知野有限，我们通过增加卷积层的数目来增大卷积核的局部感知野。当我们第一个卷积层的卷积核数目较小时，我们所得的特征图的数目也相对较少，而此时增大第二个卷积层的卷积核数目，由于第二层卷积核本身所能感知的信息较多，我们第一个卷积层提取的信息量有限，因此第二层很难得到有效训练。所以，我们认为之所以这种情况下网络性能下降较为明显，是因为网络本身训练难度增加所致。

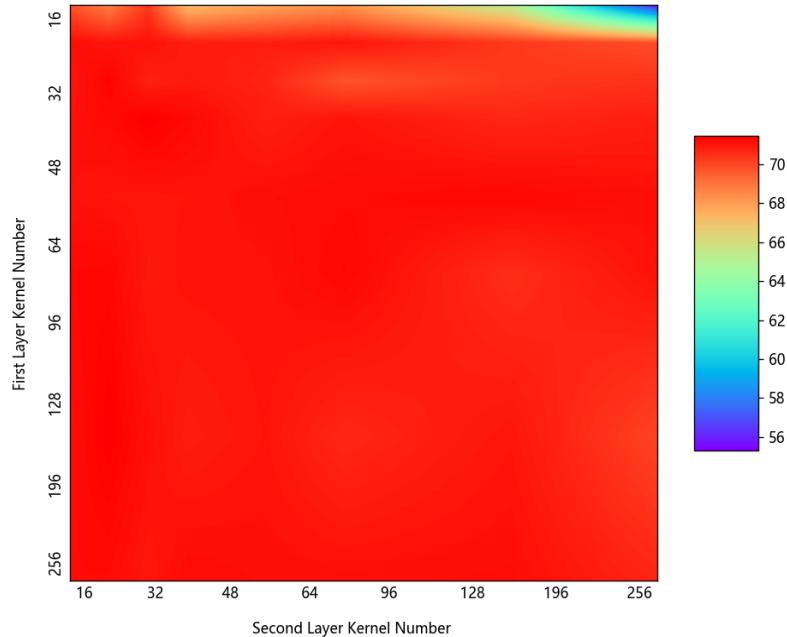


图 5.4 分类准确率随卷积核数目的变化情况

图5.5中，我们以第一个卷积层中卷积核数目为横轴，展示第一层和第二层卷积核数目变化对分类性能的影响。

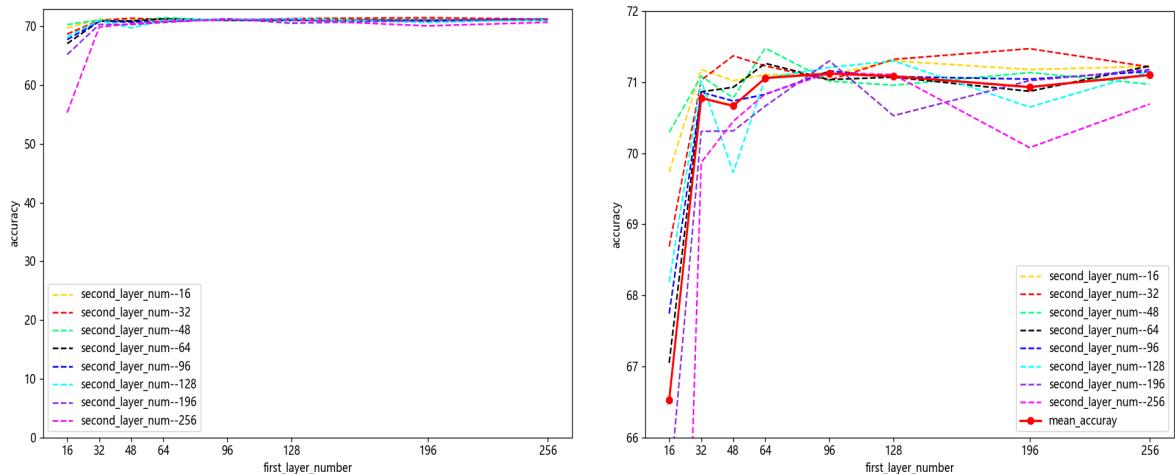


图 5.5 第一卷积层中卷积核数目对分类准确率的影响

从图5.5a我们可知第一卷积层中卷积核数目 $N_{kernel}^{(1)} \leq 32$ 时，系统性能会随着第二层中卷积核的数目变化而发生较大的波动；而当 $N_{kernel}^{(1)} \geq 48$ 时，第二个卷积层中卷积核数目变化对系统性能影响较小。这可能是由于在第一层卷积核数目较小时，偏差主导了我们模型的泛化性能。

从图5.5b我们可以发现，第一卷积层中卷积核数目 $N_{kernel}^{(1)} \leq 32$ 时，第二层卷

积核数目的变化与性能没有明确的相关关系，但是，整体而言，第二层的卷积核数目在 $N_{kernel}^{(2)} \leq 48$ 时能取得较好的性能。而在第一个卷积层数目 $N_{kernel}^{(1)} \geq 48$ 时，系统性能波动较大，此时可能是模型训练时的鞍点等因素所致。当我们以第一层卷积核数目作为自变量，第二层所有数目卷积核分类准确率的均值作为因变量时，即图5.5b中红色的粗线，我们发现其整体上准确率呈先上升后下降的趋势，系统性能在 $N_{kernel}^{(1)} = 64, N_{kernel}^{(2)} = 64$ 时取得最优。

