

# 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。学校有权保留送交论文的复印件，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、识别和最小化干扰，并实施行为策略和有效协调系统的重要工具。传统方法的研究主要集中在能量检测以及专家特征和决策准则的使用，以获取对不同信号的认知能力。这些方法依赖于信号属性、特征和决策统计的先验知识来分离已知的信号（如调制）类型，并且通常在简化的硬件资源、低干扰无线传播环境模型下得到。

随着深度学习发展，利用原始采样信号进行特征学习并对（调制）信号分类得以实现；并且，在灵敏度和准确性方面优于传统的基于专家特征提取进行统计分类决策。这为当前的无线信号识别的相关问题提供了一种全新的解决方案，但它仍然完全依赖于监督学习。在现实世界中，特别是在无线电领域，我们面临着大量的未标记的示例数据，我们的传感器感知的也只是小部分的标签数据以及大部分的无标签数据，只能获取目标的不完整低可靠性的知识。

无线信号数据易于从天线获得，但有类别标记信号样本数据通常很少，若直接使用传统的监督学习技术，仅能使用有标记数据构建模型，而无法利用未标记的数据来学习信号所包含的信息；同样由于有标记的样本信号较少，训练样本不足，学习泛化能力较差。因此，我们可以利用半监督学习技术，结合深度学习技术，进行有效的无线信号感知

在复杂电磁环境下，完成不同信噪比、不同专家特征集条件下，利用无监督数据构建原始信号的稀疏表示；针对无线信号应用场景，融合专家指纹特征与深度网络特征，利用监督数据构建无线信号的完备特征集，建立无线信号调制方式分类器，完成已知信号的调制方式识别；针对低信噪比条件，分析不同网络结构对于信号分类结果的影响，构建特征融合的最优分类器。

利用深度网络对无线信号进行特征学习，可以获得更能表征信号原始特征的稀疏特征空间。将传统无线信号特征与深度学习特征相融合，构建信号的完备特征集，并将其作为分类的特征空间。首先，对深度网络进行预训练，获取可以表征信号的深度特征空间；进而，将传统特征进行归一化等一系列的预处理，作为新的特征并以全连接的形式输入到分类器；这样整个分类器的输入特征空间相当于深度特征与传统特征的结合，在保证分类准确率的条件下，能够兼顾传统特征与深度特征，充分利用传统方法所获取的先验知识。在特征融合之后，利用监督数据对于整个网络进行训练，以获取最优分类网络。

关键词：XXX， XXX， XXX， XXX， XXX

## ABSTRACT

The Abstract is a brief description of a thesis or dissertation without notes or comments. It represents concisely the research purpose, content, method, result and conclusion of the thesis or dissertation with emphasis on its innovative findings and perspectives. The Abstract Part consists of both the Chinese abstract and the English abstract. The Chinese abstract should have the length of approximately 1000 Chinese characters for a master thesis and 1500 for a Ph.D. dissertation. The English abstract should be consistent with the Chinese one in content. The keywords of a thesis or dissertation should be listed below the main body of the abstract, separated by commas and a space. The number of the keywords is typically 3 to 5.

The format of the Chinese Abstract is what follows: Song Ti, Small 4, justified, 2 characters indented in the first line, line spacing at a fixed value of 20 pounds, and paragraph spacing section at 0 pound.

The format of the English Abstract is what follows: Times New Roman, Small 4, justified, not indented in the first line, line spacing at a fixed value of 20 pounds, and paragraph spacing section at 0 pound with a blank line between paragraphs.

**Keywords:** XXX, XXX, XXX, XXX, XXX



## 插图索引

|      |   |    |
|------|---|----|
| 2.1  | 神经元示意图 . . . . .  | 10 |
| 2.2  | <i>ReLU</i> 函数 . . . . .  | 11 |
| 2.3  | <i>Sigmoid</i> 函数 . . . . .   | 11 |
| 2.4  | 前馈神经网络 . . . . .  | 12 |
| 2.5  | 自编码器 . . . . .  | 17 |
| 2.6  | LSTM 循环网络“细胞”的框图。细胞彼此循环连接，代替一般循环网络中普通的隐藏单元。这里使用常规的人工神经元计算输入特征。如果 <i>sigmoid</i> 输入门允许，它的值可以累加到状态。状态单元具有线性自循环，其权重由遗忘门控制。细胞的输出可以被输出门关闭。所有门控单元都具有 <i>sigmoid</i> 非线性，而输入单元可具有任意的压缩非线性。状态单元也可以用作门控单元的额外输入。黑色方块表示单个时间步的延迟。 . . . . . | 18 |
| 3.1  | 信号生成框图 . . . . .  | 25 |
| 3.2  | 不同调制方式的高 SNR 样本的时域波形 . . . . .  | 28 |
| 3.3  | 不同调制方式的高 SNR 样本的功率谱 . . . . .   | 29 |
| 3.4  | 自编码器 . . . . .  | 29 |
| 3.5  | 自编码器 . . . . .  | 30 |
| 3.6  | 自编码器 . . . . .  | 30 |
| 3.7  | 自编码器 . . . . .  | 31 |
| 3.8  | 自编码器 . . . . .  | 31 |
| 3.9  | 自编码器 . . . . .  | 32 |
| 3.10 | 整体结果 . . . . .  | 35 |
| 3.11 | 0dB 条件下的结果 . . . . .  | 35 |
| 3.12 | 与其他方法的比较 . . . . .  | 35 |
| 3.13 | 训练时间 . . . . .  | 36 |
| 3.14 | 分类时间 . . . . .  | 37 |
| 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 . . . . .  | 45 |
| 4.3  | 基于 DNN 框架的特征融合框架图 . . . . .   | 46 |

|      |  |    |
|------|--|----|
| 4.4  | 基于随机森林框架的特征融合框架图 . . . . .   | 47 |
| 4.5  | 不同融合框架下的分类性能对比 . . . . .   | 48 |
| 5.1  | 改变每层滤波器的数量具有较小的影响, 在更高的 SNR 下更明显。每个网络在 2 个卷积层网络中都有 1x3 个过滤器, 并有 1 个致密层和一个 softmax 分类器。 . . . . | 54 |
| 5.2  | 改变 DNN 中卷积层的数量不会改进无线电调制识别。 . . . .   | 55 |
| 5.3  | 改变两个卷积层中的抽头数量 (滤波器大小), 较低的水龙头数量明显较差, 但性能集群随水龙头数量的增加而增加。 . . . .                                | 55 |
| 5.4  | 显示超参数优化的 CNN 和 9 层残留网络的训练损失和验证损失的训练历史记录 . . . . .  | 56 |
| 5.5  | 用于显示卷积滤波器尺寸的 RF 数据的初始模块。每个卷积层有 50 个过滤器 . . . . .   | 56 |
| 5.6  | 用于 RF 数据的 CLDNN 体系结构。在进入 LSTM 之前, 第一个 1x8 卷积层的输出与三个 1x8 卷积层的输出级联。 . . . .                      | 57 |
| 5.7  | CLDNN 的信噪比始终优于其他网络架构, 信噪比高于 -8dB . . . . .   | 57 |
| 5.8  | CLDNN 的全 SNR 混淆矩阵显示了模拟调制和高阶 QAM 之间单独混淆之间最混乱的问题。 . . . .  | 58 |
| 5.9  | 我们训练的 CLDNN 的第一卷积层中滤波器的时间和频率幅度表示。。   | 58 |
| 5.10 | 我们训练的 CLDNN 的第一卷积层中滤波器的时间和频率幅度表示。。   | 58 |
| 5.11 | 随机数据训练最大程度地激活过滤器, 看起来像 BPSK。 . . . .   | 59 |
| 5.12 | 训练最大程度激活滤波器的随机数据, 这看起来像 FM 或 FSK 调制。   | 59 |

## 表格索引

|     |                |    |
|-----|----------------|----|
| 3.1 | 自编码器 . . . . . | 34 |
|-----|----------------|----|





## 符号对照表

| 符号                 | 符号名称    |
|--------------------|---------|
| $\in$              | 属于      |
| $\mathbb{R}$       | 实数集     |
| $w$                | 权重      |
| $x$                | 样本      |
| $y$                | 标签      |
| $M$                | 特征维数    |
| $N$                | 样本数量    |
| $\eta$             | 学习率     |
| $\mathcal{F}^{-1}$ | 逆傅里叶变换  |
| $\gamma$           | 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   |
| 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 .....              | 1   |
| 1.2 研究意义 .....              | 2   |
| 1.2.1 调制识别的意义 .....         | 3   |
| 1.2.2 深度学习与无线通信的结合 .....    | 3   |
| 1.3 无线信号调制识别的发展和研究现状 .....  | 4   |
| 1.3.1 基于似然比判决理论的方法 .....    | 4   |
| 1.3.2 基于特征提取的统计机器学习方法 ..... | 5   |
| 1.3.3 基于深度学习的无线调制识别方法 ..... | 7   |
| 1.4 本文主要工作及内容安排 .....       | 8   |
| 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 调制信号 .....            | 9   |
| 2.2.2 调制信息 .....            | 9   |
| 2.2.3 信道对调制信号的影响 .....      | 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 .....          | 11  |
| 2.3.3 反向传播算法 .....          | 12  |
| 2.4 卷积神经网络 .....            | 14  |
| 2.4.1 卷积运算 .....            | 14  |
| 2.4.2 卷积特性 .....            | 15  |
| 2.4.3 卷积作为一种无限强的先验 .....    | 15  |

|       |                                  |    |
|-------|----------------------------------|----|
| 2.4.4 | 随机或无监督的特征 .....                  | 16 |
| 2.5   | 卷积自编码器 .....                     | 16 |
| 2.6   | LSTM .....                       | 18 |
| 2.7   | 神经网络优化算法 .....                   | 19 |
| 2.7.1 | 随机梯度下降优化算法 .....                 | 19 |
| 2.7.2 | RMSProp 优化算法 .....               | 20 |
| 2.7.3 | Adam .....                       | 21 |
| 2.7.4 | 选择正确的优化算法 .....                  | 21 |
| 2.8   | 本章小结 .....                       | 22 |
| 第三章   | 基于 CAE-CNN 的无线信号调制识别 .....       | 23 |
| 3.1   | 引言 .....                         | 23 |
| 3.2   | 调制信号生成 .....                     | 23 |
| 3.2.1 | 信道建模 .....                       | 24 |
| 3.2.2 | 调制信号获取 .....                     | 25 |
| 3.3   | 调制信号的表示 .....                    | 26 |
| 3.3.1 | 数据集可视化 .....                     | 26 |
| 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 .....              | 31 |
| 3.4   | 基于 CAE-CNN 的无线信号调制识别 .....       | 32 |
| 3.4.1 | CAE-CNN 网络框架 .....               | 32 |
| 3.4.2 | CAE-CNN 算法 .....                 | 33 |
| 3.4.3 | 算法运行环境及参数 .....                  | 33 |
| 3.5   | 结果及分析 .....                      | 34 |
| 3.5.1 | 学习复杂度 .....                      | 34 |
| 3.5.2 | 分类准确率与鲁棒性 .....                  | 34 |
| 3.5.3 | 训练效率以及分类效率 .....                 | 36 |
| 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 |
| 4.2.3 | 专家循环矩特征 .....                    | 42 |
| 4.3   | 特征融合理论 .....                     | 42 |

|       |  |    |
|-------|--|----|
| 4.4   | 传统特征与深度特征融合框架 .....                                | 44 |
| 4.4.1 | 基于 Softmax 回归 (Softmax Regression, SR) 的融合框架 ..... | 44 |
| 4.4.2 | 基于深度学习的融合框架 .....                                  | 45 |
| 4.4.3 | 基于随机森林的融合框架 .....                                  | 46 |
| 4.5   | 结果及分析 .....  | 47 |
| 4.5.1 | 分类性能比较 .....                                       | 47 |
| 4.5.2 | 混淆矩阵 .....   | 48 |
| 4.6   | 本章小结 .....   | 48 |
| 第五章   | 调制识别的深度框架研究 .....                                  | 51 |
| 5.1   | 引言 .....   | 51 |
| 5.2   | 超参数对 CNN 的调制识别的性能影响 .....                          | 51 |
| 5.2.1 | 卷积神经网络 .....                                       | 51 |
| 5.2.2 | GoogLenet .....                                    | 52 |
| 5.2.3 | ResNet .....                                       | 52 |
| 5.2.4 | CLDNN .....  | 52 |
| 5.3   | 结果及分析 .....  | 53 |
| 5.3.1 | 神经网络训练 .....                                       | 53 |
| 5.3.2 | 网络参数 .....   | 53 |
| 5.3.3 | 基准卷积网络 .....                                       | 54 |
| 5.3.4 | 残留网络 .....   | 55 |
| 5.3.5 | InceptionModel .....                               | 56 |
| 5.3.6 | CLDNN .....  | 56 |
| 5.4   | 本章小结 .....   | 59 |
| 第六章   | 总结与展望 .....  | 61 |
| 6.1   | 研究结论 .....   | 61 |
| 6.2   | 未来展望 .....   | 62 |
| 参考文献  | .....  | 63 |
| 致谢    | .....  | 65 |





## 第一章 绪论

### 1.1 研究背景

通信的主要目的是传达信息。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,而需要调制,基带信号可以移动到一定的频段进行有效传输。同时,对基带信号的调制还可以使通信系统具有更高的通信速率并提高频谱利用率。

随着通信技术的不断发展,用户对于信息传输速率要求的不断提高,以及人们对于无线信号便利性的需求,通信信号调制方式由传统的数字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,信号调制形式由简单到复杂等。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,可以将其分为幅度调制 (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。

通信技术的迅速发展导致了各种通信系统的共存。然而,这些通信系统的调制方式和干扰技术不同,对系统间通信造成很大的障碍。各种调制技术也导致接收机类型大幅增加,于是人们开始寻求新的通信方法,希望能够基于相同的接收机平台接收最好的调制,以减少接收机类型的快速增长,并且软件无线电技术是解决这个问题的有效方法。自动识别通信信号调制模式是软件无线电技术必须解决的关键问题之一。

调制是区分不同类型的通信信号的重要特征。所谓的调制的自动识别意味着它可以在给定未知调制信息的接收信号的情况下确定通信信号的调制模式,并在没有人为干预的情况下估计相应的调制参数。在密集和复杂的多用户频谱环境中自动识别通信信号调制。无线电信号识别是优化频谱利用,识别和最小化干扰,实施频谱策略以及实施有效的无线电意识和协调系统的重要工具。调制自动识别技术在军队和政府部门的无线电管理中也起着重要作用。在军事领域,调制识别被广泛用作通信侦察,电子战和威胁分析中的智能信号分析和处理的关键技术 [9]。在现代战争中,战场信息的传播主要依赖于无线电通信。侦察通讯信号是电子战的主要内容。在 EW 侦听和侦听接收机的设计中,获取接收到的通信信号的调制方式是侦听接收机的重要功能之一。为解调器正确选择解调算法提供参数依据,最终得到有用的情报信息。调制识别技术还有助于为电子战选择最佳干扰模式或干扰消除算法,保证友好通信,同时压制和摧毁敌方通信,达到电子战通信对抗的目的。

另外,在自适应调制系统中,传输信号的调制模式随信道状态而变化。为了正确

解调信号,接收端需要知道发送信号的调制信息。发送信令的最简单的方式是将包含调制信息的控制信号发送到分组中的接收机,其可以由接收机解调以获得信号的调制。但是,这种方法是基于有用信息的带宽。如果发送的信号不包含调制参数信息,则可以减少开销。这需要信号识别技术来检测接收到的信号并获得信号参数来解调它。

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,将成为现代战争通信和未来无线通信应用的重要组成部分。

深度学习(DL)是具有最高级分类能力的机器学习(ML)的一个分支[1]。最近引起了人们的关注,并在各个领域得到了应用。在 ImageNet 大规模视觉识别挑战赛(ILSVRC)的竞争中,许多研究团队提交了各种用于对象检测和图像分类的 DL 算法。最近 1000 个识别对象的前五位数字的准确性达到了 95%[2]。在文献[3]中,作者指出可以对 DL 网络进行培训,以便利用卫星图像进行发展中国家的准确,廉价和可扩展的经济调查。而且,生物信息学也可以从 DL 获益。可以从 DNA 序列中找到连接,可以从 X 射线图像中识别手指关节,并且可以从 EEG 信号中检测到不受控制的事件[4]。

尽管 DL 在全世界蓬勃发展,但它在通信中的使用相对较少。值得注意的是,一些传统的 ML 算法如支持向量机(SVM)和 K 近邻(KNN)已被用于媒体访问控制(MAC)协议识别[5]和调制分类[6]。在通信系统中使用 DL 有几个优点。首先,由于通信装置的数量多且通信数据速率高,因此可以在通信系统中使用 DL 所需的大量数据。其次,DL 可以自主提取特征,避免了手动特征选择这一具有挑战性的任务。第三,由于 DL 正在迅速发展,除了调制分类之外,其他通信应用将具有相当大的潜力。

另外,它也是一种不断发展的技术。新的和更复杂的信号调制方法和通信应用的出现给它们带来了越来越多的挑战,并且在这个技术领域出现了新的研究成果。这一领域的巨大潜在研究价值在这方面的研究越来越受到国内外众多学者的关注。

本文主要研究深度学习在无线信号的调制识别领域的应用。

## 1.2 研究意义

无线通信的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。需要在无线通信系统中传输或传输的基带原始信息通常是频率非常低且分量很多的频谱分量。这些低频分量信号不适合在信道中直接传输,只能调制到不同的信道。传输,所以在接收端,只有正确确定调制方式,才能使用合理的解调方法接收来自发送方的信息。如果接收信号的调制方式的误差判断包括调制类型的判断或者载波频率和波特率的错误估计,则接收机的解调结果必须使原始发送的基带信息信号被销毁

