

单位代码: 10293 密 级:           

南京邮电大学

# 专业学位硕士学位论文



论文题目: ReLU 激活函数下卷积神经网络的不同类型噪声增益研究

学 号 1220086609

姓 名 刘赞

导 师 王友国

专业学位类别 应用统计硕士

类 型 全日制

专业（领域） 应用统计

论文提交日期 二〇二三年六月

# **Study on different types of noise benefit of convolutional neural networks under ReLU activation function**

Thesis Submitted to Nanjing University of Posts and  
Telecommunications for the Degree of  
Master of Applied Statistics



By

Yun Liu

Supervisor: Prof. Youguo Wang

June 2023

# 摘要

随着计算机科技与人工神经网络理论的快速发展，卷积神经网络在图像分类、目标检测等领域得到了广泛的应用。卷积神经网络作为人工神经网络的代表性模型是非线性系统的一种，随机共振理论指出在非线性系统中加入合适的噪声可以提高系统的性能，因此随机共振辅助优化卷积神经网络的性能，是值得深入研究的问题。目前的卷积神经网络随机共振研究主要聚焦于传统的卷积网络模型，本文比较了不同激活函数对卷积神经网络的影响并研究了不同噪声下卷积神经网络中随机共振现象。论文主要研究内容如下：

(1) 比较了不同激活函数对卷积神经网络的性能影响。首先描述了 Sigmoid、Tanh 和 ReLU 激活函数的性质，并研究了对卷积神经网络训练的影响。然后在 MNIST 与 Fashion-MNIST 两个图像数据集上进行仿真。实验结果表明使用 ReLU 函数的模型在交叉熵和分类准确率上具有更好的表现，在提升模型的泛化性能方面具有优势。

(2) 提出了一种改进的噪声增强卷积神经网络 (NCNN) 算法。在 ReLU 函数作为激活函数下，通过向输出层加入均匀噪声，利用小批量随机梯度下降算法更新模型参数，研究改进 NCNN 算法的噪声增益。最后在 MNIST 和 Fashion-MNIST 数据集上进行仿真，实验结果表明在训练不充分情况下所提模型的交叉熵与分类错误率比无噪声模型更低，存在噪声增益。

(3) 将算法中加入的均匀噪声拓展到高斯噪声、拉普拉斯噪声和柯西噪声，研究不同噪声类型对卷积神经网络噪声增益的影响。实验结果表明在训练不充分的情况下，三种非均匀噪声的 NCNN 模型均存在噪声增益，在初始噪声强度相同的条件下柯西噪声的噪声增益低于另外三种噪声的噪声增益。

**关键词：**卷积神经网络，激活函数，交叉熵，图像分类，随机共振

## Abstract

With the rapid development of computer technology and artificial neural network theory, convolutional neural networks have been widely used in image classification, object detection and other fields. As a representative model of artificial neural network, convolutional neural network is a kind of nonlinear system, and stochastic resonance theory points out that adding appropriate noise in nonlinear system can improve the performance of the system, so stochastic resonance-assisted optimization of the performance of convolutional neural network is a problem worthy of in-depth study. The current stochastic resonance research of convolutional neural networks mainly focuses on traditional convolutional network models, and this thesis compares the influence of different activation functions on convolutional neural networks and studies the random resonance phenomenon in convolutional neural networks under different noises. The main research contents of the thesis are as follows:

(1) The performance effects of different activation functions on convolutional neural networks are compared. Firstly, the properties of the Sigmoid, Tanh, and ReLU activation functions are described, and the impact on convolutional neural network training is studied. Simulations are then performed on two experimental datasets, MNIST and Fashion-MNIST. Experimental results show that the model using the ReLU function has better performance in cross-entropy and classification accuracy, and has advantages in improving the generalization performance of the model. The different activation functions in the convolutional neural network are compared and studied, and the principle performance of the activation functions of Sigmoid, Tanh and ReLU and their influence on the training of convolutional neural network are analyzed, and then two experimental datasets of MNIST and Fashion-MNIST are introduced, and finally the simulation is carried out on the two datasets, and the experimental results show that compared with the other two traditional activation functions, the model using the ReLU function has better performance in cross-entropy and classification accuracy at the same time, it is found that the ReLU function has advantages in improving the generalization performance and average training time of the model, which proves that using different activation functions also has different effects on neural networks.

(2) An improved noise-enhanced convolutional neural network (NCNN) algorithm is proposed. With the ReLU function as the activation function, the noise benefit of the NCNN algorithm is improved by adding uniform noise to the output layer and updating the model parameters by using

the minibatch stochastic gradient descent algorithm. Finally, simulations are carried out on MNIST and Fashion-MNIST datasets, and the experimental results show that the cross-entropy and classification error rate of the proposed model is lower than that of the noise-free model under insufficient training, and there is noise benefit.

(3) The uniform noise added to the algorithm is extended to Gaussian noise, Laplace noise and Cauchy noise, and the influence of different noise types on the noise benefit of convolutional neural networks is studied. The experimental results show that under the condition of insufficient training, the noise benefit of the NCNN models of the three non-uniform noise is present, and the noise benefit of Cauchy noise is lower than that of the other three types of noise under the same initial noise intensity.

**Key words:** Convolutional neural networks, Activation function, Cross entropy, Image classification, Stochastic resonance

## 目 录

第一章 绪论.....	1
1.1 选题的背景和意义.....	1
1.2 国内外研究现状.....	1
1.2.1 卷积神经网络研究现状.....	2
1.2.2 随机共振研究现状.....	3
1.3 研究内容与结构框架.....	4
1.3.1 研究内容.....	4
1.3.2 结构框架.....	5
第二章 相关背景知识介绍 .....	6
2.1 卷积神经网络基本理论.....	6
2.1.1 卷积神经网络基本组成.....	7
2.1.2 反向传播算法.....	8
2.2 随机共振基本理论.....	10
2.2.1 常用测度.....	10
2.2.2 双稳态 Langevin 模型.....	10
2.3 本章小结.....	12
第三章 卷积神经网络中的激活函数比较研究 .....	13
3.1 激活函数对神经网络的影响 .....	13
3.2 常用的激活函数.....	14
3.2.1 Sigmoid 函数 .....	14
3.2.2 Tanh 函数 .....	14
3.2.3 ReLU 函数 .....	15
3.3 实验数据集介绍.....	16
3.3.1 MNIST 数据集 .....	16
3.3.2 Fashion-MNIST 数据集 .....	17
3.4 实验结果与分析.....	18
3.4.1 实验设计.....	18
3.4.2 MNIST 数据集上的实验结果与分析 .....	18
3.4.3 Fashion-MNIST 数据集上的实验结果与分析 .....	21
3.5 本章小结.....	23
第四章 ReLU 激活函数下的卷积神经网络随机共振研究 .....	25
4.1 改进的噪声增强卷积神经网络算法 .....	25
4.2 均匀噪声.....	27
4.3 实验结果分析.....	28

4.3.1 MNIST 数据集上的实验结果与分析 .....	28
4.3.2 Fashion-MNIST 数据集上的实验结果与分析 .....	30
4.4 本章小结 .....	32
第五章 不同类型噪声下的卷积神经网络随机共振研究 .....	33
5.1 常见的噪声类型 .....	33
5.1.1 高斯噪声 .....	33
5.1.2 拉普拉斯噪声 .....	34
5.1.3 柯西噪声 .....	35
5.2 实验设计 .....	35
5.3 MNIST 数据集上的实验结果与分析 .....	36
5.3.1 高斯噪声 .....	36
5.3.2 拉普拉斯噪声 .....	37
5.3.3 柯西噪声 .....	38
5.3.4 不同噪声的增益比较 .....	40
5.4 Fashion-MNIST 数据集上的实验结果与分析 .....	40
5.4.1 高斯噪声 .....	40
5.4.2 拉普拉斯噪声 .....	42
5.4.3 柯西噪声 .....	43
5.4.4 不同噪声的增益比较 .....	44
5.5 本章小结 .....	45
第六章 总结与展望 .....	46
6.1 总结 .....	46
6.2 展望 .....	47
参考文献 .....	48
附录 1 程序清单 .....	51

# 第一章 绪论

## 1.1 选题的背景和意义

随着信息技术的飞速进步，人类社会进入了大数据时代。大数据时代给人们带来了极大的生活便利，提高了人们的生活水平，让人们可以在互联网上频繁进行多种信息交互，从而在网络上产生大量的视频、图像和音频等数据。其中图像数据作为日常生活中不可或缺的信息来源，成为了研究人员的关注焦点，如何简单高效地处理图像数据也是目前技术上面临的挑战<sup>[1]</sup>。在处理图像数据方面，卷积神经网络作为人工神经网络的一个重要分支，在图像分类<sup>[2]</sup>，目标检测<sup>[3]</sup>，图像识别<sup>[4]</sup>，图像分割<sup>[5]</sup>等领域得到了相当广泛的应用。目前在图像分类问题中存在的难点主要有图像比例差异、拍摄角度差异、图像缺失以及噪声干扰等问题<sup>[6]</sup>。大量的干扰因素也导致图像分类越发困难，而卷积神经网络使用卷积核对数据进行局部特征提取，对存在噪声干扰的图像数据具有一定的鲁棒性，可以提高图像特征提取的效果，提高分类效果<sup>[7]</sup>。经典的卷积神经网络模型往往通过提高卷积核的数量或者提高网络模型层数来达到更好的性能效果，但是这种方法同时也会提高模型的复杂度，不仅会引起参数冗余和过拟合等问题，模型对硬件设施水平的要求也会随之提高，这就会导致模型不利于实际应用。

在自然环境中噪声是广泛存在的，它们通常与有效的信号相结合，因此在传统的通信领域中噪声往往被认为对系统有负面的影响，需要尽可能的抑制系统中的噪声来提高系统的性能。但在信号处理系统对噪声进行抑制的同时，有效信号的信号特征也会随着噪声被抑制而受到削弱，仅仅抑制噪声无法有效提高系统中信号传输的效率。之后随着非线性科学的发展，人们发现在非线性系统中提高噪声的强度，输出信噪比呈现不减反增的趋势，在这种情况下就存在最佳噪声强度，可以使系统产生最大输出信噪比，这一现象被称为“随机共振”<sup>[8]</sup>。此后随机共振理论常常应用于双稳态系统、阈值神经网络等非线性系统的相关研究，而卷积神经网络模型中使用的激活函数为非线性函数，因此卷积神经网络也是非线性系统的一种。将噪声加入到卷积神经网络的训练过程中并研究其中的随机共振现象，对于卷积神经网络的研究具有重要意义，也为进一步优化其在图像分类等领域的实际应用提供了新的研究方向。

## 1.2 国内外研究现状

本文结合随机共振理论对卷积神经网络进行研究，研究内容主要包括卷积神经网络以及



随机共振的理论与应用，下面分别介绍卷积神经网络与随机共振的研究现状。

### 1.2.1 卷积神经网络研究现状

最早期的卷积神经网络是 LeCun 等人在 1998 年提出的 LeNet-5 模型，从此确立了卷积神经网络模型的基本结构。作为最早出现的卷积神经网络，LeNet-5 起初仅包含输入层、卷积层、池化层与全连接层，该模型最初被用于手写数字分类，在对手写数字集进行训练之后，LeNet-5 在手写数字集上得到了很好的分类效果，进而推动了卷积神经网络在图像分类领域中的应用<sup>[9]</sup>。然而，受制于当时的科技水平，在计算机硬件设施落后的条件下 LeNet-5 处理过于复杂的图像数据时表现不佳，而且用于训练的样本图像数据也较少，因此 LeNet-5 很难在复杂的图像分类问题中取得较好的效果。

随着信息技术的进步，计算机的硬件水平有了显著提高，用于训练的样本数据集也有大量的增长，在这个时代背景下，CNN 的发展条件得到了充分的满足，因此 CNN 在短短几年时间里就取得了快速的发展，成为了图像分类领域的重点。2012 年，Krizhevsky<sup>[10]</sup>等人提出了 Alex Net，该模型采用 ReLU<sup>[11]</sup>激活函数取代了 Sigmoid 和 Tanh 函数，利用激活函数的简单性降低了模型的计算复杂度，同时提出了 Max Pooling、Dropout<sup>[12]</sup>等创新方法来缓和模型的过拟合问题。Alex Net 结构与 LeNet-5 类似，但其具有更深的网络深度和更好的分类效果。2014 年，Matthew<sup>[13]</sup>等人提出了 ZFNet，通过使用尺寸与步长更小的卷积核和提高网络深度的方法，进一步优化 Alex Net 的性能。2014 年，Simonyan 等人<sup>[14]</sup>提出了 VGG 模型。该模型在使用小尺寸卷积核的同时将网络深度提升到 19 层，证实了模型的深度能够影响卷积神经网络的分类性能。在 2014 年，谷歌公司提出了 GoogleNet<sup>[15]</sup>，首次提出了并行结构的 inception 模块，这一模块在没有增加网络计算量的同时提升了分类效果，进一步拓宽了卷积神经网络模型的优化方向，不再将研究方向限制在单纯的网络层数这一维度中。inception 模块通过使用多个小尺度卷积并行计算的方法进一步减少了模型参数的计算量。该模块通过多种尺寸的卷积运算进行多个尺度的特征提取，提升了模型对多尺度特征的适应能力和特征提取能力。2015 年，He<sup>[16]</sup>等人在 VGG 网络模型构建的基础上进行改进，将卷积网络的层数进一步提升到 152 层，提出了深度残差网络(Residual Net, ResNet)。改进后的网络模型能够综合不同深度下的特征信息，通过 residual 结构有效地克服深度网络中的网络退化问题，同时通过数据的预处理以及在网络中使用 BN (Batch Normalization) 来解决梯度消失问题。

### 1.2.2 随机共振研究现状

1981 年,为了解释远古时期地球的冰川期和暖期气候周期性交替现象,意大利科学家 Benzi 等人,将地球视为一个非线性系统,结合周期信号与随机干扰,建立了气候模型,首次提出了随机共振理论<sup>[8]</sup>。1983 年, Fauve 等人<sup>[17]</sup>使用信噪比为测度,在双稳态施密特触发电路系统中发现了信噪比随着噪声强度呈现先增后减的现象,这是第一次在物理实验中观测到随机共振的存在。在 1988 年, McNamara 等人<sup>[18]</sup>在双向环形激光器的研究中,通过加入噪声再次观察到了随机共振现象,此后随机共振就正式进入研究人员的视野。早期随机共振的研究普遍是在输入为周期性的弱信号条件下进行的,而对于非周期信号以及输入信号高于阈值的情况,通常认为在这类情况下加入噪声并不会产生随机共振现象。但在 1996 年, Collins<sup>[19]</sup>等人在研究神经网络模型时提出了非周期随机共振,该理论是将非周期信号作为输入信号,以信息理论中的平均互信息作为测度并研究其中产生的随机共振现象,从而将随机共振现象中的周期信号拓展到非周期信号的范畴。随后在 2000 年, Stocks<sup>[20]</sup>以互信息为测度,在多个阈值单元并联组成的阵列中,向每个阈值单元中加入独立同分布的高斯噪声,研究结果表明在这种情况下即使输入信号大于系统阈值,系统也会产生随机共振现象,由此提出了阈上随机共振理论,从而将随机共振理论进一步推广到阈上信号的情况。

此后,随机共振理论不断发展完善,并被广泛应用到生物医学<sup>[21]</sup>、语音信号处理<sup>[22]</sup>和微弱信号检测<sup>[23]</sup>等领域,而神经网络作为实际应用广泛的非线性系统,也有许多的学者开始研究其中的随机共振现象,并取得了大量的研究成果。2011 年, Osoba 等人以 Kullback 相对熵为测度,提出了期望最大化(Expectation-Maximization, EM)算法的噪声增益理论与噪声增益产生的一般充分条件,并通过高斯混合模型的实验结果证明了加入适当噪声可以提高 EM 算法的平均收敛速度<sup>[24]</sup>。2013 年, Osoba 等人证明了反向传播(Backpropagation, BP)算法是广义 EM 算法的一种特例,提出了 BP 算法存在噪声增益的一般充分条件,同时在 MNIST 数据集上进行实验,结果证明加入适当的噪声可以提高 BP 算法的收敛速度,为研究神经网络中的噪声增益提供了理论基础<sup>[25]</sup>。同年 Osoba 等人<sup>[26]</sup>研究发现在许多基于质心的聚类算法中,噪声可以加速算法收敛。同时在高斯混合模型上的仿真结果表明,噪声也可以提高随机无监督竞争学习、监督竞争学习和差分竞争学习算法的收敛速度。2014 年 Mitaim 等人<sup>[27]</sup>在研究 EEG 数据集的分类问题时发现将原始数据集和添加噪声的数据集串联起来训练可以提高 ESVM 的分类精度,且模型分类精度的提高量取决于噪声强度、添加噪声的数据集的数量、ESVM 集群的数量以及 ESVM 每个集群中的数据数量。2016 年, Osoba 等人<sup>[28]</sup>在高斯混合模型上进行仿真实验,将 EM 算法中的噪声增益由加性噪声推广到乘性噪声的情况。同年, Kosko

等人<sup>[29]</sup>进一步研究了卷积神经网络中的噪声增益，并提出了噪声卷积神经网络(Noise-enhanced CNN, NCNN)算法，在 MNIST 数据集上的实验结果表明，与无噪声 BP 算法训练的模型相比，NCNN 算法在前几次迭代中具有更低的交叉熵和分类错误率。在 2019 年，Brandon 等人<sup>[30]</sup>发现满足一定条件的噪声可以提高不可约马尔可夫链以及模拟退火算法的平均收敛速度，仿真实验结果表明加入合适的噪声可以提高模拟搜索深度局部最小值的广度。Adigun 等人<sup>[31]</sup>则研究了双向反向传播训练神经网络过程中的噪声增益，在 MNIST 手写数字测试集和 CIFAR-10 测试图像集上的仿真结果表明，加入合适的噪声可以提高神经网络的收敛速度和分类的准确性。同年 Kosko 等人<sup>[32]</sup>在之前的研究基础上证明了加入合适的噪声也可以加快双向反向传播算法(bidirectional backpropagation, B-BP)的收敛速度，提高模型分类和回归的准确性。2020 年，kosko 等人<sup>[33]</sup>研究了 BAM 网络以及有限 Boltzmann 机中存在的噪声增益，实验结果表明在 Sigmoid 神经元中加入合适的噪声可以加快反向传播算法在多层神经网络训练和预训练中的平均收敛速度。此外，这种噪声增益也适用于正则化回归网络。Duan 等人<sup>[34]</sup>研究了离散 Hopfield 神经网络传输二元调幅信号的随机共振效应，发现加入适当噪声可以减少模型的错误率，而且当噪声强度一定时，可以通过增加神经元数量来进一步降低模型的错误率。2021 年，Duan 等人<sup>[35]</sup>研究了阈值神经网络中的随机共振现象，发现加入合适的噪声可以提高阈值神经网络在模式分类问题和手写数字识别上的分类精度，根据仿真实验结果，验证了可以通过人为地引入随机噪声来训练具有不可微激活函数的神经网络，扩展了反向传播算法的适用性。