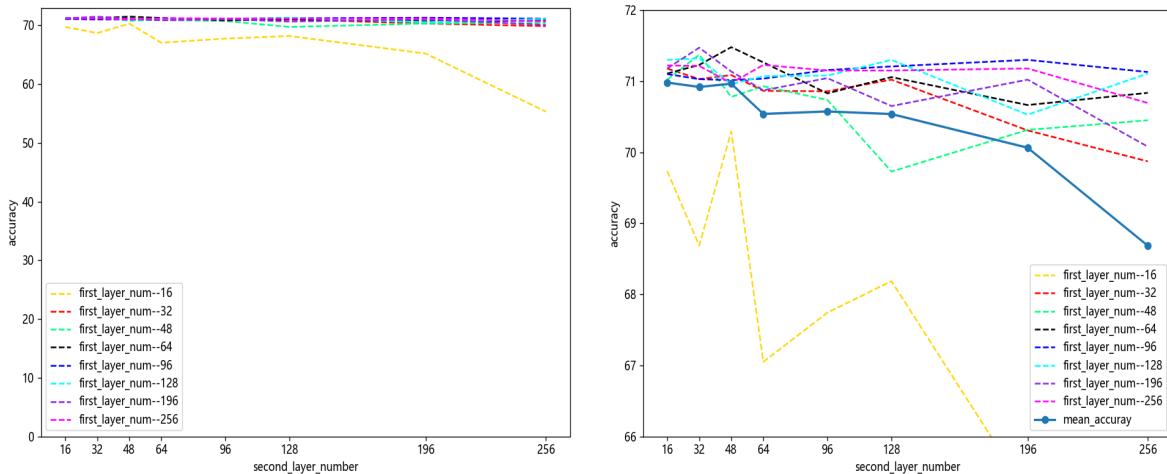


图 5.6 第二卷积层中卷积核数目对分类准确率的影响

从图5.6a中我们可以发现，分类性能整体随着第二卷积层数目的增加而下降。而且，当第一卷积层中卷积核数目为 16 时，性能下降的最为明显。图5.6b中我们可以更清楚的看到，当第一层卷积核数目较小时，系统性能随着第二卷积层中卷积核数目的增加而下降得很明显，而且波动幅度较大；从平均准确率而言，我们也可以看到图中粗的红色实线也是呈下降趋势，即系统整体性能几乎是随着第二卷积层数目的增加而下降的。

由于网络本身的深度较小，当卷积核数目很小时，受限于卷积核数目的限制，特征提取的数目也相应的较少，很难较完整地反映数据本身的特征，分类准确率的上限会出于一个较低值。当卷积核数目增加，相当于我们的特征数目也在相应的增加，这样，网络本身的拟合能力增强，卷积核所提取的卷积特征能更好地反映数据的内在特征，因此，分类性能的上界相应提高，可以获得较好的分类结果。

当卷积核的数目增大到一定的程度，网络的拟合能力过强，不仅能够拟合出数据本身的性质，同时将一些异常点也进行了拟合，这就造成了过拟合；同时，由于我们的数据量有限，也很难将网络参数训练到一个较优的水平。因此，我们的分类性能可能会随着卷积核数目的增加而降低。

由于不同卷积层中的卷积核局部感知野不同，越深度的卷积核相对而言观察的

信息范围更广，因此，更深层次的卷积层中卷积核数目应该相对较少；如果较深层的卷积核数目大于较浅层中卷积核数目，模型就容易发生过拟合现象。

5.3.3 卷积核大小对调制识别的影响

在 CNN 中，特征提取的过程从底层卷积核到高层卷积核，层层接收局部的输入，最后不断聚合。对于卷积核而言，卷积核的高度 $height$ 和宽度 $width$ 决定了卷积核感知野的范围。对于卷积层的高度而言，由于我们的样本是一个 2×128 的时间序列，我们可以看成是由两个 128 维的向量组成。其中，第一行表示信号的实部，第二行为信号的虚部。这样，我们在利用卷积核对信号进行卷积特征提取时，如果卷积核的宽度为 1，相当于是在信号的实部或者虚部进行采样值的卷积运算；如果卷积核的宽度为 2 表示不仅对信号进行实部的运算，同时考虑到了信号的虚部，这样进行卷积时，可能同时会包含信号的能量信息。在本小节我们将探究不同卷积核的高度和宽度对调制识别性能的影响。