或部分销毁。最终，它无法正确解调或接收来自发件人的原始信息。

### 1.2.1 调制识别的意义

因此，在民用通信中，对于一些非合作性的通信，如政府无线电管理部口在日常的无线电管理中，需要对民用通信信号进行监控与管理，必要时需要侦听与拦截，在这种情况下，由于管理与被管理双方通信未必事先预定，通信的调制方式类型的识别、载波频率、波特率及其它调制参数的正确估计显得尤为重要。

在军事训练中的无线电对抗、电磁干扰与反干扰等，在检测到对方无线电信号或电磁信号的同时，需要对检测到的信号或电磁信息进行调制识别包括载波频率与波特率的正确估计，以便进一步侦听破译对方通信情报信息或者对对方进行高强度的无线电干扰或者电磁干扰，使对方无线通信受到阻碍甚至瘫痪。

在认知无线领域，当前无线频谱根据具体业务的不同，相应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。同时，为了避免相邻频段或频道的互相干扰，不同的通信业务工作频段相互隔离并且预留一定隔离频带，上下频段频谱利用的极不平衡，造成频谱资源的极大浪费。显然，频谱资源利用率的不平衡性及当前有限的无线频谱资源和僵硬的频谱管理方法已经成为制约无线通信进一步发展的重要因素，因此，如何在现有无线频谱资源基础上，在频率、时间和空间等四维空间上如何有效分配频谱资源，提高频谱利用率，认知无线电

### 1.2.2 深度学习与无线通信的结合

无线电通信是一个特定的信号处理领域。在这个领域中，调制信号分类的专家特征和决策标准的使用已被广泛使用。经过多年的积累，它已经能够在特定情况下实现高识别率。在 ImageNet 竞赛中，识别深度学习的能力超出了人类专家的能力，提高了传统机器学习算法的上限。此外，通过差异，卷积和合并等一系列方法 [10]，我们还增强了模型应用的泛化能力，使得深度学习在实际生产中得到广泛应用。

深度神经网络已经在计算机视觉，自然语言处理和语音识别等领域实现了广泛的应用，并取得了很好的结果。然而，深度学习应用于图像处理 [11] 和语音识别 [18] 的趋势绝大多数是从数据学习特征而不是利用专家特征进行学习，这表明我们对于认知无线领域，需要重新考虑一下是否应该也进行相应的思路转变。同时，深度学习在通信领域的应用，也带来了新的机遇与变革，XXX 等人利用深度学习进行半监督识别，有效的利用了无监督数据，XXX 等人将深度学习应用于路由，提高了识别准确率等。

综上所述，如何有效地提取特征参数、采用不同的识别算法与分类器实现调制式的自动识别，在军用和民用通信有广泛的应用，是软件无线电、认知无线电、频谱感

知等领域研究的基础。随着通信技术的发展,通信环境日益复杂,新的调制方式不断出现,影响识别效率的因素越来越多,如何在现有自动识别技术方法基础上不断创新、提高仍然是一项颇具挑战性的研究领域。

### 1.3 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。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,而且这种方法没有功率 [20]。1969 年 4 月, C.S.Waver [3] 等人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“使用模式识别技术的调制类型的自动分类”。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论文,其中许多结果已逐渐应用到实际工作中。

对于通信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和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,对模型失配问题更为敏感,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它通常关注能量检测,并使用专家特征和判定标准来识别和分类特定的调制类型。在适当的条件下,它可以实现接近最佳的识别性能,并且其性能更稳健,因此得到更广泛的应用;深度学习其他领域取得的成果,深度学习可以与硬件自适应学习分类器相结合,具有良好的适应性,并且提高了识别的准确性和鲁棒性,因此已成为近年来调制识别的研究热点。

#### 1.3.1 基于似然比判决理论的方法

似然比决策理论方法也被称为最大似然假设检验法。其基本框架是使用概率论和复合假设检验理论来分析信号的统计特性,并根据成本函数最小化原则导出足够的统计信息进行分类。然后将其与适当的阈值进行比较,完成调制识别的自动分类功能。

对于现有的发展, Kim 和 Polydoros [4] 使用平均似然比检验来识别 BPSK 和 QPSK 信号。通过近似似然函数的级数展开,提出了近似对数相似度的准对数似然比函数。但是,为了降低功能的计算复杂度,当信噪比为 0dB 时的识别率达到 100%。该方法还可以扩展到 MPSK, QAM 等多级调制方法的识别。Schreyoegg [5] 等人。假设 QAM 信号的幅度和相位彼此大致相互独立,并且使用对数似然函数来使用振幅和相位的联合概率密度函数来识别 MQAM 信号。Wei 和 Mendel [6] 使用复数符号序列

的平均似然函数对 QAM 信号进行分类。当信噪比为 5dB 时, 识别率达到 100%, 并分析了最大似然分类器的渐近行为。它表明最大似然法可以对任何有限星座点的调制进行分类。当可用符号的数量接近无穷大时, 分类错误概率接近零。它还给出了参数与错误概率之间的关系。关系表示给定错误概率所需的 SNR 和符号的数量。巴黎将这种通用方法与繁忙均衡相结合, 研究了多径信道中的调制识别问题 [7]。Baudant 等人提出了瑞利衰落信道中 FSK 信号的调制分类方法 [8]。似然函数来自被高斯白噪声污染的 FSK 信号的包络。得到最优分类器并得到有效解决。衰落信道中的 FSK 信号调制识别问题。

似然比决策理论算法的优点是在理论上保证贝叶斯最小误判惩罚准则下分类结果最优, 通过理论分析可以得到分类性能曲线。但是他的同事有很多局限性: 首先, 他们需要比模式识别方法更多的先验知识。现有的似然比决策理论算法主要处理符号同步采样序列, 这意味着要知道信号的载波频率, 符号速率和符号时序; 其次, 未知参数的存在导致似然比分类充满了复杂的统计表达式。它有很大的计算量, 很难实时处理。第三, 似然比决策算法对模型失配和参数偏差更敏感, 即鲁棒性较差。似然比函数分类通常被建模为高斯分布, 并且诸如信噪比等参数是已知的。当实际信道噪声为非高斯时, 或存在多路径影响, 多信号干扰和 SNR 参数估计偏差时, 似乎分类性能可能急剧下降。

### 1.3.2 基于特征提取的统计机器学习方法

统计机器学习方法基于统计机器学习理论。基本框架是首先从信号中提取先前选择的特征, 然后执行模式识别, 包括特征提取子系统和模式识别子系统。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。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来区分信号的调制类型。

关于数字调制信号识别方法的最早的公开讨论是 Liedtke [10]。他提出了一种用于未知调制方法分类的通用分类器。该算法可以识别 2ASK, 2FSK, BPSK, QPSK, 8PSK 和 CW 信号, 只需要大致了解信号的载波频率和符号率。Azzouz 和 ANDI 提出了一种算法 [11-15], 它使用相位的非线性部分, 相位的非线性部分的绝对值, 归一化的瞬时振幅和频率等。调制的标准偏差使用一系列阈值或神经网络分类器来确定参数以识别调制的类型。后来研究人员在此基础上进行了更多研究 Wong 和 AK Nandi 随后对方法进行了改进 [16], 增加了信号的统计参数和训练序列, 并且在 0dB% 时识别率达到了 98%。

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。由于 Reichert J [19] 首次提出使用高阶统计量

来识别 2ASK, BPSK, 2FSK, MSK 信号, 所以基于高阶统计量的信号识别方法发展迅速。Ananthram Swami [20] [21] 使用四阶累积量来识别 BPSK, 4ASK, 16QAM 和 8PSK 信号。当信噪比为 10dB 时, 识别率为 95%。Spooner CM [22] 将基于累积量方法的前一种最大化方法打破为四阶例程, 提出用 6 阶循环累积量识别信号, 并获得良好的识别效果。当 SNR 为 9dB 时, 16QAM 总和 64QAM 的识别率分别为 81% 和 90%。QPSK 和 16QAM 信号的识别率分别为 97% 和 100%。

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。它可以使用许多现有的图像处理和模式识别算法, 适用于正交调制信号分类, 如 MPSK, MASK 和 MQAM 信号。识别 MFSK 信号。Bijian [25] 详细分析了基于星座图特征的调制识别方法。通过使用模糊逻辑方法可以很好地恢复原始信号(即接收信号的星座)的特性。在最大后验概率贝叶斯准则下, 对模板进行匹配和判断。在信噪比大于 0dB 的情况下, QPSK, 8PSK 和 16QAM 分类的正确率可以达到 90%。Donoho 等人 [26] 使用星座点处的信号计数来执行调制分类。该算法简单易行。该算法在信噪比大于 15dB 的条件下是稳定的。

近年来, 基于小波变换的多分辨率特征分析, 调制参数估计和调制识别器方兴未艾。1995 年, K C Ho [2] 首次使用小波变换来识别 MFSK 和 MPSK 信号。取得了一定的成果, 证明了该方法的可行性。1999 年, 梁红等人。[28] 使用 IF 信号的小波变换系数的方差来分类 FSK, PSK 和 QAM 信号。Haa 小波变换相位突变检测能力的主要用途是提取 CNR5dB 信号中信号的分类特征。识别率达到 97.6%, 可以很好地对 16QAM, QPSK 和 4FSK 信号进行分类。在 2000 年, 基于提取信号的小波细节特征的 KCHor, [529] 的条件下, 计算小波变换在固定尺度下的幅度值, 并根据过渡符号的峰值数量使用统计直方图, 直流电平的数量。识别 MFSK 和 MPSK 信号, 通过估计符号周期和同步时间, 进一步提高低信噪比下的分类成功率。当 CNR = 6dB 时, 当 CNR = 1dB 时, MPSK 信号的识别率为 90.2%。MFSK 信号的识别率为 91.9%。

考察上面使用的各种分类特征, 已知它是否基于时域频域中的调制信号的特征, 高阶统计量, 星座特征或小波变换, 人们很难找到调制分类。必须为每个分类问题分别考虑共同特征和方法, 并根据所需调制类型搜索特定方法和特征。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, 多普勒效应对信号产生寄生调频。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, 相位和频率参数偏离了感兴趣的通信。由于信号的实际参数特征, 使识别性能急剧下降, 稳定性差, 不能满足实际应用的需要 [34]。此外, 目前大多数识别算法都局限于简单的调制方法, 如 MPSK 信号, MFSK 信号等。只有少数算法研究了 MQAM 信号

的识别 [35]。

此外，基于特征的识别方法依赖于信号属性，特征和决策统计的先验知识来分离已知的调制，并且通常在简化的分析硬件，传播，无线电环境模型下导出。

### 1.3.3 基于深度学习的无线调制识别方法

深层神经网络是由一系列层组成的大函数近似，其中每层表示从输入到输出激活的一些变换，其基于具有一些学习权重的参数传递函数。每一层通常是已知的具有可调参数和非线性激活函数的线性函数，使得所得到的函数组成可以是高度非线性的 [3]。在深度神经网络中的函数参数通常用梯度下降优化器根据在网络输出上计算的一些损失函数进行训练。对于多类分类任务，如调制识别，目标函数通常是分类交叉熵 (两个概率分布之间的差异的量度)。利用深度学习进行分类，已被成功地用于多类视觉任务，如 **Imagenet** 数据集上的对象识别 [7]。

最近，已有学者证明了有监督的无线电分类系统的原始数据特征学习的可行性 [14]。在这种情况下，我们能够超越传统的基于专家决策统计的敏感性和准确性分类的显著差距。这是一个强有力的结果，对当前的解决方案提供了显著的性能改进，但它仍然完全依赖于监督学习和精心策划的培训数据。

半监督聚类。在现实世界中，特别是在无线电领域，我们面临着大量的未标记的示例数据可用于我们的传感器和不完整的地面真相的类标签知识。为了解决这个问题，XX 研究了无线电识别学习的替代策略，这些策略对标记的训练数据的依赖程度较低，并且能够使用无标记或较少标记的例子来理解无线电信号，这可能极大地减轻了在这种机器学习系统供开发人员和维护人员使用，并允许系统识别新的信号并扩展以了解新环境。

在半监督学习 [9] [8] [5] 中，我们试图分离和识别新的类，没有明确的类标签，这些类的例子允许我们专门学习特征来分离它们。为了解决这个问题，我们考虑将信号示例的维数降低到一个平滑的较小的空间，我们可以执行信号聚类。给定一个适当的维数降低，我们寻求一个空间，相同或相似类型的信号具有较低的距离，而不同类型的信号间隔较大的距离。理想情况下，在这样的空间中，相同或相似类型的例子形成离散且可分离的簇，彼此容易辨别。在这样的空间中信号类型的分类则成为识别聚类，将标签与每个聚类相关联（而不是每个例子，一个劳动密集度较低的任务）的问题，以及允许识别，分类和历史注释的问题新的班级和新的班级例子，即使没有标签的知识。

卷积神经网络 (CNN) [8] 是最流行的 DL 体系结构之一，被认为是调制分类。调制的信号被转换成网格状的拓扑数据，例如星座图的图像，以馈送 CNN。著名的 CNN 模型 **AlexNet** [9] 被采纳和修改用于网络训练和精度测试。

密集的多用户环境中的无线电发射器识别是优化频谱利用，识别和最小化干扰以及执行频谱策略的重要工具。无线电数据是容易获得的并且易于从天线获得，但是标记和策划的数据通常很少，使得监督式学习策略在实践中困难且耗时。我们证明，半监督学习技术可以用来扩大学习范围以外的监督数据集，允许识别和召回新的无线电信号通过使用基于非监督和监督方法的非线性特征学习和聚类方法的稀疏信号表示。

## 1.4 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，但能够直接对中频信号进行识别的还比较少，大多都需要确知信号的载波频率、码元速率等参数，通过复杂的处理把信号转换为基带信号后再进行识别，否则会造成明显的性能损失或算法失效。而在实际的非协作通信中，接收机是不可能确知这些先验信息的，因此论文重点研究了基于小波变换直接对中频信号进行识别的算法，提出了一套利用小波变换的多种参数信息对数字通信信号进行类间识别和类内识别的分类方法。论文算法在低信噪比条件下具有较高的识别率。

第一章对通信信号调制识别技术的研究背景和研究现状进行了介绍。

第二章首先介绍了现代数字通信信号的调制技术和小波变换理论，然后对数字通信信号的小波变换进行了分析，为后面基于小波变换的调制识别研究打下理论基础。

第三章对数字通信信号的类间识别进行研究。首先介绍了小波变换中最优小波尺度的选择方法，然后，针对传统基于小波变换幅度信息的类间识别算法在低信噪比下识别率低的问题，提出了一种基于信号包络方差和小波变换频率信息的类间识别算法，最后通过仿真验证了该算法在低信噪比下具有较高的识别性能。

第四章对 MPSK、MFSK 和 MQAM 信号的类内识别进行研究，分别提出了利用信号小波变换后相邻码元对应采样点的相位差信息识别 MPSK 信号，利用信号小波变换后的频率信息识别 MFSK 信号和利用信号小波变换后的归一化瞬时功率值分布特征识别 MQAM 信号的分类算法。并通过仿真验证论文算法在低信噪比下具有较高的识别性能。

第五章总结全文，并对今后的工作进行了展望。

第六章总结全文，并对今后的工作进行了展望。



## 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（DSA）中，感知周围的发射终端，以避免无线电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、本地和广域数据、语音信号、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，需要我们对信号调制方式有一定的了解。本章主要讨论信号调制的基本概念以及深度学习的基础知识。

### 2.2 调制识别

#### 2.2.1 调制信号

本文针对的调制信号有。。。

#### 2.2.2 调制信息

在无线电通信中，信号通常由定义好的和理解的基本功能上的许多调制数据位组成由这些基地形成的离散模式。信号的复合基带表示将无线电电压电平时间序列分解为其对载波频率的正弦和余弦函数的投影。通过操纵频率，幅度，相位或其总和，数据比特然后通过离散和可分离的模式调制到这个空间中，对于数字的每个不同的符号周期，或者在模拟调制的情况下连续的位置。对于 QPSK 的情况，这个相位映射如 4 所示。然后通常将脉冲整形滤波器（例如根升余弦）应用于频带限制信号，并消除这些不同模式之间的尖锐宽带瞬态，导致相邻符号在发射机处的基极以确定性和可逆性混合。在我们的模拟数据集中，我们使用一个根升余弦脉冲整形滤波器，每个数字信号的带宽为 0.35。

#### 2.2.3 信道对调制信号的影响

在传播效应的这些苛刻的现实假设下，对专家特征和决策度量的最优性进行分析建模是非平凡的，并且经常会使得简化假设。在本文中，我们将重点放在包括所有上述效应在内的苛刻的模拟传播环境中的性能的经验性测量，但不要尝试以封闭形式分析追踪其性能。

相比之下，信道效应不是确定性的，在通信系统中不是完全可逆的。真实的系统在传输的信号上经历许多影响，这使得恢复和表示具有挑战性。热噪声在接收器处产

生相对平坦的高斯白噪声，其形成本底噪声或灵敏度级别和信噪比。由于温度和其他半导体物理因发射器和接收器的不同而引起的振荡器漂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。这些效应导致信道之间的时间移位，缩放，线性混合/旋转以及基于未知时变过程的接收信号的旋转。最后，根据在接收机处发射信号的到达模式，实际信道经历随机滤波，具有不同的幅度，相位，多普勒和延迟。这是一种通常被称为多径衰落或频率选择性衰落的现象，其发生在信号可能反射出建筑物，车辆或环境中的任何形式的反射器的任何环境中。通常在接收机中通过估计时变信道响应的瞬时值并将其从接收信号解卷积来消除这种情况。