从上述研究可以发现，神经网络中的随机共振研究已经取得了一定的成果。但卷积神经网络的噪声增益研究还处于起步阶段，其中大多数研究都集中在传统卷积网络模型上，还存在许多问题值得进一步讨论，例如模型复杂度，噪声强度，噪声类型等多种因素对噪声增益的影响。

## 1.3 研究内容与结构框架

### 1.3.1 研究内容

本文主要研究卷积神经网络处理图像分类问题时可能存在的随机共振现象，具体创新点如下：

(1) 分析了三种常见的激活函数，并对以上的激活函数进行对比实验，研究比较不同激活函数对卷积神经网络的影响，发现在实验中 ReLU 函数的训练效果最优。

(2) 选用 ReLU 激活函数提出了一种改进的 NCNN 算法, 并在两个图像数据集上进行仿真, 发现改进的 NCNN 算法在训练过程中存在随机共振现象, 验证了改进算法的有效性。

(3) 在改进的 NCNN 算法的基础上研究高斯噪声、拉普拉斯噪声和柯西噪声对分类性能的影响, 发现以上三种噪声对分类性能均有不同程度的提高。

### 1.3.2 结构框架

本文的结构框架主要包括以下章节:

第一章是绪论, 首先阐述了选题的背景和意义, 然后分别叙述了卷积神经网络的国内外发展现状和随机共振理论的研究历程, 最后着重介绍了神经网络与随机共振研究相结合的研究领域中的研究进展。

第二章介绍了与本文研究内容相关的基础理论知识, 首先论述了卷积神经网络的基本特点与结构组成, 同时给出了反向传播算法的计算过程, 最后介绍了随机共振理论的经典模型: 双稳态 Langevin 模型。

第三章是卷积神经网络中的多种激活函数比较研究, 本章介绍了三种激活函数的原理性能及其对卷积神经网络训练的影响, 然后介绍了 MNIST 与 Fashion-MNIST 两个图像数据集, 并基于 Keras 框架在两个数据集上进行实验, 最后依据实验结果分析不同激活函数对卷积神经网络的影响。

第四章是基于卷积神经网络的图像分类中的噪声增益研究, 本章在已有的卷积神经网络的噪声增益研究基础上, 选用 ReLU 激活函数提出了一种改进的 NCNN 算法, 最后在两个数据集上进行实验, 研究改进后算法的噪声增益以及模型训练过程中的随机共振现象。

第五章是不同加性噪声下的卷积神经网络噪声增益研究, 本章在改进的 NCNN 算法基础上, 选择使用三种常见的噪声: 高斯噪声、拉普拉斯噪声和柯西噪声替换原本的均匀噪声, 以无噪声条件与均匀噪声条件为基准进行比较与分析, 研究了不同类型加性噪声对卷积神经网络噪声增益的影响。

第六章是总结与展望, 本章对本文的研究目的和全部研究内容进行了总结, 同时指出了在本文研究中仍未考虑到的部分方面, 提出了需要进一步完善的内容与发展方向。

## 第二章 相关背景知识介绍

### 2.1 卷积神经网络基本理论

卷积神经网络是从多层感知机模型<sup>[36]</sup>发展而来，最早由 Yann Lecun<sup>[9]</sup>受到猫的视觉皮层研究启发提出了卷积神经网络的雏形 LeNet-5。2012 年 AlexNet 取得 ImageNet 比赛的分类任务的冠军，使得卷积神经网络真正爆发。卷积神经网络是一种具有深度结构的前馈神经网络，是深度学习的代表算法之一，在图像处理和自然语言处理等领域都有广泛的应用。经典的卷积神经网络模型一般由输入层、卷积层、池化层、全连接层、激活函数和输出层构成，由于卷积层的性质使得该模型具有两个很重要的特点：局部连接和权值共享<sup>[37]</sup>。

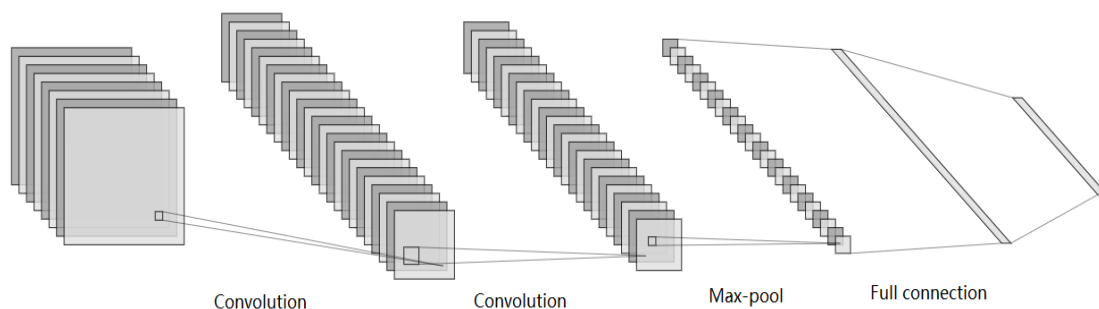


图 2.1 卷积神经网络模型结构

根据感受野理论<sup>[38]</sup>，人眼对外部环境的观测是从局部到全局的，而图像内部的空间联系也是局部间的像素相关性更高，与距离较远的像素相关性较低。因此每个神经元不需要对全局图像进行感知，只需要提取局部特征，然后在更高层将局部信息综合以实现全局特征提取。根据该理论，在 CNN 网络模型中神经元只需要提取局部区域的特征，即卷积层中的神经元与其输入层中的特征面进行局部连接，从而极大地减少了参数的计算量，图 2.1 是卷积神经网络模型结构的示意图。

此外，在卷积层中使用权值共享可以进一步控制参数的数量。每个卷积核与上一层局部连接，使每个卷积核的所有局部连接都使用同样的参数，通过权值共享就可以进一步减少网络的参数量，同时对于有重复特征的图像也可以做到有效的信息提取。

卷积神经网络具有局部连接和权值共享的特性，可以减少参数计算量使得网络易于优化，同时也降低了模型的复杂度以及过拟合问题发生的概率。当网络的输入数据维数很高时，卷积神经网络在参数计算量与模型复杂度等方面的优势就更加明显。

### 2.1.1 卷积神经网络基本组成

一般来说,卷积神经网络主要由卷积层、池化层、全连接层与输出层组成。在接收到输入图像数据之后,还需要对图像数据集进行数据标准化与数据增强<sup>[39]</sup>等预处理操作。数据预处理的目的是为了提升模型的训练效果,防止输入数据时存在图像尺寸不统一等情况,避免网络训练和模型收敛速度过慢的问题。

卷积层的功能是通过卷积运算提取输入数据的空间特征。卷积运算具体是指卷积核在输入图像上根据给定的步长移动,每次将卷积核中的元素与相对应的输入图像数据中的各个元素或像素值相乘再求和,最后输出结果,当输入图像数据的全部位置都与卷积核完成一次计算时,就完成了一次卷积运算。通常卷积核越小,卷积神经网络中的参数就越少。

若卷积层的输入数据  $X$  大小为  $w_x$ ,  $M$  个卷积核  $F_m$  大小为  $L_f$ , 步长为  $S$ , 无填充, 则输入  $X$  与卷积核  $F_m$  的卷积运算结果可表示为:

$$C_m = X \cdot F_m + B, \quad (2.1)$$

其中  $B$  为对应的偏置, 卷积层输出  $C_m$  的尺寸大小为  $\frac{(w_x - L_f)}{S} + 1$ 。

$C_m$  中第  $i$  行第  $j$  列对应的元素可以表示为:

$$C_m(i, j) = \sum_{a=1}^{L_r} \sum_{b=1}^{L_f} X(a-i, b-j) F_m(a, b) + B(i, j), \quad (2.2)$$

卷积运算得到的图像特征输出是线性的, 但实际应用中的样本通常呈非线性, 因此需要引入非线性函数来激活卷积层的输出。激活函数使得每个像素点的输出可以用激活函数值来代替, 更好的反映实际样本的特征。激活函数通常需要具有非线性、连续可微和单调性等特性。

Sigmoid 函数<sup>[40]</sup>、Tanh 函数<sup>[41]</sup>和 ReLU 函数<sup>[11]</sup>等函数均为常用的激活函数。其中 Sigmoid 函数的定义式为  $f(x) = \frac{1}{1+e^{-x}}$ , Tanh 函数的定义式为  $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , ReLU 函数的定义式为  $f(x) = \max(0, x)$ 。

sigmoid 函数单调易求导, 但输出不以 0 为中心, 且饱和时梯度非常小, 网络权值得不到有效的更新; Tanh 函数收敛速度较快但函数仍具有饱和性。当函数输入极大值或极小值时, 导数仍然趋近于 0, 因此 Tanh 函数在训练过程中仍然具有梯度消失问题; 而 ReLU 函数具有单侧抑制、稀疏激活性的特点, 即在  $x>0$  时, 梯度恒为 1,  $x<0$  时, 输出为 0, 因此在反向传

播过程中使用 ReLU 作为激活函数可以避免梯度消失问题<sup>[19]</sup>。

池化层的主要功能是进行下采样，对特征图稀疏处理，从而达到降维的目的<sup>[42]</sup>。池化层的计算过程也是使用一组带有权重的卷积核与卷积层输出的张量进行卷积运算。与卷积层不同，池化层卷积核的权重是固定的因此不需要训练。

池化层的运算方法有最大池化、均值池化等，其中最大池化是在设定卷积核大小以及步长后，对每个卷积核区域的特征值取最大值，而均值池化则是对每个卷积核区域的值求平均值。

最大池化方法对输入数据具有鲁棒性，当输入的数据存在轻微扰动时，池化运算返回的结果基本与无扰动时相同。同时池化层可以有效减少模型参数量和计算量，避免过拟合问题。

全连接层的主要作用是将经过卷积、池化操作之后得到的特征进行重新组合，从而实现特征的降维，特征经过全连接层后会被展开为特征向量。全连接层输出结果可表示为：

$$FC = C \cdot W_f + b_f, \quad (2.3)$$

其中  $FC$  为全连接层的输出， $W_f$  为全连接层权重系数矩阵， $C$  为池化层的输出， $b_f$  为对应偏置。

输出层位于全连接层之后，给出最后的分类预测结果，对于图像分类问题，输出层通常使用归一化指数（softmax）函数来输出分类标签。

当 softmax 函数用于分类时，会将卷积神经网络最终得到的输出结果映射到 (0,1) 区间，由于整个 softmax 层的输出相加结果为 1，因此输出层的输出就可以看作是输入对应原始数据中某个种类的概率。softmax 函数作为输出函数的输出结果可表示为：

$$O_k = \frac{\exp(FC \cdot U_k)}{\sum_{i=1}^n \exp(FC \cdot U_i)}, \quad (2.4)$$

其中  $O_k$  为第  $k$  个神经元的输出， $U_k$  为第  $k$  个神经元的权重矩阵， $n$  为神经元的总数。

### 2.1.2 反向传播算法

反向传播（back propagation, BP）算法是广泛应用于神经网络中的梯度计算方法，这一算法根据链式法则来计算梯度，实现损失函数的信息从后往前流动<sup>[43]</sup>。卷积神经网络的结构通常很复杂，在训练时往往需要更新大量的参数，因此需要一个快速高效的算法来提高模型的训练效率，而反向传播算法的优势就在于可以沿着网络结构从后往前依次计算模型全部参数的梯度，在每条路径上只计算一次就能得到顶点对全部下层节点的偏导数，避免了链式法则

计算方法导致的计算冗余问题，极大的减少了网络模型的参数计算量。

反向传播算法主要有三个步骤：前向传播、后向传播与权重更新。前向传播是指将训练数据送入网络，将上一层的输出作为下一层的输入，并计算下一层的输出，一直到运算到输出层得到输出结果为止；后向传播是指根据损失函数值和样本的实际标签，计算模型各层的误差；权重更新是指根据优化算法计算每个神经元参数的梯度，再将学习率乘以梯度并取反后加到权重上，迭代更新参数。

在训练过程中进行反复循环迭代以上的步骤，直到满足迭代停止条件。通常设定当网络的输出达到给定区间或者迭代次数达到预定值时迭代结束。

反向传播算法根据链式法则从后向前计算偏导，因此损失函数  $E$  对于第  $l$  层参数  $z^l$  的偏导数可以写成如下形式：

$$\frac{\partial E(\Theta)}{\partial z^l} = \frac{\partial E(\Theta)}{\partial z^L} \frac{z^L}{z^{L-1}} \frac{z^{L-1}}{z^{L-2}} \cdots \frac{z^{l+2}}{z^{l+1}} \frac{z^{l+1}}{z^l}, \quad (2.5)$$

其中  $L$  为网络的总层数， $z^l$  为网络第  $l$  层对应的参数， $E(\Theta)$  为交叉熵损失函数。

深度神经网络结构一般包括多个隐藏层，因此在反向传播的过程中就需要对各个层分别求解对应的梯度。

在训练神经网络参数的过程中，由于损失函数较为复杂且模型存在大量的超参数和待求解参数，通常很难找到解析解，所以一般采用优化算法与反向传播算法相结合的方法实现参数的迭代更新。

随机梯度下降法<sup>[44]</sup> (Stochastic Gradient Descent, SGD) 作为优化方法常用于人工神经网络模型的训练，该算法给定分组阈值并按照该阈值从样本中随机抽出一组，使用该组数据完成训练后按梯度更新一次，反复迭代以上步骤更新模型参数直到最终模型收敛。随机梯度下降法更新参数过程如下所示：

$$\Theta^{(n+1)} = \Theta^{(n)} - \eta \nabla_{\Theta} E(\Theta) |_{\Theta = \Theta^{(n)}}, \quad (2.6)$$

其中： $\eta$  为预设的学习率，一般为取值较小的正数， $\Theta^{(n)}$  为参数集合  $\Theta$  在第  $n$  次迭代时的估计值， $\nabla_{\Theta} E(\Theta) |_{\Theta = \Theta^{(n)}}$  为损失函数关于参数  $\Theta^{(n)}$  的梯度。

梯度下降算法的原理是在损失函数的给定点上，根据规定步长沿着对应梯度的反方进行迭代搜索，根据函数性质可知梯度的方向就是损失函数在给定点提升速度最大的方向，因此沿着梯度的反方向迭代就可以搜索到损失函数的极小值。损失函数值主要描述预测值不同于真实值的程度，损失函数值越小说明模型的精确度越高，常用的损失函数有交叉熵损失函数和平方损失函数等。



损失函数的表达式一般比较复杂, 在使用梯度下降算法搜寻损失函数的极小值时, 如果在每次迭代时所选择的步长过大, 则有可能偏离梯度下降的方向, 或者迭代时越过了最低点却没有停下来, 导致一直无法到达极小值点; 如果步长过小, 则需要频繁的计算梯度方向, 会非常耗时, 且有可能会陷入局部极小值点。此外, 分组样本量的大小对网络优化也有非常大的影响。通常情况下, 分组样本量的大小不会影响随机梯度的期望, 不过会影响随机梯度的方差。分组样本量大小越大, 随机梯度的方差越小, 引入的噪声也越小, 训练也越稳定。因此, 想要实现在不偏离梯度方向的同时耗时最短, 我们就需要在实际应用中选取合适的步长与分组。

## 2.2 随机共振基本理论

### 2.2.1 常用测度

互信息是随机共振研究中常用的测度, 互信息主要用来描述两个变量之间的相关性, 两个变量之间的互信息接近于 0 则两者的相关性越低。

设  $X$  与  $Y$  为两个连续型随机变量,  $p(x)$  为  $X$  的概率密度函数,  $p(y)$  为  $Y$  的概率密度函数,  $p(x,y)$  为  $X$  和  $Y$  的联合分布概率密度函数, 则  $X$  与  $Y$  的互信息可以表示为:

$$I(X,Y) = \int_Y \int_X p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dx dy, \quad (2.7)$$

### 2.2.2 双稳态 Langevin 模型

经典的随机共振理论就是在双稳态的非线性系统中传输周期信号的同时加入噪声, 在非线形系统中周期信号与随机噪声的共同作用之下, 产生协同效应, 将一部分噪声能量转化为信号能量。当噪声强度达到某个值时, 噪声的转化效率会达到最大值, 此时系统输出的信噪比也达到最大<sup>[45]</sup>。

通常随机共振是在信号、噪声和非线性系统共同作用下产生的非线性现象, 因此产生随机共振现象的必需要素为输入信号、噪声和双稳态系统。下面介绍最常用的随机共振理论分析模型: 双稳态 Langevin 模型<sup>[46]</sup>, 该系统可以用下述方程来描述, 其表达式为:

$$\frac{dx}{dt} = -\frac{\partial}{\partial x} U(x) + s(t) + n(t), \quad (2.8)$$

其中  $U(x)$  为势函数, 其表达式为:

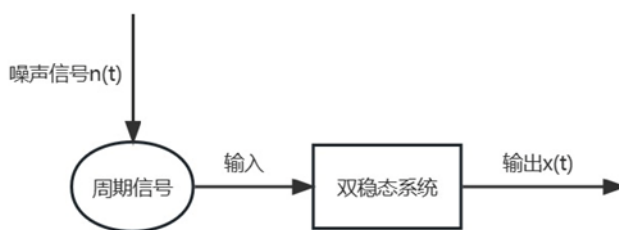
$$U(x) = -\frac{a}{2}x^2 + \frac{b}{4}x^4, \quad (2.9)$$

其中  $a>0$ ,  $b>0$ ,  $s(t)$  为周期性输入信号, 可表示为  $A\cos(\omega \cdot t)$ ,  $n(t)$  为服从高斯分布的噪声信号。

将公式(2.9)代入公式(2.8)后, 则有:

$$\frac{dx}{dt} = ax - bx^3 + A\cos(\omega \cdot t) + n(t), \quad (2.10)$$

于是根据公式(2.10), 可得到双稳态系统模型如下:



当输入信号  $s(t)$  的振幅  $A=0$  且噪声的强度为 0 时, 系统在  $x = \pm\sqrt{\frac{a}{b}}$  处存在两个固定点, 在  $x=0$  处存在一个亚稳态固定点。以上的固定点就是势函数取得极小值和极大值的位置, 且系统存在两个势阱, 阱底为  $x = \pm\sqrt{\frac{a}{b}}$ , 垒高为  $\Delta U = \frac{a^2}{4b}$ , 下图是系统结构参数  $a$  与  $b$  同时取值为 1 时, 势函数的曲线图。

