

电子科技大学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

学士学位论文

BACHELOR THESIS



论文题目 面向图像分类的衍射光神经网络设计

学科专业 “互联网+”数据科学与大数据技术

学号 2020110801009

作者姓名 程智涛

指导老师 唐泳 讲师，黎升好 副教授

学院 数学科学学院，计算机科学与工程学院

摘要

近年来，随着人工智能的不断发展，神经网络与深度学习逐渐成为人工智能的一个重点研究方向。同时，各种计算机器的计算能力也越来越强，电子计算芯片已经普及，光子计算器件也在不断发展。在此基础上，光学神经网络采用衍射、干涉、散射、光学傅里叶变换和光波分复用等方式，将成为可与电子神经网络能力相当甚至更快的计算方式。

本文主要针对衍射光神经网络的原理及架构进行分析研究，构建由多个衍射板面构成的衍射光神经网络模型。建模后，利用 Python 代码和 Pytorch 框架，在电脑中进行仿真和参数训练计算，最后使用训练好的参数进行图像分类测试。首先，建立原始的衍射光神经网络，在 MNIST, Fashion-MNIST 和 E-MNIST 数据集上分别进行了图像分类任务的训练，在测试集上分别达到了 97.64%, 87.97% 和 89.41% 的准确率。接下来利用非线性光学器件来替代传统电子神经网络中的非线性激活函数原理，仿真中使用了 ReLU 函数对原始衍射光神经网络框架进行优化，在上述三个图像识别分类任务中，新的网络在测试集上分别可以达到 98.88%, 90.15% 和 92.90% 的准确率，进一步验证了非线性激活函数的加入使得网络的学习能力得到了提升。

最后，针对传统衍射光神经网络不能处理彩色图像的问题进行研究，先使用带激活函数的衍射光神经网络对 CIFAR-10 数据集的灰度化图像进行了识别分类的训练和测试，最后准确率仅达到了 46.09%，相比电子学全连接神经网络准确率 50.56% 较差；然后本文设计了两种新型的衍射光神经网络框架，提出了位置复用网络 PDW-D2NNNet 和波分复用网络 MDW-D2NNNet 两种衍射光神经网络，再在 CIFAR-10 数据集上进行了训练和测试，两种网络分别将准确率提升到了 60.85% 和 59.27%。本文提出的模型比传统衍射光神经网络和电子全连接网络效果有较大提升，为衍射光神经网络在彩色图像识别方面发展做出了贡献。但目前结果相比更先进的电子神经网络，如沙漏网络（UNet），残差网络（ResNet）和稠密网络（DenseNet）等还有一定的差距；所以，进一步利用光学衍射原理，设计更加复杂先进的网络结构，是未来研究的一个方向。

关键词：光衍射神经网络，人工智能，图像分类识别，光智能计算，光学信号处理

ABSTRACT

In recent years, with the continuous development of artificial intelligence, neural networks and deep learning have gradually become a key research direction in AI. Simultaneously, the computational power of various computing devices has also been growing stronger, electronic computing chips have become widespread, and photonic computing devices are also evolving. Based on this, optical neural networks, which employ diffraction, interference, scattering, optical Fourier transformation, and optical wavelength division multiplexing, are poised to become a computing method comparable to or even faster than electronic neural networks.

This article primarily focuses on analyzing and studying the principles and architecture of diffractive optical neural networks, constructing a diffractive optical neural network model comprised of multiple diffractive planes. After modeling, simulations and parameter training calculations were conducted on computers using Python code and the Pytorch framework. Finally, image classification tests were performed using the trained parameters. Initially, a primitive diffractive optical neural network was established and trained on image classification tasks on the MNIST, Fashion-MNIST, and E-MNIST datasets, achieving accuracies of 97.64%, 87.97%, and 89.41% on the test sets respectively. Subsequently, nonlinear optical devices were employed to replace the principle of nonlinear activation functions in traditional electronic neural networks. In the simulation, the ReLU function was used to optimize the original diffractive optical neural network framework. In the aforementioned three image recognition and classification tasks, the new network achieved accuracies of 98.88%, 90.15%, and 92.90% on the test sets, further verifying that the addition of nonlinear activation functions has enhanced the network's learning capabilities.

Lastly, addressing the inability of traditional diffractive optical neural networks to process color images, this article first utilized a diffractive optical neural network with activation functions to train and test the grayscale images of the CIFAR-10 dataset. The final accuracy rate only reached 46.09%, which is inferior to the 50.56% accuracy rate of electronic fully connected neural networks. Then, this article designed two novel diffractive optical neural network frameworks, proposing the position division wavelength network (PDW-D2NNet) and the mode division wavelength network (MDW-D2NNNet). After

ABSTRACT

training and testing on the CIFAR-10 dataset, these two networks increased the accuracy rates to 60.85% and 59.27% respectively. The models proposed in this article significantly outperform traditional diffractive optical neural networks and electronic fully connected networks, contributing to the development of diffractive optical neural networks in color image recognition. However, compared to more advanced electronic neural networks such as UNet, ResNet, and DenseNet, there is still a certain gap. Therefore, further utilizing optical diffraction principles to design more complex and advanced network structures is a direction for future research.

Keywords: Diffraction optical neural network, Artificial intelligence, Image classification recognition, Optical intelligent computing, Optical signal processing

目 录

第一章 绪 论	1
1.1 研究背景与意义.....	1
1.2 国内外研究发展现状和未来趋势	2
1.2.1 人工智能发展简述	2
1.2.2 光计算发展简述	4
1.3 本文的主要研究内容与创新.....	10
1.4 实验设备	10
1.5 本文的结构安排.....	10
第二章 神经网络算法原理	11
2.1 神经网络	11
2.2 全连接网络	12
2.3 非线性激活函数.....	13
2.4 训练优化	15
2.5 本章小结	15
第三章 衍射光神经网络的数学建模.....	16
3.1 物理模型	16
3.1.1 亥姆霍兹公-基尔霍夫积分	16
3.1.2 基尔霍夫衍射公式	17
3.1.3 瑞利-索莫非衍射模型.....	17
3.2 数学建模	18
3.2.1 光场分布公式	18
3.2.2 衍射光全连接神经网络推导.....	18
3.3 光神经网络训练及参数优化.....	20
3.4 本章小结	21
第四章 灰度图像分类	22
4.1 灰度图像数据集.....	22
4.2 光学非线性运算	22
4.3 衍射光神经网络结构	24
4.4 灰度图像识别分类实验仿真结果	26
4.5 手写数字及字母识别软件	28

4.6 本章小结	30
第五章 彩色图像分类	33
5.1 彩色图像难点	33
5.2 彩色图像数据集	34
5.3 波分复用衍射光神经网络	34
5.4 位置复用衍射光神经网络	35
5.5 颜色分辨能力	36
5.6 彩色图像识别分类实验仿真结果	38
5.7 本章小结	38
第六章 全文总结与展望	42
6.1 全文总结	42
6.2 未来工作展望	42
致 谢	44
附录 A 相关补充	45
A.1 物理相关定理推导	45
A.2 空间光调制器	45
A.3 电子全连接神经网络设置	45
参考文献	46
外文资料原文	52
外文资料译文	57

第一章 绪论

1.1 研究背景与意义

近年来，随着以电子为计算粒子的半导体材料及相关芯片的发展^[1]，计算机的计算能力大幅度提升；进一步的，以半导体器件为载体，集成电路驱动的计算机中央处理器 CPU (Central processing unit) 和图像处理器 GPU (Graph processing unit) 等相关运算处理器已经有成熟而广泛的应用，强大的计算能力使得人工智能（AI, Artificial intelligence）^[2,3] 的兴起，其大规模网络的训练和模型部署得以实现。神经网络模型与深度学习思路^[3,4] 渐渐成为人工智能领域主要研究和发展方向。各个研究领域，如机器视觉^[5,6]、自然语言处理、语音识别^[7]、生成模型^[8-10] 等都有各类先进网络结构和模型的提出，具有强大的表达和泛化能力，也在可解释性方面有较大发展，并进一步在科学研究、教育教学、工业生产、影视娱乐、交通安全和医疗服务等众多行业中已经有着重大的突破和应用。各种先进的应用模型，如大语言模型^[11]、智能绘画模型^[12,13]、大型图像识别分类分割模型^[14] 等已经走入人们的生活之中。

但“后摩尔时代”的到来让集成电路已经渐渐进入发展的瓶颈期，随着理论研究及制造水平的提升，晶体管大小和集成电路的制程等各种电子器件的尺寸都在不断逼近物理极限。过小之后，电流泄露、信号串扰、电容寄生效应、能耗增大和传输延时等问题难以避免。因此，使用电子计算的集成电路以及相关芯片也逐渐接近瓶颈^[15]。然而人工智能，尤其是深度学习的迅猛发展，对计算能力的需求还在急剧增加。所需付出的巨大成本，以及物理本身的限制导致继续在以电子为基础的计算方式上发展难以满足人工智能发展需求。为了克服上述众多问题，结合目前光学智能计算和光子传输通信等方面的发展，人们渐渐开始使用以光子为计算基础的方法来代替以电子为计算基础的方法达到人工智能对计算能力的要求。

与电子类似，光子作为一种人类科学研究中心常见的粒子，也可以被用来存储、传输和表达信息。但不同于需要导线才能安全稳定传输的电子，光子作为一种玻色子，有着传播速度快，互不干扰，耗能低等优良性质，可以在自由空间甚至空气中稳定传播，且能比电子系统具有更高的带宽，十分适合神经网络运算模型的特点，因此目前光学神经网络 ONN (Optical neural network)^[16,17] 研究广受关注^[18-21]。而在光学神经网络中，衍射光神经网络模型由于其光衍射传播的方式与全连接网络计算的方式类似，能够很好地替代传统全连接神经网络。作为一种以光为载体的新型神经网络，其对图像处理也能达到电子系统中的神经网络的能力，而且速

度更快，耗能更低。

基于上述优点，光学神经网络 ONN 具有巨大的发展潜力，科研人员已经从光的衍射、散射、干涉及光学傅里叶变换等基础理论出发，创造了许多光学神经网络框架，并在材料学方面通过引入具有各种性质的光学晶体，研发了各类光学器件，实现了各种电子神经网络的模块和功能，如全连接层、卷积层、非线性激活函数等，大大提升了 ONN 的学习预测能力和推理精度与速度。因此，融合光计算的人工智能在未来有着巨大的科研应用价值。特别地，光神经网络因其高速、低能耗的特性，将成为未来人工智能领域的重要研究方向。

1.2 国内外研究发展现状和未来趋势

人工智能（AI, Artificial intelligence）的发展是一段波澜壮阔的历程，它从一个模糊的概念逐渐成长为当今科技领域的核心驱动力，与众多的科学领域相结合，助力各个领域的智能化发展，也广泛应用于现实世界之中。光计算则是不同于传统电子系统计算，是一种更加快速、高效且精准的计算方式；如今，算力需求在不断增加，其发展更是为人工智能大型模型计算铺垫未来的道路。这其中人工神经网络（ANN）与光神经网络（ONN）发展历程如图1-1所示。

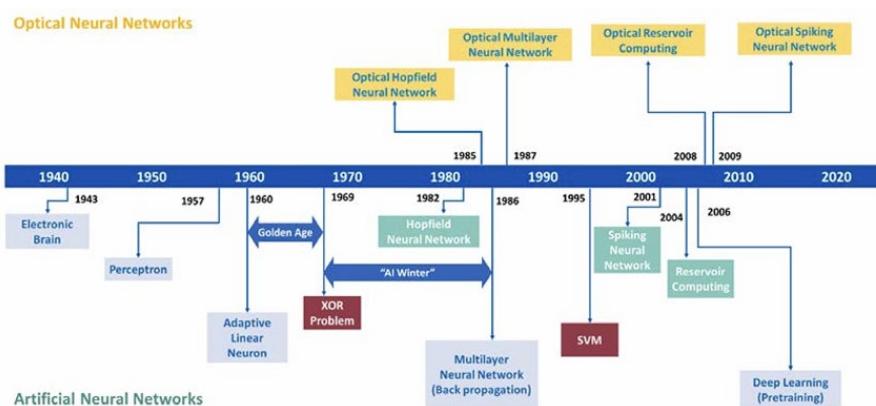


图 1-1 ANN 与 ONN 发展历程 [22]

1.2.1 人工智能发展简述

人工智能最早起源可追溯到 20 世纪 50 年代^[23]。当时，人们开始思考如何制作出类似拥有人类智能的数学模型或机器。1950 年，英国数学家艾伦·图灵提出了图灵测试^[24]，即如果人类与一台可以对话的机器聊天，但最后不能辨别出聊天对象是否是机器，那么这台机器就具有智能。这一测试成为了人工智能领域的重要标准之一。

到了 1956 年，达特茅斯学院的学者们召开了第一次人工智能会议，正式提出了“人工智能”这一概念，标志着人工智能学科的诞生。在此后的十年间，人工智能迎来了发展史上的第一个小高峰。研究者们纷纷涌入这一领域，尝试通过符号主义、连接主义等方法来实现人工智能。最早由科学家结合数学和神经学提出将大脑中神经元工作机制利用电路模拟出来，后来 1957 年又进一步提出了最基本的人工神经元模型“感知器”的模型^[25]。但早期人工智能因为该模型无法有效实现“异或”的计算方式，而陷入了第一次发展低谷。后来，随着误差传播法^[26]解决了多层感知器的训练优化问题，神经网络能以更加快捷方便地构建更复杂的模型。进一步的，随着神经网络的理论发展，深度学习渐渐兴起，各类新型网络模型，如卷积神经网络 (CNN)^[5]、图神经网络 (GNN)^[27]、循环神经网络 (RNN)^[28]、对抗式生成网络^[29] 等渐渐发展起来，使得各类复杂问题有了更好的解决方案。

但在 20 世纪末期，人工智能又再一次渐渐进入了低谷时期，因为当时的计算能力远远不能满足人工智能发展的需求，并且难以收集存储大量可供人工智能学习训练的数据。传统的模型难以解决更复杂的问题，新兴的模型又没有足够的计算环境支撑，导致人工智能的发展举步维艰。而进入 21 世纪后，随着计算机技术的飞速发展，人工智能也迎来了快速发展的转折点。计算芯片的飞速发展，计算能力的大幅度提升，再加上数据存储技术的革新，人类进入了大数据时代，而这也成为了人工智能发展的黄金时代。其中，深度学习技术的崛起为人工智能的发展注入了新的活力。深度学习通过模拟人脑神经网络的工作方式，让机器能够自动学习和识别数据中的特征，从而实现了更高级别的智能。在这一时期，人工智能在自然语言处理^[30]、机器视觉与图像识别^[31]、语音识别^[32,33] 等领域取得了突破性的进展。例如，在图像识别领域，深度学习技术使得机器能够准确地识别出图片中的物体、场景等信息；在语音识别领域，深度学习技术使得机器能够实时地将语音转换为文本；在自然语言处理领域，深度学习技术使得机器能够理解和生成自然语言文本。

随着技术的不断成熟，人工智能的应用场景也日益广泛。目前，人工智能已经广泛应用于自动驾驶、金融服务、智能家居、零售电商、工业制造等领域。在自动驾驶领域^[34]，人工智能通过感知环境、规划路径、控制车辆等方式实现了智能导航和自主驾驶；在金融服务领域^[35,36]，人工智能通过风险评估、投资管理、信用评估等方式提高了金融机构的效率且降低了风险；在智能家居领域^[37]，人工智能通过智能音响、智能家电等设备实现了智能化的居家生活；在零售电商领域^[38]，人工智能通过商品推荐、个性化营销等方式提高了用户的购物体验；在工业制造领域^[39,40]，人工智能通过智能生产、机器人自动化等方式提高了生产效率的同时

也降低了成本。

展望未来，人工智能将继续保持快速发展的态势。随着技术的不断进步和应用场景的不断拓展，人工智能将在更多领域发挥重要作用。同时，也需要关注人工智能带来的挑战和问题，如数据隐私^[41]、就业影响^[42]、伦理道德^[43]等。只有在充分考虑这些问题的前提下，才能更好地推动人工智能的健康发展。

尽管相关概念已经提出了超半个世纪，但人工智能，尤其是深度学习和神经网络在近二十年才有着飞速的发展与广泛的应用，而目前人工智能主要在电子计算机平台上有着较为成熟的实现与应用。光学神经网络 ONN 作为一种基于光子来计算的神经网络，其本质与电子的神经网络类似，有线性运算、非线性激活等模块，也需要进行训练学习来构建网络权重参数。所以利用光学原理、材料和相关器件实现线性运算、非线性激活等模块功能则是实现 ONN 的核心所在。目前网络的训练可以在电脑中仿真完成，未来也可以利用光芯片等方式实现全光上的训练。

1.2.2 光计算发展简述

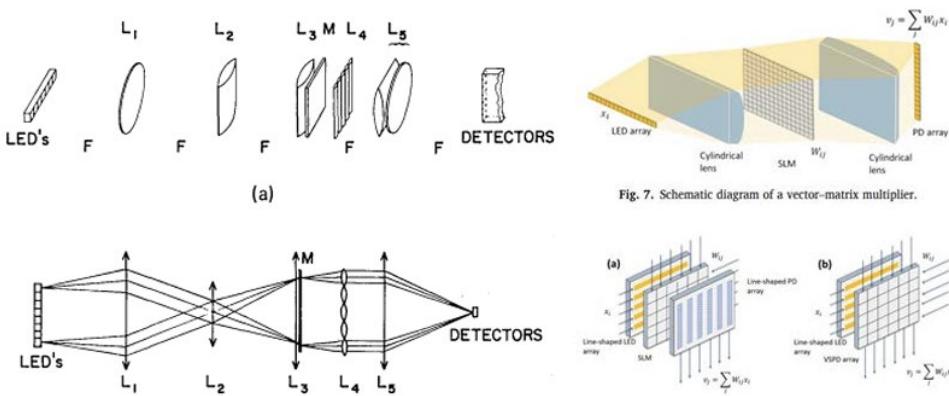


图 1-2 傅里叶变换实现光学计算矩阵乘法^[44,45]

光计算的起源可以追溯到上个世纪 40 年代，当时“现代电子计算机之父”约翰·冯·诺依曼就考虑过使用光学元件实现数字计算的可能性。然而，由于当时光学技术的发展水平较低，光计算的研究并未深入进行。直到 60 年代，激光技术的问世为光计算的研究提供了新的契机。这一时期，科学家们开始研究基于傅里叶光学的模拟光学计算，探索光信号在传播过程中的干涉、衍射等物理现象，并尝试利用这些现象来实现计算功能。在 20 世纪 60 年代左右，科学家^[44,45]最先利用各类光学透镜来模拟傅里叶变换，并采用全光的方式实现了矢量矩阵乘法运算。其中利用了发光二极管阵列作为模拟矩阵的输入，通过透镜焦面处再利用权重模板调制后实现了矢量相乘，最后的线性求和是通过一组透镜实现的。其如图1-2所示。

但电子计算方式的门槛较低，其飞速发展使得早期人们还不太关注光学计算，直到进入 70、80 年代，光学传输和非线性光学材料取得了重大进展，这为光计算的发展提供了有力支持。1985 年，有科研团队^[46]尝试利用光学办法来实现基于 Hopfield 模型的内容可寻址关联存储器，也由此提出了光学神经网络的基本概念。光神经网络的理论概念也随着神经网络理论发展而进一步深化^[47]。与此同时，电子计算机固有的某些缺陷随运算速度的提高而明显的暴露出来，如能耗高、散热难等问题。这使得光计算重新引起了科学家们的广泛注意。在这一时期，数字光计算成为了研究的热点之一。科学家们尝试利用光学元件和光学现象来实现数字信号的传输、调制和处理，以期望实现更低能耗、更高速度的计算。然而，光计算的发展也仍然面临着诸多挑战。首先，光学元件的制造精度和稳定性要求较高，这增加了光计算系统的复杂性和成本。其次，光信号在传输过程中容易受到环境因素的干扰，如温度、湿度等，这会影响光计算系统的稳定性和可靠性。此外，光计算系统的集成度和可扩展性也面临一定的挑战。

到了 1985 年，空间光调制器 (SLM, Spatial light modulator)^[48,49] 的出现使得光学计算调制变得更加方便快捷，其利用两个柱状透镜控制光传播方向，使用光电探测器 (PD, Photoelectric detector) 进行结果探测接收，并转换到电上，供现有电脑读取显示。通过对发光二极管阵列输入光信号的强度进行加乘运算来构建光学网络，最终成功利用光学傅里叶变换实现了神经网络的线性乘法与求和功能。后人首次尝试研发具有记忆和学习功能的光神经网络芯片，通过集成 SLM 和 PD 阵列，构建了可变灵敏度光电探测阵列 (VSPD)。在 1989 年，人们提出了一种基于透镜的可编程光学神经网络结构^[50]，该系统具有良好的光学并行计算能力和权重可编程能力，但因为早期光学器件性能较差，导致其神经元数量较少，难以实现大规模神经网络构建。

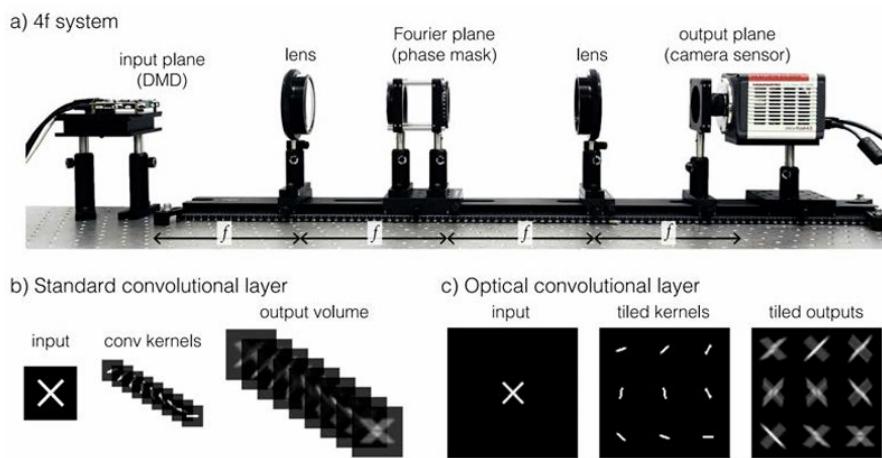


图 1-3 透镜实现光学卷积层^[51]