## 2.3 神经网络概述

### 2.3.1 神经元概述

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：

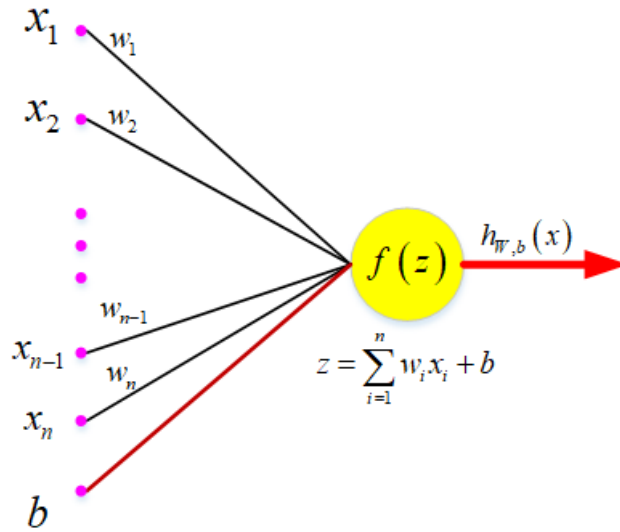


图 2.1 神经元示意图

这个“神经元”是一个以  $x_1, x_2, x_3$  及截距  $+1$  为输入值的运算单元，其输出为：

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

其中函数  $f: \mathbb{R} \mapsto \mathbb{R}$  被称为“激活函数”。在本论文中，我们一般选用  $ReLU$  函数作为激活函数  $f(\cdot)$ ：

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

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, 相对宽阔的兴奋边界, 稀疏激活性。

图2.2和图2.3分别是  $ReLU$  及  $Sigmoid$  的函数图像:

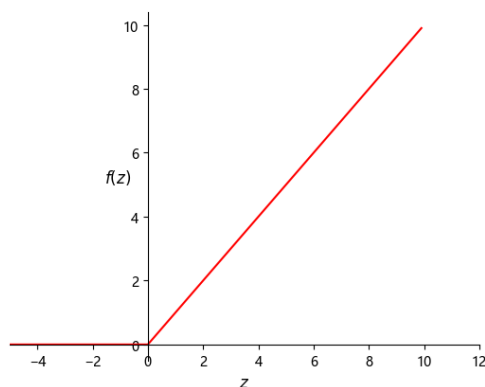


图 2.2  $ReLU$  函数

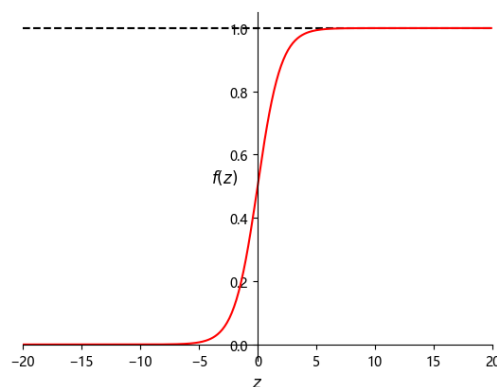


图 2.3  $Sigmoid$  函数

### 2.3.2 前馈神经网络

前馈神经网络就是将许多个单一“神经元”联结在一起, 这样, 一个“神经元”的输出就可以是另一个“神经元”的输入。例如, 图??就是一个简单的神经网络:

我们使用圆圈来表示神经网络的输入, 标上“+1”的圆圈被称为偏置节点, 也就是截距项。神经网络最左边的一层叫做输入层, 最右的一层叫做输出层 (本例中, 输出层只有一个节点)。中间所有节点组成的一层叫做隐藏层, 因为我们不能在训练样本集中观测到它们的值。同时可以看到, 以上神经网络的例子中有 3 个输入单元 (偏置单元不计在内), 3 个隐藏单元及一个输出单元。

我们用来表示网络的层数, 本例中, 我们将第层记为, 于是是输入层, 输出层是。我们用  $a_i^{(l)}$  表示第  $l$  层第  $i$  单元的激活值 (输出值)。当  $l = 1$  时,  $a_i^{(1)} = x_i$ , 也就是第  $i$  个输入值 (输入值的第  $i$  个特征)。对于给定参数集合  $W, b$ , 我们的神经网络就可以按照函数  $h_{W,b}(x)$  来计算输出结果。

我们用  $z_i^{(l)}$  表示第  $l$  层第  $i$  单元输入加权和 (包括偏置单元)

我们将上面的计算步骤叫作前向传播。回想一下, 之前我们用  $a^{(1)} = x$  表示输入层的激活值, 那么给定第  $l$  层的激活值  $a^{(l)}$  后, 第  $l + 1$  层的激活值  $a^{(l+1)}$  就可以按照

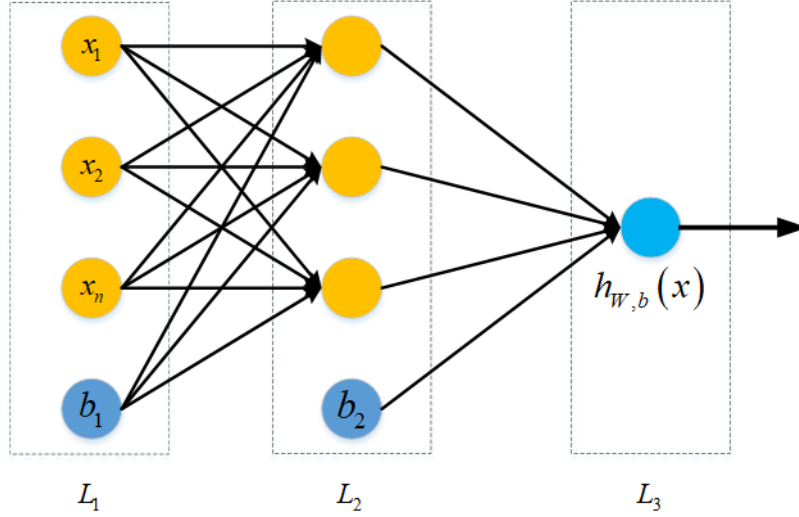


图 2.4 前馈神经网络

下面步骤计算得到：

$$z^{(l+1)} = W^{(l)}a^{(l)} + b^{(l)} = \sum_{j=1}^n W_{ij}^{(l)}x_j + b_i^{(l)} \quad (2-3)$$

$$a^{(l+1)} = f(z^{(l+1)}) \quad (2-4)$$

将参数矩阵化，使用矩阵—向量运算方式，我们就可以利用线性代数的优势对神经网络进行快速求解。

目前为止，我们讨论了一种神经网络，我们也可以构建另一种结构的神经网络（这里结构指的是神经元之间的联接模式），也就是包含多个隐藏层的神经网络。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。这种模式下，要计算神经网络的输出结果，我们可以按照之前描述的等式，按部就班，进行前向传播，逐一计算第  $L_2$  层的所有激活值，然后是第  $L_3$  层的激活值，以此类推，直到第  $L_{n_l}$  层。这是一个前馈神经网络的例子，因为这种联接图没有闭环或回路。

### 2.3.3 反向传播算法

假设我们有一个固定样本集  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ ，它包含  $m$  个样例。我们可以用批量梯度下降法来求解神经网络。具体来讲，对于单个样例  $(x, y)$ ，我们假设其代价函数  $J(W, b; x, y)$  为方差代价函数，则有：

$$J(W, b; x, y) = \frac{1}{2} \|h_{W,b}(x) - y\|^2. \quad (2-5)$$

给定一个包含  $m$  个样例的数据集，那么整体代价函数  $J(W, b)$  为：

$$\begin{aligned} J(W, b) &= \left[ \frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}, y^{(i)}) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= \left[ \frac{1}{m} \sum_{i=1}^m \left( \frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \end{aligned} \quad (2-6)$$

以上关于  $J(W, b)$  定义中的第一项是一个均方差项。第二项是一个规则化项（也叫权重衰减项），其目的是减小权重的幅度，防止过度拟合。权重衰减参数  $\lambda$  用于控制公式中两项的相对重要性。

以上的代价函数经常被用于分类和回归问题。在分类问题中，我们用  $y = 0$  或  $1$ ，来代表两种类型的标签（回想一下，这是因为 sigmoid 激活函数的值域为  $[0, 1]$ ；如果我们使用双曲正切型激活函数，那么应该选用  $-1$  和  $+1$  作为标签）。对于回归问题，我们首先要变换输出值域，以保证其范围为  $[0, 1]$ （同样地，如果我们使用双曲正切型激活函数，要使输出值域为  $[-1, 1]$ ）。

反向传播算法的思路如下：给定一个样例  $(x, y)$ ，我们首先进行“前向传导”运算，计算出网络中所有的激活值，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”  $\delta_i^{(l)}$ ，该残差表明了该节点对最终输出值的残差产生了多少影响。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。反向传播算法可表示为以下几个步骤：

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：

$$\delta^{(n_l)} = -(y - a^{(n_l)}) \bullet f'(z^{(n_l)}) \quad (2-7)$$

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

$$\delta^{(l)} = ((W^{(l+1)})^T \delta^{(l+1)}) \bullet f'(z^{(l)}) \quad (2-8)$$

计算最终需要的偏导数值：

$$\nabla_{W^{(l)}} J(W, b; x, y) = \delta^{(l+1)} (a^{(l)})^T, \quad (2-9)$$

$$\nabla_{b^{(l)}} J(W, b; x, y) = \delta^{(l+1)}. \quad (2-10)$$

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  $f'(z_i^{(l)})$ 。假设  $f(z)$  是 sigmoid 函数，并且我们已经在前向传导运算中得到了  $a_i^{(l)}$ 。最后，我们

将对梯度下降算法做个全面总结。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“ $\Delta$  与  $W^{(l)}$  相乘”。下面，我们实现批量梯度下降法中的一次迭代：

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  $i = 1$  到  $m$ ，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：

$$W^{(l)} = W^{(l)} - \alpha \left[ \left( \frac{1}{m} \Delta W^{(l)} \right) + \lambda W^{(l)} \right] \quad (2-11)$$

$$b^{(l)} = b^{(l)} - \alpha \left[ \frac{1}{m} \Delta b^{(l)} \right] \quad (2-12)$$

现在，我们可以重复梯度下降法的迭代步骤来减小代价函数  $J(W, b)$  的值，进而求解我们的神经网络。

## 2.4 卷积神经网络

卷积网络，也叫做卷积神经网络，是一种专门用来处理具有类似网格结构的数据的神经网络。例如时间序列数据（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）。卷积网络在诸多应用领域都表现优异。卷积是一种特殊的线性运算。卷积网络是指那些至少在网络的一层中使用卷积运算来替代一般的矩阵乘法运算的神经网络。

### 2.4.1 卷积运算

卷积是对两个实变函数的一种数学运算，通常我们用星号表示：

$$s(t) = (x * w)(t) = \int x(\tau)w(t - \tau)d\tau \quad (2-13)$$

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：

$$s(t) = (x * w)(t) = \sum_{\tau=-\infty}^{\infty} x(\tau)w(t\tau). \quad (2-14)$$

在机器学习的应用中，输入通常是多维数组的数据，而核通常是由学习算法优化得到的多维数组的参数。我们把这些多维数组叫做张量。因为在输入与核中的每一个元素都必须明确地分开存储，我们通常假设在存储了数值的有限点集以外，这些函数的值都为零。这意味着在实际操作中，我们可以通过对有限个数组元素的求和来实现无限求和。

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：

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - n). \quad (2-15)$$

卷积是可交换的 (commutative)，我们可以等价地写作：

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n). \quad (2-16)$$

卷积运算可交换性的出现是因为我们将核相对输入进行了翻转，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。许多神经网络库会实现一个相关的函数，称为互相关函数，和卷积运算几乎一样但是并没有对核进行翻转：

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

许多机器学习的库实现的是互相关函数但是称之为卷积，在本文中，我们将这两种运算都称之为卷积运算。

### 2.4.2 卷积特性

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 (equivariant) 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。

### 2.4.3 卷积作为一种无限强的先验

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为全连接网络，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函数只包含局部连接关系并且对平移具有等变性。

#### 2.4.4 随机或无监督的特征

通常，卷积网络训练中最耗时的部分是学习特征。输出层的计算代价通常相对不高，因为在通过若干层卷积池化之后作为该层输入的特征的数量较少。当使用梯度下降执行监督训练时，每步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。减少卷积网络训练成本的一种方式是使用那些不是由监督方式训练得到的特征。

有三种基本策略可以不通过监督训练而得到卷积核。其中一种是简单地随机初始化它们。另一种是手动设计它们，例如设置每个核在一个特定的方向或尺度来检测边缘。最后，可以使用无监督的标准来学习核。

使用无监督的标准来学习特征，允许这些特征的确定与位于网络结构顶层的分类层相分离。然后只需提取一次全部训练集的特征，构造用于最后一层的新训练集。假设最后一层类似逻辑回归或者支持向量机，那么学习最后一层通常是凸优化问题。

随机过滤器经常在卷积网络中表现得出乎意料得好<sup>[1][2][3]</sup>。说明，由卷积和随后的池化组成的层，当赋予随机权重时，自然地变得具有频率选择性和平移不变性。他们认为这提供了一种廉价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更昂贵的方法来训练整个网络。一个中间方法是学习特征，但是使用那种不需要在每个梯度计算步骤中都进行完整的前向和反向传播的方法。与多层感知机一样，我们使用贪心逐层预训练，单独训练第一层，然后一次性地从第一层提取所有特征，之后用那些特征单独训练第二层，以此类推。

### 2.5 卷积自编码器

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，使用均方误差（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，并使用随机梯度下降（SGD）算法等，以找到接近等式2-18中的最佳网络参数。

$$\arg \min_{\theta} (\sum (X - f(X, \theta))^2) \quad (2-18)$$

自编码器可以理解为一个试图去还原其原始输入的系统。它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[13] 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。如下图所示。

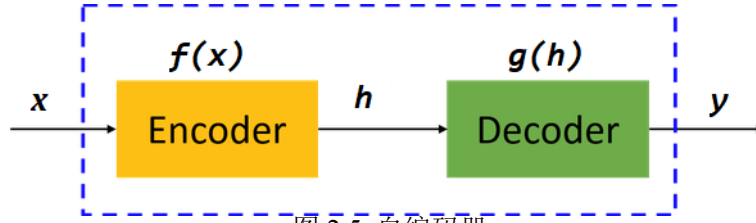


图 2.5 自编码器

图中，虚线蓝色框内就是一个自编码器模型，它由编码器（Encoder）和解码器（Decoder）两部分组成，本质上都是对输入信号做某种变换。编码器将输入信号  $x$  转换成编码信号  $y$ ，而解码器将编码信号  $y$  转换成输出信号。即：

$$\begin{aligned} y &= f(x) \\ x &= g(y) = g(f(x)) \end{aligned} \quad (2-19)$$

而自编码器的目的是，让输出尽可能复现输入  $x$ 。对于自编码器，我们往往并不关心输出是什么（反正只是复现输入），我们真正关心的是中间层的编码信号，或者说是从输入到编码的映射。在我们强迫编码信号  $y$  和输入信号  $x$  不同的情况下，系统还能够去复原原始信号  $x$ ，那么说明编码信号  $y$  已经承载了原始数据的所有信息，但以一种不同的形式，这便是特征提取。对于卷积自便器，我们中间的隐层是利用卷积网络实现的。

在自编码器的训练过程中，我们尽量减少重构均方误差（MSE），但由于我们的主要目标是获得一个良好的聚类稀疏表示，我们显着限制隐藏层维度到一个点，我们的重构作出一些可见的简化假设，在最优重构误差下降低了隐层维数。图 3 显示了两个训练样例， $2 \times 128$  输入向量的样子， $1 \times 30$  稀疏表示的样子，以及  $2 \times 128$  输出重构的样子。这给了这个网络的学习代表能力的一些直觉。

## 2.6 LSTM

循环神经网络或 RNN 是一类用于处理序列数据的神经网络。就像卷积网络是专门用于处理网格化数据的神经网络，循环神经网络是专门用于处理序列  $x^{(1)}, \dots, x^{(\tau)}$  的神经网络。

正如卷积网络可以很容易地扩展到具有很大宽度和高度的图像，以及处理大小可变的图像，循环网络可以扩展到更长的序列（比不基于序列的特化网络长得多）。大多数循环网络也能处理可变长度的序列。

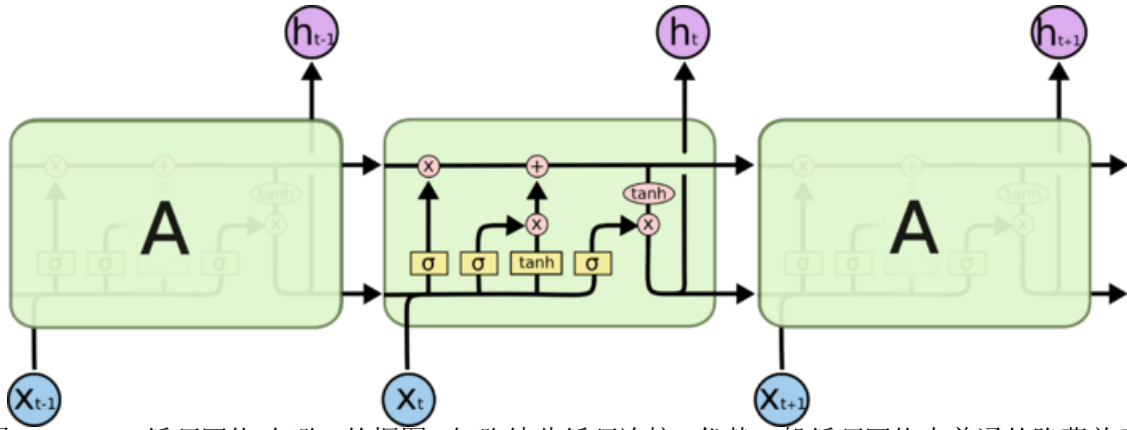


图 2.6 LSTM 循环网络“细胞”的框图。细胞彼此循环连接，代替一般循环网络中普通的隐藏单元。这里使用常规的人工神经元计算输入特征。如果 sigmoid 输入门允许，它的值可以累加到状态。状态单元具有线性自循环，其权重由遗忘门控制。细胞的输出可以被输出门关闭。所有门控单元都具有 sigmoid 非线性，而输入单元可具有任意的压缩非线性。状态单元也可以用作门控单元的额外输入。黑色方块表示单个时间步的延迟。