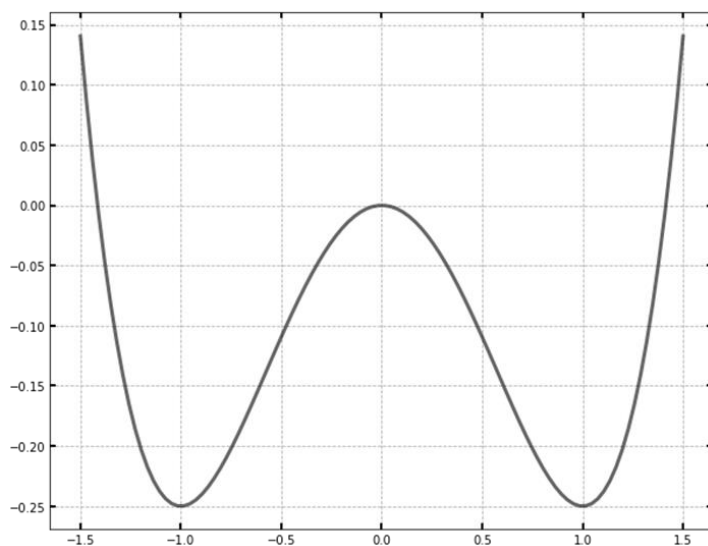


图 2.3 势函数的曲线图

当输入信号为恒定值时, 系统处于稳定状态。此时公式(2.9)中信号频率  $\omega=0$ , 双稳态系统的振幅阈值  $A_{threshold}$  可以计算求解。若系统的驱动信号振幅为  $A$ , 则势函数  $U(x)$  的极点和拐

点可表示为以下形式：

$$\frac{\partial U}{\partial x} = -ax + bx^3 + A = 0, \quad (2.11)$$

$$\frac{\partial^2 U}{\partial^2 x} = -a + 3bx = 0, \quad (2.12)$$

求解上述方程组可以得到双稳态系统的信号阈值：

$$A_{threshold} = \sqrt{\frac{4a^3}{27b}}, \quad (2.13)$$

公式 (2.13) 得到的阈值  $A_{threshold}$  是系统的信号阈值，也就是输入信号产生双稳态系统随机共振的阈值条件，只有当系统的输入大于系统输入阈值时，才有可能产生随机共振现象。

当外界输入信号  $s(t)$  的振幅  $A$  不为 0，噪声强度也不为 0 时，就会破坏系统的平衡，在输入信号  $s(t)$  和噪声的协同作用下，势阱会更加倾斜，最终克服系统输入阈值，在两个固定点  $x = \pm\sqrt{\frac{a}{b}}$  之间进行转换，从一个势阱转移到另一个势阱，输出信号发生大幅跳变。

通常双稳态系统之间的距离远大于输入信号的振幅，而噪声与输入信号的协同作用会使输出信号的振幅大于输入信号的振幅，对信号起到了放大的作用，最终使系统输出信号的信噪比得到大幅提高，产生随机共振现象。

## 2.3 本章小结

本章主要介绍了卷积神经网络和随机共振的基础理论。首先阐述了构成经典的卷积神经网络模型的多种隐藏层，分别介绍了卷积层、池化层、全连接层和输出层的计算过程，同时重点介绍了反向传播算法的特点以及其在卷积神经网络中的计算过程。最后详述了经典的随机共振系统分析模型：双稳态 Langevin 模型，论述了随机共振现象产生的基本原理。本章通过介绍相关的卷积神经网络与随机共振基础理论，为其后章节的研究内容奠定了基础。

## 第三章 卷积神经网络中的激活函数比较研究

在深度学习中, 激活函数的作用就是对隐藏层的输出进行非线性映射, 使激活后的输出具有非线性的特性, 从而提高模型特征提取和分类的性能<sup>[47]</sup>。在图像分类问题中, 激活函数的选择对模型性能会有极大的影响, 由于不同的激活函数具有不同的激活能力, 激活后的信息输出就会存在差别<sup>[48]</sup>, 研究不同激活函数对网络模型的影响也逐渐成为学者的研究重点。Hinton 等人<sup>[11]</sup>提出了 ReLU 激活函数用来解决饱和激活函数存在的梯度消失问题; He 等人<sup>[49]</sup>提出了 PReLU 激活函数解决 ReLU 函数易导致神经元坏死的问题; Glorot 等人<sup>[50]</sup>提出 Softplus 函数, 与 PReLU 函数相比, Softplus 在定义域内连续可微, 同时其稀疏性更低。因此本章对激活函数进行了深入研究, 分析了不同激活函数对卷积神经网络训练的影响及其原理性能, 并分别在 Fashion-MNIST 和 MNIST 数据集上进行对比实验。在对比实验中, 仅改变激活函数类型而其余参数与结构完全相同。最后根据实验结果比较不同激活函数对卷积神经网络的影响。

### 3.1 激活函数对神经网络的影响

神经网络中激活函数的主要作用是使神经网络对输出进行非线性变换。如果网络中不使用激活函数对输出进行激活, 网络上一层的输出与这一层的输入就会呈线性关系, 难以有效提取输入信息的数据特征, 从而影响模型的训练速度和分类效果。因此使用激活函数对输出进行非线性映射, 提高了神经网络模型对复杂信息的特征提取能力, 进而提高模型的分类准确率。在深度神经网络训练过程中, 激活函数通常会在前向传播和反向传播两个过程中对模型产生影响。

在前向传播中, 样本从输入层输入模型到输出层输出结果的整个数据传输过程都需要通过激活函数将数据进行非线性变换, 进而实现信息从前往后的传输。

反向传播过程是计算各参数梯度并更新对应权值和偏置值的过程。网络中反向传播的梯度计算公式如下:

$$\frac{\partial E(\Theta)}{\partial z^l} = \frac{\partial E(\Theta)}{\partial z^L} \frac{z^L}{z^{L-1}} \frac{z^{L-1}}{z^{L-2}} \cdots \frac{z^{l+2}}{z^{l+1}} \frac{z^{l+1}}{z^l}, \quad (2.5)$$

从公式(2.5)可以看出, 在每一次迭代更新中, 反向传播每经过一层, 参数的梯度都会乘上激活函数的导数。如果选择的激活函数具有饱和性, 那么在经过多次迭代以后神经网络模

型可能会由于参数梯度过小而无法有效更新参数。因此激活函数也是影响模型训练效果的因素。

## 3.2 常用的激活函数

### 3.2.1 Sigmoid 函数

Sigmoid 函数是经典的激活函数,其函数表达式为:  $f(x) = \frac{1}{1+e^{-x}}$ , 函数图像如图 3.1 所示。

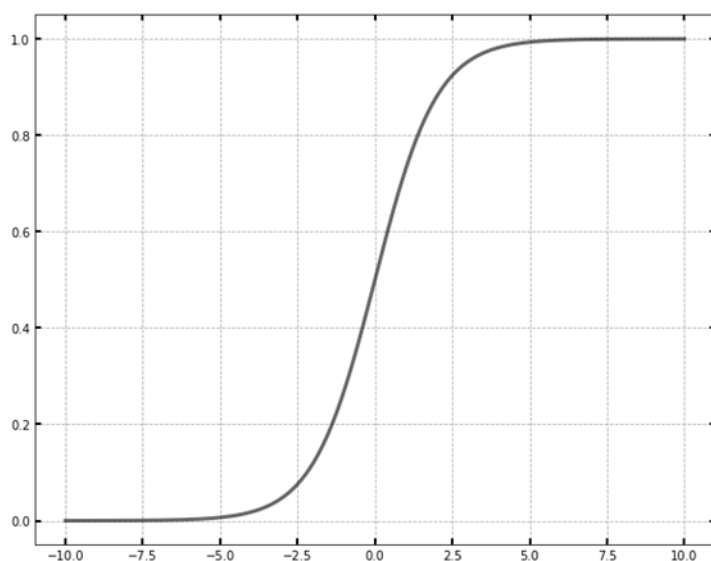


图 3.1 Sigmoid 函数图像

从图 3.1 可以看出, Sigmoid 函数的值域为  $(0,1)$ , 函数非线性。当输入  $x$  的绝对值过大时, 函数曲线趋于平缓, 表示此时的函数导数接近于 0。只有当  $x$  取值接近零点时, 导数取值较大。因此 Sigmoid 函数在零点附近的信号增益较大, 在两侧边缘区域的信号增益较小, 相比线性函数与阈值函数具有非线性的优势, 因此在一段时间内得到了广泛地应用。但是, Sigmoid 函数在实际应用中也有一定的缺陷。首先当输入非常大或者非常小时, 该函数梯度趋近于 0, 导致在反向传播过程中会使梯度变得非常小, 就会存在梯度消失问题。且该函数在反向传播过程中需要进行幂运算, 计算量较大也更加耗时。同时 Sigmoid 函数的输出不是关于原点中心对称的, 导致后面网络层的输入也不以零为中心, 可能会在一定程度上改变原始数据的分布情况。

### 3.2.2 Tanh 函数

由于 Sigmoid 函数存在各种问题, 之后又有大量的改进函数被提出来代替 Sigmoid 函数,

其中应用最广的是使用双曲正切(Tanh)函数。Tanh 函数的表达式为:  $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , 其对应的函数图像如图 3.2 所示。

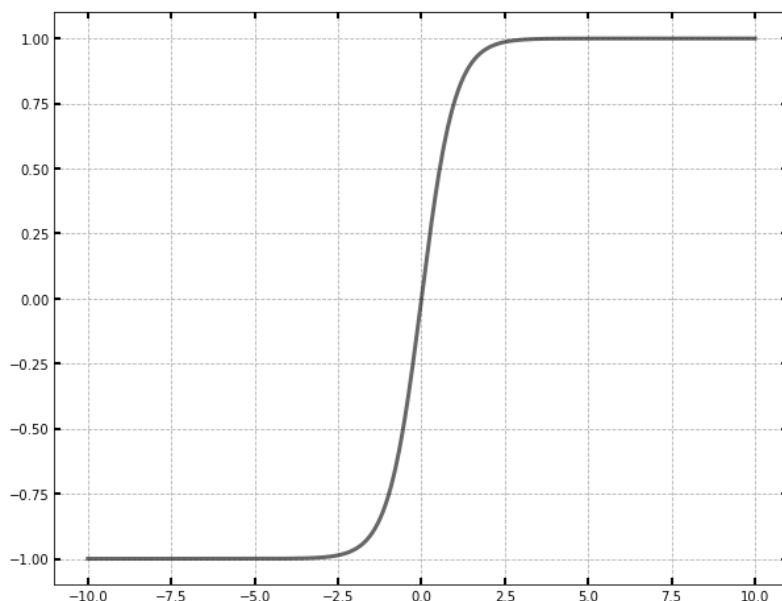


图 3.2 Tanh 函数图像

从图 3.2 可以看出, Tanh 函数也是连续可微的非线性函数。与 Sigmoid 函数不同, Tanh 的输出关于零点对称, 函数值域区间为  $(-1,1)$ 。Tanh 函数的优势在于该函数的收敛速度较快, 但从函数图像可以看到, 当函数输入非常大或者非常小时, Tanh 函数的导数值仍然接近 0, 因此该函数仍然存在梯度消失问题。同时根据其函数表达式可知, Tanh 函数在反向传播过程中也需要幂运算, 计算量较大的问题依然存在。

### 3.2.3 ReLU 函数

2001 年 Charles 等人通过使用 Softplus 函数在正数回归预测中取得了很好的效果<sup>[51]</sup>, 而 Softplus 函数可以认为是平滑处理的线性整流函数, 此后 ReLU 函数得到了进一步研究。Sigmoid 和 Tanh 两种经典的激活函数的值域是有界的, 因此会由于函数的饱和性产生梯度消失问题。而 ReLU 函数正部是线性无界的, 通过使用输入的线性映射避免了输入为正时的饱和问题。ReLU 函数在负象限输出恒为零, 保证了负部稀疏。其函数表达式为:  $f(x) = \max(0, x)$ , ReLU 函数对应的图像如图 3.3 所示。

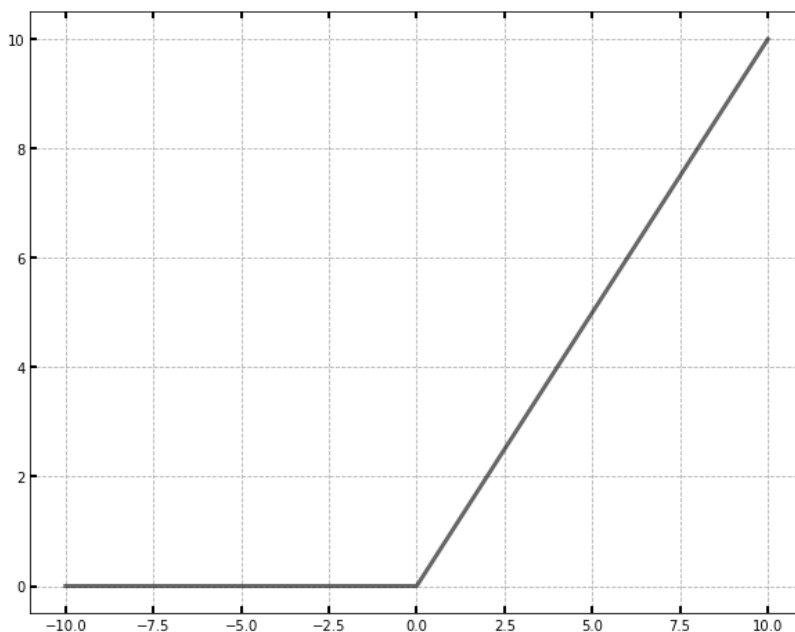


图 3.3 ReLU 函数图像

由图 3.3 可知, ReLU 函数在  $x > 0$  时可以保持梯度为常数, 因此有效的缓解了梯度消失问题, 而且其梯度运算简单, 大幅减少了反向传播过程的计算量。此外 ReLU 函数优于其他激活函数的另一特性是该函数具有稀疏性, 使模型的全部神经元不会同时被激活。当输入值为负时, ReLU 函数的输出固定为 0, 此时神经元不会被激活。ReLU 函数的稀疏性导致在模型的训练过程中只有一部分神经元会被激活, 这就使得模型训练变得高效且易于计算。

### 3.3 实验数据集介绍

#### 3.3.1 MNIST 数据集

MNIST 数据集是一个大型的手写数字或数字数据库<sup>[52]</sup>, 来自美国国家标准与技术研究所, 用于训练各种图像处理系统。该数据集还广泛用于机器学习领域的培训和测试。MNIST 数据库中的图像集是 NIST 中特殊数据库 1 和特殊数据库 3 两个数据库的组合。

MNIST 数据集的训练集由来自 250 个不同人手写的数字构成, 其中 50% 是高中学生, 50% 来自人口普查局的工作人员。测试集的手写数字数据也具有同样的构成比例, 同时保证了测试集和训练集的作者集不相交。

MNIST 数据集一共包含 7 万张手写数字图片, 其中的 6 万张划分为训练集, 另外的 1 万张划分为测试集。每张图片均是大小为  $28 \times 28$  的手写数字图片, 由 0-9 的十种数字组成。每个图片的表现形式为黑底白字, 像素值为 0 表示黑色, 白色像素则用 0-1 之间的浮点数表示, 像素值越接近 1, 其对应的区域颜色越接近白色。图 3.4 为 MNIST 数据集部分样本图片展示。

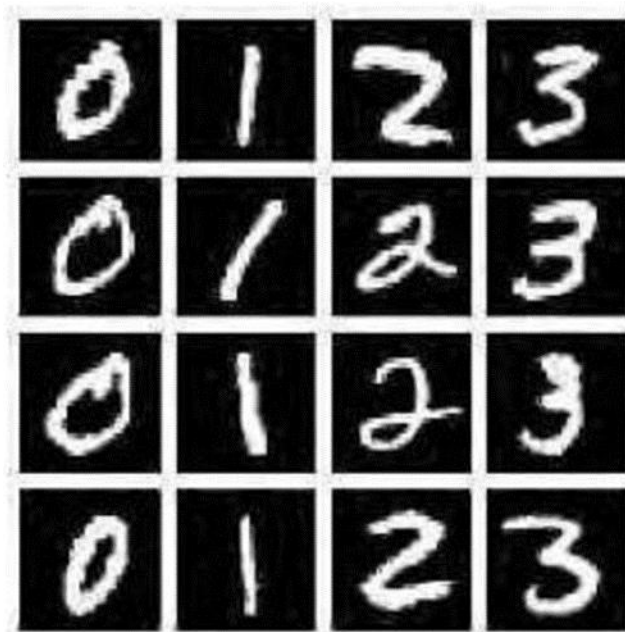


图 3.4 MNIST 数据集部分样本

### 3.3.2 Fashion-MNIST 数据集

Fashion-MNIST 数据集是德国 Zalando 公司提供的衣物图像数据集<sup>[53]</sup>, 包含 60,000 个样本的训练集和 10,000 个样本的测试集。每个样本均是大小为 28x28 的灰度图像, 与 10 个类别的标签对应。Fashion-MNIST 数据集的建立是作为原始 MNIST 数据集的替代品, 用于对机器学习算法进行测试。

与 MNIST 手写数据集不同, Fashion-MNIST 数据集包含 10 个类别的衣物图像, 分别是: t-shirt (T 恤), trouser (牛仔裤), pullover (套衫), dress (裙子), coat (外套), sandal (凉鞋), shirt (衬衫), sneaker (运动鞋), bag (包), ankle boot (短靴)。其中训练集每个类别包含 6000 个样本, 而测试集的每个类别包含 1000 个样本, 数据集一共包含 10 个类别, 因此训练数据集包含的样本量为 60000, 测试数据集包含的样本量为 10000。

在数据集中每个图像使用 28\*28 的像素值矩阵进行描述, 矩阵中每个像素的取值为 0~255 之间的无符号整数, 使用三维数组进行存储, 最后一维表示通道个数。由于样本全部为灰度图像, 故数据集通道数为 1。图 3.5 为 Fashion-MNIST 数据集部分样本的可视化展示。