进入 21 世纪以来，人们的制造水平与精度不断提升，光学技术也在不断进步，并且计算机科学和光学物理学，包括量子物理也在深入发展。这些都使得光计算取得了显著的突破和进展。一方面，科学家们通过不断优化光学元件的设计和制造工艺，提高了光计算系统的稳定性和可靠性。另一方面，他们也开始探索新的光学计算原理和算法，以期望实现更高效、更灵活的计算。光量子计算机、光计算芯片等等设备都在不断研发革新。

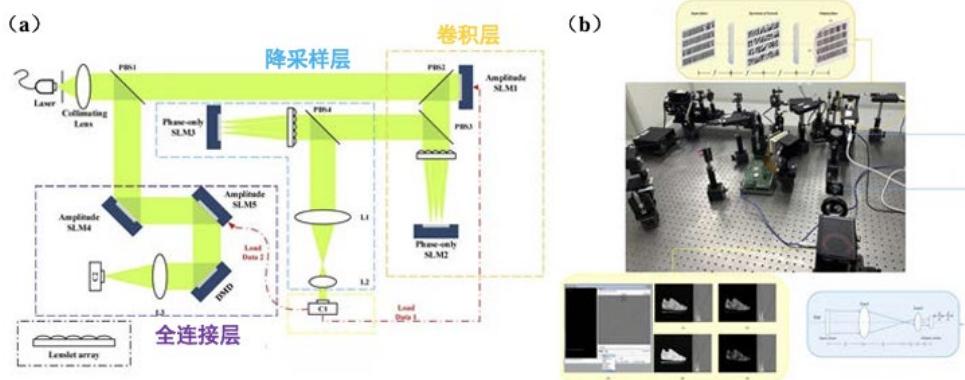


图 1-4 光电卷积神经网络^[52]

除了全连接的线性计算，神经网络中还有卷积这一类重要的运算方式。随着傅里叶光学计算理论与应用渐渐成熟，人们开始思考如何利用其实现光学上的卷积运算。2018 年，人们实现了一种光电混合卷积神经网络^[51]，其中卷积层是完全在光学层面上实现。该系统在图像输入计算机前，使用一个“4F 光学系统”进行傅里叶变换及逆变换，然后使用数字微镜器件（DMD，Digital micro-mirror device）来模拟出图像的光输入，最后使用 CMOS 探测器接收经过光学卷积层的图像，并转换到电脑中，如图1-3所示。2021 年，有研究团队提出了光电卷积神经网络（OPCNN）^[52]，如图1-4所示，其中液晶型 SLM 被用来完成卷积层，与微透镜阵列组合实现多核同步卷积特征提取；并且全连接层和降采样层功能同样也使用 SLM 实现。

2019 年，研究人员在光电混合网络结构基础上进一步研究，为了尽量避免光电转换的麻烦，最终提出了全光神经网络 (AONN, all-optical neural network)^[54] 架构，如图1-5所示。网络线性运算部分仍然由 SLM 和傅里叶透镜组完成，非线性激活函数则利用具有非线性光学特性的材料或者电磁诱导机制实现。2022 年，有团队提出不适用透镜实现光卷积层功能^[55]，利用小孔成像原理，设计不同形状的孔洞，在后方成不同的像，如图1-6所示，达到卷积效果。

在众多的光学神经网络中，通过以光的衍射为主要光信号传播方式的衍射光神经网络 (D2NN, Diffractive deep neural network) 是一种原理较为简单，功能非常

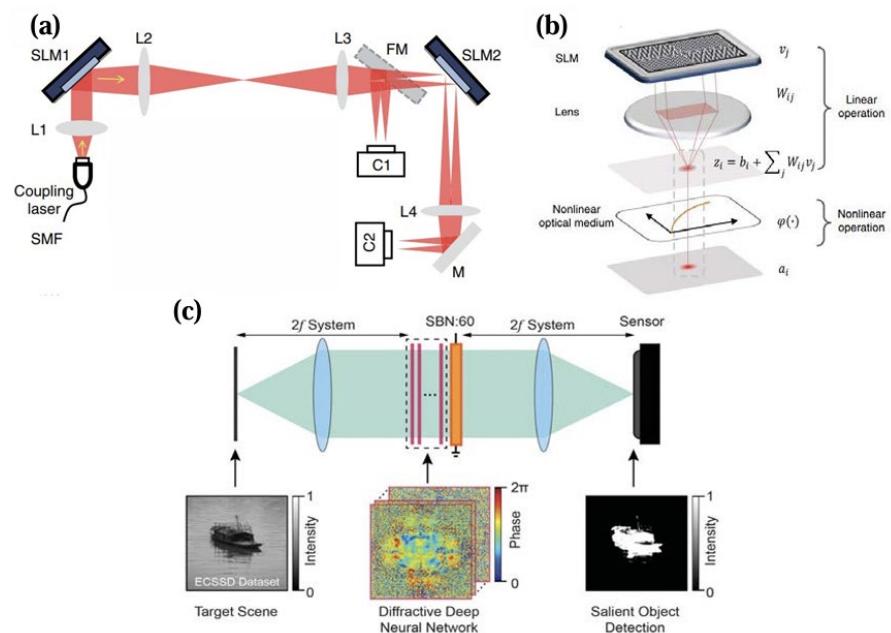


图 1-5 全光神经网络非线性结构与傅里叶空间光神经网络^[53,54]示意图

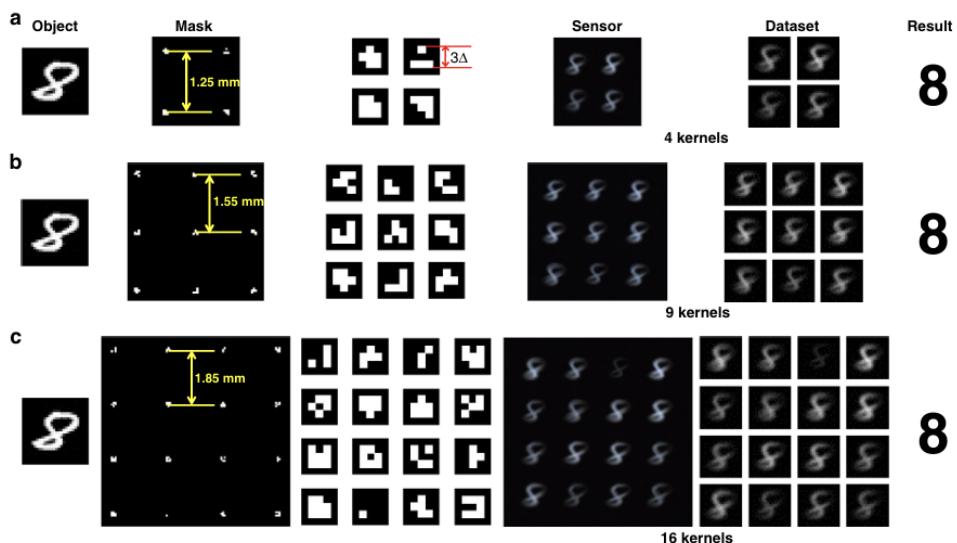


图 1-6 无透镜光学卷积层^[55]

强大的全光神经网络。其本质是基于光的衍射原理构建的计算网络，它利用光的衍射、干涉等物理现象进行信息的传输与处理。该网络由多个衍射板面构成，每个板面对应一个网络层，通过调整板面上的参数（如透明度、材料材质、材料厚度等），可以实现类似电子神经网络中的权重与偏置。与传统的电子神经网络相比，衍射光神经网络具有更高的并行度和更低的能耗。2018年，有研究团队首次提出全光衍射神经网络模型^[56]，空间衍射现象被证明可应用于类似全连接神经网络模型构建。进一步，该团队利用3D打印衍射相位板，并在现实世界中构建实物，组成神经网络物理连接层，并实现了仿真实验的图像分类效果。经过训练获得衍射板物理参数，并据此打印搭建的D2NN在现实环境中对经典的灰度图像数据集进行测试效果较好，图1-7首次展示全光衍射神经网络架构的能力并验证了网络架构的可行性。2019年，该小组提出了基于相同架构的宽带衍射光神经网络^[57]，通过使用相位参数，巧妙设计优化后，D2NN模型不再只能用单色光，可以多种频率的光同时使用，扩展了衍射神经网络的应用范围。2021年，有团队进一步将D2NN光源扩展到可见光频域^[58]，另外一团队进一步提出神经态可重构衍射计算单元（DPU）^[59]，并通过SLM与CMOS探测器组合，合成大规模空间DPU计算层，物理系统误差则被用自适应训练方式极大的避免，最终开发了新的衍射光电混合计算框架，并使其在更复杂的图像识别跟踪任务上均媲美电子计算机上神经网络的性能。

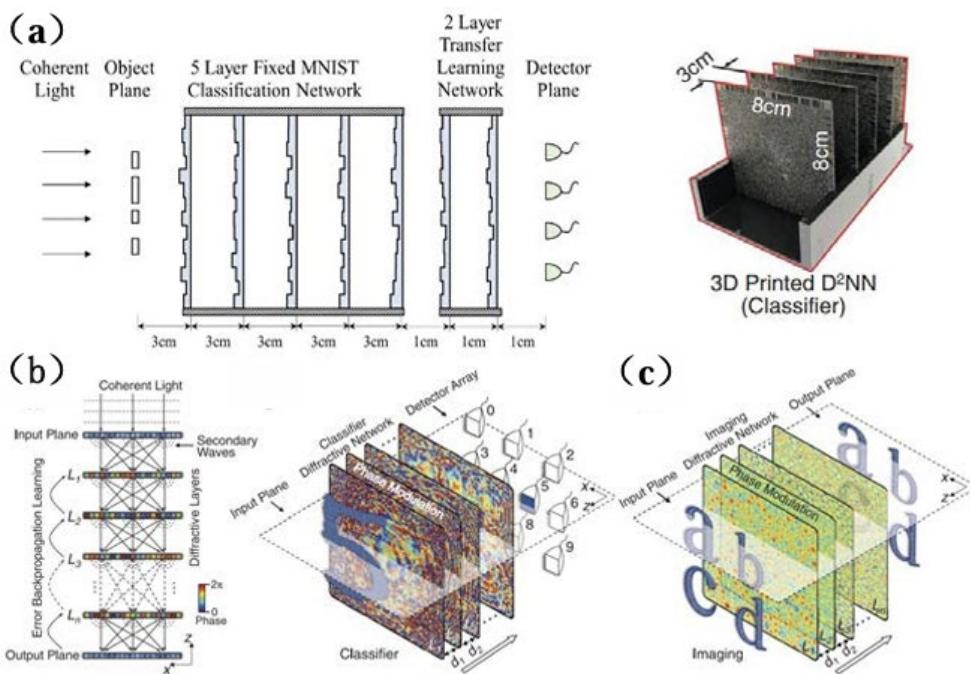


图 1-7 全光衍射深度神经网络 D2NN

在 2019 年，有团队优化衍射相位光栅的材料^[60]，并从太赫兹衍射系统换为了红外激光光源，在极大地减小了 D2NN 的尺寸和成本的同时，仍然保证其高性能，在 MNIST 数据集上仍有 90% 以上的识别能力。2022 年，有团队也对衍射光神经网络进行微型化和集成化研究^[61]，利用微纳米技术，将计算好参数的衍射超表面集成到硅基 CMOS 里面，形成片上的多层次光深神经网络，在图像光束直接照射下便能完成图像识别任务，为 ONN 进一步实际应用打下了基础。

除此之外，还有散射型光神经网络和片上光神经网络^[62-66]，如图 1-8 所示，这些网络也均有较大的发展进步，并推进了光计算的研究，使得光计算的实际应用更进一步，光子计算机在未来可能成为现实。

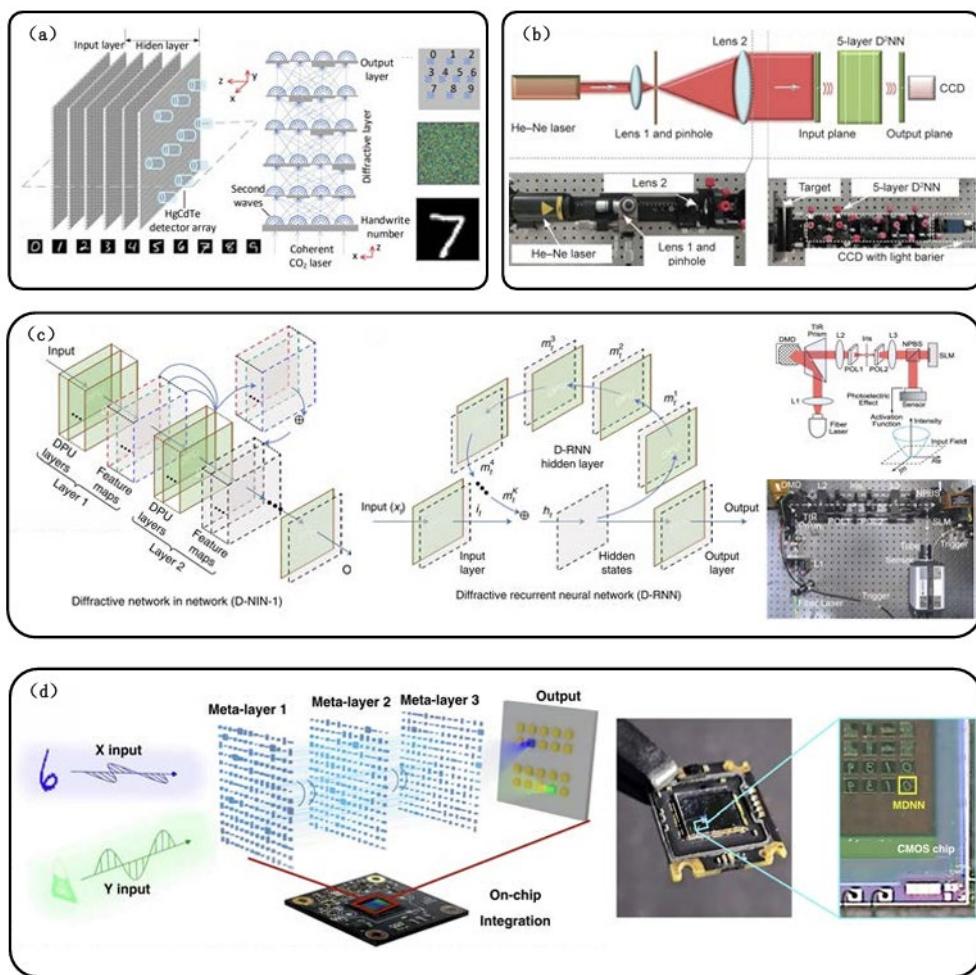


图 1-8 不同方式实现衍射神经网络示意图

随着技术的不断进步和研究的深入，光计算在未来有望取得更加显著的进展和突破。一方面，科学家们将继续优化光学元件的设计和制造工艺，提高光计算系统的性能和稳定性。另一方面，他们也将探索新的光学计算原理和算法，以期

望实现更高效、更灵活的计算。此外，随着人工智能、物联网等技术的快速发展，光计算将在这些领域发挥重要作用，推动信息技术的进一步发展。

1.3 本文的主要研究内容与创新

本文主要进行了衍射光神经网络的物理原理解释，数学原理推导以及仿真模型建模。然后在此基础上进行了灰度图像的分类识别训练和测试，并进一步引入了非线性激活函数优化框架。在灰度图像任务数据集 MNIST, Fashion-MNIST 和 E-MNIST 上对网络能力及效果进行验证后，进一步针对彩色图像进行了研究，并提出了两种新框架，分别是位置复用衍射光神经网络和波分复用衍射光神经网络，并在 CIFAR-10 数据集上进行了训练和测试，相比传统衍射光神经网络，分类效果有较大的提升。

1.4 实验设备

本文所有实验在 Windows 10 环境下，Anaconda 和 Pycharm 开发环境，使用 11th Gen Intel(R) Core(TM) i9-11900H @2.50GHz 2.50 GHz 的 CPU 和 NVIDIA GeForce RTX-3070 Laptop 版 GPU 的电脑，其存储内存 3TB，运行内存 32GB，显存 8GB 环境训练测试。

1.5 本文的结构安排

本文主要通过六个章节对相关工作进行论述，其中第一章介绍人工智能和光计算以及光神经网络相关背景知识和基础概念，然后进一步介绍近年这两个领域国内外发展现状，最后简述本文实验设备环境以及结构安排；接着第二章先介绍人工智能算法原理和神经网络相关数学原理，对全连接网络进行推导，对非线性激活函数进行描述介绍；然后第三章主要介绍物理原理和数学推导，并进行衍射光神经网络的数学建模及仿真算法描述，最后论述网络训练优化的方案及优缺点；接下来第四章对灰度图像任务实验和结果进行叙述，第五章对彩色图像任务和两种新型网络框架进行介绍，并对实验结果进行描述和分析。最后第六章对全文工作进行总结，并对未来研究发展进行展望。

第二章 神经网络算法原理

光神经网络本质上是用光学原理和各种光学器件实现与电子上的神经网络同样的功能，而衍射光神经网络则是利用全连接网络模型与衍射光理论在光上实现计算。本章主要对神经网络相关理论和算法原理进行简要介绍，特别是全连接网络和非线性激活函数的作用。

2.1 神经网络

神经网络中的神经一词来源于生物大脑中的神经，可以看出该网络的目的就是想模拟大脑的功能。而在生物的大脑中，最基本的组成单位就是神经元，通过数量庞大的神经元互相连接，存储、交流和沟通各种生物信号。基于此，大脑可以完成对声音、感觉和视觉的分析处理，并学习记忆新的知识和复杂场景与反应。其他神经元传递来的信号被该神经元用树突分支接收，若激活则通过突触传递输出信号。而在一个普通人大脑中，就有高达数千亿个神经元互相合作传递处理信息。

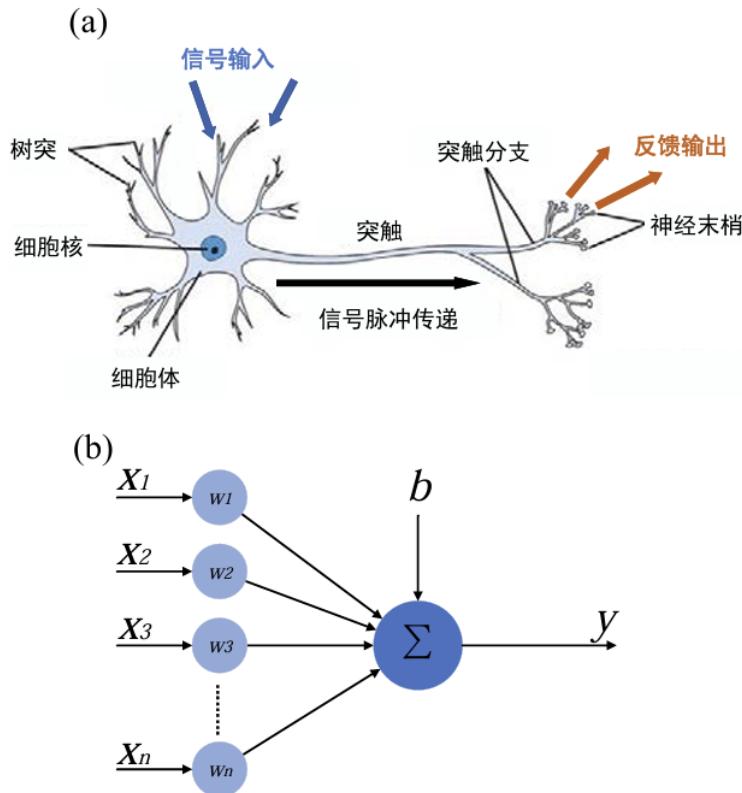


图 2-1 光学卷积层^[22]

为了模拟大脑神经元的工作原理, 如图2-1(a)所示, 从而创造类似人的智能体, 研究人员便提出了感知机模型^[67], 如图2-1(b)所示, 并尝试利用该基础神经元组成全连接神经网络(也被称作多层感知机)。一个神经元, 我们不妨将其抽象表示为一个带有权重的线性求和单元, 其中输入向量为 $\mathbf{x} = [x_1, x_2, \dots, x_n]$, 向量的每一维输入都与该神经元对应一个边权, 其也可以用向量表示为 $\mathbf{w} = [w_1, w_2, \dots, w_n]$ 。该神经元运算过程可以用函数 $f(\mathbf{x})$ 表示为:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2-1)$$

其中 b 为该神经元的偏置量, 将其展开为标量形式为:

$$f(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b \quad (2-2)$$

从式子2-2可以看出, 权重参数 \mathbf{w} 决定了该神经元的计算结果, 也就是其状态。输入的信号不同, 对应权重不同, 则该神经元的输出都会不同。这也模拟了大脑神经元在面对不同刺激下的不同反应。但是不难看出, 单个神经元的拟合(反应)能力十分有限, 只能够模拟单一情况和线性情况。

2.2 全连接网络

类比大脑众多的神经元, 为了使模型具有更强的学习表达能力, 满足各种复杂场景下的学习推理需求, 研究人员对单感知机模型, 也就是单神经元进行了增多与互连, 如图所示。在扩展后, 广度和深度均得到了加深, 在输入维度仍然不变的情况下, 其与输出之间有了更多的拟合计算单元。至此, 我们得到了最基本的全连接神经网络模型, 如图2-2所示。

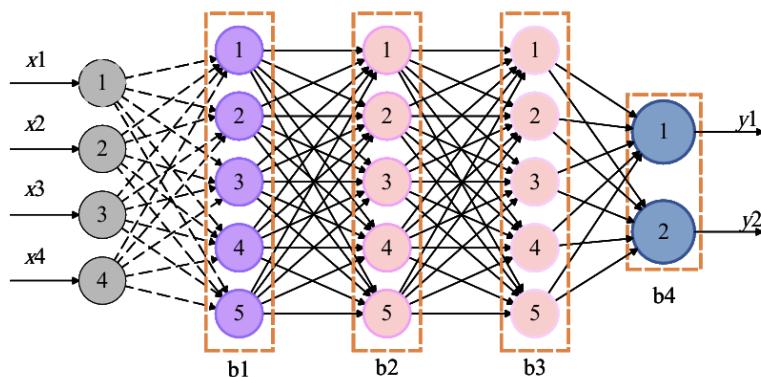


图 2-2 扩展后的感知机模型(全连接神经网络)^[22]

由于输入与输出之间神经元的计算维度进行了扩展, 所以此时不再是向量乘

积求和，而是被扩展为矩阵相乘模型，可以得到第 i 层神经元的全连接网络模型的权重矩阵 W_i 。

$$W_i = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{bmatrix} \quad (2-3)$$