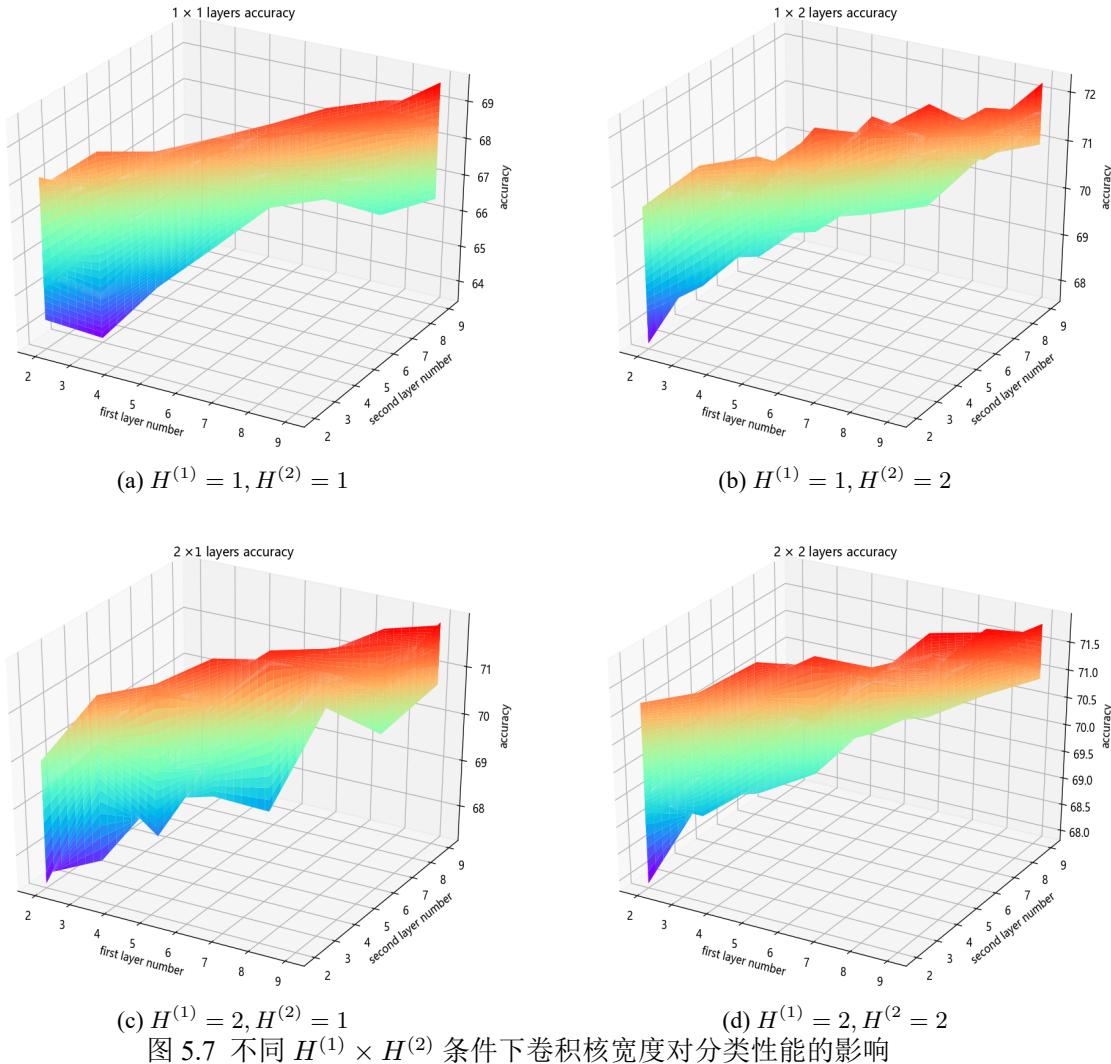
5.3.3.1 卷积核宽度对分类性能的影响

根据第三章我们采用的 CNN 结构，我们将第一个卷积层的卷积核数目固定在 64，将第二个卷积层的卷积核数目固定在 32，因为这种情况下网络具有较好的分类性能。我们假设卷积核高度 $H^{(1)}$ 和 $H^{(2)}$ 在 1 与 2 之间变化，卷积核的宽度 $W^{(1)}$ 和 $W^{(2)}$ 在 3 到 9 之间变化。则总的训练迭代次数为 $8 \times 8 \times 2 \times 2 = 256$ 。我们同样使用 Adam 优化器，并应用预停止策略，最后分类结果如图5.7所示。

从图5.7我们可以发现一个整体的趋势，即卷积核 $W^{(1)}$ 和 $W^{(2)}$ 小的卷积核性能没有其宽度大的卷积核性能强。从达到的性能上界而言，图5.7b的最值是超过 72% 的，高于其他三种情况，也就是说网络的卷积层高度为 $H^{(1)} = 1$, $H^{(2)} = 2$ 时，模型具有最优分类性能。这可能是因为随着卷积核的宽度增加，网络变得复杂，模型的容量变大，更能够捕获数据本身所携带的信息，所以呈现出较宽的卷积核具有较强的分类性能。

分类性能在两层高度 $H^{(1)} = 1$ 和 $H^{(2)} = 2$ 时达到性能上界，这可能是由于我们的数据本身表示的是信号的 $I \times Q$ 两路的信号，而 I 路信号与 Q 路信号本身呈希尔伯特变换的关系，所以，我们在第一个卷积层利用高度 $H^{(1)} = 1$ 的卷积核分别提取每一路的信息，更符合数据本身逻辑上的关系，因此也更有可能取得好的效果。