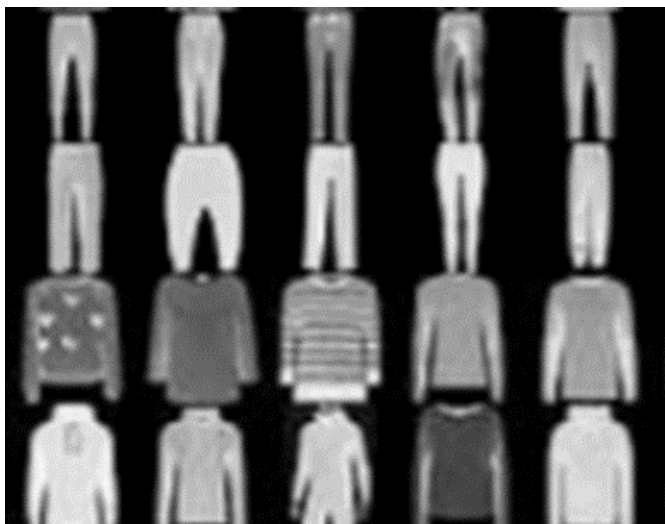


图 3.5 Fashion-MNIST 数据集部分样本

## 3.4 实验结果与分析

### 3.4.1 实验设计

为了研究激活函数对卷积神经网络的影响，实验在 MNIST 和 Fashion-MNIST 两个数据集上进行，使用 Sigmoid、Tanh 和 ReLU 激活函数进行了对比，对比结果证明了 ReLU 函数的有效性。本章实验采用 Python 语言进行编程，基于 Keras 框架实现仿真，针对两个数据集分别建立了不同结构的卷积神经网络模型，在进行对比实验时使用 Sigmoid、Tanh 和 ReLU 激活函数进行模型的训练。为了研究上述三种激活函数对模型的影响，本章在两个数据集上分别设计了三组实验，每次实验仅改变激活函数的类型，保证各实验的硬件条件与网络结构都相同。每次实验的训练次数为 1500，每 100 次训练输出模型的损失函数值与分类正确率。

### 3.4.2 MNIST 数据集上的实验结果与分析

MNIST 数据集的样本包含 10 个不同类别，选取 10000 个样本作为训练集，1000 个样本作为测试集，在训练集中再选取 20% 的数据作为验证集，其余的数据用于训练。针对 MNIST 数据集，实验设计了一个较为简单的卷积神经网络模型，模型的结构包括两个卷积层、一个最大池化层和两个全连接层。其中第一个卷积层卷积核的个数为 8，第二个卷积层卷积核的个数为 16，大小均为  $5 \times 5$ 。最大池化层窗口大小为  $2 \times 2$ 。模型的全部参数量为 209642。训练轮数 epochs 设置为 15 轮，选择 Adam 作为优化器算法，学习率为默认值。实验选取分类准确率、交叉熵损失函数值和训练耗时作为测度来分析不同激活函数对模型的影响。

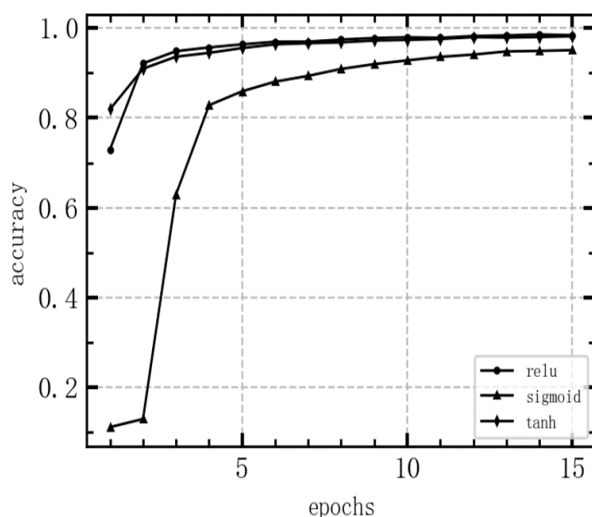


图 3.6 不同激活函数的 MNIST 训练集分类准确率

三种激活函数的 MNIST 训练集分类准确率如图 3.6 所示。由图 3.6 可知，使用 ReLU 激活函数的模型训练集分类准确率最高，且模型的收敛速度最快；使用 Tanh 激活函数的模型训练集分类准确率与 ReLU 激活函数接近，模型分类性能良好；在三种模型中，使用 Sigmoid 激活函数的模型训练集分类准确率最低，模型分类性能最差，且收敛速度不如其他的激活函数。

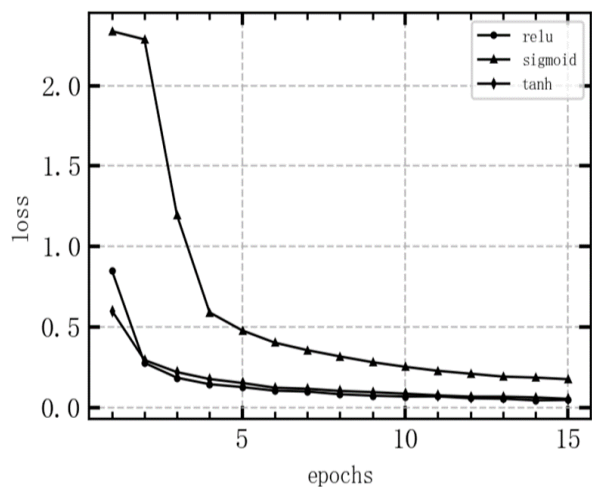


图 3.7 不同激活函数的 MNIST 训练集损失函数值

三种激活函数的 MNIST 训练集损失函数值如图 3.7 所示。由图 3.7 可知，使用 ReLU 激活函数的模型训练集损失函数值最小，且模型的收敛速度最快；与 ReLU 激活函数相比，使用 Tanh 激活函数的模型训练集损失函数值稍高，两种模型分类性能相近；在三种模型中，使用 Sigmoid 激活函数的模型训练集分类准确率最低，且收敛速度远低于其他的激活函数。

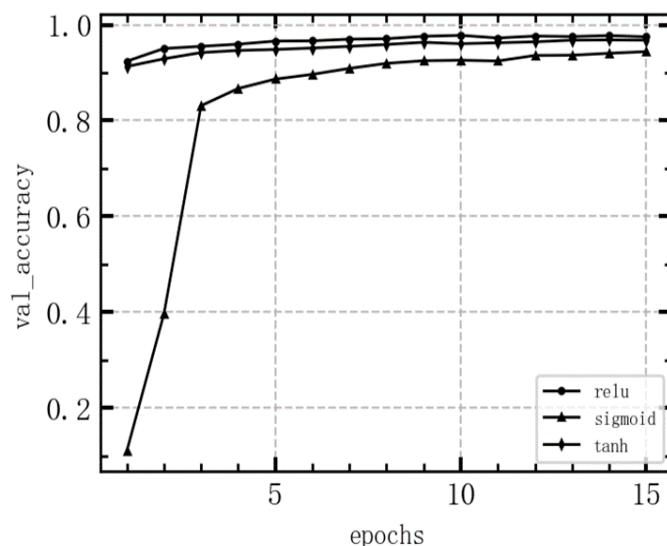


图 3.8 不同激活函数的 MNIST 验证集分类准确率

三种激活函数的 MNIST 验证集分类准确率如图 3.8 所示。由图 3.8 可知，使用 ReLU 激活函数的模型验证集分类准确率最高，在三种模型中具有最高的分类性能与泛化性能；使用 Tanh 激活函数的模型验证集分类准确率略低于 ReLU 激活函数，模型的分类性能较好；使用 Sigmoid 激活函数的模型验证集分类准确率最低，且在最初几次迭代时的验证集分类准确率过低，模型的泛化能力差。

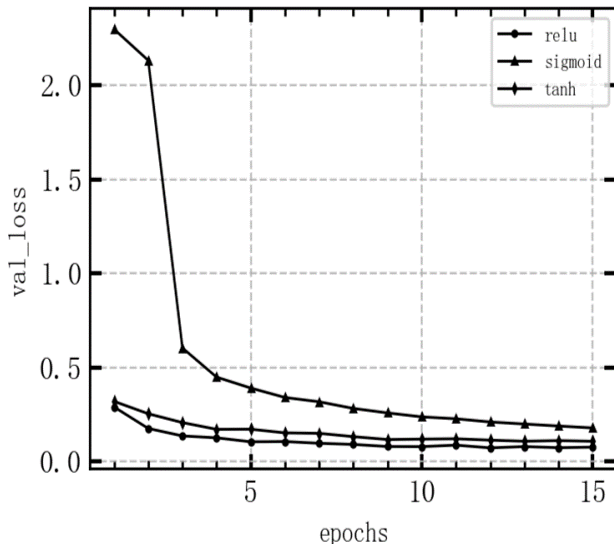


图 3.9 不同激活函数的 MNIST 验证集损失函数值

三种激活函数的 MNIST 验证集损失函数值如图 3.9 所示。由图 3.9 可知，使用 ReLU 激活函数的模型验证集损失函数值最小，模型的分类效果与泛化性能良好；使用 Tanh 激活函数的模型验证集损失函数值略高于 ReLU 激活函数，模型的泛化性能仅次于 ReLU 函数；使用 Sigmoid 激活函数的模型验证集分类准确率最低，且在最初几次迭代时的验证集分类准确率过低，分类效果不如其他的模型。

表 3.1 MNIST 测试集实验结果

	测试集分类准确率	测试集交叉熵	平均训练耗时
ReLU	0.983	0.049	141s
Sigmoid	0.950	0.143	145s
Tanh	0.973	0.078	144s

在 MNIST 测试集上不同激活函数的实验结果如表 3.1 所示。由表 3.1 可知，使用 ReLU 激活函数的模型在分类准确率、交叉熵和训练耗时三个测度上的表现均为最优。相比于 Sigmoid 和 Tanh 激活函数，ReLU 函数的分类准确率分别提高了 3.3%和 1%；交叉熵分别降低了 0.094 和 0.029；平均训练耗时分别降低了 4s 和 3s。说明在本章的实验条件下，ReLU 函数在 MNIST 数据集上取得的测试集分类效果最好，同时模型的泛化性能最优，模型参数的计算量也比其他激活函数更低。

3.4.3 Fashion-MNIST 数据集上的实验结果与分析

Fashion-MNIST 数据集的样本同样包含 10 个不同类别，选取 10000 个样本作为训练集，1000 个样本作为测试集，在训练集中再选取 20%的数据作为验证集，其余的数据用于训练。针对 Fashion-MNIST 数据集，实验设计了一个简单的卷积神经网络模型，模型的结构包括两个卷积层、一个最大池化层和两个全连接层。其中第一个卷积层卷积核的个数为 16，第二个卷积层卷积核的个数为 24，大小均为 3\*3。最大池化层窗口大小为 2\*2。模型的全部参数量为 447426。训练轮数 epochs 设置为 15 轮，选择 Adam 作为优化器算法，学习率为默认值。实验选取分类准确率、交叉熵损失函数值和训练耗时三个测度来研究不同激活函数对模型训练的影响。

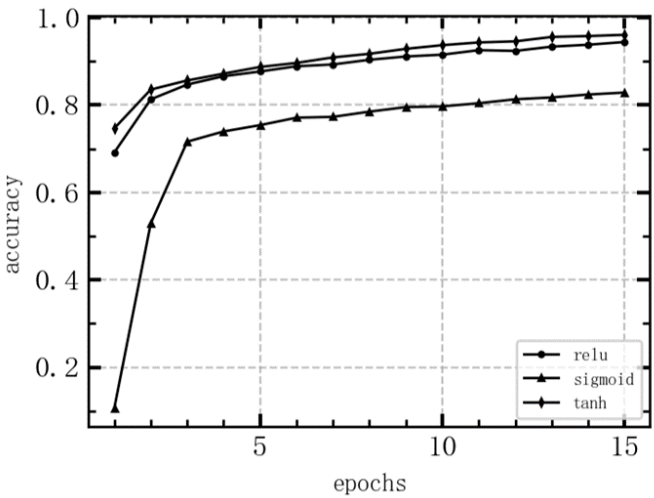


图 3.10 不同激活函数的 Fashion-MNIST 训练集分类准确率

三种激活函数的 Fashion-MNIST 训练集分类准确率如图 3.10 所示。由图 3.10 可知,使用 Tanh 激活函数的模型训练集分类准确率最高,且模型的收敛速度较快,在三种模型中的分类效果最好;使用 ReLU 激活函数的模型训练集分类准确率在最初迭代时低于 Tanh 函数,随着迭代次数增加逐渐接近于 Tanh 激活函数;使用 Sigmoid 激活函数的模型训练集分类准确率远低于其他激活函数,模型收敛速度也低于其他激活函数。

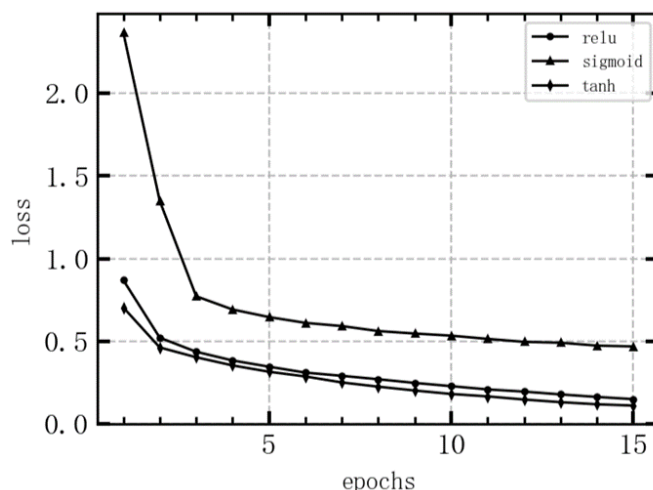


图 3.11 不同激活函数的 Fashion-MNIST 训练集损失函数值

三种激活函数的 Fashion-MNIST 训练集损失函数值如图 3.11 所示。由图 3.11 可知,使用 Tanh 激活函数的模型训练集损失函数值最小,模型的收敛速度很快;使用 ReLU 激活函数的模型训练集损失函数值在最初迭代时大于 Tanh 函数,随着迭代次数增加,其损失函数值逐渐接近于 Tanh 激活函数;使用 Sigmoid 激活函数的模型训练集损失函数值在相同迭代次数的情况下均远高于其他函数,在三种模型中使用 Sigmoid 函数的模型的训练集分类效果最差。

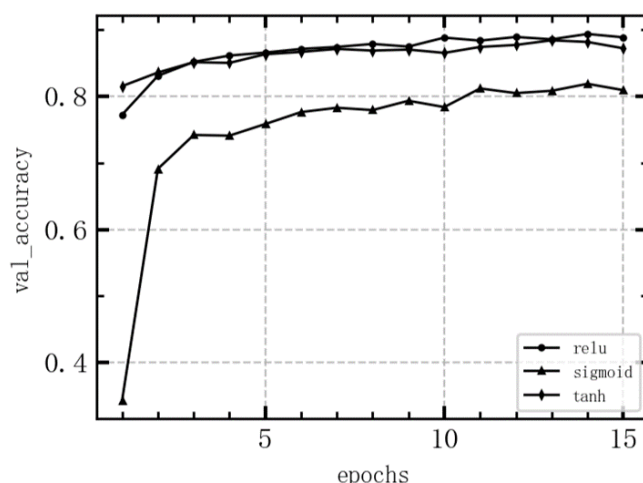


图 3.12 不同激活函数的 Fashion-MNIST 验证集分类准确率

三种激活函数的 Fashion-MNIST 验证集分类准确率如图 3.12 所示。由图 3.12 可知,使用 ReLU 激活函数的模型验证集分类准确率最高,在三种模型中具有最高的分类性能与泛化性

能；使用 Tanh 激活函数的模型验证集分类准确率略低于 ReLU 激活函数，模型的分类性能较好；使用 Sigmoid 激活函数的模型验证集分类准确率最低，且在最初几次迭代时的验证集分类准确率过低，模型的泛化能力差。

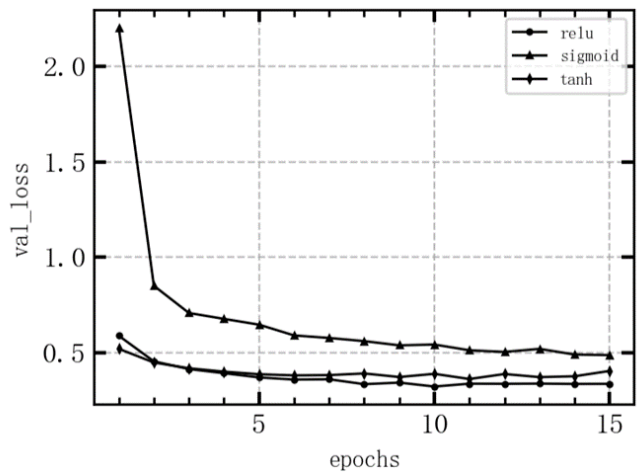


图 3.13 不同激活函数的 Fashion-MNIST 验证集损失函数值

三种激活函数的 Fashion-MNIST 验证集损失函数值如图 3.13 所示。由图 3.13 可知，使用 ReLU 激活函数的模型验证集损失函数值最小，在三种模型中具有最高的分类效果与泛化性能；使用 Tanh 激活函数的模型验证集损失函数值略高于 ReLU 激活函数，相比于 ReLU 函数模型，该模型的分类性能与泛化性能较低；使用 Sigmoid 激活函数的模型验证集分类准确率最低，且在最初几次迭代时的验证集分类准确率过低，分类效果不如其他的模型。

表 3.2 Fashion-MNIST 测试集实验结果

	测试集分类准确率	测试集交叉熵	平均训练耗时
ReLU	0.893	0.358	188s
Sigmoid	0.836	0.468	191s
Tanh	0.882	0.384	190s

在 Fashion-MNIST 测试集上不同激活函数的实验结果如表 3.2 所示。由表 3.2 可知，使用 ReLU 激活函数的模型在交叉熵、分类准确率和训练耗时三个测度上的表现均为最优。相比于 Sigmoid 和 Tanh 激活函数，ReLU 函数的分类准确率分别提高了 5.7%和 1.1%；交叉熵分别降低了 0.11 和 0.026；平均训练耗时分别降低了 3s 和 2s。说明在本章的实验条件下，ReLU 函数在 Fashion-MNIST 数据集上取得的测试集分类效果最好，同时模型的泛化性能也更优秀，模型参数的计算量也比其他激活函数更低。