LSTM 块如图 2.6 所示。在浅循环网络的架构下，相应的前向传播公式如下。更深的架构也被成功应用<sup>[2, 3]</sup>。LSTM 循环网络除了外部的 RNN 循环外，还具有内部的“LSTM 细胞”循环（自环），因此 LSTM 不是简单地向输入和循环单元的仿射变换之后施加一个逐元素的非线性。与普通的循环网络类似，每个单元有相同的输入和输出，但也有更多的参数和控制信息流动的门控单元系统。

最重要的组成部分是状态单元  $s_i^{(t)}$ ，与前一节讨论的渗漏单元有类似的线性自环。然而，此处自环的权重（或相关联的时间常数）由遗忘门（forget gate） $f_i^{(t)}$  控制（时刻  $t$  和细胞  $i$ ），由 sigmoid 单元将权重设置为 0 和 1 之间的值：

$$f_i^{(t)} = \sigma \left( b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)} \right), \quad (2-20)$$

其中  $x^{(t)}$  是当前输入向量， $h^t$  是当前隐藏层向量， $h^t$  包含所有 LSTM 细胞的输出。 $b^f, U^f, W^f$  分别是偏置、输入权重和遗忘门的循环权重。因此 LSTM 细胞内部状

态以如下方式更新，其中有一个条件的自环权重  $f_i^{(t)}$ ：

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma \left( b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)} \right), \quad (2-21)$$

其中  $b, U, W$  分别是 LSTM 细胞中的偏置、输入权重和遗忘门的循环权重。**外部输入门 (external input gate)** 单元  $g_i^{(t)}$  以类似遗忘门（使用sigmoid获得一个 0 和 1 之间的值）的方式更新，但有自身的参数：

$$g_i^{(t)} = \sigma \left( b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)} \right). \quad (2-22)$$

LSTM 细胞的输出  $h_i^{(t)}$  也可以由**输出门 (output gate)**  $q_i^{(t)}$  关闭（使用sigmoid单元作为门控）：

$$h_i^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)}, \quad (2-23)$$

$$q_i^{(t)} = \sigma \left( b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)} \right), \quad (2-24)$$

其中  $b^o, U^o, W^o$  分别是偏置、输入权重和遗忘门的循环权重。在这些变体中，可以选择使用细胞状态  $s_i^{(t)}$  作为额外的输入（及其权重），输入到第  $i$  个单元的三个门，如图 2.6 所示。这将需要三个额外的参数。

LSTM 网络比简单的循环架构更易于学习长期依赖，先是用于测试长期依赖学习能力的人工数据集<sup>[??]</sup>，然后是在具有挑战性的序列处理任务上获得最先进的表现<sup>[??]</sup>。

## 2.7 神经网络优化算法

用于深度模型训练的优化算法与传统的优化算法在几个方面有所不同。机器学习通常是间接作用的。在大多数机器学习问题中，我们关注某些性能度量  $\mathbf{P}$ ，其定义于测试集上并且可能是不可解的。因此，我们只是间接地优化  $\mathbf{P}$ 。我们希望通过降低代价函数  $J()$  来提高  $\mathbf{P}$ 。这一点与纯优化不同，纯优化最小化目标  $J$  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特化。

### 2.7.1 随机梯度下降优化算法

随机梯度下降及其变种很可能是一般机器学习中应用最多的优化算法，特别是在深度学习中。按照数据生成分布抽取  $m$  个小批量（独立同分布的）样本，通过计

算它们梯度均值，我们可以得到梯度的无偏估计。展示了如何沿着这个梯度的估计下降。

---

**Algorithm 2.1** 随机梯度下降（SGD）在第  $k$  个训练迭代的更新
 

---

**Require:** 学习率  $\epsilon_k$

**Require:** 初始参数  $\theta$

**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**

---

SGD 算法中的一个关键参数是学习率。之前，我们介绍的 SGD 使用固定的学习率。在实践中，有必要随着时间的推移逐渐降低学习率，因此我们将第  $k$  步迭代的学习率记作  $\epsilon_k$ 。

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  $\mathbf{0}$ ，因此批量梯度下降可以使用固定的学习率。保证 SGD 收敛的一个充分条件是

$$\sum_{k=1}^{\infty} \epsilon_k = \infty, \quad (2-25)$$

且

$$\sum_{k=1}^{\infty} \epsilon_k^2 < \infty. \quad (2-26)$$

实践中，一般会线性衰减学习率直到第  $\tau$  次迭代：

$$\epsilon_k = (1 - \alpha)\epsilon_0 + \alpha\epsilon_{\tau} \quad (2-27)$$

其中  $\alpha = \frac{k}{\tau}$ 。在  $\tau$  步迭代之后，一般使  $\epsilon$  保持常数。

### 2.7.2 RMSProp 优化算法

**RMSProp** 算法<sup>[2]</sup>修改 AdaGrad 以在非凸设定下效果更好，改变梯度积累为指数加权的移动平均。AdaGrad 旨在应用于凸问题时快速收敛。当应用于非凸函数训练神经网络时，学习轨迹可能穿过了很多不同的结构，最终到达一个局部是凸碗的区域。AdaGrad 根据平方梯度的整个历史收缩学习率，可能使得学习率在达到这样的凸结构前就变得太小了。RMSProp 使用指数衰减平均以丢弃遥远过去的历史，使其能够在找到凸碗状结构后快速收敛，它就像一个初始化于该碗状结构的 AdaGrad 算法实例。

RMSProp 的标准形式如算法 2.2 所示，结合 Nesterov 动量的形式如算法 ?? 所示。相比于 AdaGrad，使用移动平均引入了一个新的超参数  $\rho$ ，用来控制移动平均的长度范围。

经验上，RMSProp 已被证明是一种有效且实用的深度神经网络优化算法。目前它是深度学习从业者经常采用的优化方法之一。

---

**Algorithm 2.2** RMSProp 算法
 

---

**Require:** 全局学习率  $\epsilon$ ，衰减速率  $\rho$

**Require:** 初始参数  $\theta$

**Require:** 小常数  $\delta$ ，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**

---

### 2.7.3 Adam

Adam<sup>[?] ]</sup> 是另一种学习率自适应的优化算法，如算法 3.1 所示。“Adam”这个名字派生自短语“adaptive moments”。早期算法背景下，它也许最好被看作结合 RMSProp 和具有一些重要区别的动量的变种。首先，在 Adam 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。结合缩放的动量使用没有明确的理论动机。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 3.1）。RMSProp 也采用了（非中心的）二阶矩估计，然而缺失了修正因子。因此，不像 Adam，RMSProp 二阶矩估计可能在训练初期有很高的偏置。Adam 通常被认为对超参数的选择相当鲁棒，尽管学习率有时需要从建议的默认修改。

### 2.7.4 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率以解决优化深度模型中的难题。此时，一个自然的问题是：该选择哪种算法呢？

遗憾的是，目前在这一点上没有达成共识。<sup>[?] ]</sup> 展示了许多优化算法在大量学习任务上极具价值的比较。虽然结果表明，具有自适应学习率（以 RMSProp 和 AdaDelta 为代表）的算法族表现得相当鲁棒，不分伯仲，但没有哪个算法能脱颖而出。

**Algorithm 2.3** Adam 算法**Require:** 步长  $\epsilon$  (建议默认为: 0.001)**Require:** 矩估计的指数衰减速率,  $\rho_1$  和  $\rho_2$  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  $\delta$  (建议默认为:  $10^{-8}$ )**Require:** 初始参数  $\theta$ 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**

目前, 最流行并且使用很高的优化算法包括 SGD、具动量的 SGD、RMSProp、具动量的 RMSProp、AdaDelta 和 Adam。此时, 选择哪一个算法似乎主要取决于使用者对算法的熟悉程度 (以便调节超参数)。

## 2.8 本章小结

## 第三章 基于 CAE-CNN 的无线信号调制识别

### 3.1 引言

深层神经网络是由一系列层组成的大函数近似，其中每层代表基于具有一些学习权重集的参数传递函数的从输入到输出激活的一些变换。每一层通常是一个已知的具有可调参数和非线性激活函数的线性函数，因此得到的函数组合可以是高度非线性的 [3]。深度神经网络中的函数参数通常由梯度下降优化器根据网络输出计算出的一些损失函数进行训练。对于诸如调制识别的多类分类任务，目标函数通常是分类交叉熵（方程 1）。分类交叉熵是衡量两个概率分布之间的差异的度量。对于深度学习分类任务，概率分布通常是分类器网络输出的 softmax（等式 2），然后为了分类目的将其转换为单热编码 [3]。误差是在所谓的正向传递中计算的，并且使用链规则调整权重以找到在所谓的反向传递中的每个参数的误差贡献。这种网络输出层，优化和丢失函数已经成功用于多类视觉任务，如 Imagenet 数据集上的对象识别 [7]。

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 [13] [12][10]。[Tim Oshea] 最近证明了利用原始数据进行有监督无线调制识别 [14] 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，提出了一种基于卷积自编码器（CAE）与卷积神经网络（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 调制信号生成

无线接收端的信号实际上是信号经过信道作用得到的。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，而且受到多径等效应的影响，很难直接使用真实数据进行训练。我们利用 XXX 仿真仪，构建通信系统框架，通过多径信道和高斯信道来获取近似真实的仿真信号。

### 3.2.1 信道建模

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通过无线信道作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。

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，我们的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，以获得  $2 \times N$  的复数值向量。

$$r(t) = s(t) * c + n(t) \quad (3-1)$$

我们将接收信号用等式3-1表示。其中，将连续信号或一系列离散的时间序列信号，调制到具有变化的频率、相位、振幅、或多个变换的正弦波上，得到调制信号  $s(t)$ 。 $c$  是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。从工程的角度来看，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 3-2：

$$r(t) = e^{j * n_{Lo}(t)} \int_{\tau=0}^{\tau_0} s(n_{Clk}(t - \tau)) h(\tau) + n_{Add}(t) \quad (3-2)$$

方程3-2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3-1 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。



### 3.2.2 调制信号获取

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。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。

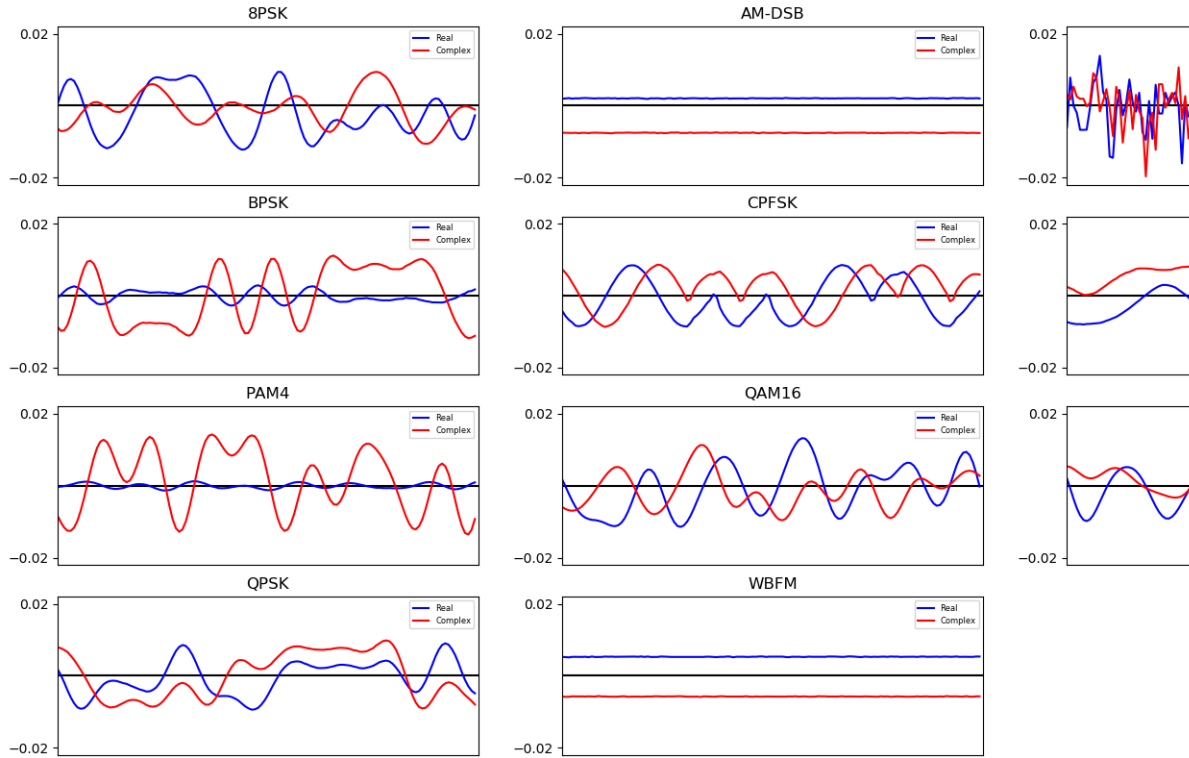


图 3.1 信号生成框图

我们使用 GNU 无线信道模型 [14] 模块在 GNU Radio[3] 中对这个数据集的生成进行建模，然后使用 128 个样本矩形窗口过程将每个时间序列信号切片成测试和训练集。总数据集大约 500 MBytes 作为一个 python pickle 文件存储与复杂的 32 位浮点样本。

这些数据有望成为该领域其他人的极大用处，并可以作为这一领域的基准。这个数据集可以在 <http://radiom.com> 上以 pickle python 格式获得，由时间窗口的例子和相应的调制类别和 SNR 标签组成。我们希望扩大调制范围和渠道现实主义作为这方面的兴趣。

我们专注于由 11 个调制组成的数据集：8 位数字调制和 3 位模拟调制，都被广泛

应用于我们周围的无线通信系统。这些包括 BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK 和 PAM4 或数字调制, 以及用于模拟调制的 WB-FM, AM-SSB 和 AM-DSB。数据以大约每个符号 8 个采样的速率进行调制, 标准化的平均发送功率为 0dB。

为了生成一个表征良好的数据集, 我们选择一系列在实践中广泛使用的调制, 并以离散二进制字母(数字调制)和连续字母(模拟调制)进行操作。我们调制每个调制解调器上的已知数据, 并使用 GNURadio 将它们分别暴露于上述频道效应。我们将数百万个样本分割成由许多短时窗组成的数据集, 其方式类似于连续的声学语音信号通常为语音识别任务而开窗。我们提取了 64 个样本的 128 个样本的步骤, 以形成我们提取的数据集。

在分割之后, 假设采样率大约为 1MSamp/sec, 则示例大致为 128 $\mu$ s。每个包含 8 到 16 个带有随机时间偏移, 缩放, 旋转, 相位, 通道响应和噪声的符号。这些例子表示关于调制数据比特的信息, 关于它们如何被调制的信息, 关于在传播期间信号通过的信息, 以及关于发送和接收器设备状态以及包含的随机过程的状态的信息。我们专注于恢复关于信号如何被调制的信息, 从而根据与调制方案相对应的离散的 11 类标签集来标记数据集。

自然图像有其固有特性, 也就是说, 图像的一部分的统计特性与其他部分是一样的。这也意味着我们在这一部分学习的特征也能用在另一部分上, 所以对于这个图像上的所有位置, 我们都能使用同样的学习特征。

我们专注于使用两种形式的卷积神经网络来学习原始采样无线电信号时间序列示例的稀疏表示。我们利用 RadioML16.04 [15] 标记的无线电调制数据集进行 11 种调制, 其中包括白噪声, 振荡器漂移, 采样时钟漂移和衰落的影响。这个数据集的例子如图 1 所示。

### 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3.2)和频域(图3.3)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, 同事也具备一定差异

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 XX 信号具备较明显的特征，而 XXX 特征在视觉上让人感觉像是噪声，很难直接判断出来。

如图(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，从视觉来说这是一个困难且繁琐的判定方法。

### 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（PCA）或独立分量分析（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。

卷积自动编码器（Convolutional Autoencoder, CAE）非常适合于减小参数空间，获取的卷及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均方误差（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。