从四种情况而言，图5.7d在第二卷积层宽度为 $W^{(2)} = 3$ 时分类性能就已经达到一个较高的水平，相对其他三种情况而言容易更容易达到好的性能；而图5.7a所示的网络最慢达到一个性能的较高水平。这可能是由于图5.7d中，卷积核 $H^{(1)} = 2$ 和 $H^{(2)} = 2$ ，卷积核宽度大，捕获数据信息能力较强，所以性能较容易在小卷积核时达

图 5.7 不同 $H^{(1)} \times H^{(2)}$ 条件下卷积核宽度对分类性能的影响

到较高水平。

5.3.3.2 卷积核宽度对分类准确率的影响

我们进一步利用热度图来观察卷积核宽度 $W^{(1)}$ 和 $W^{(2)}$ 对分类性能的影响，结果如图5.8所示。

从图5.8中我们可以发现：对卷积核的宽度而言，整体来看较小的卷积核不如较大的卷积核，但是当卷积的宽度增大到7之后，分类性能较为稳定，变化较小。从图5.8中每个子图蓝色的分布而言，我们发现图5.8d占比最小，即其在卷积核宽度更小时就可以达到较好的性能；而图5.8a蓝色占比最大，而且颜色较深，所以其只能在卷积核宽度较大时才能达到较好的性能。这说明我们卷积核的宽度越大，拟合能力可能越强，获取数据的信息越多。