3.5 本章小结

由于图像分类方向存在的不同需求，在实际应用中通常不存在统一或者固定的网络模型

参数和隐藏层数量,因此在卷积神经网络模型中,构建或选取适当的激活函数是取得良好实验效果的关键之一。本章节研究了激活函数对网络模型的影响以及传统激活函数所存在的缺陷,介绍了三种应用广泛的激活函数,并将以上的激活函数应用于灰度图像分类问题中,以分类准确率和交叉熵作为测度,通过在 MNIST 和 Fashion-MNIST 两个数据集上的对比实验,对三种激活函数进行了比较研究。在 MNIST 数据集的实验中,使用 ReLU 函数的模型在训练集、验证集和测试集上均具有最优的分类效果;在 Fashion-MNIST 数据集的实验中,使用 Tanh 函数的模型在训练集上具有最优的分类效果,但在验证集和测试集上的表现不如 ReLU 函数。这一实验结果表明相比于另外两种激活函数,使用 ReLU 函数的模型在交叉熵与分类准确率方面具有比较好的表现,同时在本章实验条件下 ReLU 函数在提升模型的泛化性能和平均训练耗时方面具有不可替代的优势,进一步证明了选择合适的激活函数对神经网络的重要性。



## 第四章 ReLU 激活函数下的卷积神经网络随机共振研究

目前针对卷积神经网络中的随机共振研究都集中在传统卷积网络模型上,这类模型通常选择 Sigmoid 函数作为激活函数且结构只包含一个卷积层和池化层,模型结构较为简单。其中 Osoba 等人<sup>[25]</sup>在 2013 年证明了反向传播算法是广义 EM 算法的一种特例,提出了 BP 算法存在噪声增益的充分条件,同时在 MNIST 数据集上的实验结果证明了加入适当的噪声可以提高 BP 算法的收敛速度。2016 年, Kosko 等人<sup>[29]</sup>提出了噪声卷积神经网络(Noise-enhanced CNN, NCNN)算法,发现与无噪声 BP 算法训练的模型相比, NCNN 算法在前几次迭代中具有更低的交叉熵和分类错误率。因此,本章首先从理论角度分析了反向传播算法在模型训练过程中产生噪声增益的原因,然后提出了一种改进的噪声增强卷积神经网络算法,结合小批量梯度下降算法,研究了使用 ReLU 函数作为激活函数的卷积神经网络中的噪声增益,最后选取交叉熵<sup>[54]</sup>和分类错误率作为测度,在 MNIST 数据集和 Fashion-MNIST 数据集上进行仿真实验,研究改进的噪声增强卷积神经网络中存在的随机共振现象。

### 4.1 改进的噪声增强卷积神经网络算法

期望最大化算法是一种迭代的极大似然估计方法<sup>[55]</sup>,该算法通过对缺失信息求条件期望来减少数据缺失带来的误差。文献<sup>[17]</sup>将相对熵作为衡量标准,以高斯混合模型为例进行仿真实验,发现加入的加性噪声在强度较低时能提高模型的平均收敛速度,强度较高时能降低算法的平均收敛速度,同时证明了公式(4.1)是加性噪声在 EM 算法中产生噪声增益的一般充分条件:

$$E_{Y,H,N|\theta^*}[\ln(\frac{f(Y+N,H|\theta_k)}{f(Y,H|\theta_k)})] \geq 0, \quad (4.1)$$

其中  $Y$  为观测变量,  $N$  为加性噪声变量,  $H$  为 EM 算法中的隐变量,  $\theta^*$  为参数极限。

文献<sup>[17]</sup>进一步证明了反向传播算法是广义期望最大化算法的一种特殊情况,说明在卷积神经网络的反向传播训练中也可以存在噪声增益,因此考虑加入合适的噪声来提高模型的性能。之后文献<sup>[29]</sup>以 MNIST 数据集为样本提出了噪声增强卷积神经网络算法,通过在反向传播算法迭代更新模型参数的过程中加入加性均匀噪声来研究卷积神经网络中的噪声增益。为了研究不同结构的卷积神经网络中的噪声增益,本文在文献<sup>[29]</sup>提出的噪声增强卷积神经网络基础上对神经网络结构与学习算法进行改进,研究了使用 ReLU 激活函数的卷积神经网络存



在的噪声增益。



图 4.1 改进的模型结构

改进后的模型结构如图 4.1 所示，改进后的卷积神经网络包括 2 个卷积层，1 个池化层和 1 个全连接层，在卷积层中使用了 8 个  $5*5$  的卷积核，步长为 1，两次卷积均不使用 padding，池化层中使用步长为 2 的  $2*2$  卷积核，模型的初始学习率设置为 0.01，采用 ReLU 函数作为激活函数。与 Sigmoid 和 Tanh 函数不同，ReLU 函数在  $x>0$  时，梯度恒为 1，在  $x<0$  时，输出为 0，这就使模型神经元在反向传播过程中具有稀疏性，在减少了模型计算量的同时使得模型训练更加高效，此外还可以避免反向传播过程中的梯度消失问题。

在反向传播更新参数的过程中，采用小批量梯度下降法对 NCNN 算法进行改进。小批量梯度下降法将全部样本随机分为  $a$  组，在每次迭代更新模型参数时只使用一组样本来进行计算，因此模型完成一次全样本的参数更新只需要  $a$  次迭代，提高了参数计算效率；当组内样本量小于训练集样本量时所需的训练时间比批量梯度下降法更少，同时该算法增加了学习过程中的噪声，有助于改善泛化误差，避免陷入局部最优<sup>[56]</sup>。

改进后的模型使用交叉熵作为损失函数。设  $t$  表示输入图片的序号，取值范围在 1 到  $T$  之间； $y$  表示输入图像  $X$  的目标标签的独热编码向量<sup>[57]</sup>，即当  $k$  为正确的类时对应的分量为 1，其他  $k-1$  个分量为 0。反向传播算法计算输出神经元的激活向量与二进制标签向量  $y$  之间的交叉熵为：

$$E(\Theta) = -\sum_{k=1}^K y_k \ln(o_k^t), \quad (4.2)$$

其中  $\Theta$  为卷积神经网络的参数， $o_k^t$  为对应的输出向量。

假设模型中  $K$  个输出神经元相互独立，在输出层中各输出神经元之间不存在相互连接。此时输出的激活向量服从伯努利分布，可得到对应的似然函数：

$$p(y|x, \Theta) = \prod_{k=1}^K p(y_k|x, \Theta), \quad (4.3)$$

若  $k$  为输入图像的正确标签分类则  $y_k = 1$ ，否则  $y_k = 0$ ，公式(4.3)中的第  $k$  个因子  $p(y_k|x, \Theta)$  可以表示为  $(o_k^t)^{y_k}$ ，此时神经网络的对数似然函数可以表示为交叉熵的函数：

$$L(\Theta) = \ln p(y|x, \Theta) = \ln \prod_{k=1}^K p(y_k|x, \Theta) = \ln \prod_{k=1}^K (o_k^t)^{y_k} = \sum_{k=1}^K y_k \ln(o_k^t) = -E(\Theta), \quad (4.4)$$

因此在上述假设下, 求解使交叉熵损失函数最小的参数等价于求解使对数似然函数最大的参数, 使用反向传播算法迭代求解损失函数的极小值在本质上与 EM 算法迭代求解对数似然函数最大值的原理是相同的, 从而可以利用含噪声的 EM 算法来优化 CNN 的反向传播训练。

根据交叉熵的公式, 可以使用梯度下降方法迭代计算参数  $\Theta$  的极大似然估计, 迭代公式为:

$$\Theta^{(n+1)} = \Theta^{(n)} - \eta \nabla_{\Theta} E(\Theta) |_{\Theta = \Theta^{(n)}}, \quad (4.5)$$

其中  $\eta$  为预设的学习率, 一般为取值较小的正数,  $\Theta^{(n)}$  为参数  $\Theta$  在第  $n$  次迭代时的估计值,

$\nabla_{\Theta} E(\Theta) |_{\Theta = \Theta^{(n)}}$  为损失函数关于参数  $\Theta^{(n)}$  的梯度。

改进后的模型在每次迭代计算交叉熵时会随机生成服从均匀分布的噪声向量  $N$ , 若噪声向量  $N$  满足公式 (4.1) 则在对应的标签向量中加入噪声, 并使用加噪后的标签向量计算交叉熵, 最后采用小批量梯度下降算法结合公式 (4.5) 更新模型参数直到收敛。

最后本章以 MNIST 和 Fashion-MNIST 数据库作为数据集结合改进的噪声增强卷积神经网络算法进行实验, 在两个实验数据集中分别随机选择 500 个样本作为训练数据, MNIST 数据集中每张图片表示 0-9 中的某一个数字, 而 Fashion-MNIST 数据集中每张图片表示 10 个不同衣物类别的其中一种, 样本中同时包含了各个训练数据相应的标签, 在原始标签集中使用 0-9 总共十种数字对图片进行分类。由于样本标签的分布是离散值而不是连续值, 且各类别的取值之间没有大小关系, 因此在训练之前将标签集中的数据统一转换为独热编码向量<sup>[57]</sup>; 为了提高模型的分类性能与计算速度, 还需要对数据集中的图片像素值做标准化处理。

## 4.2 均匀噪声

本章在反向传播算法过程中加入的噪声均服从均匀分布, 对于均匀分布的噪声, 若其均值为 0, 方差为  $\sigma_n$ , 则均匀噪声的概率密度函数表达式如下:

$$f_n(x) = \begin{cases} \frac{1}{\sqrt{12}\sigma_n}, & -\sqrt{3}\sigma_n < x < \sqrt{3}\sigma_n, \\ 0, & \text{其他} \end{cases} \quad (4.6)$$

其对应的分布函数如下式所示:

$$F_n(x) = \begin{cases} 0 & , \quad x < -\sqrt{3}\sigma_n \\ \frac{x + \sqrt{3}\sigma_n}{2\sqrt{3}\sigma_n} & , \quad -\sqrt{3}\sigma_n \leq x \leq \sqrt{3}\sigma_n \\ 1 & , \quad x > \sqrt{3}\sigma_n \end{cases} \quad (4.7)$$

服从均值为 0，标准差为 1 的均匀分布噪声的概率密度曲线如图 4.2 所示：

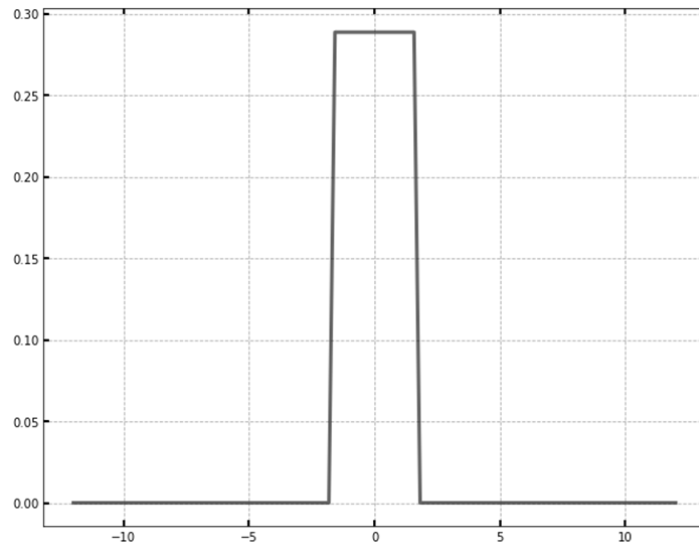


图 4.2 均值为 0，标准差为 1 的均匀噪声概率密度曲线

## 4.3 实验结果分析

### 4.3.1 MNIST 数据集上的实验结果与分析

本节通过在 MNIST 数据集上进行实验来研究卷积神经网络训练过程中的噪声增益。实验设计使用图 4.1 的网络结构进行模型训练，在对比实验中使用加入噪声的 NCNN 与无噪声 CNN 进行对比。实验中的训练集样本量为 500，初始噪声强度设置为 0.9，模型的初始学习率设置为 0.01，选择训练集交叉熵以及分类错误率作为研究测度。其中 NCNN 模型的反向传播过程中在输出层对标签向量加入满足公式的加性均匀噪声，同时采用小批量随机梯度下降算法，在每轮迭代中将样本数据随机分为 25 组进行模型训练。在输出神经元中加入的随机噪声为  $[-0.5\sqrt{N_c/d}, 0.5\sqrt{N_c/d}]$  区间上的加性均匀噪声，其中  $N_c$  为噪声强度， $d$  为迭代次数，即噪声强度随着迭代次数增加而降低。

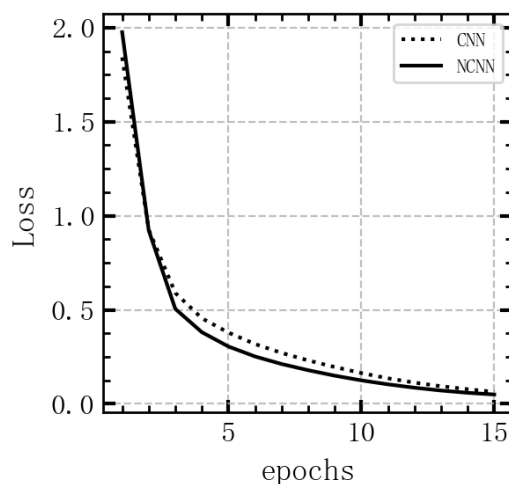


图 4.3 MNIST 数据集上的交叉熵对比

在 MNIST 数据集上卷积神经网络训练过程中交叉熵与迭代次数的变化关系如图 4.3 所示。可以发现相较于无噪声 CNN，NCNN 在训练集上的交叉熵收敛速度更快，在迭代三次之后交叉熵曲线始终处于 CNN 的交叉熵曲线下方，与标准的 CNN 相比，NCNN 在前 10 次训练迭代中平均减少了 16.3% 的交叉熵，因此 NCNN 具有更好的分类能力，验证了模型在 MNIST 数据集中存在噪声增益。

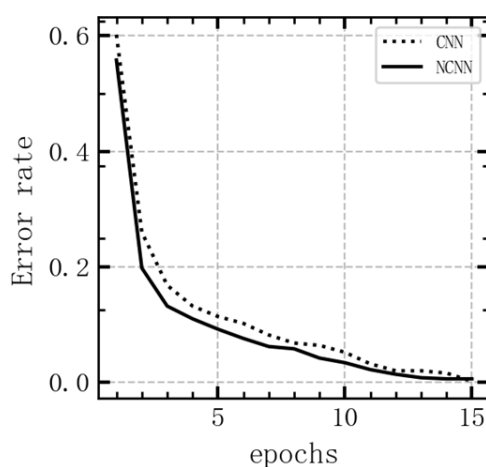


图 4.4 MNIST 数据集上的分类错误率对比

在 MNIST 数据集上卷积神经网络训练过程中分类错误率与迭代次数的变化关系如图 4.4 所示。根据图 4.4 所示，实验前几次迭代时，NCNN 在训练集上的分类错误率始终小于无噪声 CNN 的分类错误率，在前 10 次训练迭代过程中，NCNN 的平均分类错误率相较无噪声 CNN 降低了 3.1%，因此 NCNN 与无噪声 CNN 相比具有更低的分类错误率，说明噪声的加入提高了模型在 MNIST 数据集上的分类性能。

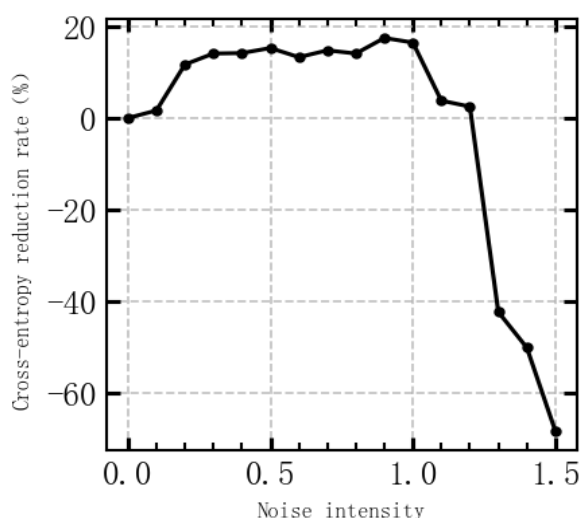


图 4.5 MNIST 训练集噪声增益变化趋势

将 MNIST 训练集交叉熵减少率作为噪声增益的衡量标准,其在不同的噪声强度下的变化趋势如图 4.5 所示。可以观察到随着噪声强度的增强,交叉熵减少率最初呈现逐渐上升的趋势,噪声增益逐渐提高,而当噪声强度超过 0.9 时交叉熵减少率则逐渐下降;当噪声强度大于 1.2 后时 NCNN 在前十次迭代时的平均交叉熵大于无噪声 CNN,此时在模型中加入噪声反而起到负面作用。图 4.5 的实验结果表明模型的最优噪声强度为 0.9,此时噪声增益最大。同时噪声增益随着噪声强度的增加呈现先增后减的趋势,出现了随机共振现象。

### 4.3.2 Fashion-MNIST 数据集上的实验结果与分析

本节通过在 Fashion-MNIST 数据集上进行实验来研究卷积神经网络训练过程中的噪声增益。实验设计使用图 4.1 的网络结构进行模型训练,在对比实验中使用加入噪声的 NCNN 与无噪声 CNN 进行对比。实验中的训练集样本量为 500,初始噪声强度设置为 0.9,模型的初始学习率设置为 0.01,选择训练集交叉熵以及分类错误率作为研究测度。其中 NCNN 模型的反向传播过程中在输出层对标签向量加入满足公式(4.1)的加性均匀噪声,同时采用小批量随机梯度下降算法,在每轮迭代中将样本数据随机分为 25 组进行模型训练。在输出神经元中加入的随机噪声为 $[-0.5\sqrt{N_c/d}, 0.5\sqrt{N_c/d}]$ 区间上的加性均匀噪声,其中 $N_c$ 为噪声强度, $d$ 为迭代次数,即噪声强度随着迭代次数增加而降低。

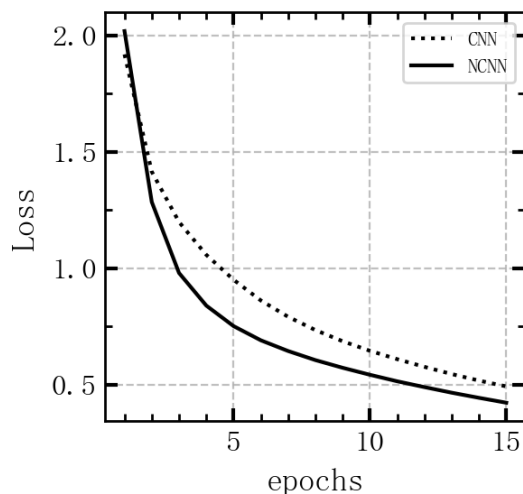


图 4.6 Fashion-MNIST 数据集上的交叉熵对比

在 Fashion-MNIST 数据集上卷积神经网络训练过程中交叉熵与迭代次数的变化关系如图 4.6 所示。可以发现相比于无噪声 CNN，NCNN 在训练集上的交叉熵收敛速度更快，在迭代两次之后 NCNN 在训练集上的交叉熵始终低于无噪声 CNN 的交叉熵。与无噪声 CNN 相比，NCNN 在前 10 次训练迭代中平均减少了 13.2% 的交叉熵，因此 NCNN 具有更好的分类性能，验证了卷积神经网络在 Fashion-MNIST 数据集上也存在噪声增益。