当输入向量为 $\mathbf{x} = [x_1, x_2, \dots, x_n]$ ，第 i 层的输出可以表示为：

$$\begin{bmatrix} o_1 \\ o_2 \\ \vdots \\ o_n \end{bmatrix} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b_i \quad (2-4)$$

那么同样的，多层级联则可以写成多个矩阵相乘的结果，例如三层全连接层相连输入输出表达式如下：

$$\mathbf{y} = W_3 \cdot [W_2 \cdot [W_1 \cdot \mathbf{x} + b_1] + b_2] + b_3 = \mathbf{W} \cdot \mathbf{x} + \mathbf{b} \quad (2-5)$$

最终，由上面推导式子可得，整个网络输入输出可以表示为 $\mathbf{y} = Net(\mathbf{x})$ ，其中 \mathbf{x} 为输入向量， \mathbf{y} 为输出向量。相比维度扩展之前，扩展后的全连接神经网络的可训练权重大大增多，且输出维数也得到了提升，可以表达更复杂信息；同时不难看出，全连接神经网络具有拟合更加复杂的函数的能力，也就意味着可以处理更多的复杂任务，应用于更多现实世界场景。

2.3 非线性激活函数

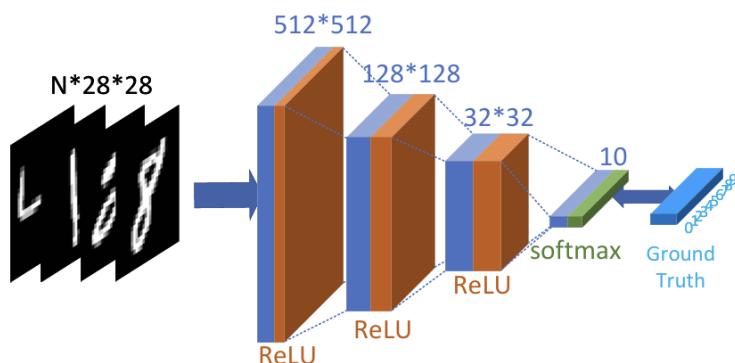


图 2-3 添加非线性激活函数的全连接神经网络 [22]

通过式子2-5可以看出，原始的全连接神经网络最后仍然是得到一个线性的结果，说明该模型对于非线性情况的拟合能力较差，所以科研人员接下来在其中类比大脑神经元的激活机制，引入了非线性函数作为激活函数，进一步提升了全连接神经网络的学习推理能力。添加如 ReLU 非线性激活函数后的全连接神经网络模型如图2-3所示。

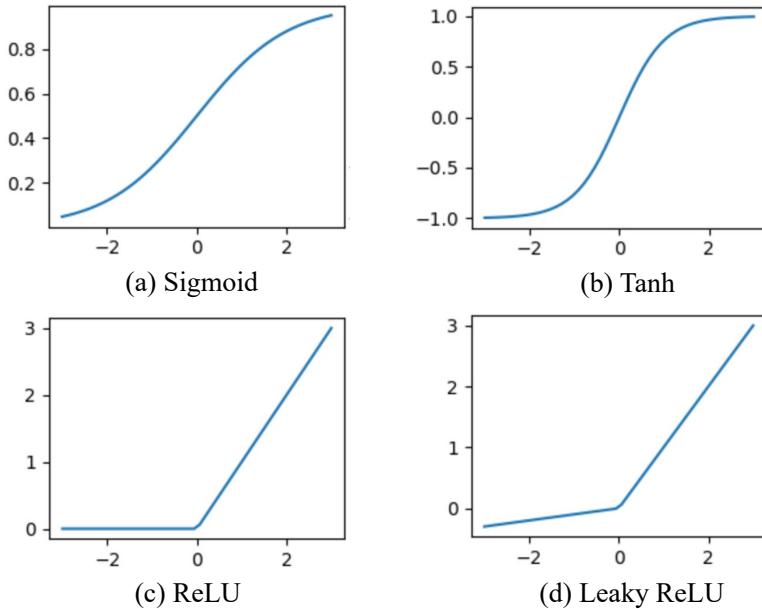


图 2-4 常用激活函数

如图2-4所示，为神经网络运算中常用的非线性激活函数，其中对应函数表达式如下：

$$\begin{aligned} \text{sigmoid}(x) &= \frac{1}{1 + e^{-x}} \\ \tanh(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \\ \text{relu}(x) &= \max(x, 0) \\ \text{leakyRelu}(x) &= \max(\alpha x, \beta x), (0 < \alpha < \beta \leq 1) \end{aligned} \quad (2-6)$$

不妨将激活函数设为 $\varphi(\cdot)$ ，则对2-1公式加入激活函数便能使全连接神经网络具有非线性拟合能力，新的映射函数如下：

$$y = \varphi(\mathbf{W}^T \mathbf{x} + b) \quad (2-7)$$

图2-4(a)是 Sigmoid 函数图像，可以将输入归一化到 $(0, 1)$ 区间，对于输出概率值等区间同样也为 $(0, 1)$ 的问题有着很好的拟合能力，并且该函数比较光滑，处处可导，十分便于使用梯度下降算法进行参数训练优化。图2-4(b)是 Tanh 函数

图像，这个函数类似 Sigmoid 函数平移后实现，可以将输入规约到 $(-1, 1)$ 范围内，在 Sigmoid 基础上对神经网络输出扩充了负数值。图2-4(c)是 ReLU 函数图像，图2-4(d)是其变体 Leaky ReLU 函数图像，这类激活函数在小于 0 的值全部为 0 或大幅度抑制，对于大于 0 的值则有较大非零梯度，具有单边抑制、激活范围更大的优势；相比 Sigmoid 函数，在较大或较小的输入时，Sigmoid 函数梯度值容易接近 0，导致训练陷入局部最小值中，网络参数难以更新，难以跳出局部最小值，学习过程难以收敛。

2.4 训练优化

为了通过梯度下降来训练得到全连接神经网络的参数，需要根据相应网络结构以及相应任务设计对应的损失函数，并通过设定学习率和学习次数来不断训练优化，使损失函数尽可能的下降，最终计算出全连接神经网络最优参数，便于后续的计算应用。损失函数通常是表现网络输出值与样本真实标签之间的差异大小，差异越小证明网络输出判断越好，所以一般都希望损失函数尽量下降。

回归模型中常用平均绝对误差（MAE）和均方误差（MSE）来估计损失，对应 L1 损失函数和 MSE 损失函数。分类模型中，最常用的是交叉熵误差，对应 CrossEntropy 损失函数。本文训练过程中，因为均为分类任务，所以使用交叉熵损失函数。

除损失函数本身之外，在其中添加正则项可以对网络训练施加约束，使得神经网络能够更好地向着期望的方向优化权重。在有些网络中，如 PINN（Physical informed neural network）^[68] 会使用一些物理法则和对应微分方程来进行约束训练，使得网络在具体现实应用中获得更好的表现。训练过程中，常见的优化算法有自适应矩估计（Adam）^[69] 及其改进^[70]、自适应梯度下降（Adagrad）、随机梯度下降算法（SGD）、动量随机梯度下降法（SGDM）等^[71]。本文中的实验均使用 Adam 优化算法，并根据输出光场的结果特点，添加检测区域外的光强信号之和作为正则项，来进一步约束神经网络。

2.5 本章小结

综上所述，通过类比大脑系统和神经元，科研人员构建出了感知机模型，并将其扩展为全连接神经网络，然后进一步加入了非线性激活函数，使得该模型不再拘泥于线性拟合，具有优良的学习表达能力，更强的泛化能力，可以完成现实世界中许多复杂的计算学习任务。

第三章 衍射光神经网络的数学建模

光的衍射现象早在 17 世纪就开始被人们观测到并开展研究。光也是一种电磁波，在穿越任何障碍物时都会发生衍射，但只有当障碍物或孔洞的尺寸与光的波长相近时，才能观察到显著的衍射效应。在衍射分析的理论框架下，用数学积分的方式可以计算无数连续子波相干叠加所形成观测区域的光的振幅和相位分布。为了模拟衍射光神经网络中的线性运算过程，我们通常采用离散近似的方法来处理这些衍射数学积分。本章将简要叙述用于近场和远场的衍射分析的基尔霍夫衍射理论和瑞利-索末菲衍射理论。

3.1 物理模型

惠更斯早在 17 世纪提出了一个关键的假设，即光波的衍射现象可以通过次级子波的传播来描述^[44]。他提出，衍射源发出的波面上的每一个点都可以被视为一个次级球面子波源，而这些子波源在随后的时刻会共同决定波面的分布情况，具体表现为这些次级子波包络面的形态。到了 19 世纪初，菲涅尔在惠更斯的子波理论基础上进行了拓展。他注意到，子波源的复振幅具有矢量形式且它们之间存在着相干关系。因此，实际上这些子波源相干叠加的结果就是光在空间中经过衍射传播后形成的光场分布。

3.1.1 亥姆霍兹-基尔霍夫积分

为了获取衍射光场的仿真表达式，亥姆霍兹-基尔霍夫积分定理被菲涅尔引入来进行更加精确的模拟计算。通过求解包围在空间中一个点的一封闭曲面上波动方程以及一阶导数的方式，能够更加精确地描述和预测光在衍射过程中的行为。不妨设 $U(P)$ 表示空间某点 P 的复振幅，假设单色光场的分布为：

$$U(P, t) = U(P)e^{-2\pi j \nu t} \quad (3-1)$$

那么 P 点光场振动满足标量波动方程：

$$\Delta^2 U - \frac{\delta^2 u}{C^2 \delta t^2} = 0 \quad (3-2)$$

综合3-1和3-2，可得时间无关的亥姆霍兹方程：

$$(\Delta^2 + k^2)U(P) = 0 \quad (3-3)$$

其中, k 为波矢量, 满足 $k = \frac{2\pi}{\lambda}$

进一步利用格林积分定理, 可以推出亥姆霍兹公-基尔霍夫积分公式:

$$U_p(x, y, z) = \frac{1}{4\pi} \oint_S \left\{ \frac{e^{ikr}}{r} \frac{\delta U}{\delta n} - U \frac{\delta}{\delta n} \left[\frac{e^{ikr}}{r} \right] \right\} dS \quad (3-4)$$

通过公式3-4可知, 用包围该点的任意封闭曲面 S 上各点扰动的边界值 U 的一阶偏导可以计算出光场中任意点 P 的复振幅分布。该积分公式为衍射理论的发展起到了重要贡献。

3.1.2 基尔霍夫衍射公式

下式即基尔霍夫光学衍射公式:

$$U(x, y, z) = \frac{1}{j\lambda} \iint_{\Sigma} U(P_0) K(\theta) \frac{e^{ikr}}{r} ds \quad (3-5)$$

$$K(\theta) = \frac{\cos(\vec{n}, \vec{r}) - \cos(\vec{n}, \vec{r}_0)}{r} = \frac{1 + \cos\theta}{2} \quad (3-6)$$

其中, $K(\theta)$ 为倾斜因子, 在菲涅尔衍射与夫琅禾费衍射近似下约等于 1, 式子中 r 为衍射点与观察点距离。

3.1.3 瑞利-索莫非衍射模型

基尔霍夫衍射公式在应对常见衍射现象时效果比较好, 但由于其边界条件的设定, 该理论自身对于一些特殊情况存在着一定的内在不一致性。忽略了衍射场中光场与物质边界之间的相互作用导致基尔霍夫衍射理论在假设中存在一定的问题。特别是当衍射发生在紧邻衍射孔径的近场区域, 也就是衍射面距离趋近于零时, 该衍射公式难以完全符合基尔霍夫边界条件。而且其边界条件明确规定了衍射屏后被阻挡部分的复振幅法向一阶导数均为零, 但根据电动力学场论的基本原理, 如果波动方程的解在任意非无限小的面元上复振幅及其法向导数同时为零, 那么这一解在整个空间都将为零, 这显然与经典电动力学中的唯一性定理相冲突^[72-76]。

索末菲巧妙地选取了适当的格林函数, 克服了基尔霍夫理论推导中的这些缺陷, 进一步避免了同时对边界上的光场复振幅及其法向导数施加边界条件的必要性。这一方法有效地改善了衍射理论的一致性和准确性。

最终可得, 瑞利-索末菲衍射公式如下:

$$U(x, y, z) = \iint_{\Sigma} \frac{z}{r} U_0(x_0, y_0, 0) \left(\frac{1}{2\pi r} + \frac{1}{j\lambda} \right) \frac{e^{i\frac{2\pi}{\lambda} r}}{r} dx_0 dy_0 \quad (3-7)$$

瑞利-索末菲衍射模型在全面考量整个衍射平面的光场分布后展现出了其在描述衍射光场，特别是近场大角度衍射下感受野能量分布的精确性。因此，基于该衍射理论，本文推导并构建了基本的衍射光神经网络模型，用来模拟和分析光在衍射过程中的行为，并进一步计算衍射光神经网络参数权重。本文所提到的所有衍射光模型都将统一采用瑞利-索末菲衍射模型进行光衍射计算与分析，以确保实验仿真结果的准确性和一致性。

3.2 数学建模

3.2.1 光场分布公式

在瑞利-索末菲衍射模型的基础上，以远场矩孔衍射为例作分析，可得单色平面光波照射边长为 a 的矩形孔洞后发生衍射，设其通过孔径平面的光波为 $U(x_0, y_0, 0)$ ，衍射板与观察屏幕之间的距离为 z ，则其三维坐标系中坐标为 $(x_0, y_0, 0)$ 和 (x_0, y_0, z) ，可得观察屏幕上光场分布为：

$$U(x, y, z) = \iint_{\Sigma} \frac{z}{r} U_0(x_0, y_0, 0) \left(\frac{1}{2\pi r} + \frac{1}{j\lambda} \right) \frac{e^{j\frac{2\pi}{\lambda} r}}{r} dx_0 dy_0 \quad (3-8)$$

其中， r 为衍射点到观察屏幕之间的直线距离：

$$r = \sqrt{z^2 + (x - x_0)^2 + (y - y_0)^2} \quad (3-9)$$

衍射傅里叶变换公式：

$$U(x, y, z) = C \cdot e^{jkz} e^{j\frac{\pi(x^2+y^2)}{\lambda z}} \cdot \text{FFT}[U_0(x_0, y_0, 0)] \Big|_{f_x=\frac{x}{\lambda z}, f_y=\frac{y}{\lambda z}} \quad (3-10)$$

再进一步推导，由公式3-10可得，最终衍射场分布为：

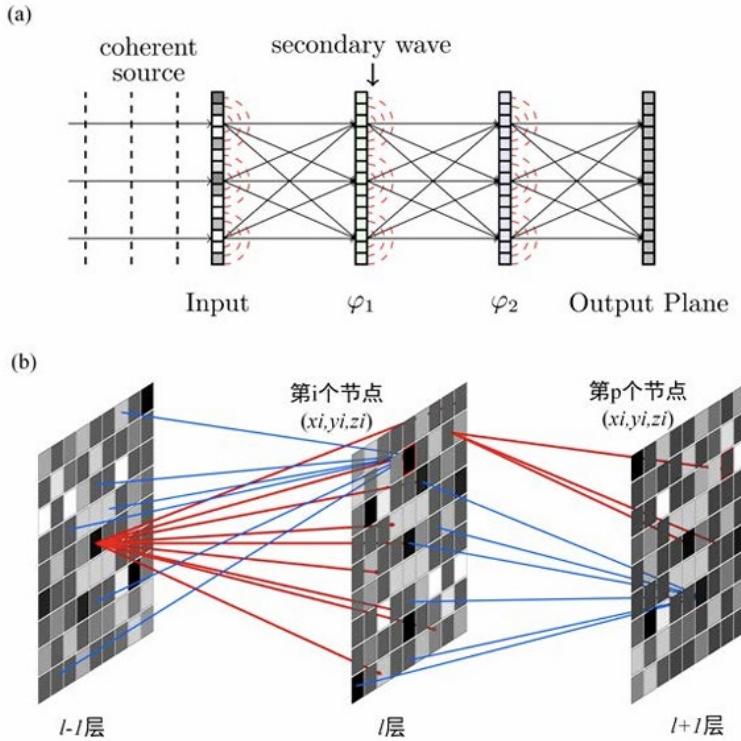
$$U'(x, y, z) = a^2 C e^{jkz} e^{j\frac{\pi(x^2+y^2)}{\lambda z}} \cdot \text{sinc}(a \frac{x - \alpha z}{\lambda z}) \text{sinc}(a \frac{y - \beta z}{\lambda z}) \quad (3-11)$$

$$I'(x, y, z) = \left(\frac{a^2}{\lambda z} \right)^2 \text{sinc}(a \frac{x - \alpha z}{\lambda z})^2 \text{sinc}(a \frac{y - \beta z}{\lambda z})^2 \quad (3-12)$$

3.2.2 衍射光全连接神经网络推导

衍射光学中，光的传播可以被描述为复振形式。所以，在处理衍射光神经网络模型时可以将其理论化为光线，然后就可以使用衍射场分布模型，进行衍射光束指向分析，推导出在衍射理论下的神经元连接计算方式。

空间平面衍射层级联来实现全连接层的功能，如图3-1所示，通过每一层衍射


 图 3-1 衍射光全连接神经网络模型示意图^[22]

板上像素单元的相位偏移和振幅调整不同，进一步实现全连接层中每个神经元的权重参数。从次级子波理论出发，根据瑞利-索末菲衍射公式3-7，可获得整个衍射网络中衍射神经元次级传播公式：

$$w_i^l(x, y, z) = \frac{z - z_i}{r^2} \left(\frac{1}{2\pi r} + \frac{1}{j\lambda} \right) e^{\frac{2j\pi r}{\lambda}} \quad (3-13)$$

其中 l 代表第 l 层衍射层， i 代表其上第 i 个神经元单位， r 为衍射层之间传播直线距离，且 $r = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}$ 。可以看出，第 l 层衍射面的第 i 个衍射神经元单位 (x_i, y_i, z_i) 的相位及振幅可以被前一层衍射面所有衍射神经元的衍射次级子波对该点的矢量叠加表示，然后与该点处的网络相位权重相乘：

$$m_i^l(x_i, y_i, z_i) = t_i^l(x_i, y_i, z_i) \cdot \sum_n n_k^{l-1}(x_i, y_i, z_i) \quad (3-14)$$

其中， $t_i^l(x_i, y_i, z_i)$ 为 l 层的第 i 个衍射单元的相位透射系数，包含了对振幅和相位的调整项，代表 l 层网络继续向前传播光时的连接权重：

$$t_i^l(x_i, y_i, z_i) = \alpha_i^l(x_i, y_i, z_i) e^{j\phi_i^l(x_i, y_i, z_i)} \quad (3-15)$$

其中 $\alpha_i^l(x_i, y_i, z_i) \in (0, 1)$ 代表了衍射层该点处的透射率，如果是用的纯相位调制系

统, α 可以根据衍射器件属性设置固定的投射率; $\varphi_i^l(x_i, y_i, z_i) \in [0, 2\pi)$, 由此保证衍射面相位分布在合理范围内。

综合上述公式, 可得最终第 l 层上的第 i 个神经元向前传播时的衍射次级子波的相位分布情况:

$$n_i^l(x, y, z) = w_i^l(x, y, z) \cdot t_i^l(x_i, y_i, z_i) \cdot \sum_n n_k^{l-1}(x_i, y_i, z_i) = w_i^l(x, y, z) \cdot m_i^l(x_i, y_i, z_i) \quad (3-16)$$

进一步的, 可以将衍射空间前向传播的神经网络模型总结如下:

$$\begin{cases} n_{u,p}^l = w_{i,p}^l \cdot m_i^l \\ m_i^l = t_i^l \cdot \sum_k n_{k,i}^{l-1} \\ t_i^l = \alpha_i^l e^{j\varphi_i^l} \\ n_{i,p}^0 = w_{i,p}^0 \cdot I_i^0 \end{cases} \quad (3-17)$$

公式3-17中, 前面三项式子已经简要叙述, 第四项式子则是表示输入层初始情况, 其中 I 表示初始输入图案的光振幅及相位分布。在同样的物理特性前提下, 其他类型衍射光系统, 如反射式和透射式, 都能通过该模型建立衍射光全连接神经网络, 并进行训练计算。

3.3 光神经网络训练及参数优化

在传统神经网络中, 其各个层的计算参数决定了整个网络的性能, 因此对网络的初始结构设置, 训练参数设置以及优化器选择等都尤为重要。在有监督的参数训练中, 通过前向传播计算出训练样本的结果, 通过其与训练样本的标签结果进行损失函数计算, 然后来反向传播, 通过梯度下降优化整个网络的权重参数。最终, 使得网络输出与实际结果之间的损失函数值最小化从而让整个网络对训练任务有较好的学习。同时添加一些正则项可以更好的使整个网络快速收敛, 收敛到一个泛化能力更强的权重参数, 避免过拟合现象出现, 使其在测试集上仍然有良好的表现。

由于衍射光传播模型较为复杂, 反向传播计算量巨大, 所以如何实时训练光神经网络参数是一个极其重要但又十分困难的点。目前绝大多数 ONN 训练均采用计算机内模拟仿真, 然后使用神经网络进行光学器件可变参数训练, 最后将训练出来的参数载入 ONN 框架进行任务推理测试。这样利用反向传播, 如图3-2所示, 通过计算机神经网络进行计算方便快捷, 但是, 计算机上仿真环境中是无法精准模拟实际物理光场分布, 再加上实际光学器件均有一定的误差以及实验环境光干

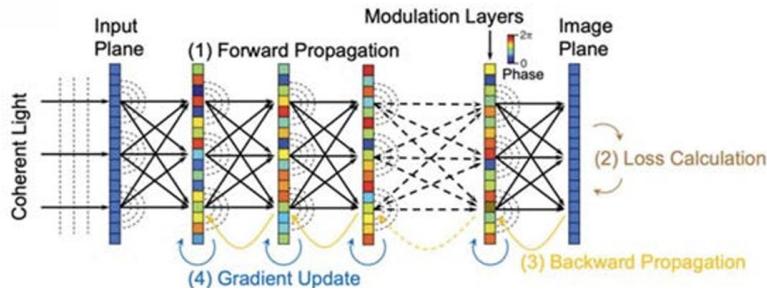


图 3-2 衍射光神经网络正向反向传播示意图 [22]