同时，我们发现图5.8中各个子图蓝色的分布情况各不相同：图5.8a和5.8d中蓝色几

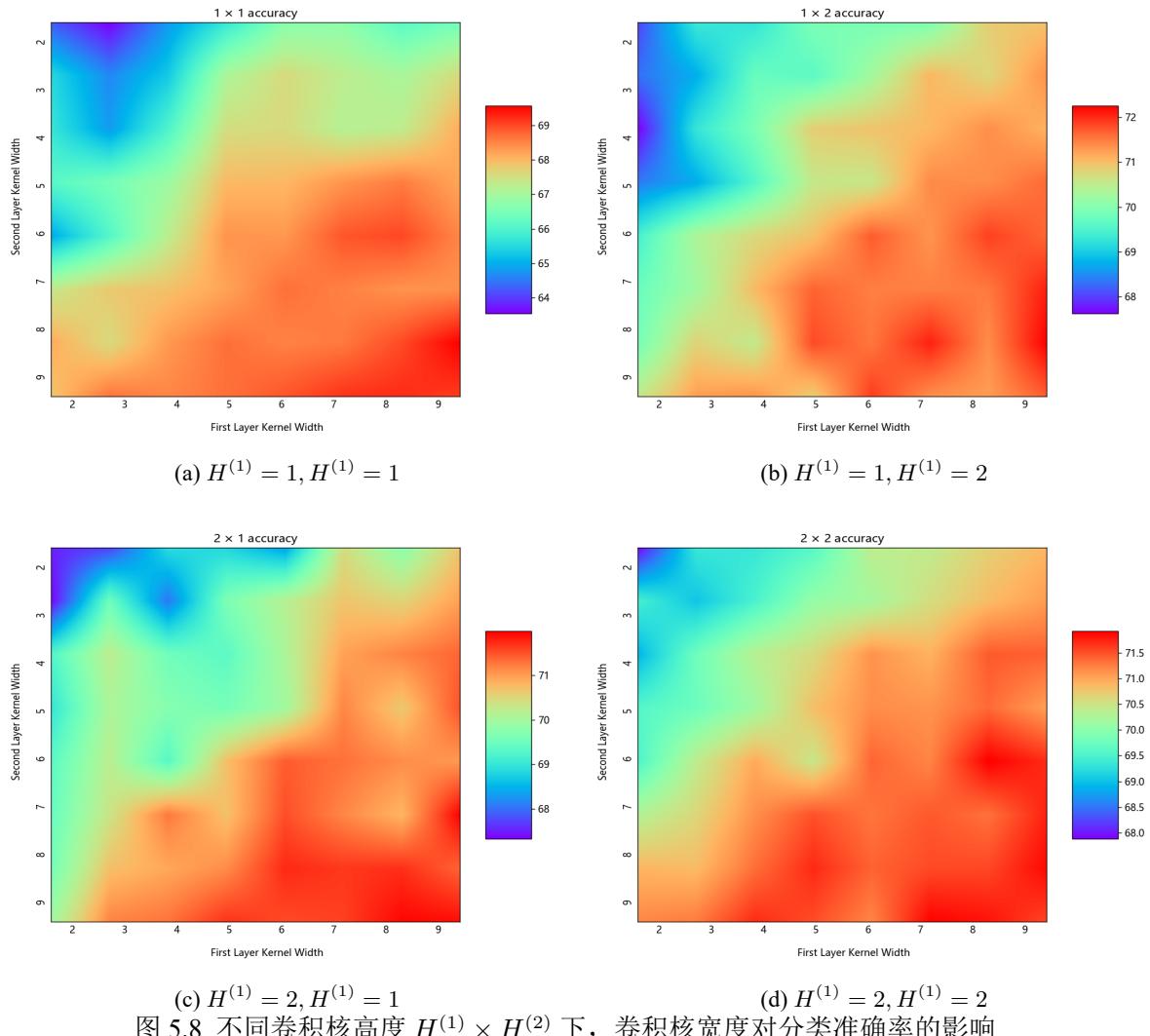


图 5.8 不同卷积核高度 $H^{(1)} \times H^{(2)}$ 下，卷积核宽度对分类准确率的影响

乎分布在对角线上，但是图5.8a中 $W^{(1)}$ 的变化对蓝色分布的影响更大一些；图5.8b中蓝色分布主要沿 $W^{(2)}$ 变化；图5.8c中蓝色分布主要沿 $W^{(1)}$ 变化。这说明，如果两个卷积层中卷积核的宽度相同，则改变两个层中卷积核的宽度，对于分类器的性能影响具有相似性；但是，可能是因为我们的应用场景针对的是 $I \times Q$ 两路信号，当 $W^{(1)} = 1, W^{(2)} = 1$ 时，改变 $W^{(1)}$ 对于分类性能影响更大一些。图5.8b中，由于 $W^{(2)} = 2$ ，在卷积核宽度较小时，由于模型处于欠拟合状态，此时偏差主导了模型的泛化误差，而 $W^{(2)}$ 的增加对于网络容量提升更快，因此，改变 $W^{(2)}$ 会取得更好的性能提升。图5.8c的现象，可能也是这种原因。

5.4 本章小结

在本章中，我们从网络底层研究网络超参数对调制识别性能的影响，并从欠拟合与过拟合以及偏差与方差的角度来理解出现这些现象的原因。CNN 在调制识别时的性能似乎不受网络深度的限制，在卷积层数目为 3 时即可达到性能的最值，此时再增大卷积层的深度，系统性能可能会发生下降。第一层卷积核数目较小时，网络的性能几乎随着第二层卷积核数目的增大而减小。而随着第一层卷积核增大，改变第二层卷积核数目，系统分类准确率变化很小，只有轻微的波动；卷积核数目 $N_{kernel}^{(1)} = 64, N_{kernel}^{(2)} = 48$ 时，系统取得所能观察到的最好性能；当 $N_{kernel}^{(1)}$ 较小时，增大 $N_{kernel}^{(2)}$ 到超过第一个卷积层的核数目时，模型的性能会下降。整体而言，卷积核宽度 $W^{(1)}$ 和 $W^{(2)}$ 小的卷积核性能没有宽度大的卷积核性能强，但是当卷积的宽度增大到 7 之后，分类性能较为稳定，变化较小；卷积核的高度 $H^{(1)}$ 和 $H^{(2)}$ 不同时，改变每一层的卷积宽度对系统造成的影响不同，相对而言，改变高度 H 大的卷积核对系统的性能影响较大。

第六章 总结与展望

6.1 研究结论

通信信号的调制识别是认知无线的重要组成部分，其在军用领域与民用领域都有重要的作用，需要我们去进一步研究实现调制的自动识别。深度学习的发展对很多行业产生了深远的影响，其在通信中的应用也获得了广泛的关注；如何将深度学习更好地用于通信领域中，需要我们做出进一步的探索。目前深度学习应用于调制识别，在学术领域已经得到了很大的关注，并有前人做出了一定的贡献。本文正是将深度学习中的算法应用于调制识别中，主要贡献如下：

首先，本文提出了一种 CAE-CNN 的深度学习框架。将 CNN 分类器的卷积层部分作为 CAE 的编码层，然后分别在其后接入解码器和类似于 DNN 的分类层，融合 CAE 的重构误差与 CNN 的分类误差，来提高分类准确率。同时，本文提出了一种 CAE-CNN 框架的训练算法，将 CAE 与 CNN 以不同轮询次数的循环进行训练，并在训练过程中以参量的形式，不断调整训练 CAE 与 CNN 的循环次数，不断降低重构误差与分类误差。最终的分类结果我们发现，在信噪比为-4dB 时，7 种信号的分类准确率仍在 85% 左右，这相对于传统的方法是具有一定的优势。

接下来，我们提出了一种传统特征与深度特征融合的框架。传统的调制识别研究已经持续数十年，相对比较成熟，对于不同的调制方式有很多特征可以进行有效区分。而我们利用深度学习也可以提取信号的一系列深度特征，而利用这些特征我们同样可以对信号进行分类。我们分别利用 Softmax、DNN、RF 三种融合框架对传统特征与深度特征进行融合，构造新的分类器。最终我们发现，使用 RF 框架的特征融合算法可以取得较高的准确率，在 0dB 时，7 种信号的分类准确率可以达到 90% 以上。同时，我们发现特征归一化在 DNN 和 Softmax 这样的融合框架中可以有效提升分类准确率。