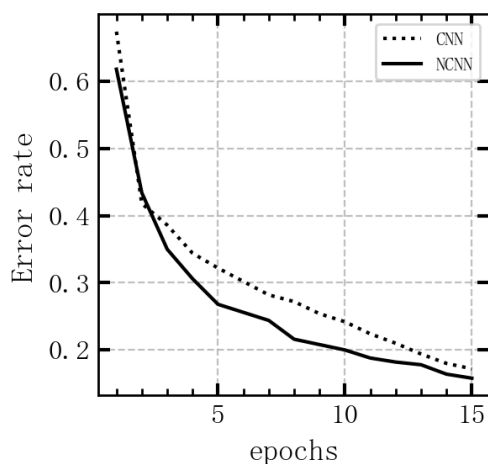


图 4.7 Fashion-MNIST 数据集上的分类错误率对比

在 Fashion-MNIST 数据集上卷积神经网络训练过程中分类错误率与迭代次数的变化关系如图 4.7 所示。从图 4.7 的实验结果可以看到，在模型前几次迭代训练时，NCNN 在训练集上的分类错误率略高于无噪声 CNN 的分类错误率，但随着迭代次数的增加 NCNN 训练集分类错误率快速下降，从第三次迭代开始 NCNN 训练集上的分类错误率始终低于 CNN，在前 10 次训练迭代过程中，NCNN 的平均分类错误率相比无噪声 CNN 模型降低了 2.4%，因此 NCNN 与无噪声 CNN 相比具有更低的分类错误率，说明噪声的加入提高了模型的分类性能。

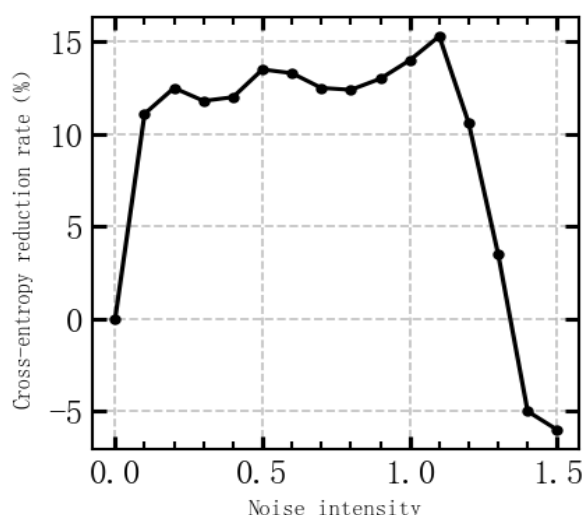


图 4.8 Fashion-MNIST 训练集噪声增益变化趋势

将 Fashion-MNIST 训练集交叉熵减少率作为噪声增益的衡量标准，其在不同的噪声强度下的变化趋势如图 4.8 所示。可以观察到随着噪声强度的增强，交叉熵减少率最初呈现逐渐上升的趋势，噪声增益逐渐提高，而当噪声强度超过 1.1 时交叉熵减少率则逐渐下降；当噪声强度大于 1.3 后时 NCNN 在前十次迭代时的平均交叉熵大于无噪声 CNN，此时在模型中加入噪声反而降低模型性能。图 4.8 的实验结果表明在 Fashion-MNIST 数据集上模型的最优噪声强度为 1.1，此时噪声增益达到最大。同时噪声增益随着噪声强度的增加呈现先增后减的趋势，出现了随机共振现象。

## 4.4 本章小结

本章在已有的卷积神经网络的噪声增益研究基础上提出了一种改进的 NCNN 算法：向输出神经元中加入满足条件的加性均匀噪声，使用 ReLU 函数作为激活函数，同时采用含噪声的小批量随机梯度下降算法更新模型参数，研究卷积神经网络这一非线性系统中可能存在的随机共振现象。最后在 MNIST 和 Fashion-MNIST 数据集上分别进行仿真实验，发现使用 ReLU 函数作为激活函数的卷积神经网络在训练过程中加入噪声，模型的交叉熵与分类错误率有所减少，说明噪声的适当加入对卷积神经网络模型具有正面影响；同时发现随着噪声强度的增加，噪声增益呈现先增后减的趋势。在两个不同的数据集上的实验均能观察到卷积神经网络训练过程中存在的随机共振现象，这一结果证明使用 ReLU 激活函数的卷积神经网络训练过程中也存在噪声增益。

## 第五章 不同类型噪声下的卷积神经网络随机共振研究

目前随机共振理论在微弱信号检测、特征信息提取和生物医学等领域中都有应用的场景。在以往关于经典随机共振系统模型的研究中,大多是在研究非线性系统中存在加性高斯白噪声的情况,而随着随机共振理论的发展,关于随机共振中的噪声类型研究也从传统的高斯噪声拓展到非高斯噪声<sup>[58]</sup>以及有色噪声<sup>[59]</sup>的情况。Mitaim 等人<sup>[60]</sup>研究了无记忆阈值神经元中的随机共振现象,发现多种类型的噪声均可以改善无记忆阈值神经元处理信号的方式并最大化互信息。2016 年 Liu 等人<sup>[61]</sup>研究了基于加性高斯噪声和四种不同乘性噪声的单阈值系统中信号检测的随机共振现象。目前,在卷积神经网络中的随机共振研究多是在加性均匀噪声的条件下进行的,其噪声类型较为单一,而对于加性非均匀噪声类型的研究则较为缺乏。因此,本章考虑针对 MNIST 与 Fashion-MNIST 两个数据集的图像分类问题,在改进的 NCNN 算法基础上,向输出神经元加入常见的非均匀噪声如高斯噪声、拉普拉斯噪声和柯西噪声,以无噪声和均匀噪声下的情况作为基准比较并研究不同类型加性噪声下的卷积神经网络噪声增益。

### 5.1 常见的噪声类型

#### 5.1.1 高斯噪声

高斯噪声的分布是一种关于均值对称的概率分布<sup>[63]</sup>,若噪声服从均值为 0,标准差为  $\sigma_n$  的高斯分布,该噪声对应的概率密度函数表达式如下:

$$f_n(x) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{x^2}{2\sigma_n^2}\right), -\infty < x < \infty, \quad (5.3)$$

其对应的分布函数的表达式如下:

$$F_n(x) = \int_{-\infty}^x f_n(t)dt = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma_n}\right)\right), -\infty < x < \infty. \quad (5.4)$$

服从均值为 0,标准差为 1 的高斯分布噪声的概率密度曲线如图 5.1 所示:



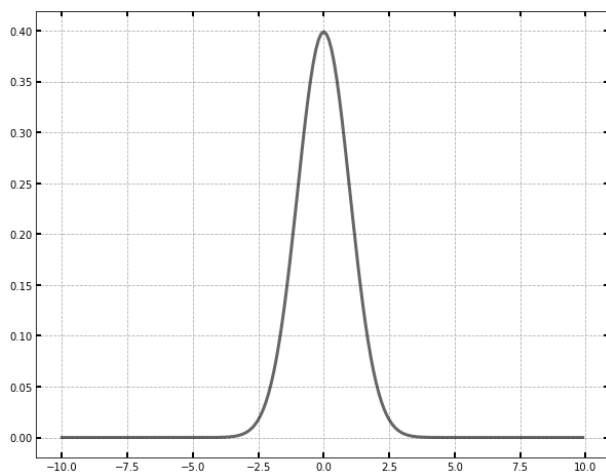


图 5.1 均值为 0，标准差为 1 的高斯噪声概率密度曲线

### 5.1.2 拉普拉斯噪声

拉普拉斯噪声的分布也是一种关于均值对称的概率分布<sup>[64]</sup>。与正态分布相比，拉普拉斯分布的尾端更加平坦，因此拉普拉斯分布出现极端值的概率会大于正态分布。若噪声服从均值为 0，标准差为  $\sigma_n$  的拉普拉斯分布，该噪声对应的概率密度函数表达式如下：

$$f_n(x) = \frac{1}{\sqrt{2}\sigma_n} \exp\left(-\frac{\sqrt{2}|x|}{\sigma_n}\right), -\infty < x < \infty, \quad (5.5)$$

其对应的分布函数的表达式如下：

$$F_n(x) = \int_{-\infty}^x f_n(t) dt = \begin{cases} \frac{1}{2} \exp\left(\frac{\sqrt{2}x}{\sigma_n}\right), & x < 0 \\ 1 - \frac{1}{2} \exp\left(-\frac{\sqrt{2}x}{\sigma_n}\right), & x > 0 \end{cases}. \quad (5.6)$$

服从均值为 0，标准差为 1 的拉普拉斯分布噪声的概率密度曲线如图 5.2 所示：

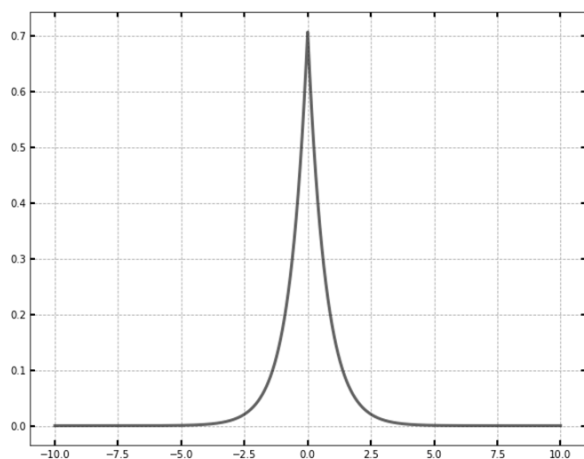


图 5.2 均值为 0，标准差为 1 的拉普拉斯噪声概率密度曲线

### 5.1.3 柯西噪声

服从柯西分布的噪声不存在均值与方差，同时噪声变量的取值范围很广。与正态分布相比，服从柯西分布的随机变量会有更高的概率可以取到极端值，因而柯西分布也被称为重尾分布<sup>[65]</sup>。

若噪声服从位置参数为 0，尺度参数为  $\gamma_n$  的柯西分布，该噪声对应的概率密度函数表达式如下：

$$f_n(x) = \frac{1}{\pi} \left[ \frac{\gamma_n}{x^2 + \gamma_n} \right], -\infty < x < \infty, \quad (5.7)$$

其对应的分布函数的表达式如下：

$$F_n(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{\gamma_n}\right), -\infty < x < \infty. \quad (5.8)$$

服从位置参数为 0，尺度参数为 1 的标准柯西分布噪声的概率密度曲线如图 5.3 所示：

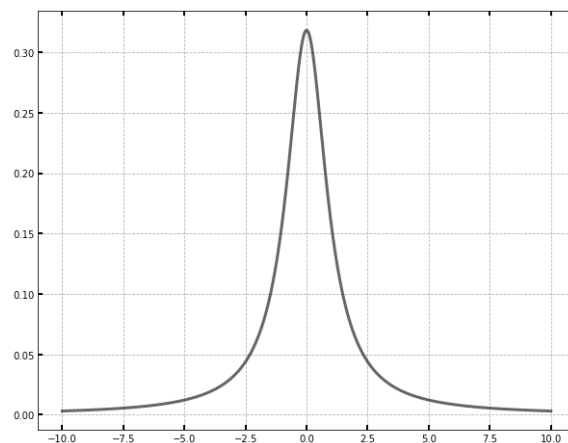


图 5.3 位置参数为 0，尺度参数为 1 的标准柯西分布噪声的概率密度曲线

## 5.2 实验设计

为了研究不同类型噪声下卷积神经网络训练过程中的噪声增益，本章实验在 MNIST 和 Fashion-MNIST 两个数据集上进行。实验设计使用包含两个卷积层和一个池化层的网络结构在两个数据集上进行实验，在对比实验中将无噪声 CNN 与加入加性均匀噪声的 NCNN 作为基准，与加入不同类型加性噪声的 NCNN 进行对比。实验中训练集样本量为 500，加性均匀噪声强度设置为 0.9，模型的初始学习率设置为 0.01，选择训练集交叉熵以及分类准确率作为研究测度。其中 NCNN 模型的反向传播过程中在输出神经元对标签向量加入满足公式 (4.1) 的加性均匀噪声，同时采用小批量随机梯度下降算法，在每轮迭代中将样本数据随机分为 25

组进行模型训练。输出层中加入的噪声是均值为 0，标准差为  $N_c/d$  的高斯噪声、拉普拉斯噪声和位置参数为 0，尺度参数为  $N_c/d$  的柯西噪声，其中  $N_c$  为噪声强度， $d$  为迭代次数。

## 5.3 MNIST 数据集上的实验结果与分析

### 5.3.1 高斯噪声

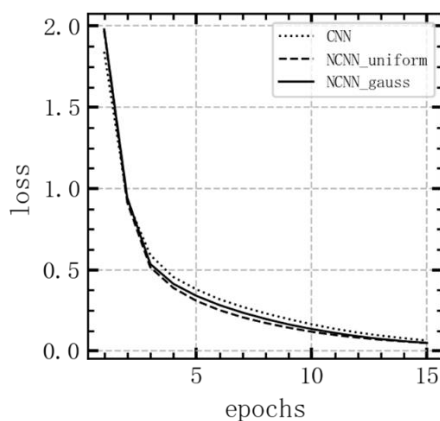


图 5.4 MNIST 数据集上的交叉熵对比

在 MNIST 数据集上卷积神经网络训练过程中不同类型噪声的交叉熵变化曲线如图 5.4 所示。可以发现在迭代三次之后 NCNN 的交叉熵曲线始终处于 CNN 的交叉熵曲线下方，与标准的 CNN 相比，加入高斯噪声的 NCNN 在前 10 次训练迭代中平均减少了 15.3% 的交叉熵。因此在本章实验条件下，加入高斯噪声的模型同样产生了噪声增益，实验结果表明加入高斯噪声可以提高模型训练过程中的交叉熵收敛速度并提升模型的分类性能。加入高斯噪声的 NCNN 在训练过程中的交叉熵与加入均匀噪声的 NCNN 相近，证明在本章实验条件下高斯噪声带来的噪声增益接近均匀噪声的噪声增益。

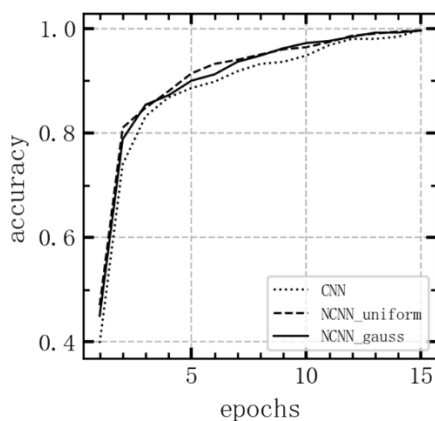


图 5.5 MNIST 数据集上的分类准确率对比

在 MNIST 数据集上卷积神经网络训练过程中不同类型噪声分类准确率与迭代次数的变

化关系如图 5.5 所示。根据图 5.5 可知, 在实验的迭代过程中, 加入噪声的 NCNN 在训练集上的分类准确率始终大于无噪声 CNN 的分类准确率, 在前 10 次训练迭代过程中, 加入高斯噪声的 NCNN 的平均分类准确率相较无噪声 CNN 提高了 2.9%, 因此以平均分类准确率为测度, 加入高斯噪声的模型同样产生了噪声增益, 实验结果表明加入高斯噪声的 NCNN 与无噪声 CNN 相比具有更高的分类准确率, 高斯噪声的加入同样提高了模型在 MNIST 数据集上的分类性能。加入高斯噪声的 NCNN 在前 10 次训练迭代中的平均分类准确率与加入均匀噪声的 NCNN 之间的差距只有 0.2%, 同样证明了在本章实验条件下高斯噪声带来的噪声增益接近均匀噪声的噪声增益。

### 5.3.2 拉普拉斯噪声

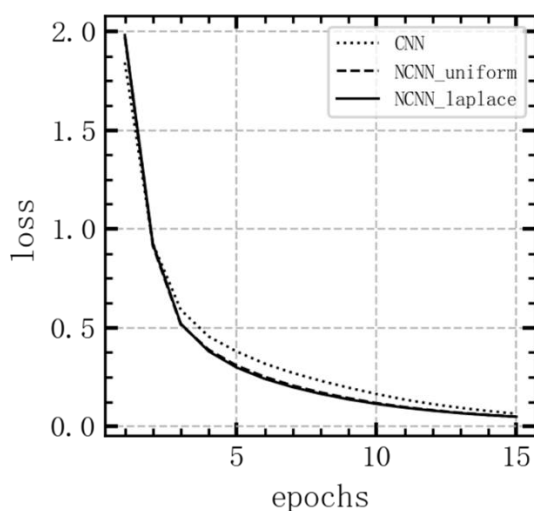


图 5.6 MNIST 数据集上的交叉熵对比

在 MNIST 数据集上卷积神经网络训练过程中不同类型噪声的交叉熵变化曲线如图 5.6 所示。可以发现相较于无噪声 CNN, 加入噪声的 NCNN 在训练集上的交叉熵收敛速度更快且两者的交叉熵曲线很接近, 在迭代三次之后 NCNN 的交叉熵曲线始终处于 CNN 的交叉熵曲线下, 与标准的 CNN 相比, 加入拉普拉斯噪声的 NCNN 在前 10 次训练迭代中平均减少了 18.3% 的交叉熵。因此在本章实验条件下, 加入拉普拉斯噪声的模型同样产生了噪声增益, 实验结果表明加入拉普拉斯噪声可以提高模型训练过程中的交叉熵收敛速度并提升模型的性能。此外, 加入拉普拉斯噪声的 NCNN 在训练过程中的交叉熵与加入均匀噪声的 NCNN 相比相差不大, 证明在本章实验条件下拉普拉斯噪声带来的噪声增益与均匀噪声的噪声增益相近。