扰等影响^[47,77]，所以训练好的网络模型不能与实际的光学神经网络系统做到严格匹配，在实际测试推理中，由于这些干扰因素影响，导致目前计算机仿真的神经网络推理预测能力仍然大于光学计算性能，并且误差随着网络的连接广度以及连接深度扩展而积累，最终会导致网络因为误差过大而完全失去表达能力，该类神经网络的规模扩展也因此受到了极大的限制。所以光网络推理能力的提升以及其实际应用效果仍然需要深入研究解决方案。虽然，现在有一些研究已经在 ONN 实时训练上有一定的进展，比如光自适应训练，尽量减小了仿真与实际之间的误差，但问题的根本还是难以消除。

3.4 本章小结

本章简要介绍了衍射光神经网络中的物理原理与数学模型，神经网络的大规模扩展可以由光学面逐层连接并利用光在空间中传播的计算方式较为便捷地实现。然而，这类空间计算方式往往因其用途单一、光电转换数据刷新率不高和空间带宽等物理因素影响而受限。同时，由于空间光学器件的体积较大，光神经网络系统的硬件集成面临挑战。片上光神经网络矩阵运算模型受限于当前的制造工艺，误差的积累导致片上光神经网络集成度较低，成本高昂。总体来看，尽管相关理论早在半个世纪前已被提出，但国内外光神经网络的研究尚处于初期阶段，直到近年来才有团队不断将其转化为现实应用，目前这些技术尚还未成熟，难以实现真正的工程化应用。然而，光学方式替代电子完成神经网络计算的巨大潜力持续激励着人们不断探索和研究光神经网络的实用化途径。

第四章 灰度图像分类

衍射光神经网络利用光强来携带图像信息，所以其本身天然具有灰度图像处理计算能力，本章节主要介绍了所使用的灰度图像数据集和任务，以及衍射光神经网络框架和非线性函数原理，还有仿真实验结果。

4.1 灰度图像数据集

现有常用的公开灰度图像数据集为 MNIST 手写数字数据集^[78]，其图像为 28×28 像素大小的灰度图像，包含 10 种类别，分别是数字 0 到 9；训练集有 60000 张图像，测试集有 10000 张，每种类别 1000 张。另一个数据集为 Fashion-MNIST 服装数据集^[79]，其图像也为 28×28 像素大小的灰度图像，包含 10 种类别，分别是 t-shirt, trouser, pullover, dress, coat, sandal, shirt, sneaker, bag, ankle boot 这 10 种，同样的，其训练集有 60000 张图像，测试集有 10000 张，每种类别 1000 张。最后一个是 E-MNIST 手写字母数据集^[80]，其图像仍为 28×28 像素大小的灰度图像，其中手写字母是其一个子数据集，包含 26 个类别，分别为字母 a 到 z（不区分大小写），训练集有 124800 张图像，测试集有 20800 张图像，每种类别 800 张。

在后续实验中，均将 28×28 大小图像插值扩充为 128×128 大小，方便契合衍射光神经网络输入尺寸，也使得网络具有更多的可学习单元。

4.2 光学非线性运算

在神经网络中，非线性函数激活是其非常重要的一个组成部分，对应在人类神经学中为神经元节点是否能够受到激活并发射传递信号的功能。激活函数为模型添加非线性成分，使得模型具有更强的学习泛化能力，可以应用于更加复杂场景的学习和推理计算。

当前，大多数框架利用的非线性激活功能为电子或光电子系统实现，从光转换成电子计算必定会增加器件，并且降低网络效率，增大网络耗能。但利用特殊材料的非线性光学器件会因为材料本身性质所造成的误差，难以实现理想的效果^[81]。为实现更好的光学非线性运算功能，研究人员从各种非线性光学性质或者理论入手，在 1967 年，有团队提出利用饱和吸收器模型或电子模块以实现光学非线性运算^[82]；进一步，在 2018 年，有团队提出量子点材料的粒子数反转机制可实现 ReLU 函数激活功能^[83]；同一时间，有团队^[84]也有提出利用等离子中 Fano 共振引起的可逆透明灵敏度和 Buckyball (C₆₀) 膜的异饱和吸收来实现快速高效的

全光非线性传输，以此提高神经网络吞吐量，降低时延和功耗；在 2019 年，有团队^[85]研究全光尖峰神经突触网络时，结合波分复用技术，提出利用相变材料 PCM 和微环振荡器 MRR 构成光脉冲神经元，并将其结合，所得的峰值函数作为非线性激活函数，由此实现权重非线性调制。同一时间，也有团队^[54]提出的全光学非线性激活是通过具有电磁诱导透明性质（EIT，electromagnetic induced transparency）材料实现，并用激光冷却在暗线二维磁光阱（MOT，magneto-optical trap）的 ^{85}Rb 原子实现 EIT 激活功能。共振峰的传输和带宽由耦合激光光强控制，最终共振探测激光束的输出可表示为：

$$I_{p,out} = I_{p,in} e^{-OD \frac{4\gamma_{12}\gamma_{13}}{\Omega_c^2 + 4\gamma_{12}\gamma_{13}}} = \Phi(\Omega_c^2) \quad (4-1)$$

由公式4-1可得，MOT 可以将经过线性传输计算的光束调制，激活的输出光束具有非线性传输特性。以上非线性例子如图4-1所示。

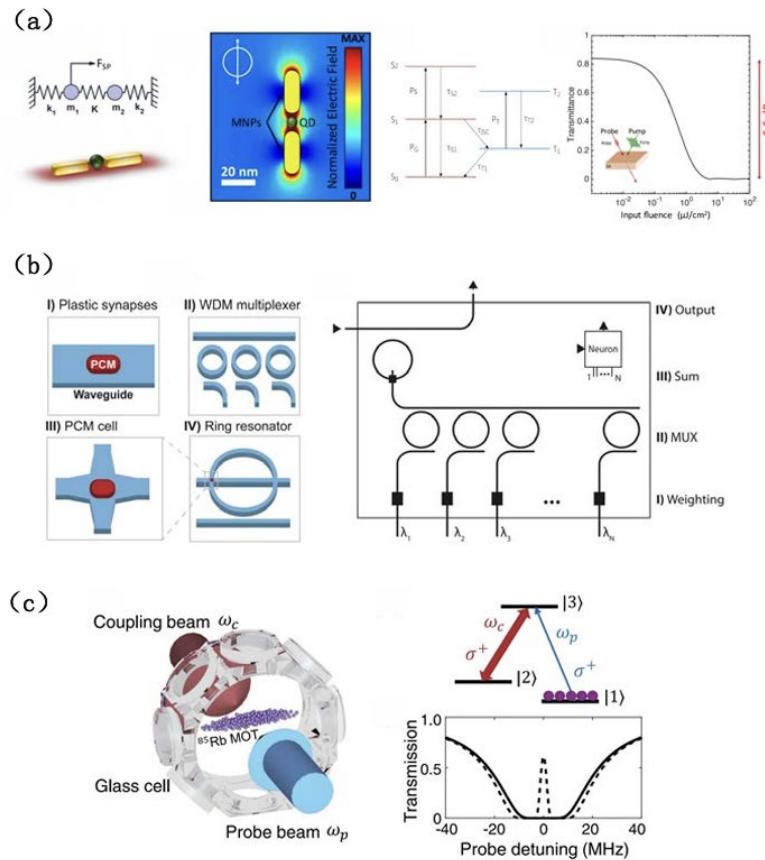


图 4-1 非线性激活函数的光学实现例子^[22]

所以实际物理光学器件可以实现非线性激活函数的功能，在仿真中便使用 ReLU 函数来作为非线性激活函数。

4.3 衍射光神经网络结构

衍射光神经网络由输入层，中间层和最后输出层（光子探测器区域）以及分类结果组成，如图4-2所示。

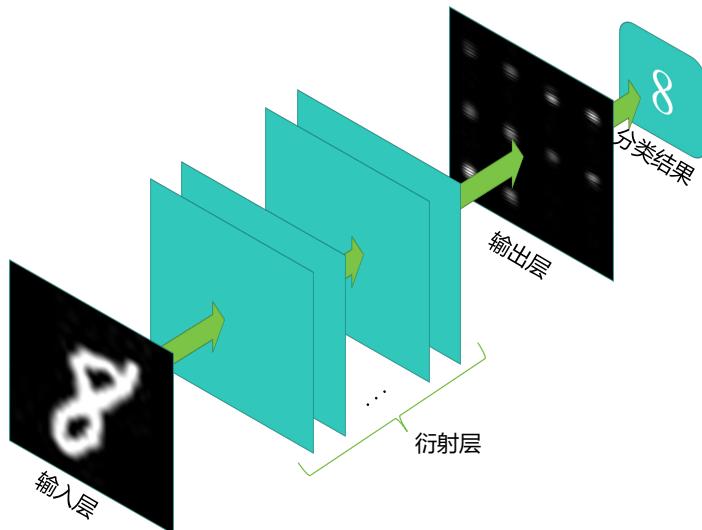


图 4-2 基本衍射光神经网络结构示意图

衍射层伪代码如4-1：

算法 4-1 单层衍射层前向传播算法 Dlayer

Data: 包含相位和强度信息的复数二维矩阵输入光场 w

Result: 包含相位和强度信息的复数二维矩阵输出光场 w_o

- 1 **初始化** (f 为光的频率, c 为光速, i 为虚数单位, d 为板间距离, s 为板边长大小, s_n 为板边长对应像素数量)；
- 2 $\varphi_{(x,y)} = (x - \frac{s_n}{2s})^2 + (y - \frac{s_n}{2s})^2;$
- 3 $w_n = \frac{2\pi f}{c};$
- 4 $w_l = \frac{c}{f};$
- 5 $h = fftshift(e^{w_n di} \cdot e^{-w_l \pi d i \varphi});$
- 6 **前向传播:**；
- 7 $t = fft_2(w);$
- 8 $k_{space} = complex(h.real \cdot t.real - h.imag \cdot t.imag, h.real \cdot t.imag + h.imag \cdot t.real);$
- 9 $w_o = ifft_2(k_{space})$

衍射光神经网络伪代码如4-2：

算法 4-2 衍射光神经网络算法 D²NN

Data: 包含灰度图信息二维矩阵输入光场 w_I

Result: 包含强度信息的二维矩阵输出光场 w_O

```

1 初始化 ( $p$  为可训练相位参数) ;
2  $p_i = 2\pi \cdot \text{random}(0, 1), i \in (0, n - 1);$ 
3  $l_i = DLayer, i \in (0, n - 1);$ 
4  $x.real = w_I, x.imag = 0;$ 
5  $i = 0;$ 
6 while  $i < n$  do
7    $x = p_i l_i(x);$ 
8    $i = i + 1;$ 
9 end
10  $w_O = |x|;$ 

```

网络层数经过多次实验，最终选择为 1 层输入层，4 层中间层和 1 层输出层。在实际仿真代码中，设置光速为 $3 \times 10^8 m/s$ ，使用光的频率为 $4 \times 10^1 Hz$ ，拟定单层衍射板大小为 $0.08m \times 0.08m$ ，板间距离设置为 $0.03m$ 。傅里叶计算则采用 Pytorch 包 fft 模块中的 fft2 函数进行正向变换，ifft2 函数进行逆向变换。像素单元数量则设置与输入图像像素大小一致，在整个 D2NN 中，对每一层还要添加可训练相位权重，并在传播中利用其计算下一层光场情况。除输入为只含强度的单色光外，传播过程中均使用复数进行计算，实部表示振幅，虚部表示相位，最后输出取其复数模长作为光子检测器的强度检测。检测区域为均匀设定的若干正方形小区域，如图4-2中输出层所示，对于手写数字分类，则设置 10 个正方形区域，检测区域大小为 10×10 个像素，从左往右，从上往下依次为 0 到 9 数字十类的结果。

网络损失函数 $Loss_{D2NN}$ 包含一个交叉熵损失函数和一个检测外部区域光场和的正则化项，不妨将输出光场记为 I_o ，检测区域的遮罩为 d_m ，则可以看出 d_m 为一个和输出大小一样的 01 矩阵，检测区域像素对应为 1，其余为 0。那么可得检测外部区域光场 I_{ex} 为：

$$I_{ex} = I_o \cdot (I_{one} - d_m) \quad (4-2)$$

其中， I_{one} 为与输出光场矩阵等大的全 1 矩阵，乘法为矩阵点乘。假设交叉熵损失函数为 $CE(\cdot)$ ，输出区域归一化概率为 y ，标签为 y_{true} ，那么 $Loss_{D2NN}$ 可以表示为：

$$Loss_{D2NN} = \alpha \cdot CE(y, y_{true}) + \beta \cdot \sum I_{ex} \quad (4-3)$$

其中, $\alpha + \beta = 1, \beta = 1 \times 10^{-5}$, $\sum I_{ex}$ 表示将 I_{ex} 矩阵中所有元素求和。添加此正则项后, 可以使网络输出光场更快收敛, 更快聚集到对应分类检测区域之中。添加非线性激活函数 ReLU 后的网络结构如图4-3所示

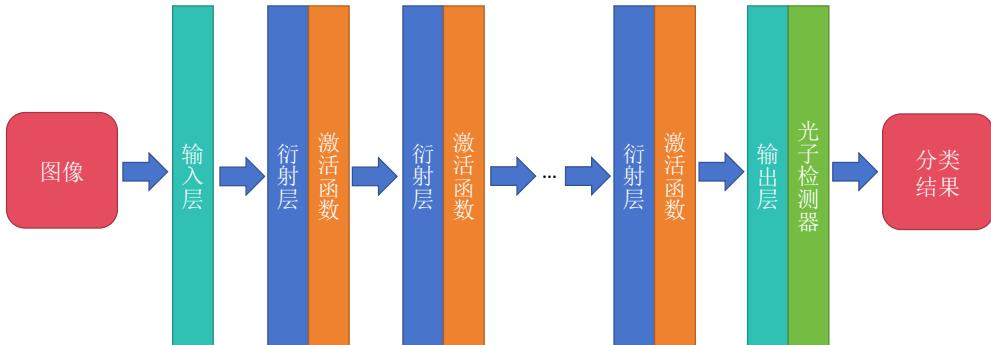


图 4-3 带非线性激活函数的衍射光神经网络结构

4.4 灰度图像识别分类实验仿真结果

实验中, 设置 Epoch 数为 40, 耐心值为 8 (若超过 8 个 Epoch 准确率没有提升则提前停止训练), 学习率初始为 1×10^{-2} , 然后每 5 个 Epoch 以 60% 比例衰减, 使用 Adam 优化器, betas 区间为 $(0.99, 0.999)$, 权重衰减和精度均为 1×10^{-5} , 随机数种子固定为 42, batch 大小为 64, 图像插值扩大到 128×128 像素, 然后进行归一化。

表 4-1 各灰度图数据集测试准确率结果

模型	MNIST	Fashion-MNIST	E-MNIST
D2NN	0.9764	0.8797	0.8941
D2NN+ReLU	0.9888	0.9015	0.9290

所有实验结果如表4-1所示, 可以看出使用 ReLU 激活函数的模型明显比未使用的传统衍射光神经网络模型准确率均有提高, 其中 MNIST 数据集识别准确率最高, 在 Fashion-MNIST 和 E-MNIST 上结果稍差, 但使用非线性激活函数后的准确率均达到了 90% 以上。

不同数据集在不含激活函数的 D2NN 上实验的训练过程中测试集准确率变化曲线, loss 变化曲线和最终测试集混淆矩阵结果分别如图4-4(a)、(b) 和 (c), 图4-5(a)、(b) 和 (c), 图4-6(a)(b) 和 (c) 所示。

其中可视化结果如图4-7所示, 其中每一个例子从左往右, 从上往下有七张图, 第一张图表示输入层图像, 倒数第二张图为输出层图像, 最后一张为光子检测器

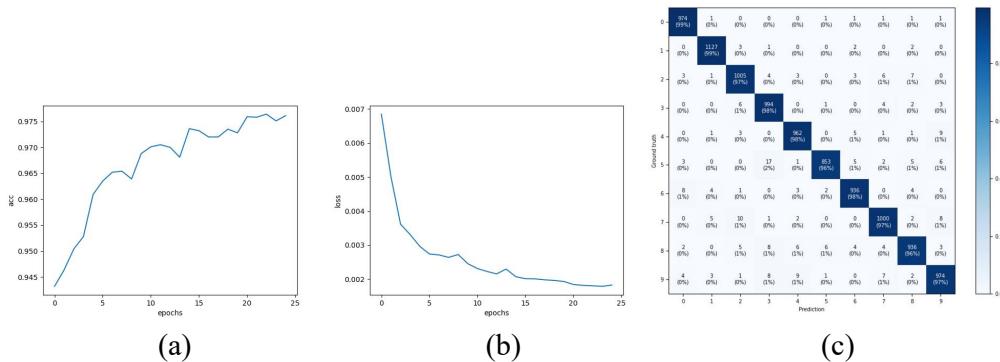


图 4-4 基础衍射光神经网络在 MNIST 数据集上实验结果

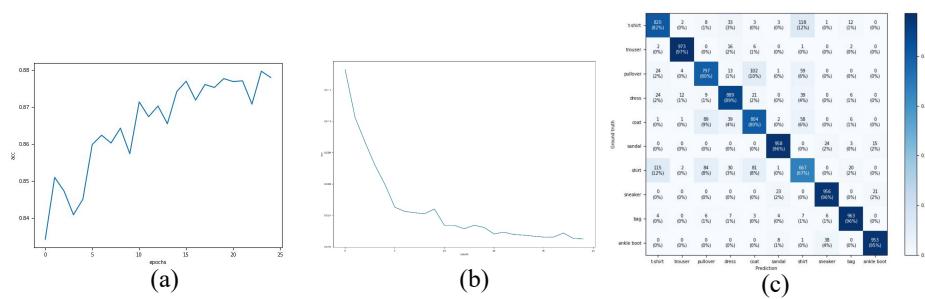


图 4-5 基础衍射光神经网络在 Fashion-MNIST 数据集上实验结果

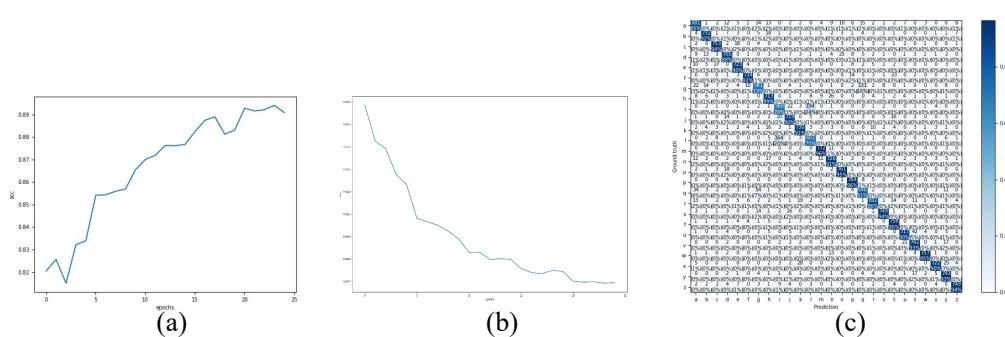


图 4-6 基础衍射光神经网络在 E-MNIST 数据集上实验结果

所检测区域图像，其余分别是经过每一层衍射层后的输出图像。

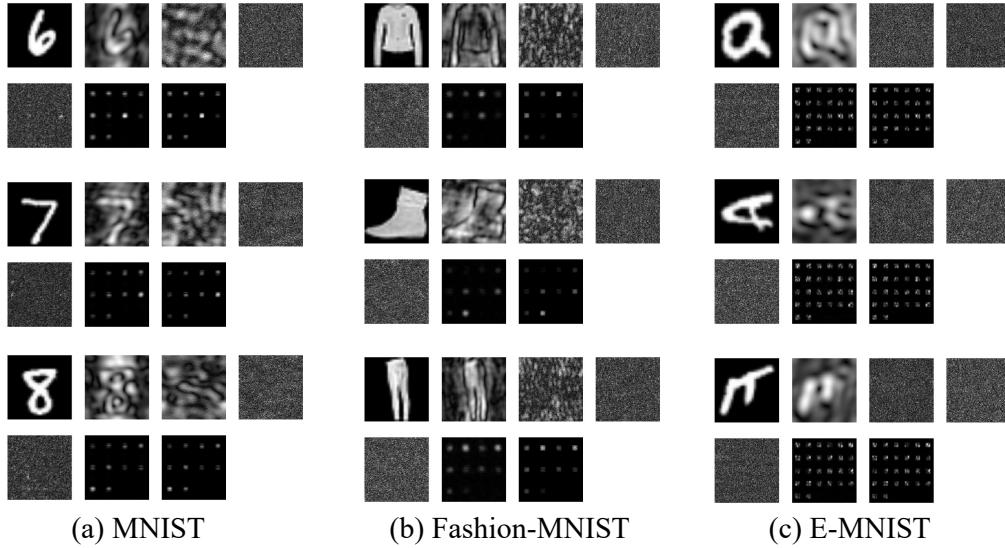


图 4-7 基础衍射光神经网络在灰度图数据集上可视化

4-9, 图4-10所示。

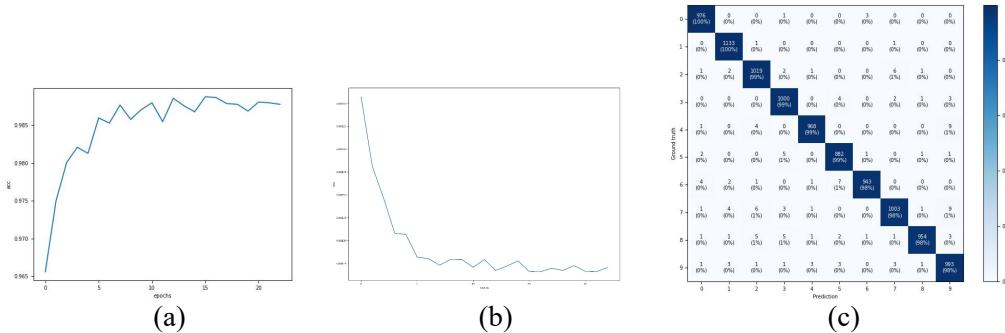


图 4-8 带非线性激活函数衍射光神经网络在 MNIST 数据集上实验结果

其中可视化结果如图4-11所示，每个例子的图像含义和图4-7一样。不同数据集在含有激活函数的 D2NN 上实验的训练过程中相关结果分别如图4-8，图

4.5 手写数字及字母识别软件

在训练好权重参数后，利用 Python 语言和 Pygame 包实现了手写数字和手写字母的 D2NN 传播的可视化可操作软件。软件界面图如下所示，右下三个按钮从上往下依次是数字识别功能，字母识别功能，绘图区清空按钮。然后最左边黑框为绘图区，可以使用鼠标左键绘制数字或者字母，右键擦除。右边六个黑框则是可视化了 D2NN 中输入、中间层和输出层光场衍射情况。最后，绘图区下方则是展示了网络输出判断结果，初始为 ‘NULL’，图4-12为软件初始界面。