最后，们从网络底层研究网络超参数对调制识别性能的影响，并从欠拟合与过拟合以及偏差与方差的角度来理解出现这些现象的原因。CNN 在调制识别时的性能似乎不受网络深度的限制，在卷积层数目为 3 时即可达到性能的最值，此时再增大卷积层的深度，系统性能可能会发生下降。第一层卷积核数目较小时，网络的性能几乎随着第二层卷积核数目的增大而减小。而随着第一层卷积核增大，改变第二层卷积核数目，系统分类准确率变化很小，只有轻微的波动。整体而言，卷积核宽度 $W^{(1)}$ 和 $W^{(2)}$ 较大的网络性能较好，但是当卷积的宽度增大到 7 之后，分类性能较为稳定，仅仅发生微小的波动。

无线信号调制识别是一种不断发展的无线认知技术，更多复杂的调制方式可能会出现，这给现有的调制识别系统带来了很大的挑战，需要我们进行进一步的研究。深度学习的发展很大程度上提高了机器学习算法性能的上界，而其在调制识别领域得到的应用还相对较少，需要我们进行包括但不限于算法融合、特征融合、数据融合、深度网络框架等相应的研究。

6.2 未来展望

本文的算法是在高斯白噪声条件下进行的，如何在瑞利、多径信道条件下，利用深度学习实现较好性能的调制信号识别，仍然需要继续深入研究。随着数字调制技术的发展，新的、更复杂的调制方式不断出现，如何通过深度学习的方法，去发现位置类别的信号，将会是我们后续的研究点。随着深度学习的发展，各种网络框架层出不穷，网络的底层结构也有诸如 LSTM、CNN、CLDNN、InceptionNet、GoogleNet 等框架，如何将这些框架应用到调制识别中，什么样的网络框架可以提高我们分类识别的性能，我们目前正在探索，这也将是后续研究的重点。我们希望通过研究深度学习在调制识别中的应用，能够提高调制识别的性能，并促进深度学习在通信领域的发

展。

参考文献

- [1] 樊昌信, 曹丽娜. 通信原理 [M]. [S.I.]: 国防工业出版社, 2006.
- [2] NANDI A, AZZOUZ E E. Automatic analogue modulation recognition[J]. Signal processing, 1995, 46(2): 211–222.
- [3] 程汉文, 朱雷, 吴乐南, et al. 基于累计量的干扰信号调制识别算法 [J]. 电子与信息学报, 2009, 31(7): 1741–1745.
- [4] 张春磊, 杨小牛. 认知电子战初探 [J]. 通信对抗, 2013(2): 1–4.
- [5] HINTON G E, OSINDERO S, TEH Y-W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527–1554.
- [6] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. nature, 2015, 521(7553): 436.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C] // Advances in neural information processing systems. 2012 : 1097–1105.
- [8] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper With Convolutions[C] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.
- [9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016 : 770–778.
- [10] OSAKO K, SINGH R, RAJ B. Complex recurrent neural networks for denoising speech signals[C] // Applications of Signal Processing to Audio and Acoustics (WASPAA), 2015 IEEE Workshop on. 2015 : 1–5.
- [11] GUPTA A, WANG H, GANAPATHIRAJU M. Learning structure in gene expression data using deep architectures, with an application to gene clustering[C] // Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on. 2015 : 1328–1335.
- [12] GATYS L A, ECKER A S, BETHGE M. A neural algorithm of artistic style[J]. arXiv preprint arXiv:1508.06576, 2015.
- [13] WEAVER C, COLE C, KRUMLAND R, et al. THE AUTOMATIC CLASSIFICATION OF MODULATION TYPES BY PATTERN RECOGNITION.[R]. [S.I.]: STANFORD UNIV CALIF STANFORD ELECTRONICS LABS, 1969.
- [14] KIM K, POLYDOROS A. Digital modulation classification: the BPSK versus QPSK case[C] // Military Communications Conference, 1988. MILCOM 88, Conference record. 21st Century Military Communications-What's Possible? 1988 IEEE. 1988 : 431–436.
- [15] HWANG C, POLYDOROS A. Advanced methods for digital quadrature and offset modulation clas-