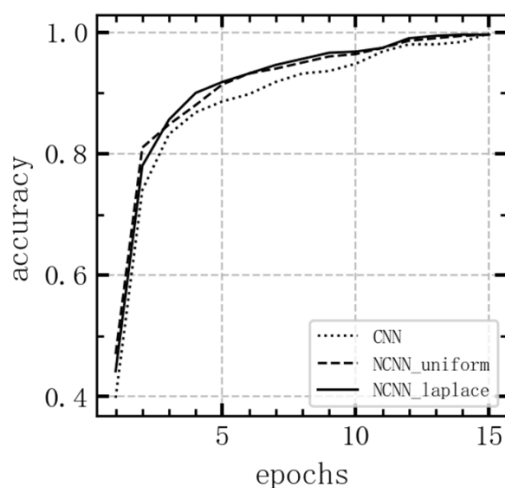


图 5.7 MNIST 数据集上的分类准确率对比

在 MNIST 数据集上卷积神经网络训练过程中不同类型噪声分类准确率与迭代次数的变化关系如图 5.7 所示。根据图 5.7 可知，在实验的迭代过程中，加入噪声的 NCNN 在训练集上的分类准确率始终高于无噪声 CNN 的分类准确率，在前 10 次训练迭代过程中，加入拉普拉斯噪声的 NCNN 的平均分类准确率相较无噪声 CNN 提高了 3.1%，因此以平均分类准确率为测度，加入拉普拉斯噪声的模型同样产生了噪声增益，实验结果表明加入拉普拉斯噪声的 NCNN 与无噪声 CNN 相比具有更高的分类准确率，拉普拉斯噪声的加入同样提高了模型在 MNIST 数据集上的分类性能。此外，加入拉普拉斯噪声的 NCNN 在前 10 次训练迭代中的平均分类准确率与加入均匀噪声的 NCNN 相同，同样证明了在本章实验条件下拉普拉斯噪声带来的噪声增益与均匀噪声的噪声增益相等。

### 5.3.3 柯西噪声

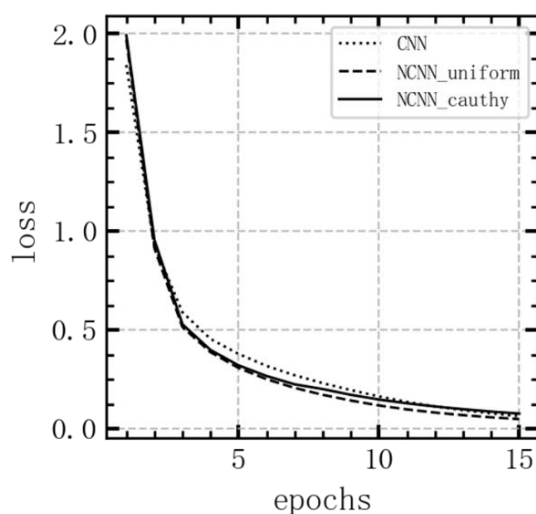


图 5.8 MNIST 数据集上的交叉熵对比

在 MNIST 数据集上卷积神经网络训练过程中不同类型噪声的交叉熵变化曲线如图 5.8 所示。可以发现相较于无噪声 CNN，加入噪声的 NCNN 在训练初期的训练集交叉熵收敛速度更快，与标准的 CNN 相比，加入柯西噪声的 NCNN 在前 10 次训练迭代中平均减少了 9.4% 的交叉熵，但随着迭代次数增加，加入柯西噪声的 NCNN 交叉熵收敛速度逐渐变慢，在模型训练完毕时的交叉熵反而高于 CNN。因此在本章实验条件下，加入柯西噪声的模型在模型训练不充分时存在一定噪声增益，实验结果表明加入柯西噪声可以降低模型训练初期过程中的交叉熵。与加入均匀噪声的 NCNN 相比，加入柯西噪声的 NCNN 模型在训练过程中的交叉熵更高，证明在本章实验条件下柯西噪声带来的噪声增益低于均匀噪声的噪声增益。

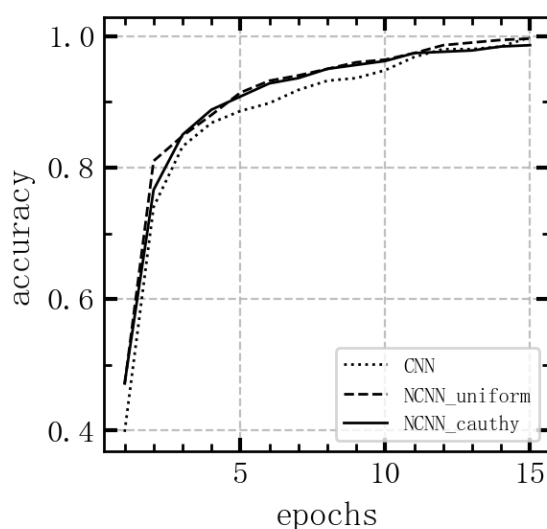


图 5.9 MNIST 数据集上的分类准确率对比

在 MNIST 数据集上卷积神经网络训练过程中不同类型噪声分类准确率与迭代次数的变化关系如图 5.9 所示。根据图 5.9 可知，在实验的迭代过程中，加入柯西噪声的 NCNN 的训练集分类准确率在训练初期高于无噪声 CNN 的分类准确率，但随着迭代次数增加，加入柯西噪声的 NCNN 模型收敛速度逐渐变慢，在模型训练完毕时的分类准确率反而低于 CNN，在前 10 次训练迭代过程中，加入柯西噪声的 NCNN 的平均分类准确率相较无噪声 CNN 提高了 2.6%，因此以平均分类准确率为测度，加入柯西噪声的模型在模型训练不充分时也具有一定噪声增益，实验结果表明加入柯西噪声可以提高模型训练初期过程中的分类准确率。此外，加入柯西噪声的 NCNN 在前 10 次训练迭代中的平均分类准确率与加入均匀噪声的 NCNN 相比降低了 0.5%，证明了以平均分类准确率为测度时柯西噪声带来的噪声增益同样低于均匀噪声的噪声增益。

5.3.4 不同噪声的增益比较

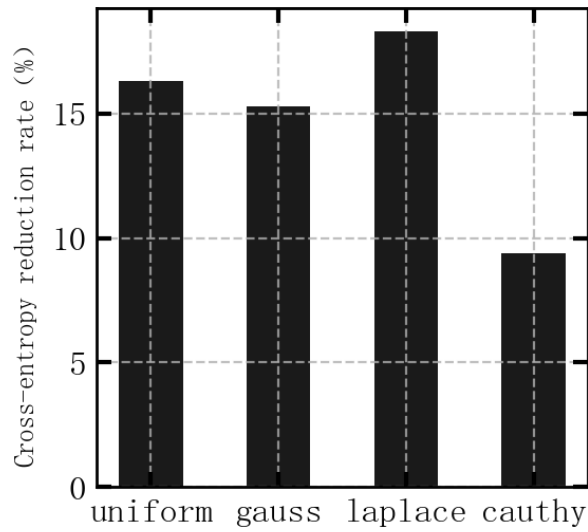


图 5.10 MNIST 数据集上不同噪声的增益比较

图 5.10 表示 MNIST 数据集上不同噪声的增益比较，可以看到在本章的实验条件下，拉普拉斯噪声的噪声增益最高，与无噪声模型相比其在训练集前 10 次迭代过程中的平均交叉熵降低了 18.3%。而柯西噪声的噪声增益最低，平均交叉熵仅降低了 9.4%。

5.4 Fashion-MNIST 数据集上的实验结果与分析

5.4.1 高斯噪声

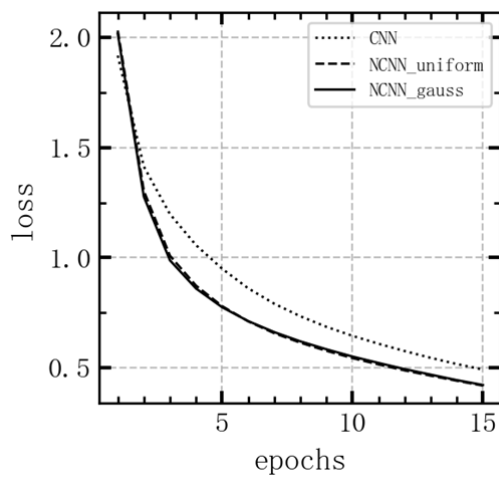


图 5.11 Fashion-MNIST 数据集上的交叉熵对比

在 Fashion-MNIST 数据集上卷积神经网络训练过程中不同类型噪声的交叉熵变化曲线如图 5.11 所示。可以发现相较于无噪声 CNN，加入噪声的 NCNN 在训练集上的交叉熵收敛速

度都会更快,在迭代两次之后 NCNN 的交叉熵曲线始终处于 CNN 的交叉熵曲线下方,与标准的 CNN 相比,加入高斯噪声的 NCNN 在前 10 次训练迭代中平均减少了 13.7% 的交叉熵。因此在本章实验条件下,加入高斯噪声的模型同样产生了噪声增益,实验结果表明加入高斯噪声可以提高模型训练过程中的交叉熵收敛速度并提升模型的性能。此外,可以看到加入高斯噪声的 NCNN 在训练过程中的交叉熵与加入均匀噪声的 NCNN 交叉熵十分接近,证明在本章实验条件下高斯噪声带来的噪声增益与均匀噪声的噪声增益相近。

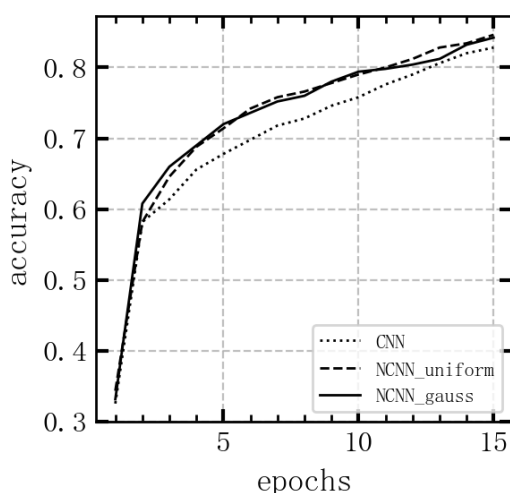


图 5.12 Fashion-MNIST 数据集上的分类准确率对比

在 Fashion-MNIST 数据集上卷积神经网络训练过程中不同类型噪声分类准确率与迭代次数的变化关系如图 5.12 所示。根据图 5.12 可知,在实验的迭代过程中,加入噪声的 NCNN 在训练集上的分类准确率始终高于无噪声 CNN 的分类准确率,在前 10 次训练迭代过程中,加入高斯噪声的 NCNN 的平均分类准确率相较无噪声 CNN 提高了 3.2%,因此以平均分类准确率为测度,加入高斯噪声的模型同样产生了噪声增益,实验结果表明加入高斯噪声的 NCNN 与无噪声 CNN 相比具有更高的分类准确率,高斯噪声的加入同样提高了模型在 Fashion-MNIST 数据集上的分类性能。加入高斯噪声的 NCNN 在前 10 次训练迭代中的平均分类准确率与加入均匀噪声的 NCNN 相差不大,同样证明了在本章实验条件下高斯噪声带来的噪声增益接近于均匀噪声的噪声增益。



### 5.4.2 拉普拉斯噪声

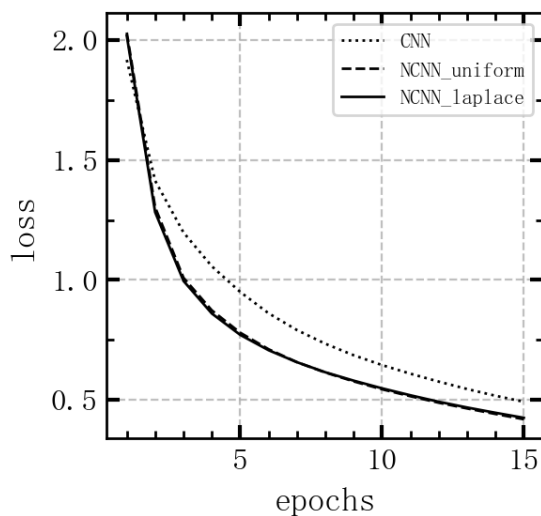


图 5.13 Fashion-MNIST 数据集上的交叉熵对比

在 Fashion-MNIST 数据集上卷积神经网络训练过程中不同类型噪声的交叉熵变化曲线如图 5.13 所示。可以发现相较于无噪声 CNN，加入噪声的 NCNN 在训练集上的交叉熵收敛速度更快且两者的交叉熵曲线很接近，在迭代三次之后 NCNN 的交叉熵曲线始终处于 CNN 的交叉熵曲线下方，与标准的 CNN 相比，加入拉普拉斯噪声的 NCNN 在前 10 次训练迭代中平均减少了 13.9% 的交叉熵。因此在本章实验条件下，加入拉普拉斯噪声的模型同样产生了噪声增益，实验结果表明加入拉普拉斯噪声可以提高模型训练过程中的交叉熵收敛速度并提升模型的性能。此外，加入拉普拉斯噪声的 NCNN 在训练过程中的交叉熵与加入均匀噪声的 NCNN 相比差距不大，证明在本章实验条件下拉普拉斯噪声带来的噪声增益与均匀噪声的噪声增益相近。

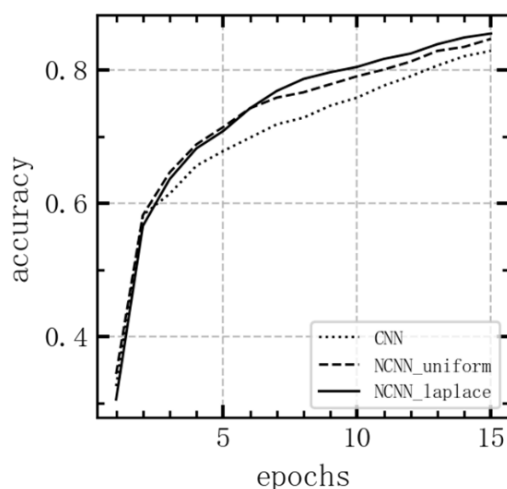


图 5.14 Fashion-MNIST 数据集上的分类准确率对比

在 Fashion-MNIST 数据集上卷积神经网络训练过程中不同类型噪声分类准确率与迭代次数的变化关系如图 5.14 所示。根据图 5.14 可知，在三次迭代之后，加入噪声的 NCNN 在训练集上的分类准确率始终高于无噪声 CNN 的分类准确率，在前 10 次训练迭代过程中，加入拉普拉斯噪声的 NCNN 的平均分类准确率相较无噪声 CNN 提高了 2.9%，因此以平均分类准确率为测度，加入拉普拉斯噪声的模型同样产生了噪声增益，实验结果表明加入拉普拉斯噪声的 NCNN 与无噪声 CNN 相比具有更高的分类准确率，拉普拉斯噪声的加入同样提高了模型在 Fashion-MNIST 数据集上的分类性能。此外，加入拉普拉斯噪声的 NCNN 在前 10 次训练迭代中的平均分类准确率与加入均匀噪声的 NCNN 接近，同样证明了在本章实验条件下拉普拉斯噪声带来的噪声增益与均匀噪声的噪声增益相近。

### 5.4.3 柯西噪声

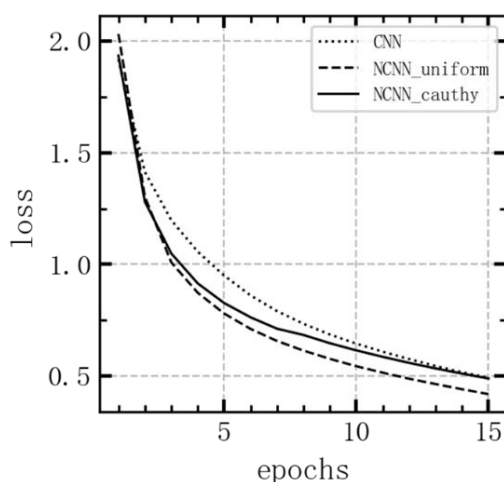


图 5.15 Fashion-MNIST 数据集上的交叉熵对比

在 Fashion-MNIST 数据集上卷积神经网络训练过程中不同类型噪声的交叉熵变化曲线如图 5.15 所示。可以发现相较于无噪声 CNN，加入噪声的 NCNN 在训练初期的训练集交叉熵收敛速度更快，与标准的 CNN 相比，加入柯西噪声的 NCNN 在前 10 次训练迭代中平均减少了 8.5% 的交叉熵，但随着迭代次数增加，加入柯西噪声的 NCNN 交叉熵收敛速度逐渐变慢，在模型训练完毕时的交叉熵接近于 CNN。因此在本章实验条件下，加入柯西噪声的模型在模型训练不充分时存在一定噪声增益，实验结果表明加入柯西噪声可以降低模型训练初期过程中的交叉熵。可以看到与加入均匀噪声的 NCNN 相比，加入柯西噪声的 NCNN 模型在训练过程中的交叉熵更高，证明在本章实验条件下柯西噪声带来的噪声增益低于均匀噪声的噪声增益。

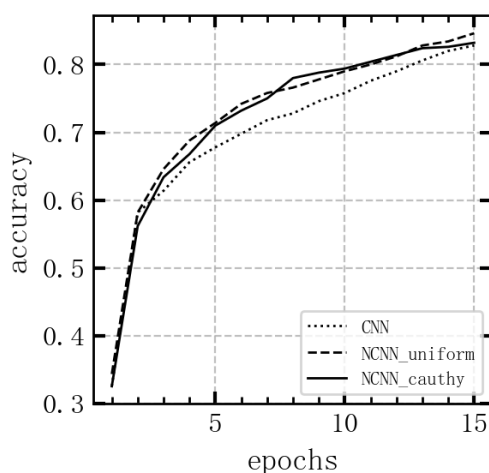


图 5.16 Fashion-MNIST 数据集上的分类准确率对比

在 Fashion-MNIST 数据集上卷积神经网络训练过程中不同类型噪声分类准确率与迭代次数的变化关系如图 5.16 所示。根据图 5.16 可知,在实验的迭代过程中,加入柯西噪声的 NCNN 的训练集分类准确率在训练初期高于无噪声 CNN 的分类准确率,但随着迭代次数增加,加入柯西噪声的 NCNN 模型收敛速度逐渐变慢,在模型训练完毕时的分类准确率接近于 CNN,在前 10 次训练迭代过程中,加入柯西噪声的 NCNN 的平均分类准确率相较无噪声 CNN 提高了 2.4%,因此以平均分类准确率为测度,加入柯西噪声的模型在模型训练不充分时也具有一定噪声增益,实验结果表明加入柯西噪声可以提高模型训练初期过程中的分类准确率。此外,加入柯西噪声的 NCNN 在前 10 次训练迭代中的平均分类准确率与加入均匀噪声的 NCNN 相比降低了 0.6%,证明了以平均分类准确率为测度时柯西噪声带来的噪声增益同样低于均匀噪声的噪声增益。