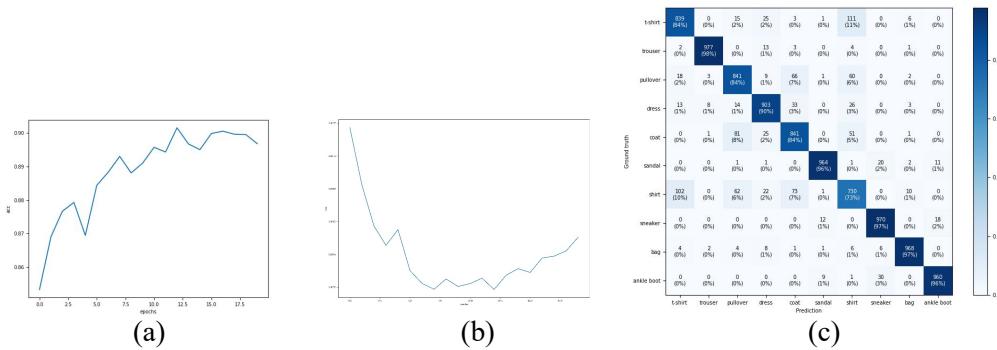


图 4-9 带非线性激活函数衍射光神经网络在 Fashion-MNIST 数据集上实验结果

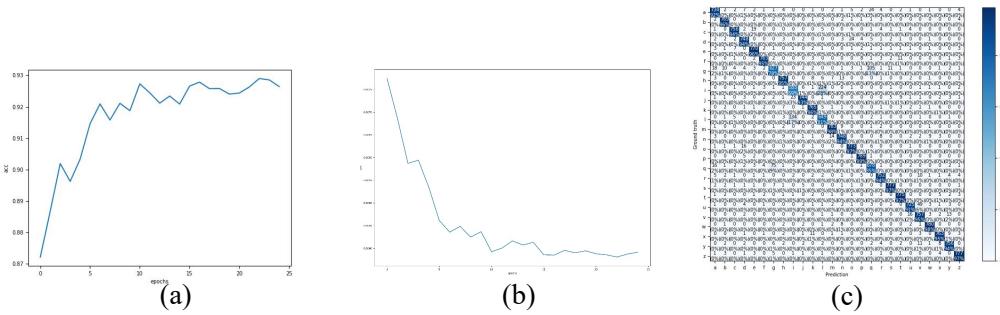


图 4-10 带非线性激活函数衍射光神经网络在 E-MNIST 数据集上实验结果

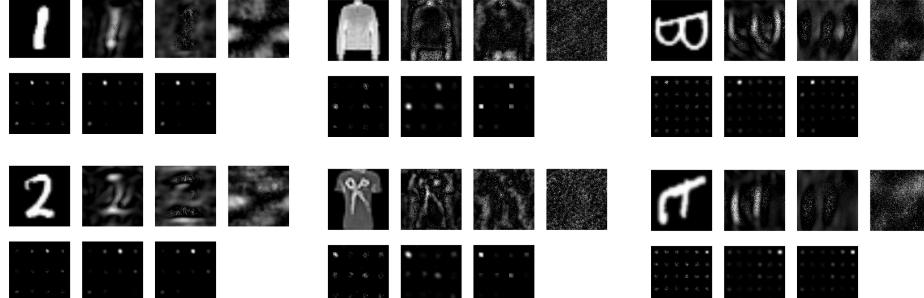


图 4-11 带非线性激活函数衍射光神经网络在灰度图数据集上可视化

该软件伪代码如4-3所示。

算法 4-3 软件伪代码

```

Data: 屏幕获取手写矩阵  $M_{in}$ 
Result: 识别结果  $R$ 

1 initialDrawBoard();
2  $Net_n.load(weight_n);$ 
3  $Net_l.load(weight_l);$ 
4 while Not exit do
5   detectMouseEvent();
6   if select number detection then
7     task = 'number';
8     result =  $Net_n(M_{in});$ 
9   else
10    task = 'letter';
11    result =  $Net_l(M_{in});$ 
12  end
13  if Click on 'Clear' button then
14    result = 'NULL';
15    clearDrawBoard();
16  else
17    updateDLayerResult();
18  end
19  showDLayerResult();
20 end
```

图4-13为该软件实际操作截图，分布测试了三种手写数字和三种手写字母，均判断正确。

4.6 本章小结

本章构建了传统的衍射光神经网络和添加了激活函数的衍射光神经网络，对常用灰度图数据集 MNIST, Fashion-MNIST 和 E-MNIST 进行了实验，构建了相应的两类衍射光神经网络，均得到了较高的准确率结果。而其中带有激活函数的模型效果明显好于不带激活函数的模型，验证了在衍射光神经网络中，添加激活函

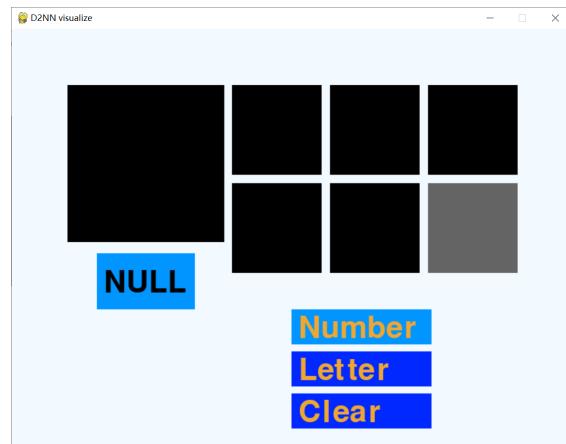


图 4-12 衍射光神经网络可视化操作软件初始界面图

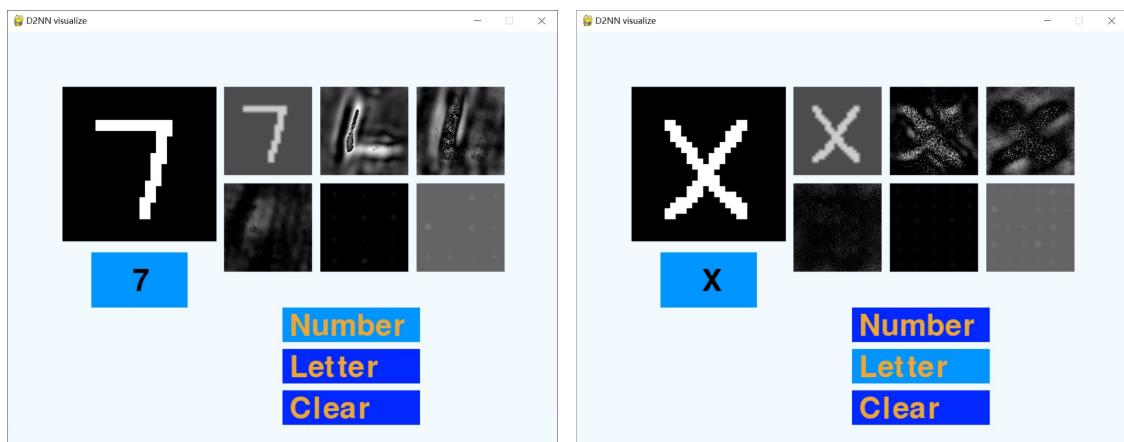


图 4-13 衍射光神经网络可视化操作软件实际操作图

数后能够提升其在图像识别分类任务中的学习表达和泛化能力。最后编写了手写数字和字母的识别软件，进行了一定的应用拓展。

第五章 彩色图像分类

彩色图像的分类识别任务是计算机视觉中更为普遍存在的任务，现实世界中绝大多数生物包括人类都有识别不同颜色的能力，所以人工智能也提出了许多处理彩色图像的神经网络，比如多通道卷积神经网络等。电子上的神经网络在彩色图像分类任务里已经有非常出色的表现^[86]，但光学神经网络对彩色图像研究较少，所以本章则以 CIFAR-10 数据集为例，对衍射光神经网络的颜色识别能力和彩色图像分类能力进行研究讨论。

5.1 彩色图像难点

彩色图像相对灰度图像，在信息维度上更高，一般以 RGB 三通道形式的彩色图像相比单通道灰度图而言，包含了红色、绿色和蓝色三原色的强度信息，借此组合出不同色彩的效果。而衍射光神经网络，由于基于瑞利-索莫非衍射模型，使用单色光，如图5-1所示，只能以光强度记录单通道的信息，所以对灰度图处理有天然优势但对于多通道的彩色图像处理则有诸多限制，若单纯将彩色图像灰度化，则会丢失有关颜色的许多信息，这样对于数据集而言是一种浪费。

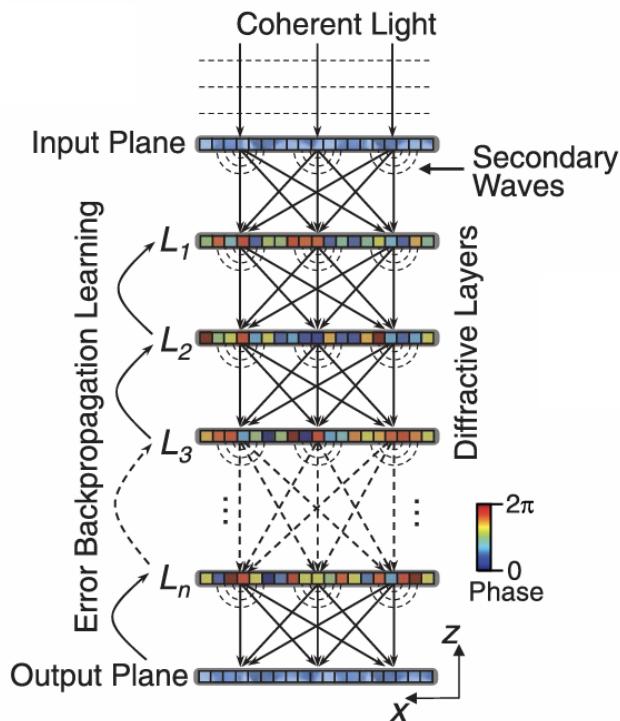


图 5-1 衍射光神经网络单色光输入图示^[56]

所以针对上述问题，本文将提出两种新型衍射光神经网络结构，保存彩色图像原有的三通道信息，并提升模型对其识别分类能力。

5.2 彩色图像数据集

本文选取了经典的彩色图像分类数据集 CIFAR-10^[87]，该数据集包含三通道彩色图像，共有 airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck 十种类别，训练集 50000 张，测试集 10000 张，各类别在训练集和测试集中均等数量出现，图像大小为 32×32 像素。相比上一章使用的灰度图数据集，该数据集因包含背景，所以分类难度更大，再加上彩色信息对于衍射光神经网络来说本身是一个难点，所以效果难以达到灰度图数据集和一些较为先进的电子神经网络的同等准确率。

在后续实验中，均将 32×32 大小图像插值扩充为 128×128 大小，方便契合衍射光神经网络输入尺寸，也使得网络具有更多的可学习单元。

5.3 波分复用衍射光神经网络

已有团队提出多频率和多通道的衍射光神经网络^[88,89]，所以利用此特性，本文提出波分复用衍射神经网络（MDW-D2NNet），其结构如图5-2所示，其将输入彩色图像的三个通道信息分开，利用不同频率光传入输入衍射层，然后将其加权求和得到融合光场信息，最后再通过分类衍射层输出，得到最终分类结果。不妨

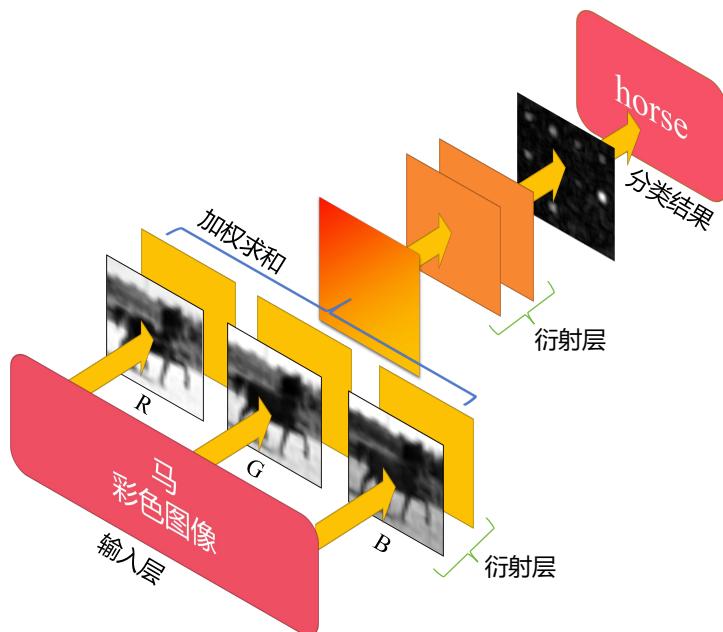


图 5-2 波分复用衍射神经网络结构示意图

假设输入图像为 $I = (r, g, b)$, 输入层映射为 f , 则融合光场 I' 可得为:

$$I' = \frac{w_r f(r) + w_g f(g) + w_b f(b)}{w_r + w_g + w_b} \quad (5-1)$$

其中, w_r, w_g, w_b 为每一通道的可学习权重参数, 初始均为 1.0, 随着网络训练而自适应变化。经过实验, 最终将三通道处理模块设置为两层衍射层, 分类模块设置为 3 层衍射层。检测区域中每一个分类对于检测块大小设置为 20×20 像素。

5.4 位置复用衍射光神经网络

位置复用衍射光神经网络 (PDW-D2NNet) 结构如图5-3所示, 其将输入彩色图像的三个通道信息与其灰度化信息分别拼接起来, 原输入大小插值扩充到 128×128 像素后, 再拼接为 256×256 大小的位置复用模块, 将颜色信息通过位置复用后形成新的光场信号输入, 然后再通过 4 层衍射层进行分类判断, 最后输出层由光子检测器检测后输出最终分类结果。

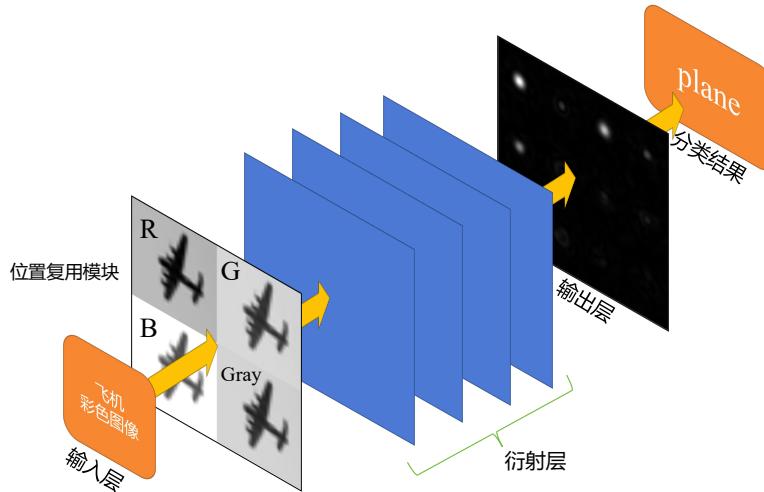


图 5-3 位置复用衍射光神经网络结构示意图

不妨假设输入图像为 $I = (r, g, b)$, 其灰度化结果为 I_{gray} , 则经过位置复用模块 P 后, 新的光场 I' 输入可表示为:

$$I' = \begin{bmatrix} r & g \\ b & I_{gray} \end{bmatrix} \quad (5-2)$$

同波分复用, 最后输出检测区域的每一类的区块大小设置为 20×20 像素。

5.5 颜色分辨能力

为了验证传统网络与新的两种架构网络在颜色分辨能力上的区别，排除在灰度图中仍然能体现图像特征的形状、大小等影响，本文利用纯色块，构造了含有RGB信息的数据集，其中类别标签是按照RGB中值最大的色彩通道作为结果，所以为一个三分类任务。训练集通过随机RGB三个值的方式，构建了30000张图像（类别均衡），测试集通过同样的方式，构建了6000张图像（类别均衡）。训练集与测试集图像大小均为 128×128 像素。

彩色图像可以通过三通道加权求和的方式转换成灰度图，转换公式如下：

$$M_{gray} = 0.299M_r + 0.587M_g + 0.114M_b \quad (5-3)$$

其中 M_{gray} 表示最后的灰度图矩阵， M_r, M_g, M_b 分别表示原图的RGB三个通道矩阵。

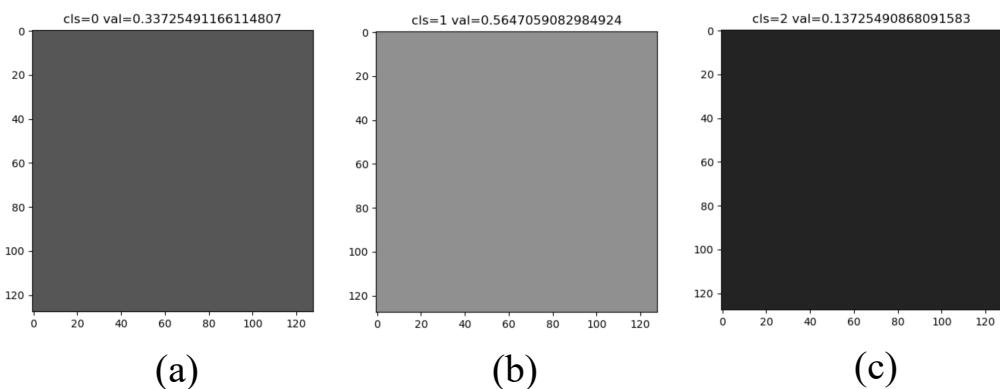


图 5-4 不同的彩色图在灰度化后示例

虽然绝大多数情况下两张颜色不同的彩色图在灰度化后会有区别，如图5-4所示，为偏红、偏绿、偏蓝三种颜色灰度化后结果，但由灰度图值计算公式5-3可以看出，颜色不一样时灰度值会存在完全一样的情况，例子如图所示，其中图5-4(a)色彩RGB值为(255, 0, 55)，由公式计算推导可得，与其相似的图5-4(c)色彩RGB值为(179, 0, 255)，这二者对应灰度化图像图5-4(b)和图5-4(d)的RGB值均为(81, 81, 81)，所以理论上传统网络只使用灰度图是会出现无法分辨的情况。

实际实验结果混淆矩阵如图5-6所示，可以看出传统网络结果如图5-6(a)准确率在33.33%，完全没有分辨能力，而两种新网络框架结果如图5-6(b)和(c)，都在100.00%准确率，说明新型框架均有较好的色彩识别分辨能力。

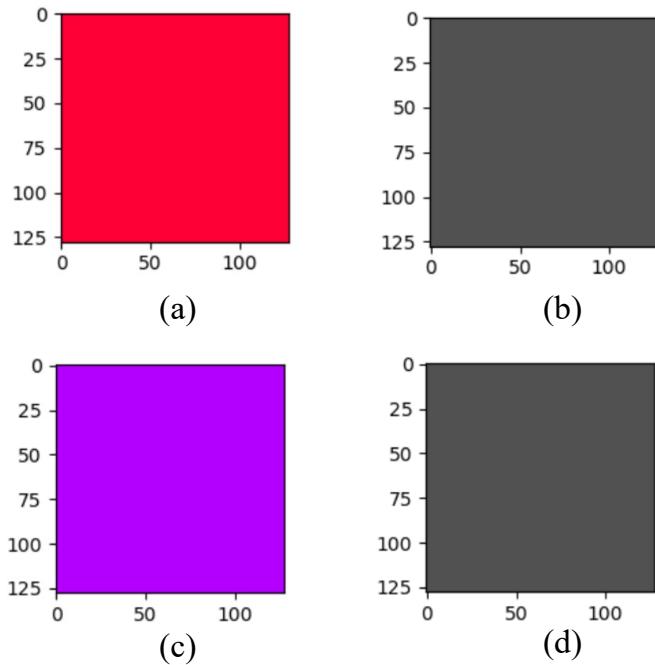


图 5-5 不同的彩色图在灰度化后相同的示例

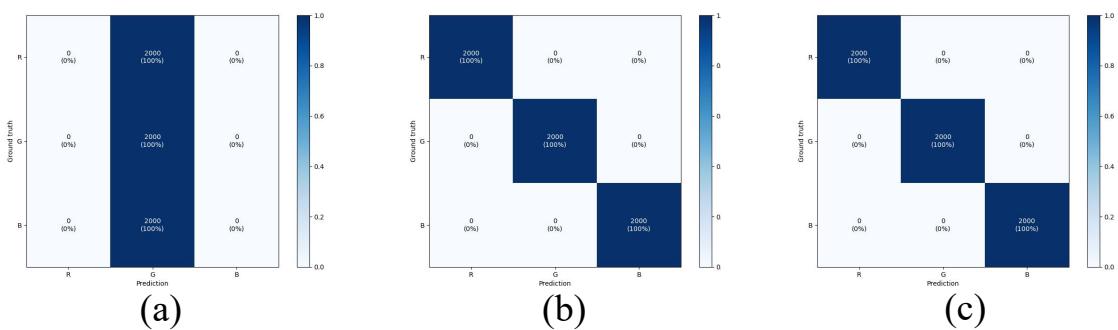


图 5-6 色彩识别数据集训练结果

5.6 彩色图像识别分类实验仿真结果

实验中，设置 Epoch 数为 40，耐心值为 8（若超过 8 个 Epoch 准确率没有提升则提前停止训练），学习率初始为 1×10^{-3} ，然后每 5 个 Epoch 以 90% 比例衰减，使用 Adam 优化器，betas 区间为 $(0.99, 0.999)$ ，权重衰减和精度均为 1×10^{-5} ，随机数种子固定为 42，batch 大小为 64，图像插值扩大到 128×128 像素，然后进行归一化。

表 5-1 不同模型在 CIFAR-10 数据集上测试结果

模型	准确率
D2NN+ReLU	0.4609
FCNet	0.5056
PDW-D2NNNet	0.6085
MDW-D2NNNet	0.5927

实验结果如表5-1所示，其中 D2NN 为灰度图实验中所用效果最好的模型，将 CIFAR-10 数据集灰度化后输入进行训练测试，准确率只有 46.09%，表现最差，验证了原始的衍射光神经网络在彩色数据集上分辨能力确实较差。FCNet 为电子上的全连接神经网络，具体设置与详细参数详见附录A.3，其效果稍好于原始衍射光神经网络，但也只有 50.56%。