图3.3显示了我们的卷积自动编码器使用的体系结构。

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，使用均方误差（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，并使用随机梯度下降（SGD）算法等，以找到接近等式3-4中的最佳网络参数。

$$\arg \min_{\theta} \left( \sum (X - f(X, \theta))^2 \right) \quad (3-3)$$

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

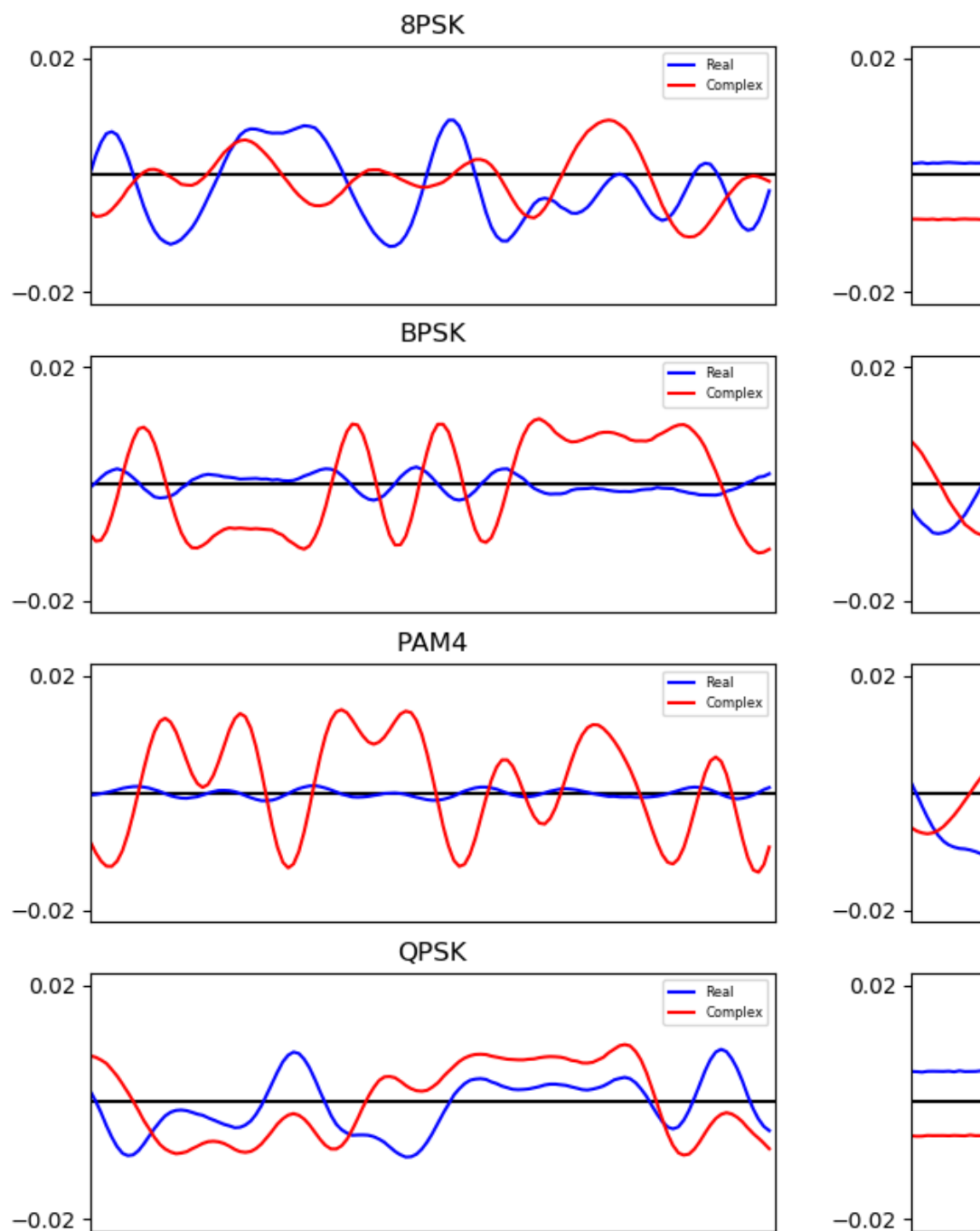


图 3.2 不同调制方式的高 SNR 样本的时域波形

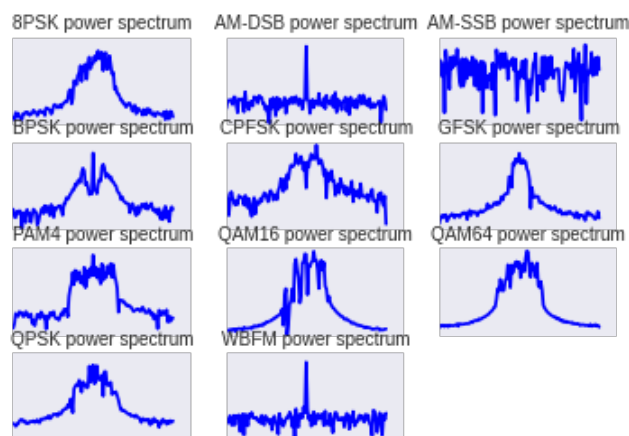


图 3.3 不同调制方式的高 SNR 样本的功率谱

无线电时间序列信号表示。我们使用 `dropout`[13] 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。图2.5显示了我们的卷积自动编码器使用的体系结构。

»»»»> master

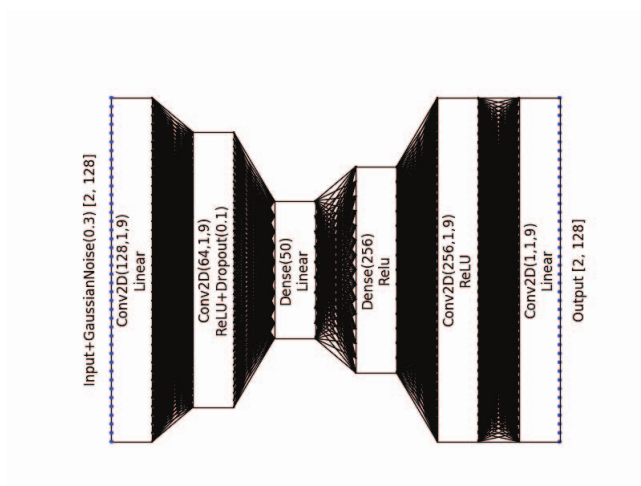


图 3.4 自编码器

我们使用 `RMSProp`[11] 和 `Adam`[12] 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。

最终，我们训练卷积自编码器的网络参数如表 XXX 所示：

图3.5显示了两个输入维度为  $2 \times 128$  的训练样本，中间隐层维度为  $1 \times 30$ ，以及输出维度为  $2 \times 128$  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。

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

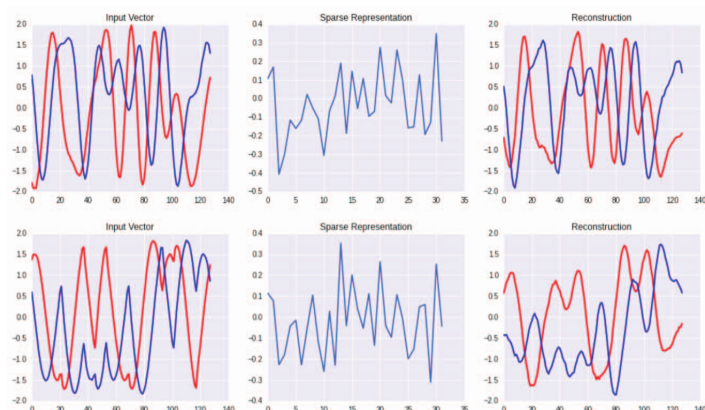


图 3.5 自编码器

将数据的低维嵌入特征，利用 t-分布随机邻接嵌入 (t-asdfasdfa, t-SNE) [6] 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 X:

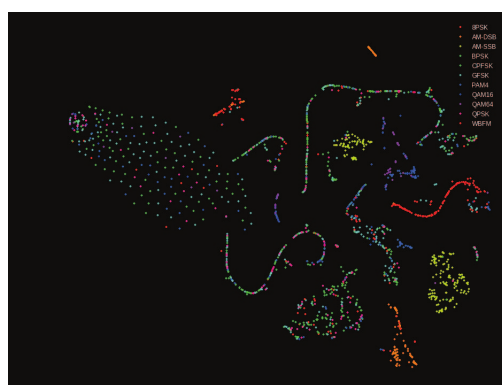


图 3.6 自编码器

在这种情况下，我们看到几个类如 WBFM, AMDSB, AM-SSB 和 QPSK 已经形成了独立的、大部分可分离的簇，可以利用 DBSCAN 等聚类方法形成单独的类别；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是很好，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。

### 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。TIM Shead 在他的工作 [14] 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。图 X 中显示了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。



图 3.7 自编码器

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：

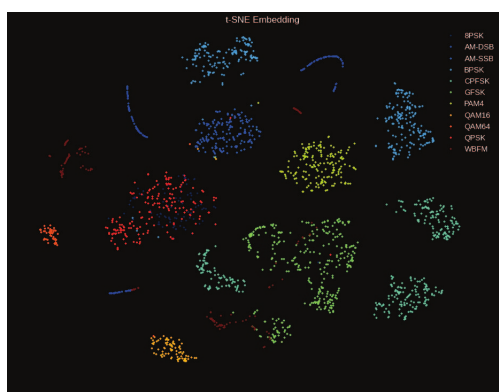


图 3.8 自编码器

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 X 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。

### 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在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3.9展示了我们网络的结构：

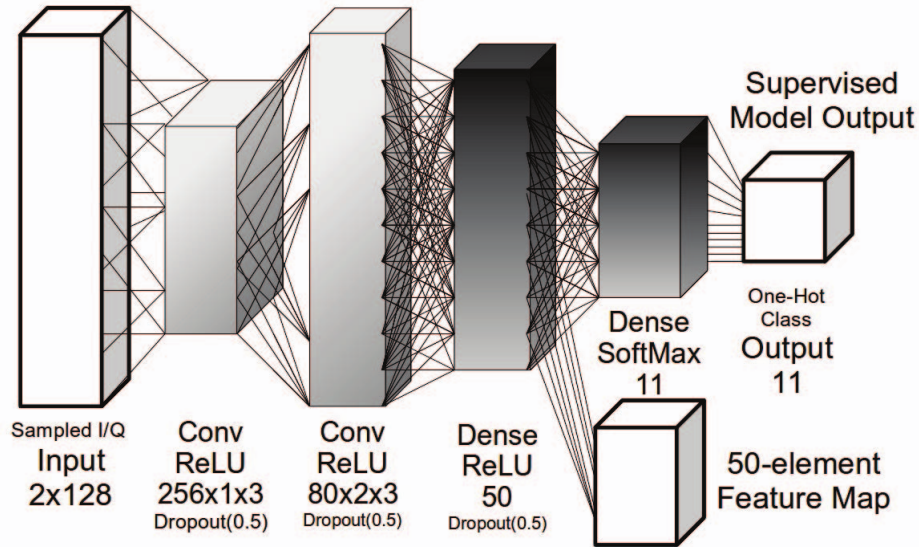


图 3.9 自编码器

在 CAE-CNN 中的卷积层中，我们使用了 Dropout，降低模型的过拟合；并在卷积权重增加了权重值  $W$  的 2 范数作为惩罚项，使权重尽量小；同时，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 [5][10]。



那么我们有 CAE 的损失函数：

$$r(t) = s(t) * c + n(t) \quad (3-4)$$

由于 CNN 的损失函数为重构误差与分类误差之和，那么我们有 CNN 的损失函数：

$$r(t) = s(t) * c + n(t) \quad (3-5)$$

### 3.4.2 CAE-CNN 算法

---

#### Algorithm 3.1 Adam 算法

---

**Require:** 步长  $\epsilon$  （建议默认为：0.001）

**Require:** 矩估计的指数衰减速率， $\rho_1$  和  $\rho_2$  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  $\delta$  （建议默认为： $10^{-8}$ ）

**Require:** 初始参数  $\theta$

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**

---

### 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 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。我们的神经网络参数如表3.1所示：

图 3 显示了 CNN 架构的一个例子.CNN2 是相同的但是较大的，在层 1 和层 2 中包含 256 和 80 个过滤器，在层 3 中包含 256 个神经元。评估的 DNN 包含 4 个密度层，大小分别为 512,256,128，和  $n$  级神经元。

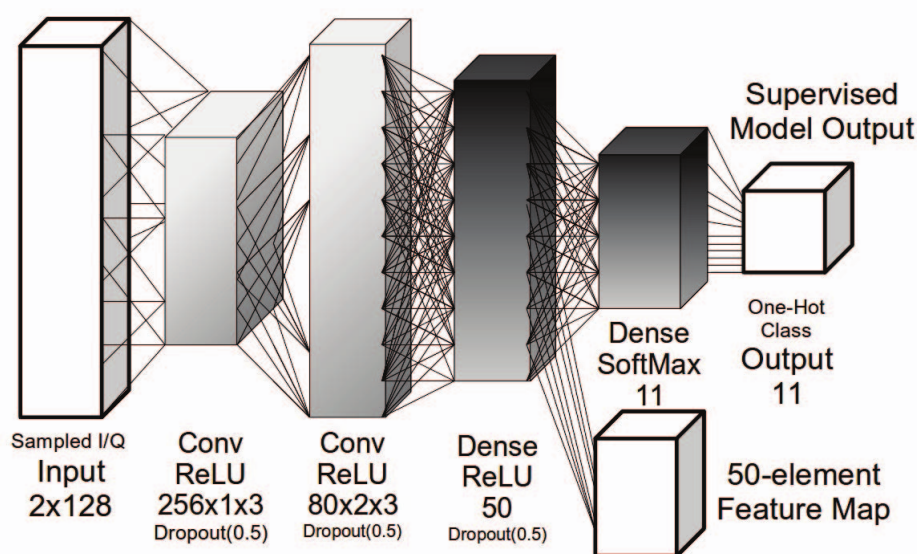


表 3.1 自编码器

## 3.5 结果及分析

### 3.5.1 学习复杂度

我们使用 Adam 求解器训练了大约 23 分钟的最高复杂度模型，批量大小为 1024 的样本训练集大约需要 15 秒。我们确实观察到一些过度拟合，尽管没有正规化，但验证损失确实没有显著变化，我们保持最佳的验证损失模型进行评估。

绘制学习的特征有时可以让我们直觉了解网络正在学习的底层表示。在这种情况下，我们在下面绘制卷积层 1 和卷积层 2 的滤波器权重。在图 5 中，第一层，我们有 64 个  $1 \times 3$  的过滤器。在这种情况下，我们只需获得一组边缘和梯度检测器，它们在每个 I 和 Q 通道上进行操作。

在卷积层 2 中，如图 6 所示的权重，我们将这个第一层特征图组合成  $64 \times 16 \times 2 \times 3$  较大的特征图，其包括在 I 和 Q 通道上同时出现的情况。这些特征图与在包括 2D 学习边缘检测器和 Gabor 滤波器的图像转换网络的较低层处所看到的特征图看起来没有太大的不同。

### 3.5.2 分类准确率与鲁棒性

#### 分类结果

为了评估分类器的性能，我们看一下测试数据集的分类性能。我们调制方式总共有 11 中，每一类调制信号在噪声信噪比为  $-20dB$   $18dB$  之间每隔  $2dB$  进行数据采样，每个信噪比下采样 1000 个样本，每个样本共包含 128 个采样值；我们将采样值进行希尔伯特变化，这样每个采样值便分为实部与虚部。我们将 80% 的样本作为训

训练集，10% 的样本作为验证集，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 +20dB 的 SNR 中，并被标记以便我们可以评估特定子集上的性能。



图 3.10 整体结果

对于我们最高的 SNR 情况下的 CNN2 (0.6) 分类，我们在图 8 中显示了一个混淆矩阵。在 +18dB SNR 时，在混淆矩阵中我们有一个干净的对角线，可以看到我们剩下的差异是 8PSK 误分类为 QPSK，WBFM 误分类作为 AM-DSB。这两个都可以在基础数据集中解释。由于 QPSK 星座点由 8PSK 点跨越，所以包含特定比特的 8PSK 符号从 QPSK 难以分辨。在 WBFM /AM-DSB 的情况下，模拟语音信号具有只有载波音调存在的静默时段，这使得这些示例不可见。因此，即使在这个数据集的高信噪比下，也不可能获得 100% 的准确度，并使得重新合理的混淆被合理地容忍。

### 0dB 误分结果



图 3.11 0dB 条件下的结果

### 与其他算法的比较



图 3.12 与其他方法的比较

在训练之后，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 87.4%，但要理解这个意义，我们必须检查这个分类精度如何在不同训练样本的 SNR 值之间进行分解，以及它与现有的基于专家特征的分类器的性能进行比较。绘制测试集调制分类精度，作为每个分类器的示例信噪比的函数 7。实线表示直接在无线电时间序列数据上进行深度特征学习训练的分类器，而虚线表示使用前面描述的专家特征作为输入的分类器。这种观点是检验结果的关键方法，因为在低信噪比影响范围和覆盖范围的性能，我们可以有效地使用分类器。我们从具有大量丢失正则化（0.6）的大卷积神经网络（CNN2）中获得显著更好的低 SNR 分类准确性性能。在低信噪比情况下，最佳 CNN 模型的性能比基于专家特征的系统的信噪比高 2.5-5dB，而 +5dB SNR 性能相似。这是一个显著的性能改进，可能至少是传感系统有效覆盖面积的两倍。

### 3.5.3 训练效率以及分类效率

#### a) 训练时间

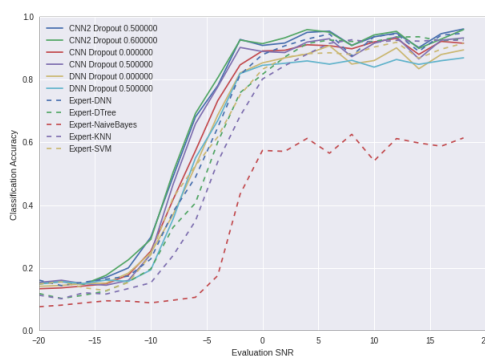


图 3.13 训练时间

为了更好地理解性能如何随信噪比而变化，我们检查了不同信噪比级别的几个分类器的混淆矩阵。

在非常低的信噪比（-6dB）的情况下，在图 9,10,11 和 12 中，我们看到一个有趣的情况，其中  $\pm 20\%$  内的所有精度都在 50% 左右。在这种情况下，CNN2 分类器上的清洁器对角线比其他 3 种情况显著更明显，在这个区域学习到的特征具有显著的性能优势。

现在，所有 4 个分类器的信噪比（0dB）略高但仍然很低，现在有一个明确的对角线，但是我们发现在 8PSK 情况下发生的非对角线误分类更少。

#### b) 分类时间

许多无线电系统中的一个重要考虑因素是训练和分类运行时间，由于计算复杂性。深度学习的一个普遍批评是对大量计算资源的需求，然而在本文中，我们的网络

相对紧凑，数据集相对较小。我们比较下面每个模型的训练和分类运行时间。在图 17 中我们可以看到，我们的 CNN 模型确实需要大量的训练时间，但是比 SVM 训练案例所需要的时间要少。

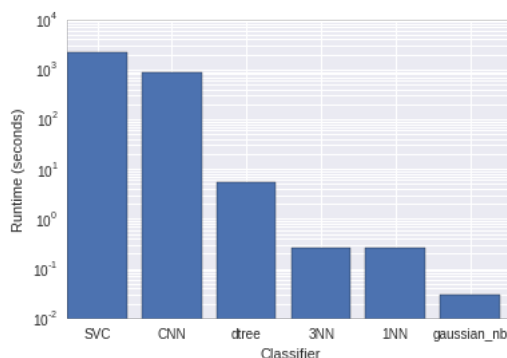


图 3.14 分类时间

在图 18 中显示，使用 Keras 编译 python 的这个模型的分时间比使用 scikit-learn 的最近邻和 SVM 模型的大多数其他模型显著更快。只有决策树和 GaussianNB 模型获得更快的分类运行时间。在这两种情况下，基于 ConvNet 的这种规模的这种数据集分类模型提出了一个有吸引力的选择这个任务时，分类性能考虑。