#### 5.4.4 不同噪声的增益比较

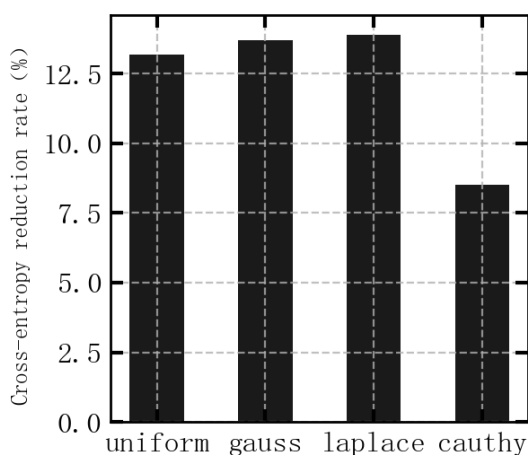


图 5.17 Fashion-MNIST 数据集上不同噪声的增益比较

图 5.17 表示 Fashion-MNIST 数据集上不同噪声的增益比较,可以看到在本章的实验条件下,拉普拉斯噪声的噪声增益最高,与无噪声模型相比其在训练集前 10 次迭代过程中的平均交叉熵降低了 13.9%。而柯西噪声的噪声增益最低,平均交叉熵仅降低了 8.5%。

## 5.5 本章小结

本章选取 MNIST 与 Fashion-MNIST 两个不同数据集分别进行实验,研究不同类型加性噪声下的卷积神经网络随机共振现象:以交叉熵和分类准确率作为评价指标,在改进的 NCNN 算法基础上,向输出神经元分别加入高斯噪声、拉普拉斯噪声和柯西噪声,以无噪声情况与均匀噪声情况为基准进行实验结果的比较与分析。其中在同一数据集上进行的实验仅改变噪声类型,所用的模型结构与模型初始参数均相同,加入的噪声强度随着迭代次数的增加而逐渐降低。

在 MNIST 和 Fashion-MNIST 数据集上的实验结果表明,高斯噪声、拉普拉斯噪声、柯西噪声与均匀噪声类似,都能在 NCNN 的训练过程中产生随机共振现象,其在交叉熵与分类准确率两个评价指标上的表现均优于无噪声 CNN。此外在本文实验条件下,使用高斯噪声和拉普拉斯噪声的模型在评价指标上的表现均与使用均匀噪声的 NCNN 相接近,而柯西噪声产生的噪声增益综合来看低于其他三种噪声的噪声增益。

## 第六章 总结与展望

### 6.1 总结

在线性系统中，人们通常认为噪声是对系统性能有负面影响的存在，需要尽量消除其中的噪声才能获得更多的有效信息。但是随机共振理论则指出在一些非线性系统中，适量的噪声反而可能会提高系统的性能。而卷积神经网络作为实际应用广泛的一种非线性系统，研究其中的噪声增益与随机共振现象就成为了本文的关注重点。本文首先详细介绍了选题背景和意义，接着叙述了国内外卷积神经网络及随机共振的研究历程与发展现状并对卷积神经网络和经典的随机共振理论进行了详细的阐述。然后本文分析了三种不同的激活函数对卷积神经网络模型的影响以及传统激活函数所存在的缺陷，通过仿真实验，研究使用 ReLU 函数以及另外两种传统的激活函数的模型在分类准确率和交叉熵方面的效果。之后，本文在已有的卷积神经网络的噪声增益研究基础上，根据随机共振理论提出了一种改进的 NCNN 算法来进一步研究卷积神经网络中可能存在的噪声增益，同时在数据集上进行改进算法的仿真实验，发现使用 ReLU 函数作为激活函数的卷积神经网络在训练过程中加入噪声可以减少模型的交叉熵与分类错误率，验证了使用改进算法的卷积神经网络训练过程中存在噪声增益。最后，将改进的 NCNN 算法中使用的噪声类型由加性均匀噪声拓展到加性高斯噪声、加性拉普拉斯噪声和加性柯西噪声，研究不同噪声类型下模型的噪声增益，在 MNIST 数据集和 Fashion-MNIST 数据集上的仿真实验结果表明，使用三种非均匀噪声的模型均存在不同程度的噪声增益。

具体的研究内容如下：

(1) 以分类准确率和交叉熵作为测度，通过在 MNIST 和 Fashion-MNIST 两个数据集上对三种激活函数进行了对比实验。在 MNIST 数据集的实验中，使用 ReLU 函数的模型在训练集、验证集和测试集上均具有最优的分类效果；在 Fashion-MNIST 数据集的实验中，使用 ReLU 函数的模型在验证集和测试集上具有最优的分类效果。综合来看使用 ReLU 函数的模型在交叉熵与分类准确率方面具有比较好的表现，同时 ReLU 函数在提升模型的泛化性能和平均训练耗时方面具有不可替代的优势。

(2) 在已有的卷积神经网络的噪声增益研究基础上，根据随机共振理论提出了一种改进的 NCNN 算法：向输出神经元中加入满足条件的加性均匀噪声，使用 ReLU 函数作为激活函数，同时采用含噪声的小批量随机梯度下降算法更新模型参数，研究卷积神经网络这一非线性

性系统中可能存在的噪声增益。在 MNIST 和 Fashion-MNIST 数据集上分别进行算法的仿真实验,发现使用 ReLU 函数作为激活函数的卷积神经网络在训练过程中加入噪声,模型的交叉熵与分类错误率有所减少;同时发现随着噪声强度的增加,噪声增益呈现先增后减的趋势。在两个不同的数据集上进行的实验均能观察到卷积神经网络训练过程中存在的随机共振现象,验证了使用 ReLU 激活函数的卷积神经网络训练过程中同样存在噪声增益。

(3) 将改进的 NCNN 算法中使用的噪声类型由加性均匀噪声拓展到加性高斯噪声、加性拉普拉斯噪声和加性柯西噪声,研究不同噪声类型下模型的噪声增益。在 MNIST 和 Fashion-MNIST 数据集上的实验结果表明,高斯噪声、拉普拉斯噪声、柯西噪声与均匀噪声类似,都能在 NCNN 的训练过程中产生随机共振现象,其在交叉熵与分类准确率两个评价指标上的表现均优于无噪声 CNN。此外在本文实验条件下,使用高斯噪声和拉普拉斯噪声的模型在评价指标上的表现均与使用均匀噪声的 NCNN 相接近,而柯西噪声产生的噪声增益综合来看低于其他三种噪声的噪声增益。

## 6.2 展望

与卷积神经网络中的噪声增益有关的研究目前仍然处于起步阶段,在理论创新和实际应用等方面还有许多的问题值得进一步的思考与研究。受本人的研究时间与学术水平所限,本文只对以上的几点内容进行了研究,在此基础之上还可以从以下几个角度进行深入的探讨:

(1) 本文仅考虑了不同激活函数和噪声类型下的卷积神经网络噪声增益,而已有的研究表明样本量同样会对模型的噪声增益产生影响,因此可以考虑在改进的 NCNN 研究基础之上,讨论样本量大小的变化对模型噪声增益产生的影响。

(2) 本文仅讨论了四种不同类型的加性噪声在加入输出层的神经元情况下产生的噪声增益,对卷积层、池化层和全连接层中加入不同类型的加性噪声是否存在噪声增益还缺乏进一步的讨论。因此之后可以考虑向卷积神经网络的其他层中加入噪声并研究其中的噪声增益。

(3) 本文仅研究了基于卷积神经网络图像分类的随机共振现象,而卷积神经网络也常应用于自然语言处理和语音信号处理等任务,噪声对其他类型的任务具有怎样的影响也值得研究。此外,除了卷积神经网络,还有循环神经网络以及长短期记忆网络等神经网络模型也是非线性系统,在其他神经网络模型中是否存在随机共振现象也有待进一步讨论。

## 参考文献

- [1] 张弛, 郭媛, 黎明. 人工神经网络模型发展及应用综述[J]. 计算机工程与应用, 2021, 57(11): 57-69.
- [2] 陈耀文, 谭晓玲. 基于注意力机制的卷积神经网络在图像分类中的应用[J]. 科学技术创新, 2021(34): 166-169.
- [3] Liu Z, Waqas M, Yang J. A multi-task CNN for maritime target detection[J]. IEEE Signal Processing Letters, 2021 (28): 434-438.
- [4] 靳晶晶, 王佩. 基于卷积神经网络的图像识别算法研究[J]. 通信与信息技术, 2022(2): 76-81.
- [5] Mortazi A, Bagci U. Automatically designing CNN architectures for medical image segmentation[C]. International Workshop on Machine Learning in Medical Imaging, 2018: 98-106.
- [6] 席挺. 基于深度神经网络的图像去噪算法研究与优化[D]. 南京大学, 2021.
- [7] 张荣, 李伟平, 莫同. 深度学习研究综述[J]. 信息与控制, 2018, 47(4): 385-397.
- [8] Benzi R, Sutera A, Vulpiani A. The mechanism of stochastic resonance [J]. Journal of Physics A: Mathematical and General, 1981, 14: 453-457.
- [9] Lecun Y, Bottou L. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [10] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [11] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]. International Conference on Machine Learning (ICML-10), 2010: 807-814.
- [12] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [13] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]. European conference on computer vision. Springer, Cham, 2014: 818-833.
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv Preprint ArXiv, 2014: 1-14.
- [15] C.Szegedy, W.Liu, Y.Jia, et al. Going deeper with convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 1-9.
- [16] 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.
- [17] Fauve S, Heslot F. Stochastic resonance in a bistable system[J]. Physics Letters A, 1983, 97(1): 5-7.
- [18] Mcnamara B, Wiesenfeld K, Roy R. Observation of stochastic resonance in a ring laser[J]. Physical Review Letters, 1988, 60(25): 2626-2629.
- [19] Collins J, Chow C, Imhoff T. Aperiodic stochastic resonance in excitable systems [J]. Physical Review E Statistical Physics Plasmas Fluids & Related Interdisciplinary Topics, 1995, 52(4): R3321.
- [20] Stocks, Nigel G. Suprathreshold stochastic resonance in multilevel threshold systems [J]. Physical Review Letters, 2000, 84(11): 2310-2313.
- [21] 李国芳, 孙晓娟. 小世界神经网络随机共振现象:混合突触和部分时滞的影响[J]. 物理学报, 2017, 66(24): 16-27.
- [22] 王杰, 王友国, 翟其清. 语音信号传输过程中的阈上随机共振现象[J]. 计算机技术与发展, 2021, 31(2): 155-160.
- [23] 郑文秀, 文心怡, 杨威, 姚引娣. 基于混合智能算法的随机共振微弱信号检测[J]. 计算机仿真, 2021, 38(6): 469-474.
- [24] Mitaim, Kosko. Noise benefits in the expectation-maximization algorithm: Nem theorems and models[C]. International Joint Conference on Neural Networks. IEEE, 2011, 3178-3183.

- [25] Audhkhasi K, Osoba O, Kosko B. Noise benefits in backpropagation and deep bidirectional pre-training[C]. The 2013 International Joint Conference on Neural Networks (IJCNN). IEEE, 2013: 1-8.
- [26] Osoba O, Kosko B. Noise-enhanced clustering and competitive learning algorithms[J]. Neural Networks, 2013, 37: 132-140.
- [27] Sampanna R, Mitaim S. Noise benefits in motor imagery classification using ensemble support vector machine[C]. 2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings. IEEE, 2014: 53-56.
- [28] Osoba O, Kosko B. The noisy expectation-maximization algorithm for multiplicative noise injection[J]. Fluctuation and Noise Letters, 2016, 15(1): 1650007.
- [29] Audhkhasi K, Osoba O, Kosko B. Noise-enhanced convolutional neural networks[J]. Neural Networks, 2016, 78: 15-23.
- [30] Brandon, Franzke, Bart, et al. Noise can speed Markov chain Monte Carlo estimation and quantum annealing.[J]. Physical Review E, 2019, 100(5):53309.
- [31] Adigun O, Kosko B. Noise-boosted bidirectional backpropagation and adversarial learning[J]. Neural Networks, 2019, 120:9-31.
- [32] Kosko B. Noise Benefits in Feedback Machine Learning: Bidirectional Backpropagation[C]. Proceedings of the 5th International Conference on Applications in Nonlinear Dynamics. Springer International Publishing, 2019: 267-275.
- [33] Kosko B, Audhkhasi K, Osoba O. Noise can speed backpropagation learning and deep bidirectional pretraining[J]. Neural Networks, 2020, 129: 359-384.
- [34] Duan L, Duan F, Chapeau-Blondeau F, et al. Stochastic resonance in Hopfield neural networks for transmitting binary signals[J]. Physics Letters A, 2020, 384(6): 126143.
- [35] Liu X, Duan L, Duan F, et al. Enhancing threshold neural network via suprathreshold stochastic resonance for pattern classification[J]. Physics Letters A, 2021, 403: 127387.
- [36] 王之仓, 邓伟. 多层感知器学习算法研究[J]. 青海师范大学学报(自然科学版), 2007(1): 37-39.
- [37] 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1): 1-17.
- [38] Seriès Peggy, Lorenceau Jean, Frégnac Yves. The "silent" surround of V1 receptive fields: theory and experiments.[J]. Journal of Physiology, 2003, 97(4-6): 453-474.
- [39] 高友文, 周本君, 胡晓飞. 基于数据增强的卷积神经网络图像识别研究[J]. 计算机技术与发展, 2018, 28(8): 62-65.
- [40] Ito Y. Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory[J]. Neural Networks, 1991, 4(3): 385-394.
- [41] Fan E. Extended tanh-function method and its applications to nonlinear equations[J]. Physics Letters A, 2000, 277(4-5): 212-218.
- [42] 李旭冬, 叶茂, 李涛. 基于卷积神经网络的目标检测研究综述[J]. 计算机应用研究, 2017, 34(10): 2881-2886+2891.
- [43] Hecht-Nielsen R. Theory of the backpropagation neural network[M]. Neural Networks for Perception. Academic Press, 1992: 65-93.
- [44] 王功鹏, 段萌, 牛常勇. 基于卷积神经网络的随机梯度下降算法[J]. 计算机工程与设计, 2018, 39(02): 441-445+462.
- [45] 郑仕谱. 基于随机共振的弱信号提取方法研究[D]. 浙江大学, 2014.
- [46] Gammaitoni L, Hänggi P, Jung P, et al. Stochastic resonance[J]. Reviews of Modern Physics, 1998, 70(1): 223.
- [47] Yuen B, Hoang M T, Dong X, et al. Universal activation function for machine learning[J]. Scientific Reports, 2021, 11(1): 18757.



- [48] 刘海, 刘波, 胡瑜. 多斜率自适应卷积神经网络激活函数[J]. 空间控制技术与应用, 2020(3): 28-35.
- [49] Kaiming. H, Xiangyu. Z, Shaoqing. R, et al. Delving Deep into Rectifiers: Surpassing Human Level Performance on ImageNet Classification[C]. International Conference on Computer Vision (ICCV), 2015, 1026-1034.
- [50] Glorot, Bor De S A, Bengio Y. Deep Sparse Rectifier Neural Networks[C]. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS), 2011: 315-323.
- [51] Chapados N, Bengio Y, Vincent P, et al. Estimating car insurance premia: A case study in high-dimensional data inference[J]. Advances in Neural Information Processing Systems, 2001. 1369-1376.
- [52] Deng L. The mnist database of handwritten digit images for machine learning research [J]. IEEE Signal Processing Magazine, 2012, 29(6): 141-142.
- [53] Xiao H, Rasul K, Vollgraf R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms[J]. ArXiv Preprint ArXiv, 2017, 08: 07747.
- [54] De Boer P T, Kroese D P, Mannor S, et al. A tutorial on the cross-entropy method[J]. Annals of Operations Research, 2005, 134: 19-67.
- [55] Moon T K. The expectation-maximization algorithm[J]. IEEE Signal Processing Magazine, 1996, 13(6): 47-60.
- [56] He T, Zhang Z, Zhang H, et al. Bag of tricks for image classification with convolutional neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 558-567.
- [57] 梁杰, 陈嘉豪, 张雪芹. 基于独热编码和卷积神经网络的异常检测[J]. 清华大学学报(自然科学版), 2019, 59(07): 523-529.
- [58] Wu D, Zhu S. Stochastic resonance in a bistable system with time-delayed feedback and non-Gaussian noise[J]. Physics Letters A, 2007, 363(3): 202-212.
- [59] Lu L, Jia Y, Ge M, et al. Inverse stochastic resonance in Hodgkin-Huxley neural system driven by Gaussian and non-Gaussian colored noises[J]. Nonlinear Dynamics, 2020, 100: 877-889.
- [60] Mitaim S, Kosko B. Adaptive stochastic resonance in noisy neurons based on mutual information [J]. IEEE Transactions on Neural Networks, 2004, 15(6): 1526-1540.
- [61] Liu J, Wang Y, Zhai Q. Stochastic resonance of signal detection in mono-threshold system using additive and multiplicative noises[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2016, 99(1): 323-329.
- [62] Ries D T. The influence of noise type and level upon stochastic resonance in human audition[J]. Hearing Research, 2007, 228(1-2): 136-143.
- [63] Galdi V, Pierro V, Pinto I M. Evaluation of stochastic-resonance-based detectors of weak harmonic signals in additive white Gaussian noise[J]. Physical Review E, 1998, 57(6): 6470.
- [64] Kumar S, Jha R K. Characterization of supra-threshold stochastic resonance for uniform distributed signal with laplacian and gaussian noise[C]. 2017 International Conference on Noise and Fluctuations (ICNF). IEEE, 2017: 1-4.
- [65] Kosko B, Mitaim S. Stochastic resonance in noisy threshold neurons[J]. Neural Networks, 2003, 16(5-6): 755-761.

## 附录 1 程序清单

### 第三章 卷积神经网络中的激活函数比较算法程序

`mnist_activationfunction.py`: MNIST 数据集上的激活函数比较

`fashionmnist_activationfunction.py`: Fashion-MNIST 数据集上的激活函数比较

### 第四章 改进的噪声增强卷积神经网络算法程序

`mnist_cnn.py`: MNIST 数据集上的无噪声模型

`mnist_cnn_noise.py`: MNIST 数据集上的均匀噪声模型

`fashionmnist_cnn.py`: Fashion-MNIST 数据集上的无噪声模型

`fashionmnist_cnn_noise.py`: Fashion-MNIST 数据集上的均匀噪声模型

### 第五章 不同加性噪声下的卷积神经网络算法程序

`Disturbtions.py`: 噪声分布图

`mnist_cnn_gauss.py`: MNIST 数据集上的高斯噪声模型

`mnist_cnn_laplace.py`: MNIST 数据集上的拉普拉斯噪声模型

`mnist_cnn_cauthy.py`: MNIST 数据集上的柯西噪声模型

`fashionmnist_cnn_gauss.py`: Fashion-MNIST 数据集上的高斯噪声模型

`fashionmnist_cnn_laplace.py`: Fashion-MNIST 数据集上的拉普拉斯噪声模型

`fashionmnist_cnn_cauthy.py`: Fashion-MNIST 数据集上的柯西噪声模型