- sification[C] // Military Communications Conference, 1991. MILCOM'91, Conference Record, Military Communications in a Changing World., IEEE. 1991 : 841 – 845.
- [16] HUAN C-Y, POLYDOROS A. Likelihood methods for MPSK modulation classification[J/OL]. IEEE Transactions on Communications, 1995, 43(2/3/4) : 1493 – 1504. <http://dx.doi.org/10.1109/26.380199>.
- [17] SCHREYOGG C, REICHERT J. Modulation classification of QAM schemes using the DFT of phase histogram combined with modulus information[C/OL] // MILCOM 97 MILCOM 97 Proceedings : Vol 3. 1997 : 1372 – 1376 vol.3. <http://dx.doi.org/10.1109/MILCOM.1997.644992>.
- [18] WEI W, MENDEL J M. Maximum-likelihood classification for digital amplitude-phase modulations[J]. IEEE transactions on Communications, 2000, 48(2) : 189 – 193.
- [19] BOITEAU D, LE MARTRET C. A general maximum likelihood framework for modulation classification[C] // Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on : Vol 4. 1998 : 2165 – 2168.
- [20] LIEDTKE F. Computer simulation of an automatic classification procedure for digitally modulated communication signals with unknown parameters[J]. Signal processing, 1984, 6(4) : 311 – 323.
- [21] AZZOUZ E E, NANDI A K. Automatic identification of digital modulation types[J]. Signal Processing, 1995, 47(1) : 55 – 69.
- [22] WONG M D, NANDI A K. Automatic digital modulation recognition using spectral and statistical features with multi-layer perceptrons[C] // Signal Processing and its Applications, Sixth International, Symposium on. 2001 : Vol 2. 2001 : 390 – 393.
- [23] 杨发权. 无线通信信号调制识别关键技术与理论研究 [D]. [S.l.]: 西安电子科技大学, 2015.
- [24] REICHERT J. Automatic classification of communication signals using higher order statistics[C] // Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on : Vol 5. 1992 : 221 – 224.
- [25] SWAMI A, SADLER B M. Hierarchical digital modulation classification using cumulants[J]. IEEE Transactions on communications, 2000, 48(3) : 416 – 429.
- [26] MOBASSERI B G. Constellation shape as a robust signature for digital modulation recognition[C] // Military Communications Conference Proceedings, 1999. MILCOM 1999. IEEE : Vol 1. 1999 : 442 – 446.
- [27] MOBASSERI B G. Digital modulation classification using constellation shape[J]. Signal processing, 2000, 80(2) : 251 – 277.
- [28] O'SHEA T J, CORGAN J, CLANCY T C. Convolutional radio modulation recognition networks[C] // International Conference on Engineering Applications of Neural Networks. 2016 : 213 – 226.
- [29] O'SHEA T J, WEST N. Radio machine learning dataset generation with gnu radio[C] // Proceedings

- of the GNU Radio Conference : Vol 1. 2016.
- [30] MENDIS G J, WEI J, MADANAYAKE A. Deep learning-based automated modulation classification for cognitive radio[C] // Communication Systems (ICCS), 2016 IEEE International Conference on. 2016 : 1 – 6.
- [31] ALI A, YANGU F. Automatic Modulation Classification Using Deep Learning Based on Sparse Autoencoders With Nonnegativity Constraints[J]. IEEE Signal Processing Letters, 2017, 24(11) : 1626 – 1630.
- [32] RAJENDRAN S, MEERT W, GIUSTINIANO D, et al. Distributed deep learning models for wireless signal classification with low-cost spectrum sensors[J]. arXiv preprint arXiv:1707.08908, 2017.
- [33] TANG B, TU Y, ZHANG S, et al. Digital Signal Modulation Classification with Data Augmentation Using Generative Adversarial Nets in Cognitive Radio Networks[J/OL]. IEEE Access, 2018, PP(99) : 1 – 1. <http://dx.doi.org/10.1109/ACCESS.2018.2815741>.
- [34] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain.[J]. Psychological review, 1958, 65(6) : 386.
- [35] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C] // Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. 2011 : 315 – 323.
- [36] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models[C] // Proc. icml : Vol 30. 2013 : 3.
- [37] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. nature, 1986, 323(6088) : 533.
- [38] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4) : 541 – 551.
- [39] SAXE A M, KOH P W, CHEN Z, et al. On Random Weights and Unsupervised Feature Learning.[C] // ICML. 2011 : 1089 – 1096.
- [40] RUDER S. An overview of gradient descent optimization algorithms[J]. arXiv preprint arXiv:1609.04747, 2016.
- [41] BOTTOU L. Large-scale machine learning with stochastic gradient descent[G] // Proceedings of COMPSTAT’2010. [S.l.] : Springer, 2010 : 177 – 186.
- [42] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [43] TIELEMANS T, HINTON G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude[J]. COURSERA: Neural networks for machine learning, 2012, 4(2) : 26 – 31.
- [44] SCHAUL T, ANTONOGLOU I, SILVER D. Unit tests for stochastic optimization[J]. arXiv preprint arXiv:1312.6055, 2013.