### 3.6 本章小结

一旦示例在嵌入空间中形成相对可分的簇，我们可以使用任意数量的聚类算法来将它们中的每一个分组并将其分配给类标签。在图 9 中，我们展示了一个使用 DBSCAN [4] 聚类算法的例子，将聚类分成一组未知但不同的调制类。我们发现这种聚类方法比较适合在我们的压缩空间中形成的不明确形状的聚类。聚类之后的数据处理可以包括标记许多示例的集群而不是每个单独的示例，从而提供从数据保存时间的效率提高的数量级。在包含许多类和例子的非常大的数据集上，这使得管理大规模学习任务比其他方法更容易处理。

虽然这些集群并非没有错误，但是在这个例子中，通常我们可以找到从发现的类集合到不同的真实命名类的一对一或多对一的映射。这保证了这样一种方法可以在将来用于快速组织和标记大量的无线电发射，并利用关于发射器类别特征的先前知识，但是仍然允许随着时间的推移识别系统能力，特征和类别标签的缩放最大限度地减少这样做所需的人力劳动。

虽然这些结果并不是现有的最好的基于专家特征的调制分类器的全面比较，但是它们证明了，与相对专业的被认为的方法相比，时间序列无线电信号数据上的盲卷积网络是可行的并且工作得很好。在图 7 中，我们比较了几种分类器策略的准确性和信噪比，并且认为对于低信噪比和短时间的示例（128 个复杂采样），这代表了调制

分类的最先进的精确度方法。这种方法有可能容易地扩展到额外的调制类别，并且应该被视为依赖于无线电发射器的稳健的低 SNR 分类的 DSA 和 CR 系统的有力候选。

我们的结果与当前最好的专家系统方法的合理近似相比较，但是由于在无线电领域新兴的机器学习领域不存在强大的竞争数据集，所以很难直接比较性能和当前的现有技术状态。我们希望在以后的工作中进一步评估这一点，并将特征学习和专家方法从目前的水平上进行改进。CNN2 网络体系结构上的性能改进是不可避免的，我们花费了一些努力来优化它，但并没有做到这一点。较大的过滤器，不同的体系结构和池化层可能会显著影响性能，但是在这项工作中没有充分考虑其适用性。许多附加技术可以应用于这个问题，包括引入附加通道引起的效应的不变性，例如膨胀，I/Q 不平衡，相位偏移等等。空间变换网络 [17] 已经证明了学习图像数据的这种不变性的强大能力，并且可以作为一个有趣的候选者，使得能够改善对这些效应的不变性学习。序列模型和递归层 [13] 可能能够表示信号序列嵌入，并且在更长时间表示中几乎肯定会证明是有价值的，但是我们还没有完全调查这个区域。这个应用领域已经成熟，可以进一步研究和应用，这将大大影响无线信号处理和认知无线电领域的技术发展水平，并将其转向机器学习和数据驱动方法。

## 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，使用特征融合方法解决分类或者回归等问题。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，来提高模型的精度和鲁棒性。

在上一章中，我们分别展示了 CAE 以及 CNN 对调制信号所提取特征的类可分性，即我们可以利用 CNN 提取数据样本的类别特征。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

统计模式识别分类方法的研究重点是特征提取方法和分类准则的选择与训练。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、循环平稳特性等。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， $y(t)$  为实信号的希尔伯特（Hilbert）变换。

通信信号的瞬时幅度、瞬时频率和瞬时相位等时域特征包含了丰富的调制信息，合理选择和构造它们的各阶统计量是获得调制识别特征的有效途径。

##### 1) 零中心归一化瞬时幅度谱密度的最大值

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

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  $A(t)$  的平均值，用平均值对瞬时幅度进行归一化的目的

是消除信道影响。 $\gamma_{max}$  表征了信号瞬时幅度的变化情况，反映调制信号包络的变化特性，可以通过一定的判别门限，区分恒定包络和非恒定包络的信号。

## 2) 零中心归一化非弱信号瞬时幅度标准差 $\delta_{da}$

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

其中， $C$  是全部  $N$  个采样数据中属于非弱信号值的个数，非弱信号是指幅度大于幅度判决门限电平  $a_t$  的信号。 $\delta_{da}$  表征一个符号区间内信号的幅度变化信息，可以用来区分一个符号区间内归一化中心瞬时幅度的调制方式和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 $\delta_{aa}$

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

$\delta_{aa}$  表征了信号的绝对幅度信息，可区分不具备归一化的绝对幅度信息的调制方式，以及具备归一化的绝对幅度信息的调制方式。

## 4) 零中心归一化瞬时幅度的四阶紧致性 $\mu_{42}^a$

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

$\mu_{42}^a$  是用来度量“瞬时幅度分布的密集性”的特征值，可以用来区分瞬时幅度高密集分布信号和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 $R_f$

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

式中， $u_f$  和  $d_f$  分别代表信号零中心归一化瞬时频率均值和方差。此参数可用来判断信号是否含有频率信息。

## 6) 零中心非弱信号段归一化瞬时频率绝对值的标准差 $\delta_{af}$

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

$\delta_{af}$  表征信号的绝对频率信息，可用来区分归一化中心瞬时频率绝对值为常数的调制方式和具有绝对、直接频率信息的调制方式。



### 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，适合于区分幅度相位调制方式，并具有抗噪声等优点，因此基于高阶统计量特征的调制识别算法也是一类有效的分类算法。

对于零均值的高斯白噪声信号，三阶及三阶以上的高阶累积量为零。根据这一特性，求取接收信号  $r(t)$  的高阶累积量就是求取发送信号  $s(t)$  的高阶累积量。从而高阶累积量在对零均值高斯白噪声具有有效的抑制作用以及其它一定程度上的干扰抑制。对于不同数字调制信号的高阶累积量的值不同，而且不受零均值高斯白噪声的影响，因此可以有效提取数字调制信号的高阶累积量为特征值，进而区分识别不同信号的调制方式。

对于零均值的平稳随机过程  $x(t)$ ，其  $k$  阶矩定义为：

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

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。

根据文献 [9] 中的高阶矩与高阶累积量的定义，我们由零均值的平稳随机过程  $x(t)$  的各阶累积量的公式。二阶累积量：

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

四阶累积量：

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

六阶累积量：

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

四阶累积量与二阶累积量之比：

$$R_{mn} = \left| \frac{C_{42}}{C_{21}^2} \right| \quad (4-13)$$

### 4.2.3 专家循环矩特征

传统的信号分析模型都是以平稳随机过程为基础的，而通信信号经过调制、周期性采样、编码等具有周期平稳性。循环统计量可以很好地刻画通信信号的这种周期性。基于综合循环矩特征 [1] 目前广泛流行的调制识别和形成分析导出的决策树调制分类到不同的类别。一般来说，它们采用方程 4-14 给出的形式。

$$snm = f_m(x^n(t) \dots x^n(t + \tau)) \quad (4-14)$$

通过计算关于瞬时或时间延迟的接收信号  $r(t)$  的  $n$  次方的第  $m$  阶统计量，我们可以获得一组统计量，该统计量在给定特征的决策过程的情况下将其与其他调制方式唯一地分开。对于我们的专家功能集，我们计算 32 个功能。这些由 0 和 8 个样本的循环时间滞后组成。复数接收信号的前 2 个功率的前 2 个时刻，每个时滞的幅度，相位和相位的绝对值。

## 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的某些方向特征，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。这样，纯粹的基于人工特征提取训练的分类器就无法输出正确的分类。

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，它既可以直接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。

深度学习理论是在人工神经网络的基础上发展起来的机器学习理论，在多层神经网络中加入了更多隐层单元。基于深度学习的特征融合算法，主要是对深度神经网络模型引入特征融合的思想，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：

- 1) 多特征融合通常比单特征具备更好的分类性能；
- 2) 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，我们实际上可以用更小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：

$$h(x) = \bigcup_{0 < i < N} w_i * h_i(x) \quad (4-15)$$

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：

$$f(x) = \bigcup_{0 < i < |Z|} \mu_i * f_i(x) \quad (4-16)$$

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  $Z$ ：

$$Z = F \cup H = \{z | (z \in H) \vee (z \in F)\} \quad (4-17)$$

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：

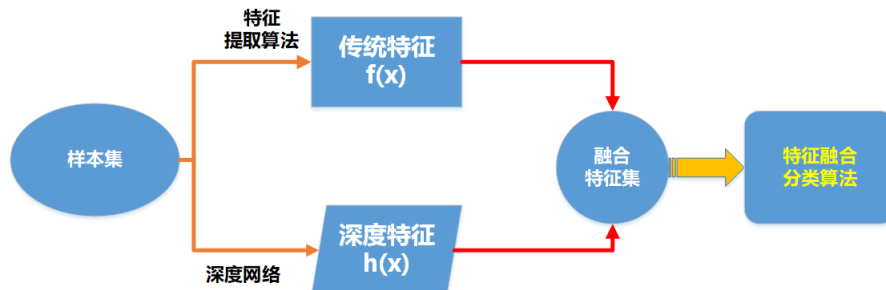


图 4.1 特征融合框架图

其中，XX 表示我们利用的不同融合算法。在下一节中，我们将对不同融合算法的效果进行研究。

## 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也称为基于特征的调制识别方法。这种方法把通信信号的调制识别视为一个统计模式识别问题，整个调制识别系统由两个子系统组成：特征提取子系统和模式分类子系统。特征提取子系统的作用是从原始测数据中提取事先定义好的能表征信号调制类型的特征，可以看作是从输入信号所在的观测空间到选定的特征空间的一个映射。

### 4.4.1 基于 Softmax 回归 (Softmax Regression, SR) 的融合框架

LR 是用线性回归模型的预测结果去逼近真实标记的对数几率的一种二分类算法。SVM 是一种基于核函数的方法，它通过某些核函数把特征向量映射到高维空间。SVM 是一种基于核函数的方法，它通过某些核函数把特征向量映射到高维空间，然后建立一个线性判别函数对数据进行分类。由于 SVM 训练时间较长，而线性核函数的 SVM 与 LR 的结果基本一样【吴恩达】，因此，为了兼顾算法的有效性和高效性，我们选择 LR 框架作为线性特征融合的框架。

根据【吴恩达】，我们有样本为正样本的概率：

$$h_{\theta}(x) = p(y = 1|x) = \frac{1}{1 + e^{-(\theta^T x)}} \quad (4-18)$$

我们用  $Cost(h_{\theta}(x^{(i)}), y^{(i)})$  表示样本预测错误时的损失，根据【1】，在此我们使用逻辑损失函数：

$$Cost(h_{\theta}(x^{(i)}), y^{(i)}) = \begin{cases} -\log(h_{\theta}(x^{(i)})) & y = 1 \\ -\log(1 - h_{\theta}(x^{(i)})) & y = 0 \end{cases} \quad (4-19)$$

最终，我们有 LR 的损失函数：

$$\begin{aligned} J(\theta) &= -\frac{1}{m} \sum_{i=1}^m Cost(h_{\theta}(x^{(i)}), y^{(i)}) \\ &= -\frac{1}{m} \sum_{i=1}^m \left[ y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \right] \end{aligned} \quad (4-20)$$

由于，我们传统的 LR 算法针对的是二分类问题，因此，对于调制识别这样的多分类问题，我们需要对算法进行改进。在此，我们使用的是 LR 多分类算法中的 1VM 版本【1】，即假设我们的调制类别共有  $M$  类，则我们需要训练  $M$  个分类器，分别判断样本属于每一个类别的概率值，取概率值最高的那一个类别作为我们的分类结果。

最终的基于 LR 的特征融合框架，如图4.2所示：

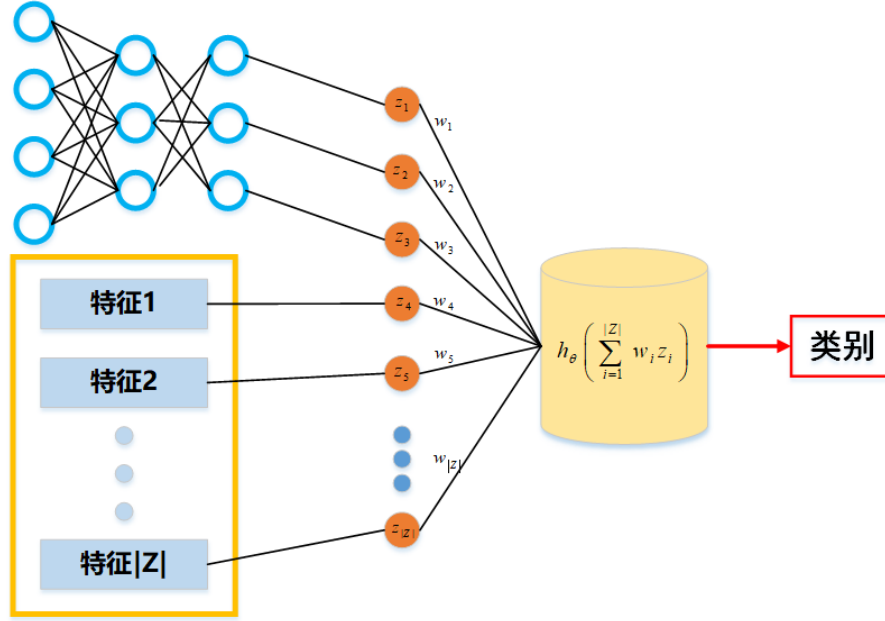


图 4.2 基于 LR 框架的特征融合框架图

#### 4.4.2 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。Simonyan 等 [16] 首先提出了一种使用双流架构的深度卷积神经网络模型，可解决视频中的动作识别问题。Feichtenhofer 等 [15] 则在这个基础上改进了网络融合方法，提出了空间特征融合方法和时间特征融合算法。

“如果我们的深度网络达到一定的条件，可以拟合任意的函数。”，XXX 说。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，来输出我们想要的结果。

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  $F$ 。

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：

$$\hat{Z} = \{f_{norm}(z) | z \in H \cup F\} \quad (4-21)$$

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。XXX 等人提出了可以利用卷积层、池化层等对不同的特征进行融合处理。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：

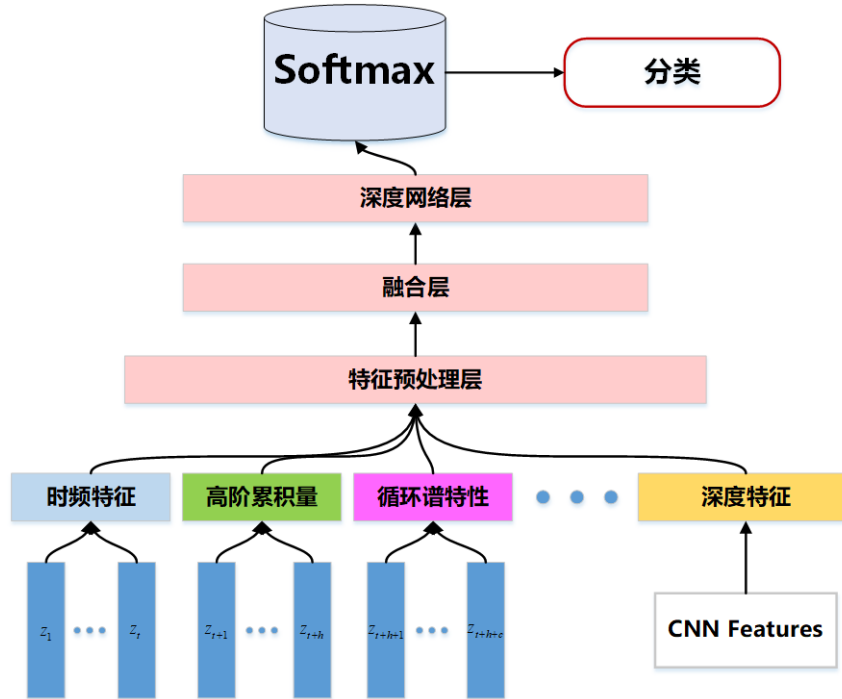


图 4.3 基于 DNN 框架的特征融合框架图

#### 4.4.3 基于随机森林的融合框架

集成学习通过构建并结合多个基学习器来完成学习任务，按照算法的思想它主要分为两类：Bagging 族算法和 Boosting 族算法。

Bagging [Breiman, 1996a] 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 算法的思想是先训练简单的基学习器，再根据基学习器的表现对训练样本分布进行调整，使得先前基学习器做错的训练样本在后续受到更多关注，然后基于调整后的样本分布来训练下一个基学习器，直至基学习器数目达到事先指定的值  $T$ ，最终将这  $T$  个基学习器进行加权结合。Boosting 主要通过过个学习器的加性组合降低偏差。

随机森林 (Random Forest, RF) [Breiman, 2001a] 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。它通过自助法 (bootstrap) 重采样技术，从原始训练样本集  $N$  中有放回地重复随机抽取  $k$  个样本生成新的训练样本集合，然后根据自助样本集生成  $k$  个分类树组成随机森林，新数据的分类结果按分类树投票多少形成的分数而定。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 [1] 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、staggering 学习法等方法, 平均法主要是针对回归问题, 而 staggering 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??所示:

$$C(x) = \arg \max_c \sum_{i=1}^K I(T_i(x), c) \quad (4-22)$$

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, 可通过个体学习器之间差异度的增加而进一步提升。

图??展示了我们基于随机森林算法的特征融合框架图:

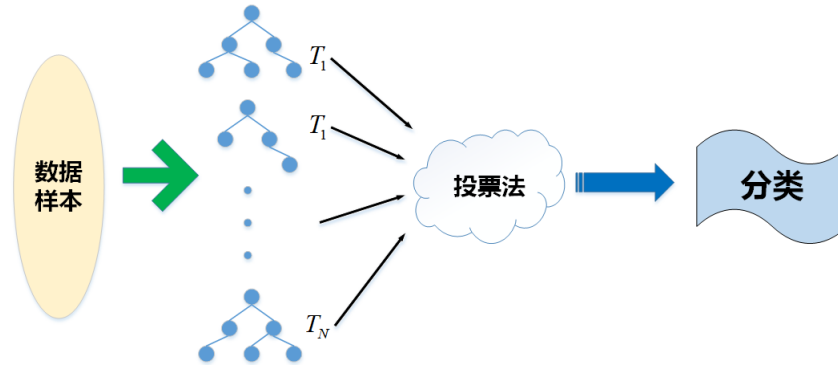


图 4.4 基于随机森林框架的特征融合框架图

。随机森林 (RandomForest, RF) 基于划分与模型融合的方法。降低了泛化误差。

## 4.5 结果及分析

### 4.5.1 分类性能比较

我们利用 11 类数据对三种框架的模型进行训练, 对于不同框架模型的超参数如表 1 所示: ..... 最后的分类结果如图 2 所示:

通过上图, 我们可以发现在高信噪比条件下, 三种框架的差别并不太大, 但是利用深度 CNN 的融合框架更好一些, 可能是因为这种情况的拟合能力更强一些, 由于训练数据本身具有较小的噪声, 所以对于强拟合能力的模型而言, 泛化误差影响相对较小; 在 0dB 时候, 三种框架的性能有: A》B》C; 在信噪比低于》》》时, XX 的性

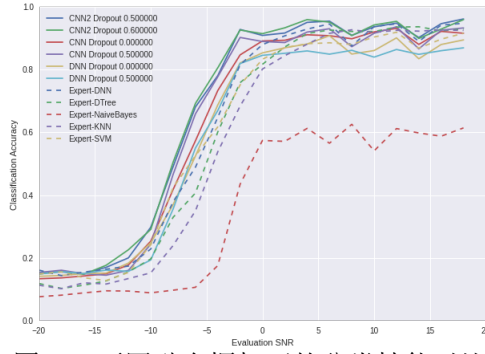


图 4.5 不同融合框架下的分类性能对比

能更好一些，因为 RF 是通过随机的样本采样和特征采样来降低泛化误差，对噪声数据相对不敏感一些。

## 4.5.2 混淆矩阵

### 4.5.2.1 RF 混淆矩阵

### 4.5.2.2 DNN 混淆矩阵

### 4.5.2.3 SR 混淆矩阵

仔细检查这些结果，我们可以看到 BPSK 不幸与 QPSK 和 8PSK 调制（它是其中的一个子集）相当混杂。然而，QAM16 是一个以前看不见的类，这些特征在 QAM64 附近紧密聚集在一个相当明确的可分离的嵌入空间区域。这些结果虽然非常初步和定性，但确实支持这一事实，即自举特征映射具有显著的泛化能力，能够识别、聚类 and 辨别新的未知的或以前看不见的调制类型，但是并不保证在所有情况下都具有清晰的可分离性。我们希望随着类别和特征数量的增加，这种泛化将会得到改善。推进这一研究领域的挑战之一将是确定和量化特征总结能力的衡量标准，并着力于改进这一指标。

## 4.6 本章小结

在这项工作中，我们已经证明，使用原始采样无线电时间序列数据上的卷积神经网络学习的低级别时间序列特征可以用于有效地聚类许多无线电信号调制类型，而没有明确标记的训练数据。我们已经表明，通过利用从差异特征映射和压缩重构空间学习的压缩表示，我们可以开始组织和构造复杂的无线电信号数据集与未标记或标记不佳的起点。这是一个强有力的结果，因为它展示了一个潜在的前进方向，可以学习区分、推理、回忆和描述新的和未知的无线电信号，而无需手动指导或专家指导。这是一个关键的要求，因为我们试图建立随着时间的推移而从经验中扩展能力的系统。将压缩的特征空间基础推广到新的信号类型仍然是这个领域的一个关键挑战，但



是我们在这项工作中已经表明，在某些情况下这种特征泛化确实发生。展望未来，量化和优化这种效应的尝试将是重要的。

特征融合方法是模式识别领域的一种重要方法。计算机视觉领域的图像识别问题作为一种特殊的模式分类问题，仍然存在很多挑战。特征融合方法能够综合利用多种图像特征，实现多特征的优势互补，获得更加鲁棒和准确的识别结果。笔者基于信息融合理论分析了特征融合方法的原理，介绍了特征融合方法的研究现状，讨论了特征融合与 3 类主流基础理论相结合的方法，其中基于贝叶斯理论的特征融合算法可以实现多特征的融合决策，基于稀疏表示理论的特征融合算法能够得到多特征的联合稀疏表示，基于深度学习理论的特征融合算法能够强化深度神经网络模型的特征学习过程。



## 第五章 调制识别的深度框架研究

### 5.1 引言

目前存在几种较完善的网络，比如包括多层感知机、RNN、CNN 及其许多变体等等。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，比如谷歌提出的长短期深度神经网络（CLDNN）、Inception 模型，微软的残留网络（ResNet）等。

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，从网络底层结构单元的层面，研究当前在各个行业表现优异的不同的网络底层结构对调制识别性能的影响。

### 5.2 超参数对 CNN 的调制识别的性能影响

#### 5.2.1 卷积神经网络

在大部分现有的深度神经网络中，卷积层成了它们共同的元素。Lecun 成功地将卷基层应用到手写数字识别 [1] 中，以保证特征的平移不变性 [8]。卷积层由  $N_f$  个卷积滤核组成，每个卷积核通常非常小（1x1 到 5x5 是图像处理中的常见尺寸）。在传统的 DSP 应用中，滤波器通常设计得非常宽（很多分支/高阶），而非设计成多层很深的结构（小分支/级联）。而计算机视觉领域的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功能和层次特征关系 [2]，[1]。

标准卷积层 [9] 的传递函数在等式 3 中给出，其中  $y_i$  是第  $i$  个滤波器的输出特征图， $b$  和  $k$  代表学习偏差和滤波器权重参数， $x_i$  代表输入激活， $*$  表示卷积运算，并且  $f(\cdot)$  表示诸如整流线性单元（ReLU）或 S 形的（通常非线性的）激活函数：

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

深度网络使得可以从原始数据中更容易地学习更复杂的功能，而不是使用相同数量参数的浅层网络 [1]，[18]；然而，人们普遍认为神经网络中的深度受不稳定梯度的限制，不稳定梯度在网络中的早期层或后期层中爆炸或消失。近年来，通过在优化器中使用梯度归一化以及不会加剧消失梯度问题的非线性（如整流线性单位（ReLU）），近年来该问题得到了改善。因此，几个重要的架构已被用于赢得竞争，如 ImageNetby 增加深度，我们将着眼于提高无线电调制识别。

### 5.2.2 GoogLenet

GoogLenet [17] 中使用的初始架构是一种成功的方法，可以提高网络深度，并能够将不同规模的特性推广到一般管理复杂性。该网络由重复启动模块组成。每个启动模块（如图 1 所示）包含四条并行路径，输出是四个并行输出的串联。第一条路径是一组沿着选定信息转发的  $1 \times 1$  卷积。 $1 \times 1$  卷积是一种选择性高速公路网络，它只是简单地向前传递信息而不进行变换。第二和第三路径是  $1 \times 1$  卷积，接着是一组  $3 \times 3$  和  $5 \times 5$  卷积以提供多个比例的特征检测。最后，最后的平行路径是一个  $3 \times 3$  池化层，接着是  $1 \times 1$  卷积。网络中的中间启动模块连接到 softmax 分类器，这会导致网络全球培训损失。这些分类器被认为有助于防止消失梯度。

### 5.2.3 ResNet

增加深度的另一种方法是使用跨层转发信息的体系结构。迄今为止赢得 ImageNet 2015 的最佳方法是残余网络 [4]。残余网络将一层的输出添加到更深层的两层输出中（如图 2 所示）。这被称为残留网络，因为转发的信息迫使网络学习残差函数作为特征提取的艺术。残余网络作者认为消失梯度可以通过广泛采用的归一化技术来解决，而深度网络的深度则受限于深度网络的训练复杂度，这可以通过残差函数来简化。

### 5.2.4 CLDNN

CLDNNs 是一种语音处理方法，它可以在原始时域波形上进行操作，而不是专家语音功能，如 log-mel cepstrums [16], [15]。该体系结构使用两个卷积层，随后是两个由长期短期内存（LSTM）单元组成的循环层。LSTM 是一种常见的经常性网络架构，由几个门控制，历史维持多久 [5]。CLDNN 也可以具有绕过层的连接，旨在为提取的功能提供更长的时间上下文。例如，原始 CLDNN 在 LSTM 层之前转发具有卷积层输出的原始样本 [15]。

受使用专业知识指导网络体系结构（如卷积网络和 CLDNN）的启发，我们尝试使用卷积网络，我们将其称为卷积匹配滤波器。相当简单的想法是采用典型通信接收器的通用架构，并构建具有类似部分的神经网络架构。通信接收机有一个滤波器（通常是采样器），通常前置滤波器抽取每个符号的少量采样，用于执行相移以找到最佳采样点的同步器，然后采样器切片成比特或发射音频以进行模拟调制。与此类似的神经网络体系结构是一个卷积层，后面跟着一个 LSTM。

## 5.3 结果及分析

### 5.3.1 神经网络训练

网络的超参数（如学习速率，每层过滤器/特征映射的数量，过滤器大小以及某种程度上的层数）都会影响网络规模并且难以优化。最近的研究尝试将超参数优化为可以用反向传播和梯度下降（如网络权重和偏差）进行训练的常规参数。对于这项研究，我们忽略训练超参数，并使用 adam 优化器 [6]，它提供梯度归一化和动量，降低超参数的重要性，如学习率。

在工作的指导下，深度比特特征图的数量更重要 [2]，我们将建立一个类似于无线电卷积调制网络 [13] 中使用的基线卷积网络。我们的第一步是调整每个滤波器的滤波器数量和抽头数量，并将其视为不重要的超参数，用于余下的实验以测试不同架构对 RF 数据的适用性。

### 5.3.2 网络参数

网络的超参数（如学习速率，每层过滤器/特征映射的数量，过滤器的大小以及层的数量）都会影响网络规模，难以优化。最近的研究尝试将超参数优化为可以用反向传播和梯度下降（如网络权重和偏差）进行训练的常规参数。对于这项研究，我们忽略训练超参数，并使用 adam 优化器 [6]，它提供了梯度归一化和动量，降低了像学习率这样的超参数的重要性。在工作的指导下，深度比特特征图的数量更重要 [2]，我们将建立一个类似于无线电卷积调制网络 [13] 中使用的基线卷积网络。我们的第一步是调整每个滤波器的滤波器数量和抽头数量，并将其视为不重要的超参数，用于剩余的实验，以测试不同架构对 RF 数据的适用性。

我们使用 RadioML2016.10a 数据集 [12] 作为评估调制识别任务的基础。目标是使用 128 个样本的复数（基带 I / Q）时域矢量来识别 11 种可能类别中的调制方案。128 个样本以 2x128 向量馈入网络，其中复数时间样本的实部和虚部被分开。该数据集使用功率延迟分布，频率选择性衰落，本地振荡器偏移和加性高斯白噪声以及这些效应的细节 [12]。数据集标有调制类型和 SNR 地面实况。我们使用全 SNR 的 top-1 分类精度作为单个数字基准，并且比较技术显示 SNR 高于 1 的精度。所有模型和培训均使用 Kera 深度学习库，使用 Nano GTX 1070 GPU 使用 theano 后端。我们从一个类似于 CNN2 网络的网络开始 [13]。这是选择的基线，因为 [13] 的结果显示专家方法的显着改进；任何进一步的改进都应该被认为是最先进的。主要区别在于我们将在每个图层上使用尺寸为 1xtaps 的 nfilt 滤镜。我们将做一个简单的超参数优化

- a. 为 RF 调制识别找到最佳滤波器数量和滤波器大小

## b. 从网络深度和过滤器大小的其他领域获得测试假设

### 5.3.3 基准卷积网络

基线卷积网络在 softmax 分类器之前有两个卷积层和一个致密层。每个隐藏层具有整流线性单元 (ReLU) 激活功能和 50% 的丢失。第一个超参数优化是卷积层的大小。每个图层都会有  $1 \times 3$  的过滤器，我们将改变过滤器的数量以找出需要的数量。从 [1], [18], [2] 我们预计在过度配合发生之前，过滤器数量的较大范围会带来类似的性能。

正如预期的那样，每层大约有 30 到 70 个过滤器，而且性能非常相似，这个窗口相当大。图 10 显示了 10-滤波器增量的 20-90 滤波器的前 1 类分类精度。对于其余的实验，我们将使用每层 50 个滤波器。接下来，我们优化每个过滤器的大小。[2] 表明，滤波器的大小也有最小的影响，但基于无线电领域和数据集的专业知识，我们预期 8 抽头滤波器是最优的。对于这个实验，我们使用一个具有单个隐藏密集层的双卷积层网络，接着是 softmax 分类器。卷积层每个都有 50 个滤波器，滤波器大小为  $1 \times \text{ntaps}$ ，其中  $\text{ntaps}$  从 3 到 12 变化。。

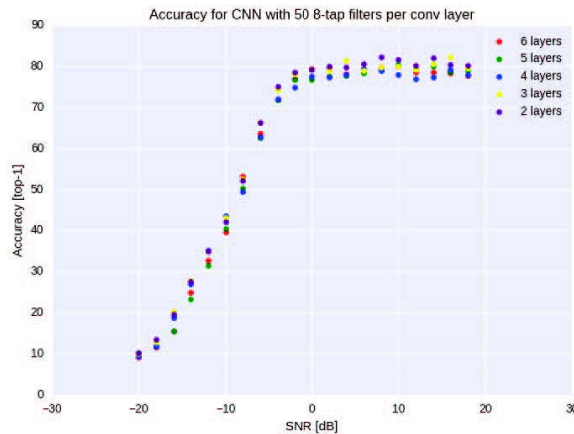


图 5.1 改变每层滤波器的数量具有较小的影响，在更高的 SNR 下更明显。每个网络在 2 个卷积层网络中都有  $1 \times 3$  个过滤器，并有 1 个致密层和一个 softmax 分类器。

每个卷积层的滤波器尺寸变化的结果表明，较小的滤波器不如较大的滤波器。我们根据数据集的专家知识假设 8 抽头滤波器是最好的。图 4 中每个 SNR 图的结果很难区分清楚的赢家；然而，整个数据集的分类准确性表明，7-12 次水龙头都具有相似的性能约 61%，差异在统计上不显著。

最后，对于卷积网络，我们尝试增加网络深度。对于这个实验，我们使用带  $1 \times 8$  滤波器的 50 抽头卷积层。卷积层之后，我们使用一个隐藏的密集层，然后使用最后一个密集的 softmax 分类器。我们从一个 2 卷积层网络开始，并添加卷积层。深度学习的趋势表明，增加更多图层应该会改善分类性能，直到渐变变得不稳定。

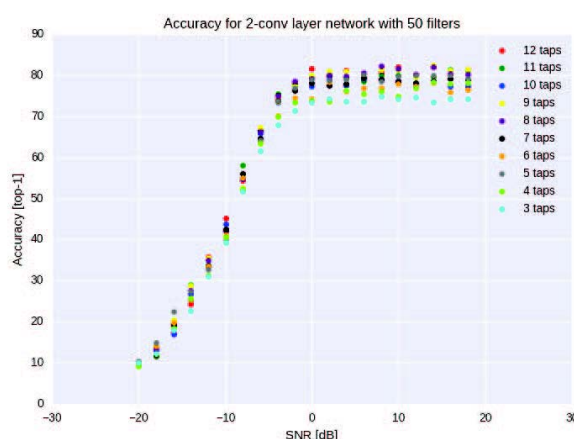


图 5.2 改变 DNN 中卷积层的数量不会改进无线电调制识别。

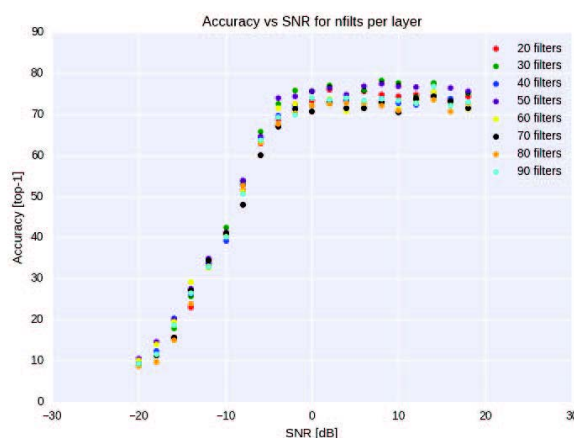


图 5.3 改变两个卷积层中的抽头数量（滤波器大小），较低的水龙头数量明显较差，但性能集群随水龙头数量的增加而增加。

改变卷积层的数量在分类准确性方面几乎没有改善。图 5 显示了该任务的信噪比精度。这表明我们的网络没有更多的特征深度可以学习。由于调制数据通常只改变复数正弦曲线的幅度，频率或相位，因此数据的起始层次不高；然而，令人惊讶的是，添加更多的卷积层似乎并不能在较低的信噪比下帮助降低噪声的影响。

### 5.3.4 残留网络

尽管添加更多卷积层不会提高分类准确性并不奇怪，但令人惊讶的是，分类和丢失只要 2 或 3 卷积层即可达到平稳。最初的研究结果表明，较深的网络会导致较高的训练损失，这表明较高的训练难度而不是过度训练。图 6 显示，我们的超参数优化 CNN 和 9 层残留网络达到了相似的损失，验证损失和未显示的准确性；然而，剩余网络在更少的时期学习。我们还试验了 5-9 层的残留网络，它们都具有相似的性能和训练时间。这与我们针对普通 CNN 深度的超参数搜索相结合，表明我们不受无线电学