两种新型网络相比之前方法，准确率都有较大提升，其中位置复用效果略好于波分复用，高出约 1.58%。分析其原因，波分复用在融合通道信息时是简单的加权相加，本质上还是类似于灰度化操作，虽然有三个通道信息的处理，但在此步骤会丢失一定的信息，导致效果略差；而位置复用不仅保留了三通道的所有信息，还附加了灰度图信息，所以效果稍好。

实验的训练过程中测试集准确率变化曲线，loss 变化曲线和最终测试集混淆矩阵结果分别如图5-7(a)、(b) 和 (c)，图5-8(a)、(b) 和 (c)，图5-9(a)、(b) 和 (c)，图5-10(a)、(b) 和 (c) 所示。

具体测试例子可视化如图5-11和图5-12所示，其中图5-12(a)表示 MDW-D2NNNet 的三通道衍射处理结果，图5-12(b) 则为将前面 RGB 三通道结果加权融合后，输入分类器衍射模块所得到结果。

5.7 本章小结

本章描述了本文新提出的两种针对彩色图像的衍射光神经网络模型，并通过纯色块的对比，验证其颜色学习能力。然后在 CIFAR-10 数据集上的训练测试，和

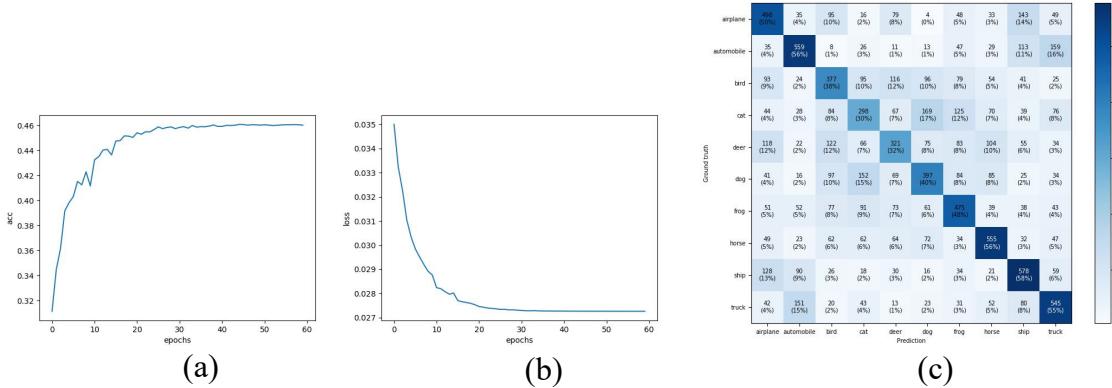


图 5-7 D2NN+ReLU 在 CIFAR-10 数据集上实验结果

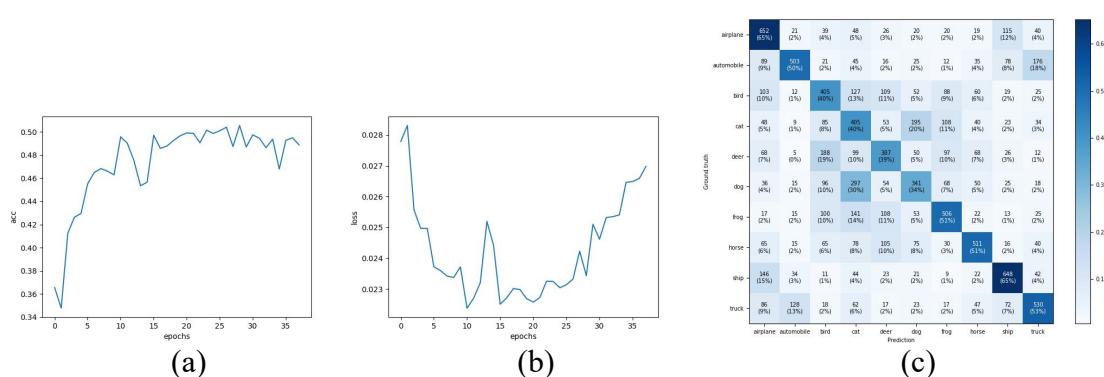


图 5-8 FCNet 在 CIFAR-10 数据集上实验结果

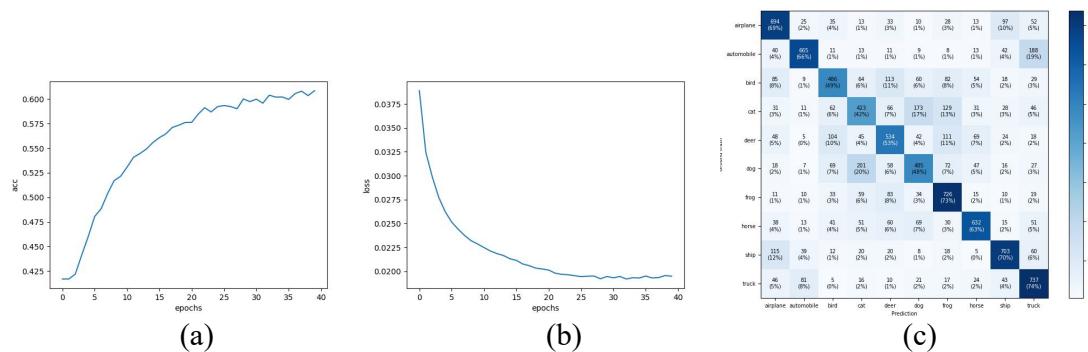


图 5-9 PDW-D2NNet 在 CIFAR-10 数据集上

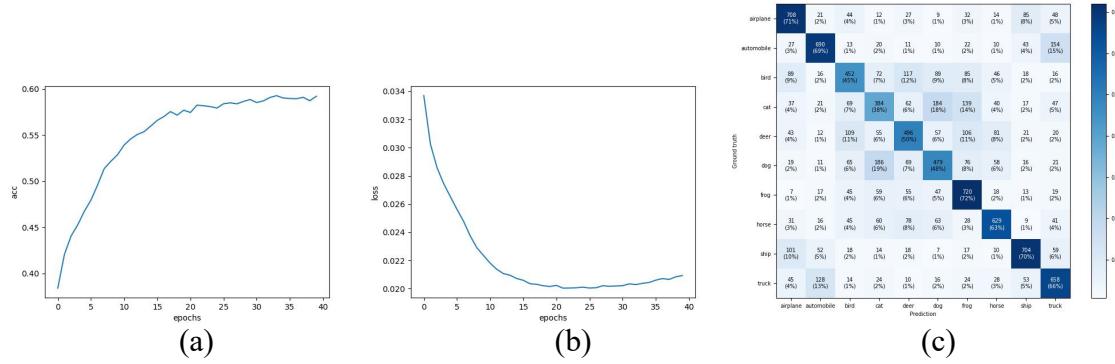


图 5-10 MDW-D2NNet 在 CIFAR-10 数据集上

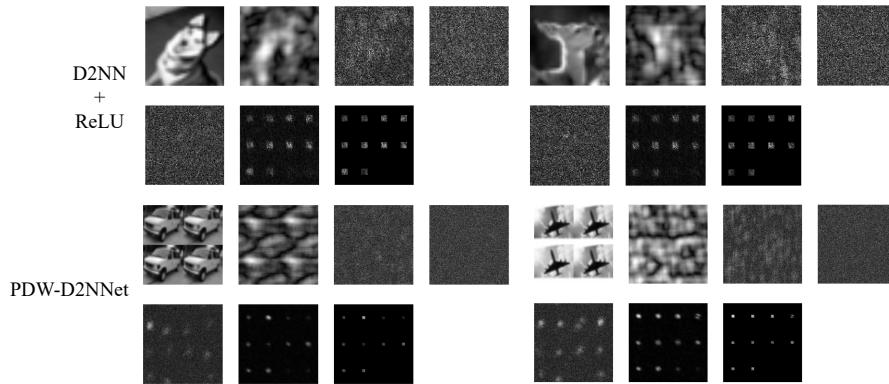


图 5-11 D2NN 与 PDW-D2NNet 的可视化例子

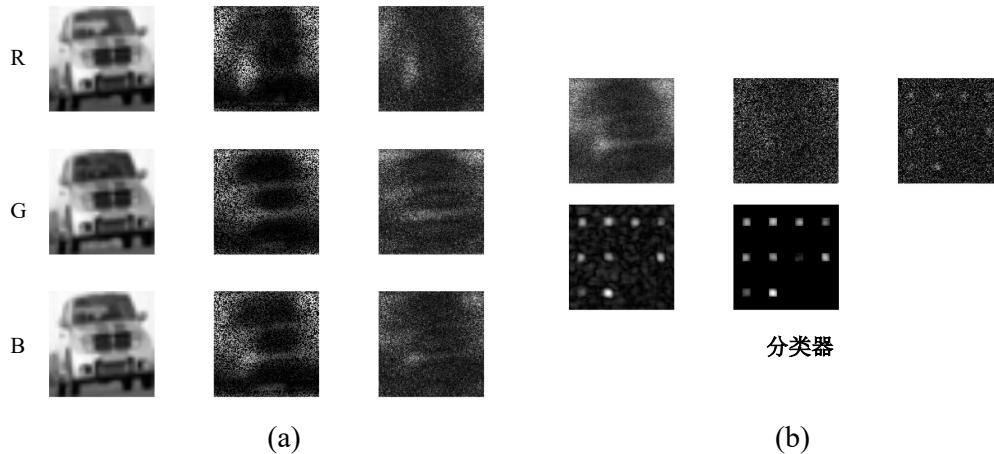


图 5-12 MDW-D2NNet 的可视化例子

传统衍射光神经网络和全连接神经网络相比，准确率有一定的提升，也证明了新的模型在彩色图像方面有更强的处理能力。

第六章 全文总结与展望

6.1 全文总结

光计算理论优势在如今电子计算系统发展瓶颈时期更加凸显出来，光学神经网络 ONN 更是具有低功耗、高并行、快速度等诸多优点，但目前因为物理理论和光学元器件等方面还存在许多困难，导致实际应用仍然较少。

为进一步研究光学智能计算，本文结合目前 AI 理论和电子上 AI 中已有的神经网络框架，主要研究了衍射光神经网络，通过理论分析推导，模型设计、仿真实验来对图像分类任务的数据集进行了训练测试，并取得了较好的结果。分别在 MNIST, Fashion-MNIST, E-MNIST 数据集上进行了实验，对比了是否加入非线性激活函数的效果，验证了加入非线性激活函数后模型的学习表达能力更强。

在完成传统灰度图像识别分类，验证了衍射光神经网络的可行性之后，本文转向目前衍射光神经网络较为困难的彩色图像识别分类任务，通过位置复用和波分复用两种方式，提出了 PDW-D2NNNet 和 MDW-D2NNNet 两种新网络结构，保留了彩色图像三通道的信息。先初步使用自己构建的纯颜色块数据集进行了颜色识别能力的验证。然后在 CIFAR-10 彩色图像识别分类数据集上进行实验测试后，和原始 D2NN 网络与电子上的全连接网络效果进行对比，有较大的提升，验证了新型网络模型的结构对于彩色图像识别分类任务有一定的优势。

6.2 未来工作展望

本文初步对衍射光神经网络进行了实验探索，在非线性激活函数和彩色图像识别方面重点做了模拟仿真实验，对比了新老网络模型和光电网路模型之间的区别。但由于资金与条件限制，无法进行实际物理环境实验测试；也由于设备和时间等限制，无法进行更大规模网络计算测试；本文主要侧重于电脑上的模拟仿真实验，对于一些物理与光学的原理和光学元器件及光学材料的讨论较少。

未来，光智能计算与人工智能仍有许多地方可以借鉴融合，互相促进发展；本文也有一些可以进一步研究的工作，在这里简述如下：

- (1) 实际物理仿真实验：将仿真实验与实际物理网络搭建相结合，互相验证，推进衍射光神经网络的实际应用发展。
- (2) 更复杂的网络结构：不仅类比全连接神经网络，更要与现有人工智能中，电子上的复杂网络结构相融合，如 ResNet 的残差模块^[90]，构建更加复杂，学习表达能力更强的衍射光神经网络模型。

- (3) 更多光学建模方式：不仅衍射，还有更多光学性质可以进行研究，应用于光智能计算之中。
- (4) 光智能芯片与光学器材：微型化、便携化，并且进一步降低成本，使其大规模应用能够实现。

总的来说，光计算与人工智能相结合是未来发展的必然趋势，传统电子计算系统也将逐渐会被光子计算系统所替代，运算更快、减少能耗，为人类社会文明发展提供更大的助力。

致 谢

在攻读本科学位期间，首先衷心感谢我的指导老师唐泳和黎升好教授，他们在我学习生活中给予了我很大帮助；也感谢为我解答疑惑的肖师姐，在毕业设计中，师姐为我耐心地解答了很多疑难问题，并给了我许多好的建议；也感谢答辩老师所提出的宝贵修改建议；然后感谢大学四年给我授课解惑的老师们，也感谢在学习生活中帮助过我的学院老师、辅导员和同学们；最后感谢一直在后面默默支持我的父母和亲人们。

附录 A 相关补充

A.1 物理相关定理推导

亥姆霍兹公-基尔霍夫积分公式求解，基尔霍夫衍射推理和瑞利-索莫非衍射模型具体推导，相关定理公式在原文中已经写出，具体推导细节详见参考文献^[22]。

A.2 空间光调制器

空间光调制器（SLM）^[91,92] 是一种较为先进的光学器件。可以在主动控制下，通过调整其内在属性来调制输入光的振幅、相位等参量。它通过精确量化及定向控制光波的波前和波束，确保在最终接收处能够产生预期的光场分布。SLM 由众多微小单元构成，每个单元均由独立的光信号或电信号进行精准控制，所以无论是一维还是二维阵列，均可以根据调制信号的变化来调整自身参数，实现逐像素级别的光场调控。

由于 SLM 具备高空间分辨率、可编程控制以及高度集成化等优点，所以衍射光神经网络中主要使用 SLM 进行实际环境测试，但本文并未进行实际环境实验，所以具体相关原理推导以及实验测试细节，详见参考文献^[22]。

A.3 电子全连接神经网络设置

在彩色图像 CIFAR-10 的实验中，FCNet 的结构为 5 层全连接层，每层后接上 ReLU 激活函数，5 层的参数分别为 $128 \times 128 \times 3, 1024, 512, 256, 128$ ，最后输出长度为 10 的向量，表示十分类结果概率。

参考文献

- [1] Franklin A. D. Nanomaterials in transistors: From high-performance to thin-film applications[J]. Science, 2015, 349(6249): aab2750.
- [2] Jain A. K, Mao J, Mohiuddin K. M. Artificial neural networks: A tutorial[J]. Computer, 1996, 29(3): 31-44.
- [3] Schmidhuber J. Deep learning in neural networks: An overview[J]. Neural networks, 2015, 61: 85-117.
- [4] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [5] Krizhevsky A, Sutskever I, Hinton G. E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.
- [6] Voulodimos A, Doulamis N, Doulamis A, et al. Deep learning for computer vision: A brief review[J]. Computational intelligence and neuroscience, 2018, 2018.
- [7] Young T, Hazarika D, Poria S, et al. Recent trends in deep learning based natural language processing[J]. ieee Computational intelligenCe magazine, 2018, 13(3): 55-75.
- [8] Harshvardhan G, Gourisaria M. K, Pandey M, et al. A comprehensive survey and analysis of generative models in machine learning[J]. Computer Science Review, 2020, 38: 100285.
- [9] Croitoru F.-A, Hondu V, Ionescu R. T, et al. Diffusion models in vision: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
- [10] Yang L, Zhang Z, Song Y, et al. Diffusion models: A comprehensive survey of methods and applications[J]. ACM Computing Surveys, 2023, 56(4): 1-39.
- [11] Wu T, He S, Liu J, et al. A brief overview of chatgpt: The history, status quo and potential future development[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(5): 1122-1136.
- [12] Zhang C, Zhang C, Zhang M, et al. Text-to-image diffusion model in generative ai: A survey[J]. arXiv preprint arXiv:2303.07909, 2023.
- [13] Chen Z, Li S, Haque M. A. An overview of openai's sora and its potential for physics engine free games and virtual reality[J]. EAI Endorsed Transactions on AI and Robotics, 2024, 3.
- [14] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, 4015-4026.
- [15] Markov I. L. Limits on fundamental limits to computation[J]. Nature, 2014, 512(7513): 147-154.

- [16] Sui X, Wu Q, Liu J, et al. A review of optical neural networks[J]. IEEE Access, 2020, 8: 70773-70783.
- [17] Caulfield H. J, Kinser J, Rogers S. K. Optical neural networks[J]. Proceedings of the IEEE, 1989, 77(10): 1573-1583.
- [18] 成骏伟, 江雪怡, 周海龙, et al. 光电智能计算研究进展与挑战 [J]. 中国激光, 2022, 49(12): 13.
- [19] 项水英, 宋紫薇, 高爽, et al. 光神经形态计算研究进展与展望 (特邀)[J]. 光子学报, 2021, 50(10): 17.
- [20] 马婷, 李万杰, 冯佳楠, et al. 光脉冲神经网络研究进展 [J]. 光学与光电技术, 2022, 20(4): 16.
- [21] 陈宏伟, 于振明, 张天, et al. 光子神经网络发展与挑战 [J]. 中国激光, 2020, 47(5): 12.
- [22] 王子荣. 面向图像分类识别应用的衍射光神经网络系统设计与实现 [D]. 中国科学院大学 (中国科学院长春光学精密机械与物理研究所), 2023, .
- [23] McCulloch W. S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5: 115-133.
- [24] Saygin A. P, Cicekli I, Akman V. Turing test: 50 years later[J]. Minds and machines, 2000, 10(4): 463-518.
- [25] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological review, 1958, 65(6): 386.
- [26] Rumelhart D. E, Hinton G. E, Williams R. J. Learning representations by back-propagating errors[J]. nature, 1986, 323(6088): 533-536.
- [27] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. IEEE transactions on neural networks and learning systems, 2020, 32(1): 4-24.
- [28] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: Lstm cells and network architectures[J]. Neural computation, 2019, 31(7): 1235-1270.
- [29] Aggarwal A, Mittal M, Battineni G. Generative adversarial network: An overview of theory and applications[J]. International Journal of Information Management Data Insights, 2021, 1(1): 100004.
- [30] Cambria E, White B. Jumping nlp curves: A review of natural language processing research[J]. IEEE Computational intelligence magazine, 2014, 9(2): 48-57.
- [31] Khan A. A, Laghari A. A, Awan S. A. Machine learning in computer vision: a review[J]. EAI Endorsed Transactions on Scalable Information Systems, 2021, 8(32): e4-e4.

- [32] Gaikwad S. K, Gawali B. W, Yannawar P. A review on speech recognition technique[J]. International Journal of Computer Applications, 2010, 10(3): 16-24.
- [33] Chandolikar N, Joshi C, Roy P, et al. Voice recognition: A comprehensive survey[C]. 2022 International Mobile and Embedded Technology Conference (MECON), 2022, 45-51.
- [34] Ayoub J, Zhou F, Bao S, et al. From manual driving to automated driving: A review of 10 years of autoui[C]. Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications, 2019, 70-90.
- [35] Cao L. Ai in finance: A review[J]. Available at SSRN 3647625, 2020.
- [36] Cao L. Ai in finance: challenges, techniques, and opportunities[J]. ACM Computing Surveys (CSUR), 2022, 55(3): 1-38.
- [37] Guo X, Shen Z, Zhang Y, et al. Review on the application of artificial intelligence in smart homes[J]. Smart Cities, 2019, 2(3): 402-420.
- [38] Margaje S. C, Rasal V. A review on smart retail shop using artificial intelligence (ai) and internet of things (iot)[J]. International Research Journal of Modernization in Engineering Technology and Science, 2021, 3(11): 949-954.
- [39] Wan J, Yang J, Wang Z, et al. Artificial intelligence for cloud-assisted smart factory[J]. IEEE Access, 2018, 6: 55419-55430.
- [40] Edvards J. Building a smart factory with ai and robotics[J]. Robotics Business Review, 2022.
- [41] Elliott D, Soifer E. Ai technologies, privacy, and security[J]. Frontiers in Artificial Intelligence, 2022, 5: 826737.
- [42] Petropoulos G. The impact of artificial intelligence on employment[J]. Praise for Work in the Digital Age, 2018, 119: 121.
- [43] Savulescu J, Maslen H. Moral enhancement and artificial intelligence: moral ai?[J]. Beyond artificial intelligence: The disappearing human-machine divide, 2015, 79-95.
- [44] Goodman J. W. Introduction to fourier optics[M]. Roberts and Company publishers, 2005.
- [45] Goodman J. W, Dias A, Woody L. Fully parallel, high-speed incoherent optical method for performing discrete fourier transforms[J]. Optics Letters, 1978, 2(1): 1-3.
- [46] Farhat N. H, Psaltis D, Prata A, et al. Optical implementation of the hopfield model[J]. Applied optics, 1985, 24(10): 1469-1475.
- [47] Xu R, Lv P, Xu F, et al. A survey of approaches for implementing optical neural networks[J]. Optics & Laser Technology, 2021, 136: 106787.

- [48] Psaltis D, Farhat N. Optical information processing based on an associative-memory model of neural nets with thresholding and feedback[J]. Optics Letters, 1985, 10(2): 98-100.
- [49] Collins D. R, Sampsell J. B, Hornbeck L. J, et al. Deformable mirror device spatial light modulators and their applicability to optical neural networks[J]. Applied optics, 1989, 28(22): 4900-4907.
- [50] Lu T, Wu S, Xu X, et al. Two-dimensional programmable optical neural network[J]. Applied optics, 1989, 28(22): 4908-4913.
- [51] Chang J, Sitzmann V, Dun X, et al. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification[J]. Scientific reports, 2018, 8(1): 1-10.
- [52] Gu Z, Gao Y, Liu X. Optronic convolutional neural networks of multi-layers with different functions executed in optics for image classification[J]. Optics Express, 2021, 29(4): 5877-5889.
- [53] Yan T, Wu J, Zhou T, et al. Fourier-space diffractive deep neural network[J]. Physical review letters, 2019, 123(2): 023901.
- [54] Zuo Y, Li B, Zhao Y, et al. All-optical neural network with nonlinear activation functions[J]. Optica, 2019, 6(9): 1132-1137.
- [55] Shi W, Huang Z, Huang H, et al. Loen: Lensless opto-electronic neural network empowered machine vision[J]. Light: Science & Applications, 2022, 11(1): 121.
- [56] Lin X, Rivenson Y, Yardimci N. T, et al. All-optical machine learning using diffractive deep neural networks[J]. Science, 2018, 361(6406): 1004-1008.
- [57] Luo Y, Mengu D, Yardimci N. T, et al. Design of task-specific optical systems using broadband diffractive neural networks[J]. Light: Science & Applications, 2019, 8(1): 112.
- [58] Chen H, Feng J, Jiang M, et al. Diffractive deep neural networks at visible wavelengths[J]. Engineering, 2021, 7(10): 1483-1491.
- [59] Zhou T, Lin X, Wu J, et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit[J]. Nature Photonics, 2021, 15(5): 367-373.
- [60] Lu L, Zhu L, Zhang Q, et al. Miniaturized diffraction grating design and processing for deep neural network[J]. IEEE Photonics Technology Letters, 2019, 31(24): 1952-1955.
- [61] Luo X, Hu Y, Ou X, et al. Metasurface-enabled on-chip multiplexed diffractive neural networks in the visible[J]. Light: Science & Applications, 2022, 11(1): 158.
- [62] Dong J, Gigan S, Krzakala F, et al. Scaling up echo-state networks with multiple light scattering[C]. 2018 IEEE Statistical Signal Processing Workshop (SSP), 2018, 448-452.