- [45] MIGLIORI B, ZELLER-TOWNSON R, GRADY D, et al. Biologically inspired radio signal feature extraction with sparse denoising autoencoders[J]. arXiv preprint arXiv:1605.05239, 2016.
- [46] SCHMIDT M, BLOCK D, MEIER U. Wireless interference identification with convolutional neural networks[J]. arXiv preprint arXiv:1703.00737, 2017.
- [47] OSHEA T J, CORGAN J, CLANCY T C. Convolutional Radio Modulation Recognition Networks[J]. arXiv: Learning, 2016 : 213 – 226.
- [48] LE Q V. Building high-level features using large scale unsupervised learning[C]// Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. 2013 : 8595 – 8598.
- [49] 刘渭滨, 邹智元, 邢薇薇. 模式分类中的特征融合方法 [J]. 北京邮电大学学报, 2017, 40(4) : 1 – 8.
- [50] 杨杰. 通信信号调制识别: 原理与算法 [M]. [S.l.]: 人民邮电出版社, 2014.
- [51] 张利, 李青. 基于高阶累积量的调制识别算法的研究 [J]. 信息工程大学学报, 2017(4) : 403 – 408.
- [52] 李世平, 陈方超. 基于小波和高阶累积量的数字调制识别算法 [J]. 计算机应用, 2011(11) : 2926 – 2928,2935.
- [53] 韩钢, 张文红, 李建东, et al. 基于高阶累积量和支撑矢量机的调制识别研究 [J]. 系统工程与电子技术, 2003(8) : 1007 – 1011.
- [54] 陆凤波, 黄知涛, 易辉荣, et al. 一种基于高阶累积量的数字调相信号识别方法 [J]. 系统工程与电子技术, 2008(9) : 1611 – 1615.
- [55] 马永军, 薛永浩, 刘洋, et al. 一种基于深度学习模型的数据融合处理算法 [J]. 天津科技大学学报, 2017, 32(4) : 71 – 74.
- [56] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C] // Advances in neural information processing systems. 2014 : 568 – 576.
- [57] FEICHTENHOFER C, PINZ A, ZISSERMAN A. Convolutional two-stream network fusion for video action recognition[J], 2016.
- [58] 周志华. 机器学习 [M]. [S.l.]: Qing hua da xue chu ban she, 2016.
- [59] BREIMAN L. Bagging predictors[J]. Machine learning, 1996, 24(2) : 123 – 140.
- [60] FREUND Y, SCHAPIRE R, ABE N. A short introduction to boosting[J]. Journal-Japanese Society For Artificial Intelligence, 1999, 14(771-780) : 1612.
- [61] FREUND Y, SCHAPIRE R E, OTHERS. Experiments with a new boosting algorithm[C] // Icml : Vol 96. 1996 : 148 – 156.
- [62] LIAW A, WIENER M, OTHERS. Classification and regression by randomForest[J]. R news, 2002, 2(3) : 18 – 22.
- [63] BREIMAN L. Classification and regression trees[M]. [S.l.]: Routledge, 2017.

致谢

时光荏苒，如月如梭。

转眼之间，研究生三年已接近尾声，回想起过去的这三年，有成长当然也有遗憾，但在此还是要感谢研究生陪伴我三年的老师、同学以及朋友们。

首先，衷心地感谢我的导师杨清海教授。在读研期间，他给我提供了很多好的科研条件。从研一开始，他便每周陪我们一起做一次周报，指导我们如何读论文，如何去做 PPT，如何去把自己学到的东西给大家讲出来。这对于我的研究生生涯是一个很大的帮助，让我懂得了如何去学习并将自己的成果展现出来。生活上，杨老师也给了我们很多无微不至的关怀。他经常会询问我们的近况，给予我们一定的帮助与指导；每年的春天，也会让我们走出实验室，体会大自然的风光，增强了我们团队的凝聚力。是他教会了我，做人善事然后科研，让我懂得了一些为人处世的基本原则，这对于我以后的帮助也是非常大的。真心挺感谢杨老师这三年的谆谆教诲，临别之际，他虽远在大洋彼岸，也望风儿能传信至杨老师心里。

我还要感谢沈中、刘明骞和王勇超老师。在我的科研项目中，沈老师给予了很多支持，并给予正确的指导，才让我有了一定的进步。在我的毕设过程中，刘老师对算法逻辑合理性、数据合理性给予了很大的帮助，让我能够顺利完成我的毕业设计。数据的获取是我毕设的关键，王老师在整个过程中，从理论上给予了我很大的帮助，并指导我如何去使用我们的设备，对于我的毕设的数据基础提供了非常大的帮助。在此，我对帮助我的各位老师们致以诚挚的敬意并表示衷心的感谢。

还要感谢在我论文成稿中给我提供了许多帮助的秦剑、张劭、李万、唐靖旋、刘祥等人。在项目中我们互帮互助，在论文的写作仿真中我也受到启发，这很大程度上促进了我的进展并提高了效率。

特别感谢我的父母，虽然远隔千里，虽然我早已成人，但他们仍是我精神上的支柱，是我在遇到困难时继续前进的动力。

最后，衷心地感谢为评阅本论文而付出辛勤劳动的专家和教授们！

赵纪伟

2018 年 4 月于西安电子科技大学

作者简介

1. 基本情况

赵纪伟，男，陕西西安人，1982年8月出生，西安电子科技大学XX学院XX专业2008级硕士研究生。

2. 教育背景

2001.08 ~ 2005.07, 西安电子科技大学, 本科, 专业: 电子信息工程

2008.08 ~ , 西安电子科技大学, 硕士研究生, 专业: 电磁场与微波技术

3. 攻读硕士学位期间的研究成果

3.1 发表学术论文

- [1] XXX, XXX, XXX. Rapid development technique for drip irrigation emitters[J]. RP Journal, UK., 2003, 9(2): 104-110. (SCI: 672CZ, EI: 03187452127)
- [2] XXX, XXX, XXX. 基于快速成型制造的滴管快速制造技术研究 [J]. 西安交通大学学报, 2001, 15(9): 935-939. (EI: 02226959521)
- [3] ...

3.2 申请（授权）专利

- [1] XXX, XXX, XXX 等. 专利名称: 国别, 专利号 [P]. 出版日期.
- [2] ...

3.3 参与科研项目及获奖

- [1] XXX 项目, 项目名称, 起止时间, 完成情况, 作者贡献。
- [2] XXX, XXX, XXX 等. 科研项目名称. 陕西省科技进步三等奖, 获奖日期.
- [3] ...