习任务网络深度的限制，尽管我们受限于纯粹 CNN 架构可以学习的功能。

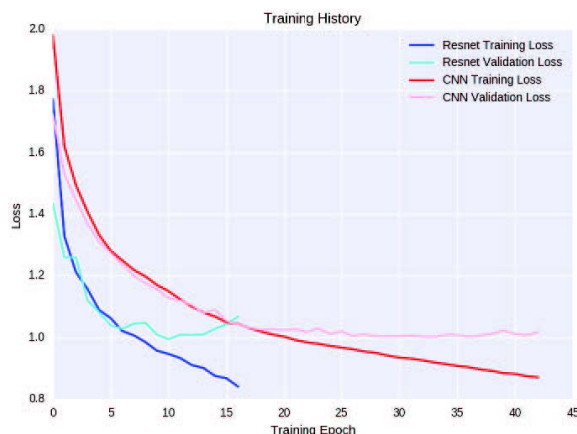


图 5.4 显示超参数优化的 CNN 和 9 层残留网络的训练损失和验证损失的训练历史记录

### 5.3.5 InceptionModel

先启模块在我们的实验中也没有改进无线电调制分类，使用为我们的数据集调整的初始模块。每个模块使用的三个分支是 50 个  $1 \times 1$  滤波器，50 个  $1 \times 3$  滤波器和 50 个  $1 \times 8$  滤波器。 $1 \times 3$  和  $1 \times 5$  滤波器分支前面也有 50 个  $1 \times 1$  滤波器，如图 7 所示。网络中 1-4 个启动模块的结果并没有显示出我们的超参数优化 CNN 的改进。同样，这表明我们不受深度的限制，也不受限于滤波器的规模。

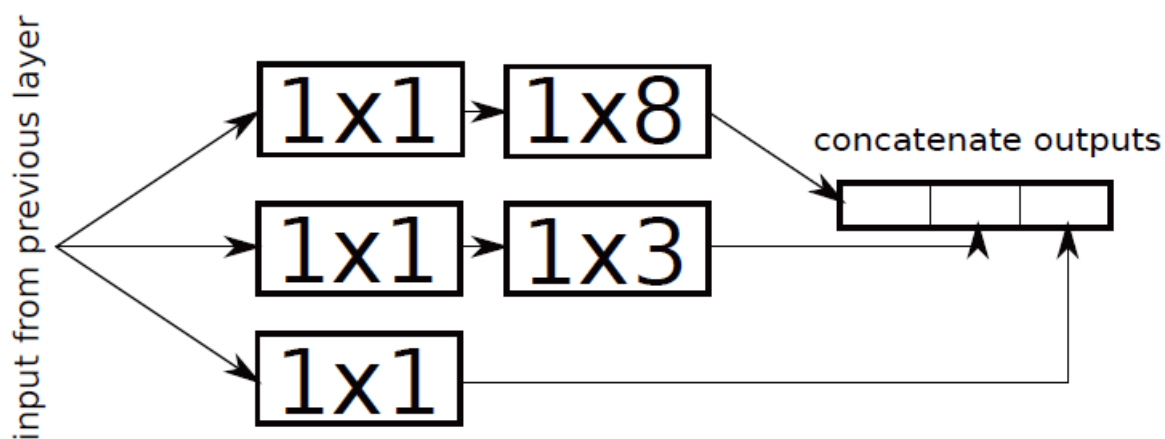


图 5.5 用于显示卷积滤波器尺寸的 RF 数据的初始模块。每个卷积层有 50 个过滤器

### 5.3.6 CLDNN

作为我们测试的最终架构添加了经常性网络层，即那些由 LSTM 单元组成的层，用于对时间特征进行建模。这种方法广泛用于时间序列应用，我们预计调制基带时间序列可能同样适用。我们测试了两层和三层卷积，然后在 CLDNN 类型体系结构中重



复出现层，在重复层之前有和没有前向/旁路连接。我们发现前向连接作为原始波形和卷积输出的连接，如图 8 所示，与其他体系结构相比，分类精度更高，梯度下降更稳定。使用会创建类似前面描述的卷积匹配滤波器检测器的体系结构的合并层不利于分类。

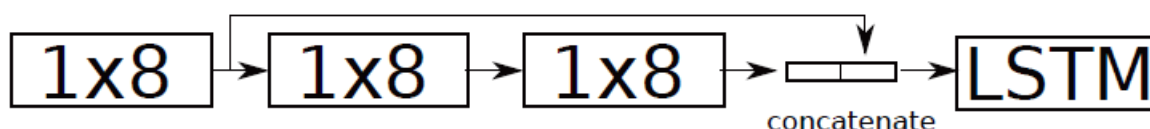


图 5.6 用于 RF 数据的 CLDNN 体系结构。在进入 LSTM 之前，第一个 1x8 卷积层的输出与三个 1x8 卷积层的输出级联。

为了进一步理解什么限制了分类精度，我们查看了图 10 中所示的 CLDNN 的混淆矩阵。有两个主要混淆领域。一个在模拟调制之间，另一个在高阶 QAM 之间。模拟调制将很难解决，但是 QAM 可以通过更好的同步和减少信道损伤来改善。

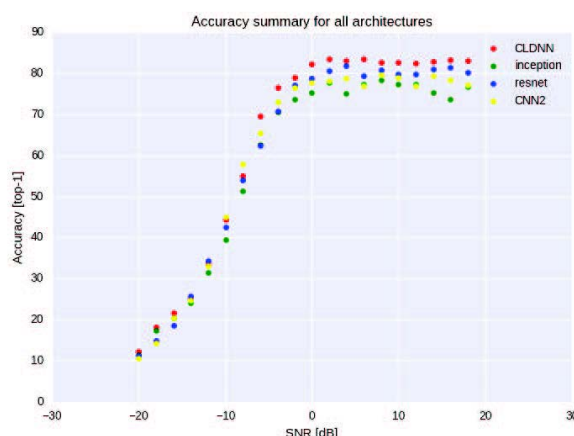


图 5.7 CLDNN 的信噪比始终优于其他网络架构，信噪比高于 -8dB

为了进一步理解什么限制了分类的准确性，我们看一下图 8 所示的 CLDNN 的混淆矩阵。有两个主要的混淆领域。一个在模拟调制之间，另一个在高阶 QAM 之间。模拟调制将很难解决，但是 QAM 可以在更好的同步和减少信道损伤的情况下得到改善。

对 CLDNN 在每一层学习的内容有直觉，对指导未来的工作很重要。为此，我们绘制了一些滤波器抽头的时间和频率表示。对于频率响应，滤波器抽头用零填充 100 个零点以获得 128 点 FFT。图 11a 和 12a 示出了来自第一层的两个选择滤波器。专家的眼睛看起来并不特别熟悉时域表示；然而频率响应确实显示出成形的低通滤波器。未示出的其他滤波器具有频率选择性组件，DC 阻断器和类似 sinc 的频谱形状。

另一种可视化这些滤波器的方法是将随机数据应用于它们，并对特定滤波器的

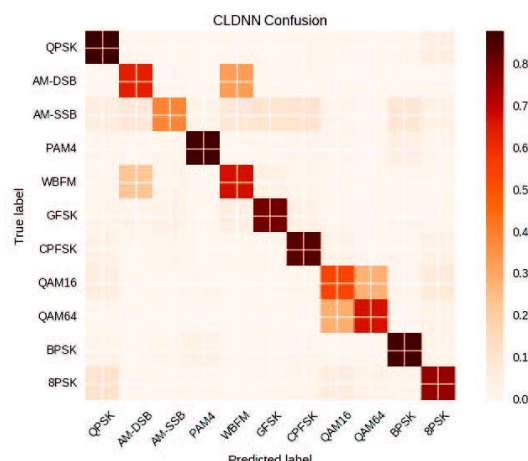


图 5.8 CLDNN 的全 SNR 混淆矩阵显示了模拟调制和高阶 QAM 之间单独混淆之间最混乱的问题。

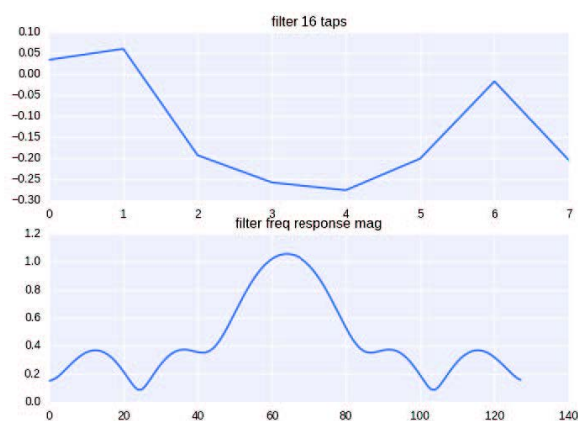


图 5.9 我们训练的 CLDNN 的第一卷积层中滤波器的时间和频率幅度表示。。

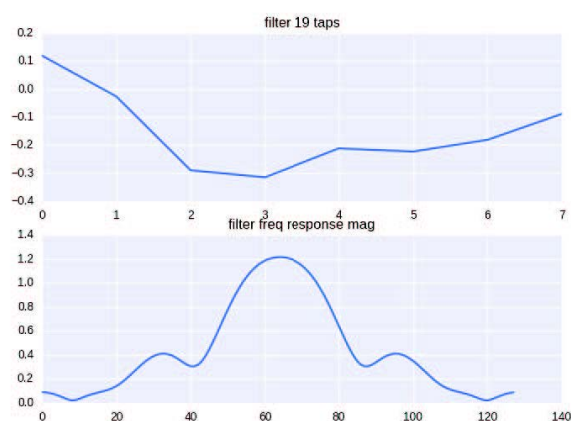


图 5.10 我们训练的 CLDNN 的第一卷积层中滤波器的时间和频率幅度表示。。

输出执行梯度上升，该滤波器将收敛到最能激活卷积神经元的数据上 [11]。选定的两个过滤器的结果如图 2 所示。11b 和 12b。所得到的载体看起来有点像粗 PSK 和 FM

/FSK 调制到专家眼睛。由于模拟的信道模型，矢量还显示出存在于我们的数据集中的一些恒定的相位旋转。需要注意的是，选择这两种过滤器可视化并非所有过滤器都对专家有意义。

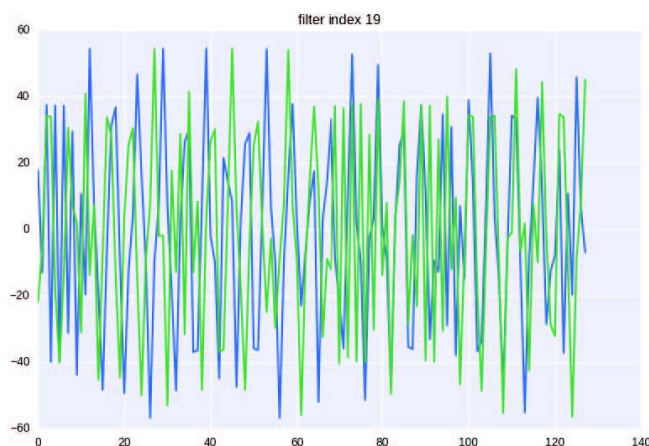


图 5.11 随机数据训练最大程度地激活过滤器，看起来像 BPSK。

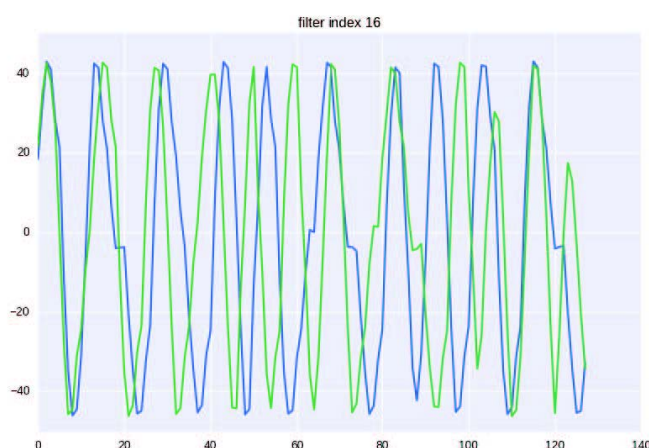


图 5.12 训练最大程度激活滤波器的随机数据，这看起来像 FM 或 FSK 调制。

## 5.4 本章小结

深度神经网络在无线电领域的性能似乎不受网络深度的限制。虽然我们的实验将调制识别作为基准任务，但我们期望其他无线机器学习任务能够使用类似的网络架构。无线电任务深度学习的进一步发展可能来自改进的训练方法和网络架构，这些架构可以学习转换射频数据以消除无线信道的影响，而这些神经网络架构并非为此设计的。目前正在探索的一个例子是使用空间变换来均衡和同步输入波形 [14]。

这些实验还着重于名义上带宽归一化的数据集，这是对从真实无线电传输中捕获的信号的不良假设。将来在实际应用中使用的网络需要学习对信号进行重新采样

以获得带宽规格化，或者学习许多带宽的特性。可重新采样，同步和消除非线性信道失真的网络都是该领域未来令人兴奋的工作。我们相信，随着无线电环境变得越来越复杂，将调制，多调制协议的不同时间行为和多个无线电发射机组合在一个频带内进行互操作，这些深层网络中的层次结构的许多概念将越来越重要，使我们的网络可以有效地应对复杂性，正如在复杂的多物体场景中的视觉域中所显示的一样。

## 第六章 总结与展望

### 6.1 研究结论

通信信号调制识别是通信系统中信号解调、信息提取和信号检测的关键技术，因此这是一个很有研究价值的课题，近年来取得了很多成果，在电子对抗，软件无线电等领域都有了广泛的应用。在近一年的时间里，查阅了大量的文献，广泛学习了通信信号调制识别领域的相关研究成果。相对于其它调制识别算法，由于小波变换对瞬时信息具有良好的检测性能，并且在进行通信信号调制类型识别时不需要任何的先验知识，可以在中频上直接对信号进行识别，因此本文针对基于小波变换的调制识别技术进行了深入的研究。下面是本文的主要工作：

提出了一套基于小波变换的通信信号调制识别算法。能够在低信噪比条件下直接对所接收到的各种中频数字信号进行识别，与其它同类调制类型算法相比具有较高的识别性能，有一定的实用价值。对于类间识别，首先，针对传统的基于小波变换幅度信息的调制识别技术进行研究，分析其不足及在低信噪条件下识别率低的问题。然后，提出了一种利用信号包络方差和小波变换频率信息的类间识别算法：首先，利用信号包络的方差分类 MQAM 信号和 MFSK、MPSK 信号，然后采用最优尺度对信号进行小波变换，再提取小波变换相位信息，第二次采用小波变换从小波变换的相位信息中提取瞬时频率信息，根据所提取的频率信息统计方差特征值以区分 MFSK 信号和 MPSK 信号。

对于类内识别，针对传统小波变换算法利用相邻码元处的小波变换幅度信息进行 MPSK 信号识别和码元区间内小波变换幅度信息进行 MFSK 信号识别，在低信噪比条件下受噪声影响显著的问题，提出了利用小波变换后相邻码元对应采样点的相位差信息对 MPSK 信号进行类内识别，利用小波变换从小波变换相位信息中提取频率信息进行 MFSK 信号的类内识别，经过仿真验证，本文方法在相同仿真条件下，具有较高的识别性能，且稳定性好，实现了低信噪比条件下的调制类型识别。由于不同的 QAM 信号瞬时功率均值归一化后其分布特征具有很大差别，本文利用该特点，对小波变换后的瞬时功率进行均值归一化，以识别不同的 QAM 信号。

通信信号调制识别是一项不断发展的技术，新的、更复杂的信号调制方式及通信应用的出现，为其提出了越来越多的挑战，使得该技术领域不断有新的研究成果涌现，具有巨大的潜在研究价值。基于小波变换的数字信号调制识别算法的研究已经比较普遍，但是不论在理论方面还是在应用方面都还存在一些值得研究的问题。由于作者在这方面的研究时间较短，因此还有许多相关的问题需要进一步更深入的研究：

在调制识别中，如何充分利用小波变换方法和其它支撑矢量机或神经网络方法的优点，将它们有机结合，进一步提高在各种信噪比下的正确识别率。

## 6.2 未来展望

本文的算法只是在高斯白噪声条件下进行的，如何在瑞利、多径信道条件下实现调制信号的识别，仍然值得继续深入研究。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，进一步研究与其它新的调制信号的分类问题等。希望今后能把通信信号调制识别技术深入的研究下去，做出自己的贡献。

## 参考文献





## 致谢

时光荏苒，日月如梭，转眼间将近三年的研究生时光即将结束，在此，我要向在研究生学习期间帮助过我的各位老师、同学好友及家人表示诚挚的谢意和衷心的祝福！

首先，我要衷心地感谢我的导师杨清海教授。杨老师为我们提供了良好的学习和科研环境，给予我悉心的指导和耐心的帮助。在我读研期间，受到了杨老师的方方面面的教导和影响。学习工作中，杨老师以他丰富的经验和学识，指导我不断的完善和充实自己，并引导我在通信的前沿领域中求知和探索。在他的指引下，我参与了一些项目的研究和实现，培养了我分析问题与解决问题的能力，在其过程中真正学会了如何将所学转化为所用。另外，杨老师严谨的治学态度，一丝不苟的治学精神，坚韧的毅力，只争朝夕的勇气，锐意进取创业精神，严于律己的作风，都深深的感动和影响了我，是我学习的目标和榜样。

我还要感谢沈中和刘明骞老师，在我的研究生学习中，他们给了我悉心的指导和无私的帮助，在此特表示深深的敬意和衷心的感谢！真诚感谢在我论文撰写中给我提供了许多帮助的秦剑、张劲、李万、唐靖旋、刘祥，每次与他们的交流都能给我不同的启发，同时他们的友谊也在支持和帮助着我。

我要特别感谢辛勤养育我多年的父母。在我成长过程中的每一步，都离不开他们的关心与支持。

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1月于西安电子科技大学