- [63] 刘世杰, 陈臣, 张志刚, et al. 基于透明介质成像的光学散射神经网络 [J]. 光学技术, 2022, 48(2): 159.
- [64] Reck M, Zeilinger A, Bernstein H. J, et al. Experimental realization of any discrete unitary operator[J]. Physical review letters, 1994, 73(1): 58.
- [65] Shen Y, Harris N. C, Skirlo S, et al. Deep learning with coherent nanophotonic circuits[J]. Nature photonics, 2017, 11(7): 441-446.
- [66] Xu S, Wang J, Shu H, et al. Optical coherent dot-product chip for sophisticated deep learning regression[J]. Light: Science & Applications, 2021, 10(1): 221.
- [67] Dong S, Wang P, Abbas K. A survey on deep learning and its applications[J]. Computer Science Review, 2021, 40: 100379.
- [68] Cai S, Mao Z, Wang Z, et al. Physics-informed neural networks (pinns) for fluid mechanics: A review[J]. Acta Mechanica Sinica, 2021, 37(12): 1727-1738.
- [69] Kingma D. P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [70] Zhang Z. Improved adam optimizer for deep neural networks[C]. 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS), 2018, 1-2.
- [71] Haji S. H, Abdulazeez A. M. Comparison of optimization techniques based on gradient descent algorithm: A review[J]. PalArch's Journal of Archaeology of Egypt/Egyptology, 2021, 18(4): 2715-2743.
- [72] 胡继刚. 非傍轴衍射理论的研究及其应用 [D]. 合肥工业大学, 2004, .
- [73] 李俊昌. 角谱衍射公式的快速傅里叶变换计算及在数字全息波面重建中的应用 [J]. 光学学报, 2009, 5.
- [74] 郭福源, 李连煌. 标量衍射积分公式比较分析 [J]. 光学学报, 2013, 7.
- [75] Wolf E, Marchand E. Comparison of the kirchhoff and the rayleigh–sommerfeld theories of diffraction at an aperture[J]. JOSA, 1964, 54(5): 587-594.
- [76] 樊丽晶. 广义瑞利-索末菲衍射与分数傅里叶变换关系探究 [D]. 山西师范大学, 2018, .
- [77] Liu J, Wu Q, Sui X, et al. Research progress in optical neural networks: theory, applications and developments[J]. PhotoniX, 2021, 2: 1-39.
- [78] Deng L. The mnist database of handwritten digit images for machine learning research [best of the web][J]. IEEE signal processing magazine, 2012, 29(6): 141-142.
- [79] Xiao H, Rasul K, Vollgraf R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms[J]. arXiv preprint arXiv:1708.07747, 2017.

- [80] Cohen G, Afshar S, Tapson J, et al. Emnist: Extending mnist to handwritten letters[C]. 2017 international joint conference on neural networks (IJCNN), 2017, 2921-2926.
- [81] Destras O, Beux S. L, Magalhães F. G. D, et al. Survey on activation functions for optical neural networks[J]. ACM Computing Surveys, 2023, 56(2): 1-30.
- [82] Selden A. Pulse transmission through a saturable absorber[J]. British Journal of Applied Physics, 1967, 18(6): 743.
- [83] George J, Amin R, Mehrabian A, et al. Electrooptic nonlinear activation functions for vector matrix multiplications in optical neural networks[C]. Signal Processing in Photonic Communications, 2018, SpW4G-3.
- [84] Miscuglio M, Mehrabian A, Hu Z, et al. All-optical nonlinear activation function for photonic neural networks[J]. Optical Materials Express, 2018, 8(12): 3851-3863.
- [85] Feldmann J, Youngblood N, Wright C. D, et al. All-optical spiking neurosynaptic networks with self-learning capabilities[J]. Nature, 2019, 569(7755): 208-214.
- [86] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J]. OpenReview, 2020.
- [87] Krizhevsky N, Sutskever I. Hinton[J]. The cifar-10 dataset. online: <http://www.cs.toronto.edu/kriz/cifar.html>, 2009.
- [88] Chen Y, Zhu J. An optical diffractive deep neural network with multiple frequency-channels[J]. arXiv preprint arXiv:1912.10730, 2019.
- [89] Xu Z, Yuan X, Zhou T, et al. A multichannel optical computing architecture for advanced machine vision[J]. Light: Science & Applications, 2022, 11(1): 255.
- [90] Dou H, Deng Y, Yan T, et al. Residual d 2 nn: training diffractive deep neural networks via learnable light shortcuts[J]. Optics Letters, 2020, 45(10): 2688-2691.
- [91] Savage N. Digital spatial light modulators, 2009.
- [92] 郭怀文. 液晶空间光调制器的幅度调制特性及其应用 [D]. 中国工程物理研究院, 2020, .

外文资料原文

1 All-optical machine learning using diffractive deep neural networks

RESEARCH

OPTICAL COMPUTING

All-optical machine learning using diffractive deep neural networks

Xing Lin^{1,2,3*}, Yair Rivenson^{1,2,3†}, Nezih T. Yardimeci^{1,3}, Muhammed Veli^{1,2,3}, Yi Luo^{1,2,3}, Mona Jarrahi^{1,3}, Aydogan Ozcan^{1,2,3,4†}

Deep learning has been transforming our ability to execute advanced inference tasks using computers. Here we introduce a physical mechanism to perform machine learning by demonstrating an all-optical diffractive deep neural network (D²NN) architecture that can implement various functions following the deep learning-based design of passive diffractive layers that work collectively. We created 3D-printed D²NNs that implement classification of images of handwritten digits and fashion products, as well as the function of an imaging lens at a terahertz spectrum. Our all-optical deep learning framework can perform, at the speed of light, various complex functions that computer-based neural networks can execute; will find applications in all-optical image analysis, feature detection, and object classification; and will also enable new camera designs and optical components that perform distinctive tasks using D²NNs.

Deep learning is one of the fastest-growing machine learning methods (1). This approach uses multilayered artificial neural networks implemented in a computer to digitally learn data representation and abstraction and to perform advanced tasks in a manner comparable or even superior to the performance of human experts. Recent examples in which deep learning has made major advances in machine learning include medical image analysis (2), speech recognition (3), language translation (4), and image classification (5), among others (1, 6). Beyond some of these mainstream applications, deep learning methods are also being used to solve inverse imaging problems (7–13).

Here we introduce an all-optical deep learning framework in which the neural network is physically formed by multiple layers of diffractive surfaces that work in collaboration to optically perform an arbitrary function that the network can statistically learn. Whereas the inference and prediction mechanism of the physical network is all optical, the learning part that leads to its design is done through a computer. We term this framework a diffractive deep neural network (D²NN) and demonstrate its inference capabilities through both simulations and experiments. Our D²NN can be physically created by using several transmissive and/or reflective layers (14), where each point on a given layer either transmits or reflects the incoming wave, representing an artificial neuron that is connected to other neurons of the following layers through optical diffraction (Fig. 1A). In accordance with the Huygens-Fresnel principle, our terminology is

based on each point on a given layer acting as a secondary source of a wave, the amplitude and phase of which are determined by the product of the input wave and the complex-valued transmission or reflection coefficient at that point [see (14) for an analysis of the waves within a D²NN]. Therefore, an artificial neuron in a D²NN is connected to other neurons of the following layer through a secondary wave modulated in amplitude and phase by both the input interference pattern created by the earlier layers and the local transmission or reflection coefficient at that point. As an analogy to standard deep neural networks (Fig. 1D), one can consider the transmission or reflection coefficient of each point or neuron as a multiplicative “bias” term, which is a learnable network parameter that is iteratively adjusted during the training process of the diffractive network, using an error back-propagation method. After this numerical training phase, the D²NN design is fixed and the transmission or reflection coefficients of the neurons of all layers are determined. This D²NN design—once physically fabricated using techniques such as 3D-printing or lithography—can then perform, at the speed of light, the specific task for which it is trained, using only optical diffraction and passive optical components or layers that do not need power, thereby creating an efficient and fast way of implementing machine learning tasks.

In general, the phase and amplitude of each neuron can be learnable parameters, providing a complex-valued modulation at each layer, which improves the inference performance of the diffractive network (fig. S1) (14). For coherent transmissive networks with phase-only modulation, each layer can be approximated as a thin optical element (Fig. 1). Through deep learning, the phase values of the neurons of each layer of the diffractive network are iteratively adjusted (trained) to perform a specific function by feeding training data at the input layer and then computing the network’s output through optical

diffraction. On the basis of the calculated error with respect to the target output, determined by the desired function, the network structure and its neuron phase values are optimized via an error back-propagation algorithm, which is based on the stochastic gradient descent approach used in conventional deep learning (14).

To demonstrate the performance of the D²NN framework, we first trained it as a digit classifier to perform automated classification of handwritten digits, from 0 to 9 (Figs. 1B and 2A). For this task, phase-only transmission masks were designed by training a five-layer D²NN with 55,000 images (5000 validation images) from the MNIST (Modified National Institute of Standards and Technology) handwritten digit database (15). Input digits were encoded into the amplitude of the input field to the D²NN, and the diffractive network was trained to map input digits into 10 detector regions, one for each digit. The classification criterion was to find the detector with the maximum optical signal, and this was also used as a loss function during the network training (14).

After training, the design of the D²NN digit classifier was numerically tested using 10,000 images from the MNIST test dataset (which were not used as part of the training or validation image sets) and achieved a classification accuracy of 91.75% (Fig. 3C and fig. S1). In addition to the classification performance of the diffractive network, we also analyzed the energy distribution observed at the network output plane for the same 10,000 test digits (Fig. 3C), the results of which clearly demonstrate that the diffractive network learned to focus the input energy of each handwritten digit into the correct (i.e., the target) detector region, in accord with its training. With the use of complex-valued modulation and increasing numbers of layers, neurons, and connections in the diffractive network, our classification accuracy can be further improved (figs. S1 and S2). For example, fig. S2 demonstrates a Lego-like physical transfer learning behavior for the D²NN framework, where the inference performance of an already existing D²NN can be further improved by adding new diffractive layers—or, in some cases, by peeling off (i.e., discarding) some of the existing layers—where the new layers to be added are trained for improved inference (coming from the entire diffractive network: old and new layers). By using a patch of two layers added to an existing and fixed D²NN design ($N = 5$ layers), we improved our MNIST classification accuracy to 93.39% (fig. S2) (14); the state-of-the-art convolutional neural network performance has been reported as 99.6% to 99.77% (16–18). More discussion on reconfiguring D²NN designs is provided in the supplementary materials (14).

Following these numerical results, we 3D-printed our five-layer D²NN design (Fig. 2A), with each layer having an area of 8 cm by 8 cm, followed by 10 detector regions defined at the output plane of the diffractive network (Figs. 1B and 3A). We then used continuous-wave illumination at 0.4 THz to test the network’s inference performance (Figs. 2, C and D). Phase values of

Downloaded from https://www.science.org at University of Electronic Science & Technology on July 21, 2022

*Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095, USA. †Department of Bioengineering, University of California, Los Angeles, CA 90095, USA. ²California NanoSystems Institute (CNSI), University of California, Los Angeles, CA 90095, USA.

⁴Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: ozcan@ucla.edu

RESEARCH | REPORT

each layer's neurons were physically encoded using the relative thickness of each 3D-printed neuron. Numerical testing of this five-layer D²NN design achieved a classification accuracy of 91.75% over ~10,000 test images (Fig. 3C). To quantify the match between these numerical testing results and our experiments, we 3D-printed 50 handwritten digits (five different inputs per digit), selected among the same 91.75% of the test images for which numerical testing was successful. For each input object that is uniformly illuminated with the terahertz source, we imaged the output plane of the D²NN to map the intensity distribution for each detector region that is assigned to a digit. The results (Fig. 3B) demonstrate the success of the 3D-printed diffractive neural network and its inference capability: The average intensity distribution at the output plane of the network for each input digit clearly reveals that the 3D-printed D²NN was able to focus the input energy of the beam and achieve a maximum signal at the corresponding detector region assigned for that digit. Despite 3D-printing errors, possible alignment issues, and other experimental error sources in our setup (14), the match between the experimental and numerical testing of our five-layer D²NN design was found to be 88% (Fig. 3B). This relatively small reduction in the performance

of the experimental network compared to our numerical testing is especially pronounced for the digit 0 because it is challenging to 3D-print the large void region at the center of the digit. Similar printing challenges were also observed for other digits that have void regions; e.g., 6, 8, and 9 (Fig. 3B).

Next, we tested the classification performance of D²NN framework with more complicated image dataset—i.e., the Fashion-MNIST dataset (19), which includes 10 classes, each representing a fashion product (t-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots; see fig. S3 for sample images). In general, for a coherently illuminated D²NN, we can use the amplitude and/or phase channels of the input plane to represent data to be classified or processed. In our digit classification results reported earlier, input objects were encoded by using the amplitude channel, and to demonstrate the utility of the phase channel of the network input, we encoded each input image corresponding to a fashion product as a phase-only object modulation (14). Our D²NN inference results (as a function of the number of layers, neurons, and connections) for classification of fashion products are summarized in figs. S4 and S5. To provide an example of our performance, a phase-only and a complex-

valued modulation D²NN with $N = 5$ diffractive layers (sharing the same physical network dimensions as the digit classification D²NN shown in Fig. 2A) reached an accuracy of 81.13 and 86.33%, respectively (fig. S4). By increasing the number of diffractive layers to $N = 10$ and the total number of neurons to 0.4 million, our classification accuracy increased to 86.60% (fig. S5). For convolutional neural net-based standard deep learning, the state-of-the-art performance for Fashion-MNIST classification accuracy has been reported as 96.7%, using ~8.9 million learnable parameters and ~2.5 million neurons (20).

To experimentally demonstrate the performance of fashion product classification using a physical D²NN, we 3D-printed our phase-only five-layer design and 50 fashion products used as test objects (five per class) on the basis of the same procedures employed for the digit classification diffractive network (Figs. 2A and 3), except that each input object information was encoded in the phase channel. Our results are summarized in Fig. 4, revealing a 90% match between the experimental and numerical testing of our five-layer D²NN design, with five errors out of 50 fashion products. Compared with digit classification (six errors out of 50 digits; Fig. 3), this experiment yielded a slightly better match

Downloaded from https://www.science.org at University of Electronic Science & Technology on July 21, 2022

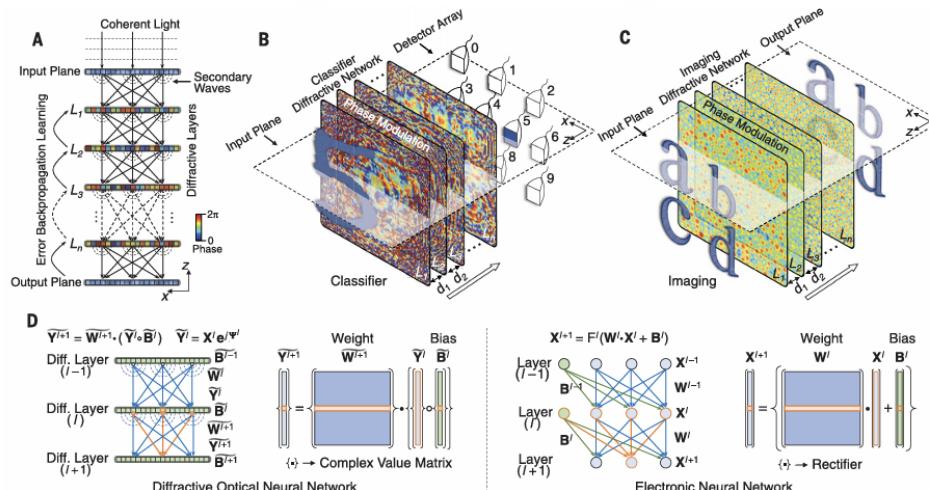


Fig. 1. Diffractive deep neural networks (D²NNs). (A) A D²NN comprises multiple transmissive (or reflective) layers, where each point on a given layer acts as a neuron, with a complex-valued transmission (or reflection) coefficient. The transmission or reflection coefficients of each layer can be trained by using deep learning to perform a function between the input and output planes of the network. After this learning phase, the D²NN design is fixed; once fabricated or 3D-printed, it performs the learned function at the speed of light. L_i , layer. (B and C) We trained and experimentally implemented different types of D²NNs: (B) classifier (for handwritten digits and fashion products) and (C) imager.

d. distance. (D) Comparison between a D²NN and a conventional neural network (14). Based on coherent waves, the D²NN operates on complex-valued inputs, with multiplicative bias terms. Weights in a D²NN are based on free-space diffraction and determine the interference of the secondary waves that are phase- and/or amplitude-modulated by the previous layers. “ \odot ” denotes a Hadamard product operation. “Electronic neural network” refers to the conventional neural network virtually implemented in a computer. \mathbf{Y} , optical field at a given layer; Ψ , phase of the optical field; \mathbf{X} , amplitude of the optical field; F , nonlinear rectifier function [see (14) for a discussion of optical nonlinearity in D²NN].

RESEARCH | REPORT

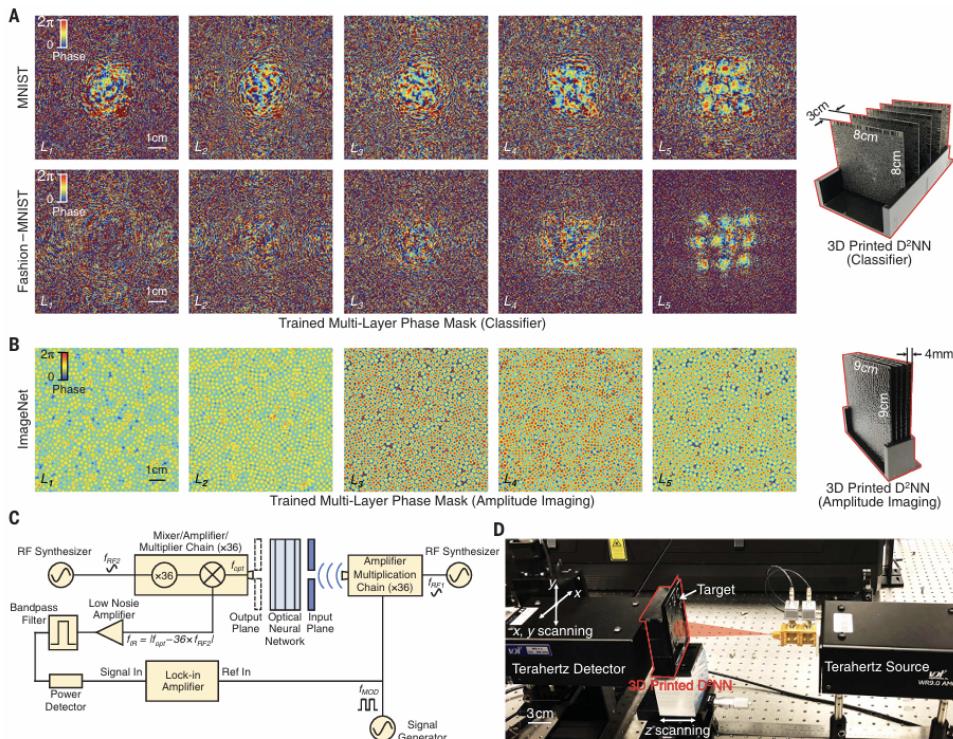


Fig. 2. Experimental testing of 3D-printed D²NNs. (A and B) After the training phase, the final designs of five different layers (L_1, L_2, \dots, L_5) of the handwritten digit classifier, fashion product classifier, and the imager D²NNs are shown. To the right of the network layers, an illustration of the corresponding 3D-printed D²NN is shown. (C and D) Schematic (C) and photo (D) of the experimental terahertz setup. An amplifier-multiplexer chain was used for the detection at the output plane of the network. RF, radio frequency; f , frequency.

Downloaded from https://www.science.org at University of Electronic Science & Technology on July 21, 2022

between the experimental and numerical testing results (despite the more challenging nature of Fashion-MNIST dataset), perhaps because we used the phase channel, which does not suffer from the challenges associated with 3D-printing of void regions [such as in digits 0, 6, 8, and 9 (Fig. 3)], to encode input image information for fashion products.

Next, we tested the performance of a phase-only D²NN, composed of five 3D-printed transmission layers to implement amplitude imaging (Fig. 2B). This network was trained using the ImageNet database (27) to create a unit-magnification image of the input optical field amplitude at its output plane (~9 cm by 9 cm)—that is, the output image has the same physical size as the input object (14). As illustrated in fig. S6, A and C, the trained network initially connects every amplitude point at the input plane to various neurons and features of the fol-

lowing layers, which then focus the light back to a point at the output (i.e., image) plane, which is, as expected, quite different than the case of free-space diffraction (i.e., without the presence of the diffractive network), illustrated in fig. S6, B and D.

After training and blind testing, which served to numerically prove the imaging capability of the network (figs. S6 and S7), we then 3D-printed this designed D²NN. Using the same experimental setup shown in Fig. 2, C and D, we imaged the output plane of the 3D-printed D²NN for various input objects that were uniformly illuminated by continuous-wave radiation at 0.4 THz. Figure S8 summarizes our experimental results achieved with this 3D-printed D²NN, which successfully projected unit-magnification images of the input patterns at the output plane of the network, learning the function of an imager, or a physical auto-encoder. To evaluate the point spread

function of this D²NN, we imaged pinholes with different diameters (1, 2, and 3 mm), which resulted in output images, each with a full width at half maximum of 1.5, 1.4, and 2.5 mm, respectively (fig. S8B). Our results also revealed that the printed network can resolve a linewidth of 1.8 mm at 0.4 THz (corresponding to a wavelength of 0.75 mm in air), which is slightly worse in resolution compared with the numerical testing of our D²NN design, where the network could resolve a linewidth of ~1.2 mm (fig. S7C). This experimental degradation in the performance of the diffractive network can be due to factors such as 3D-printing errors, potential misalignments, and absorption-related losses in the 3D-printed network (14).

Optical implementation of machine learning in artificial neural networks is promising because of the parallel computing capability and power efficiency of optical systems (22–24). Compared

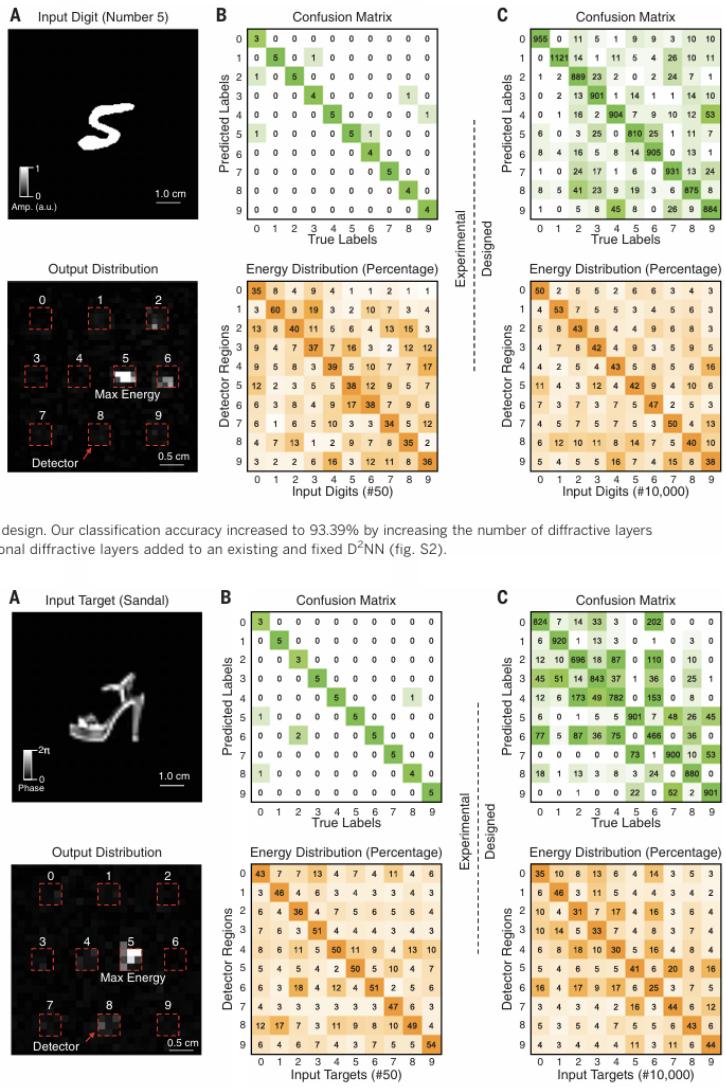
RESEARCH | REPORT

Fig. 3. Handwritten digit classifier D²NN.

(A) A 3D-printed D²NN successfully classifies handwritten input digits (0, 1, ..., 9) on the basis of 10 different detector regions at the output plane of the network, each corresponding to one digit. As an example, the output image of the 3D-printed D²NN for a handwritten input of "5" is demonstrated, where the red dashed squares represent the trained detector regions for each digit. Other examples of our experimental results are shown in fig. S9. (B) Confusion matrix and energy distribution percentage for our experimental results, using 50 different handwritten digits (five for each digit) that were 3D-printed, selected among the images for which numerical testing was successful. (C) Same as (B), except summarizing our numerical testing results for 10,000 different handwritten digits (~1000 for each digit), achieving a classification accuracy of 91.75% using a five-layer design. Our classification accuracy increased to 93.39% by increasing the number of diffractive layers to seven, using a patch of two additional diffractive layers added to an existing and fixed D²NN (fig. S2).

Fig. 4. Fashion product classifier D²NN.

(A) As an example, the output image of the 3D-printed D²NN for a sandal input (Fashion-MNIST class 5) is demonstrated. The red dashed squares represent the trained detector regions for each fashion product. Other examples of our experimental results are shown in fig. S10. (B) Confusion matrix and energy distribution percentage for our experimental results, using 50 different fashion products (five per class) that were 3D-printed, selected among the images for which numerical testing was successful. (C) Same as (B), except summarizing our numerical testing results for 10,000 different fashion products (~1000 per class), achieving a classification accuracy of 81.13% using a five-layer design. By increasing the number of diffractive layers to 10, our classification accuracy increased to 86.60% (fig. S5).



Downloaded from https://www.science.org at University of Electronic Science & Technology on July 21, 2022

with previous optoelectronics-based learning approaches (22, 25–27), the D²NN framework provides a distinctive all-optical machine learning engine that efficiently operates at the speed of light using passive components and optical diffraction. An important advantage of D²NNs is

that they can be easily scaled up using various high-throughput and large-area 3D-fabrication methods (such as soft lithography and additive manufacturing), as well as wide-field optical components and detection systems, to cost-effectively reach tens to hundreds of millions of

neurons and hundreds of billions of connections in a scalable and power-efficient manner. For example, integration of D²NNs with lens-free on-chip imaging systems (28, 29) could provide extreme parallelism within a cost-effective and portable platform. Such large-scale D²NNs may

be transformative for various applications, including image analysis, feature detection, and object classification, and may also enable new microscope or camera designs that can perform specific imaging tasks using D²NNs. To achieve these new technologies, nonlinear optical materials (14) and a monolithic D²NN design that combines all layers of the network as part of a 3D-fabrication method would be desirable. Among other techniques, laser lithography based on two-photon polymerization (30) can provide solutions for creating such D²NNs.

REFERENCES AND NOTES

- Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**, 436–444 (2015).
- G. Litjens et al., *Med. Image Anal.* **42**, 60–88 (2017).
- A. Graves, A. Mohamed, G. Hinton, in Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE, 2013), pp. 6645–6649.
- K. Cho et al., arXiv:1406.1078 [cs.CU] (3 June 2014).
- A. Krizhevsky, I. Sutskever, G. E. Hinton, in Advances in Neural Information Processing Systems 25 (NIPS 2012), F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger, Eds. (Curran Associates, 2012), pp. 1097–1105.
- D. Silver et al., *Nature* **529**, 484–489 (2016).
- U. S. Kamlov et al., *Optica* **2**, 517 (2015).
- Y. Rivenson et al., *Optica* **4**, 1437 (2017).
- K. H. Jin, M. T. McCann, E. Froustey, M. Unser, *IEEE Trans. Image Process.* **26**, 4509–4522 (2017).
- Y. Rivenson, Y. Zhang, H. Gunaydin, D. Teng, A. Ozcan, *Light Sci. Appl.* **7**, 17141 (2018).
- A. Sinha, J. Lee, S. U. G. Barbastathis, *Optica* **4**, 1117 (2017).
- K. Hammerink et al., *Magn. Reson. Med.* **79**, 3055–3071 (2018).
- Y. Rivenson et al., *ACS Photonics* **5**, 2354–2364 (2018).
- Materials and methods are available as supplementary materials.
- Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *Proc. IEEE* **86**, 2278–2324 (1998).
- D. Ciregan, U. Meier, J. Schmidhuber, in Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE, 2012), pp. 3642–3649.
- C.-Y. Lee, P. W. Gallagher, Z. Tu, in Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS, 2016), pp. 464–472.
- M. A. Ranzato, C. Poulton, S. Chopra, Y. LeCun, in Advances in Neural Information Processing Systems 20 (NIPS 2007), J. C. Platt, D. Koller, Y. Singer, S. T. Roweis, Eds. (Curran Associates, 2007), pp. 1137–1144.
- H. Xiao, K. Rasul, R. Vollgraf, arXiv:1708.07747 [cs.LG] (25 August 2017).
- S. Zeng, B. Zhang, J. Gou, arXiv:1802.07589 [cs.CV] (21 February 2018).
- ImageNet database, www.image-net.org.
- Y. Shen et al., *Nat. Photonics* **11**, 441–446 (2017).
- D. Psaltis, D. Brady, X.-G. Gu, S. Lin, *Nature* **343**, 325–330 (1990).
- K. H. Wagner, in Proceedings of Frontiers in Optics 2017 (OSA, 2017), FW2C.1.
- B. J. Shastri et al., arXiv:1801.00016 [cs.ET] (29 December 2017).
- M. Hermans, M. Burn, T. Van Vaerenbergh, J. Dambre, P. Bientman, *Nat. Commun.* **6**, 6729 (2015).
- D. Brunner, M. C. Soriano, C. R. Mirasso, I. Fischer, *Nat. Commun.* **4**, 1364 (2013).
- A. Greenbaum et al., *Nat. Methods* **9**, 889–895 (2012).
- A. Ozcan, E. McLeod, *Annu. Rev. Biomed. Eng.* **18**, 77–102 (2016).
- M. Emrani et al., *Opt. Mater. Express* **2**, 942–947 (2012).

ACKNOWLEDGMENTS

We thank D. Mengu, Z. Wei, X. Wang, and Y. Xiao of UCLA for assistance with coding. **Funding:** The Ozcan Group at UCLA acknowledges the support of the National Science Foundation and the Howard Hughes Medical Institute. **Author contributions:** A.O., X.L., and Y.R. conceived of the research; X.L., N.T.Y., Y.L., Y.R., M.V., and M.J. contributed to the experiments; X.L., N.T.Y., M.V., and Y.R. processed the data; A.O., X.L., M.V., N.T.Y., Y.R., Y.L., and M.J. prepared the manuscript; and A.O. initiated and supervised the research. **Competing interests:** A.O., X.L., and Y.R. are inventors of a patent application on D²NNs. **Data and materials availability:** All data and methods are present in the main text and supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6406/1004/suppl/DC1
Materials and Methods
Figs. S1 to S16
References (31–46)

6 April 2018; accepted 12 July 2018
Published online 26 July 2018
10.1126/science.aat8084

外文资料译文

1 基于衍射深度神经网络的全光机器学习

1.1 摘要

深度学习已经改变了我们使用计算机执行高级推理任务的能力。在这里，我们介绍了一种物理机制来执行机器学习，即通过展示一种全光衍射深度神经网络(D2NN)架构，该架构能够基于被动衍射层共同工作的深度学习设计来实现各种功能。我们创建了3D打印的D2NN，用于实现手写数字和时尚产品的图像分类，以及太赫兹光谱下的成像透镜功能。我们的全光深度学习框架能够以光速执行基于计算机的神经网络可以执行的各种复杂功能；将在全光图像分析、特征检测和对象分类中找到应用；并且还将使用D2NN启用新的相机设计和光学组件，以执行独特的任务。

1.2 内容

深度学习是发展最快的机器学习方法之一(1)。这种方法使用在计算机中实现的多层次人工神经网络，以数字方式学习数据表示和抽象，并以与人类专家相当甚至优于人类专家的方式执行高级任务。最近深度学习在机器学习方面取得重大进展的例子包括医学图像分析(2)、语音识别(3)、语言翻译(4)和图像分类(5)等(1,6)。除了这些主流应用之外，深度学习方法还被用于解决逆成像问题(7-13)。在这里，我们引入了一个全光深度学习框架，其中神经网络由多层次衍射表面物理形成，这些衍射表面协同工作，以光学方式执行网络可以统计学习的任意函数。物理网络的推理和预测机制都是光的，而导致其设计的学习部分是通过计算机完成的。我们将该框架称为衍射深度神经网络(D2NN)，并通过仿真和实验证明了其推理能力。我们的D2NN可以通过使用几个透射层和/或反射层来物理创建(14)，其中给定层上的每个点要么传输要么反射入射波，代表一个人工神经元，通过光学衍射与以下层的其他神经元连接(图1A)。根据惠更斯-菲涅耳原理，我们的术语是基于给定层上的每个点作为波的次级源，其振幅和相位由输入波和该点处的复值透射或反射系数的乘积决定[见(14)D2NN内波的分析]。因此，D2NN中的人工神经元通过由前一层产生的输入干涉图样和该点的局部透射或反射系数调制的振幅和相位的二次波与下一层的其他神经元连接。与标准深度神经网络(图1D)类似，我们可以将每个点或神经元的传输或反射系数视为一个乘法“偏置”项，该偏置项是一个可学习的网络参数，在衍射网络的训练过程中使用误差反向传播方法迭

代调整。在这个数值训练阶段之后，D2NN 的设计是固定的，并且确定了各层神经元的传输或反射系数。这种 D2NN 设计-一旦使用 3d 打印或光刻等技术进行物理制造-然后可以以光速执行其训练的特定任务，仅使用光学衍射和无源光学元件或不需要功率的层，从而创建一种高效快速的实现机器学习任务的方式。

一般来说，每个神经元的相位和幅度都是可学习的参数，在每一层提供复值调制，从而提高了衍射网络的推理性能 (图 S1)(14)。对于纯相位调制的相干传输网络，每一层都可以近似为一个薄光学元件 (图 1)。通过在输入层输入训练数据，然后通过光学计算网络的输出，迭代调整 (训练) 衍射网络每层神经元的相位值以执行特定的功能。基于相对于目标输出的计算误差，由期望函数确定，通过误差反向传播算法优化网络结构及其神经元相位值，该算法基于传统深度学习中使用的随机梯度下降方法 (14)。为了展示 D2NN 框架的性能，我们首先将其训练为数字分类器，以执行从 0 到 9 的手写数字的自动分类 (图 1B 和 2A)。对于这项任务，通过训练来自 MNIST (Modified National Institute of Stang 数据库) 的 55000 张图像 (5000 张验证图像) 的五层 D2NN 来设计纯相位传输掩模 (15)。将输入数字编码为输入场的振幅，并将衍射网络训练成将输入数字映射到 10 个检测器区域，每个检测器区域对应一个数字。分类标准是寻找光信号最大的检测器，这也是网络训练时的损失函数 (14)。

训练结束后，使用 MNIST 测试数据集 (未作为训练或验证图像集的一部分) 中的 10,000 张图像对 D' nn 数字分类器的设计进行了数值测试，取得了 91.75% 的分类精度 (图 3C 和图 S1)。除了对该网络的分类性能进行分析，我们还分析了相同 10000 个测试数字在网络输出平面上的能量分布 (图 3C)，结果清楚地表明，衍射网络学习到将每个手写数字的输入能量聚焦到正确 (即目标) 探测器区域，与训练结果一致。随着复数调制的使用以及衍射网络中层数、神经元和连接数的增加，我们的分类精度可以进一步提高。S1 和 S2)。例如，图 S2 展示了一种类似于乐高的 D' nn 框架的物理迁移学习行为，其中已经存在的 D2NN 的推理性能可以通过添加新的衍射层来进一步提高，或者在某些情况下，通过剥离 (即丢弃) 一些现有的层，其中添加的新层是经过训练以改进推理 (来自整个衍射网络: 旧层和新层)。通过使用添加到现有和固定的 D2NN 设计 ($N = 5$ 层) 的两层补丁，我们将 MNIST 分类精度提高到 93.39%(图 S2) (14); 据报道，最先进的卷积神经网络性能为 99.60 至 99.77% (16-18)。在补充材料 (14) 中提供了关于重新配置 D2NN 设计的更多讨论。根据这些数值结果，我们 3d 打印了我们的五层 D2NN 设计 (图 2A)，每层的面积为 $8\text{cm} \times 8\text{cm}$ ，随后在衍射网络的输出平面上定义了 10 个探测器区域 (图 1B 和 3A)。然后，我们使用 0.4 THz 的连续波照明来测试网络的推理性能 (图 2,C 和

D)。利用 3d 打印神经元的相对厚度对每层神经元的相位值进行物理编码; 这五层神经网络设计在 10000 张测试图像上取得了 91.75% 的分类精度(图 3C)。为了量化这些数值测试结果与实验之间的匹配, 我们 3d 打印了 50 个手写数字(每个数字有 5 个不同的输入), 从数值测试成功的 91.75% 的测试图像中选择。对于每个被太赫兹光源均匀照射的输入物体, 我们对 D2NN 的输出平面进行成像, 以映射分配给一个数字的每个探测器区域的强度分布。结果(图 3B)证明了 3d 打印的衍射神经网络及其推理能力的成功: 网络输出平面上每个输入数字的平均强度分布清楚地表明, 3d 打印的 D'NN 能够聚焦光束的输入能量, 并在分配给该数字的相应探测器区域实现最大信号。尽管在我们的设置中存在 3d 打印错误、可能的对齐问题和其他实验误差来源(14), 我们发现五层 D2NN 设计的实验和数值测试之间的匹配度为 88%(图 3B)。与我们的数值测试相比, 实验网络的性能相对较小的下降对数字 0 的影响尤其明显, 因为对数字中心的大空白区域进行 3d 打印是一项挑战。对于其他具有空白区域的数字, 也观察到了类似的打印挑战;、6、8、9(图 3B)。

然后, 在一个更复杂的图像数据集上测试了 D2NN 框架的分类性能。其中, fashion-mnist 数据集(19)包含 10 个类别, 每个类别代表一种时尚产品(t 恤、裤子、套头衫、连衣裙、外套、凉鞋、衬衫、运动鞋、包和短靴; 示例图像见图 S3)。一般来说, 对于一个相干光照的 D2NN, 我们可以使用输入平面的振幅和/或相位通道来表示要分类或处理的数据。在我们之前报告的数字分类结果中, 输入对象通过使用幅度通道进行编码, 并且为了演示网络输入的相位通道的效用, 我们将每个与时尚产品对应的输入图像编码为纯相位对象调制(14)。我们用于时尚产品分类的 D2NN 推理结果(作为层数、神经元和连接数的函数)总结在图中 S4 和 S5。举个例子在我们的性能中, 一个相位调制 D2NN 和一个 N=5 衍射层的复数调制 D2NN(与图 2A 所示的数字分类 D2NN 共享相同的物理网络尺寸)分别达到了 81.13 和 86.33% 的精度(图 S4)。将衍射层数增加到 N=10, 神经元总数增加到 40 万个, 分类精度提高到 86.60%(图 S5)。对于基于卷积神经网络的标准深度学习, 据报道, 使用约 890 万个可学习参数和约 250 万个神经元(20), Fashion-MNIST 分类精度的最先进性能为 96.7%。

为了使用物理 D2NN 实验证明时尚产品分类的性能, 我们基于数字分类衍射网络所采用的相同程序(图 2A 和 3), 将我们的仅相位的五层设计和 50 个时尚产品作为测试对象(每个类别 5 个)3d 打印, 除了每个输入对象信息都在相位通道中编码。我们的结果总结在图 4 中, 显示了我们的五层 D2NN 设计的实验和数值测试之间 90% 的匹配, 50 个时尚产品中有 5 个错误。与数字分类相比(50 个数字中有 6 个错误; 图 3), 这个实验在实验和数值测试结果之间产生了稍好的匹配(尽管

fashion - mnist 数据集具有更大的挑战性), 这可能是因为我们使用了相位通道来编码时尚产品的输入图像信息, 它不受 3d 打印空白区域(如数字 0、6、8 和 9(图 3))的挑战。

接下来, 测试了由五个 3d 打印传输层组成的纯相位 D2NN 的性能, 以实现振幅成像(图 2B)。该网络使用 ImageNet 数据库(21)进行训练, 以创建输入光场振幅在其输出平面(约 $9 \text{ cm} \times 9 \text{ cm}$)的单位放大图像, 即输出图像与输入对象具有相同的物理尺寸(14)。如图 S6, A 和 C 所示, 经过训练的网络最初将输入平面上的每个振幅点连接到各种神经元和以下层的特征, 然后将光线聚焦回输出平面的一点(即图像), 正如预期的那样, 这与自由空间衍射的情况(即没有衍射网络的存在)有很大的不同, 如图 S6,B 和 D 所示。

经过训练和盲测, 从数值上证明了网络的成像能力(S6 和 S7), 然后我们 3d 打印这个设计好的 D2NN。使用图 2、C 和 D 所示的相同实验设置, 我们对在 0.4 THz 连续波辐射均匀照射下的各种输入物体的 3d 打印 D2NN 的输出平面进行了成像。图 S8 总结了我们用这个 3d 打印的 D2NN 实现的实验结果, 它成功地在网络的输出平面上投影了输入模式的单位放大图像, 学习了图像器或物理自编码器的功能。为了评估该 DNN 的点扩展函数, 我们对不同直径(1、2 和 3 mm)的针孔进行了成像, 从而产生了输出图像, 每个半最大值的全宽度分别为 1.5、1.4 和 2.5 mm(图 S8B)。我们的结果还显示, 印刷网络在 0.4 太赫兹时可以分辨 1.8 mm 的线宽(在空气中对应 0.75 mm 的波长), 而在 1.2 mm 线宽的分辨率稍差(图 S7C)。衍射网络性能的这种实验退化可能是由于 3D 打印错误、潜在的不对齐和 3D 打印网络中的吸收相关损失等因素造成的(14)。

与光学系统的数值测试效率相比, 由于其并行计算能力和功率分辨率, 人工神经网络中机器学习的光学实现是有希望的(22-24)。与之前基于光电的学习方法(22,25-27)相比, D2NN 框架提供了一个独特的全光学机器学习引擎, 利用无源组件和光学衍射以光速高效运行。D2NN 的一个重要优势是, 它们可以使用各种高通量和大面积的 3d 制造方法(如软光刻和增材制造), 以及宽视场光学组件和检测系统轻松地进行扩展, 以可扩展和低功耗的方式, 成本有效地达到数千万至数亿个神经元和数千亿个连接。例如, 将 D2NN 与无透镜片上成像系统(28,29)集成, 可以在一个具有成本效益的便携式平台上提供极端的并行性。这种大规模的 D2NN 可能对各种应用具有变革性, 包括图像分析、特征检测和对象分类, 还可能使新的显微镜或相机设计成为可能, 使用 d2nn 执行特定的成像任务。为了实现这些新技术, 非线性光学材料(14)和单片 D2NN 设计将网络的所有层作为 3d 制造方法的一部分是理想的。在其他技术中, 基于双光子聚合的激光光刻(30)可以为创建这

样的 D2NN 提供解决方案。