

基于 10.6 微米全光深度神经网络衍射光栅的设计与实现

牛海莎¹, 于明鑫¹, 祝博飞², 姚齐峰¹, 张乾坤¹, 鹿利单¹, 钟国舜³, 祝连庆^{1*}

(1. 光电测试技术与仪器教育部重点实验室, 北京信息科技大学, 北京 100192;

2. 北京紫溪智芯科技有限公司, 北京 100876;

3. 中国电子科技集团公司第十一研究所, 北京 100015)

摘要:光子人工智能芯片以光速执行运算, 且具有低功耗、延迟低、抗电磁干扰的优势。小型化与集成化是实现这一技术革新的关键步骤。本文将光刻技术运用于衍射光栅的制作, 提出一种基于 10.6 微米激光的全光衍射深度学习神经网络光栅设计及实现方法。由于光源波长由毫米波向微米波进化, 神经元的特征尺度缩小至 20 微米, 与现有光衍射神经网络相比, 深度学习神经网络特征尺寸缩小了 80 倍, 为进一步实现光子计算芯片大规模集成奠定了基础。

关键词:光子芯片; 衍射光栅; 深度学习; 神经网络

中图分类号: TP183 **文献标识码:** A

Design and implementation of diffraction grating based on 10.6 μm all-optical depth neural network

NIU Hai-Sha¹, YU Ming-Xin¹, ZHU Bo-Fei², YAO Qi-Feng¹, ZHANG Qian-Kun¹, LU Li-Dan¹,
ZHONG Guo-Shun³, ZHU Lian-Qing^{1*}

(1. Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science & Technology University, Beijing 100192, China;

2. Beijing ZX Intelligent Chip Technology Co., Ltd., Beijing 100876, China;

3. The 11th Research Institute of China Electronic Science & Technology Group Inc., Beijing 100015, China)

Abstract: The photonic artificial intelligent chip performs calculations at the speed of light, and has the advantages of low power consumption, low delay, and anti-electromagnetic interference. Miniaturization and integration are the key steps to realize this technological innovation. In this paper, lithography is applied to the fabrication of diffraction gratings. A design and implementation method of all-optics diffraction deep learning neural network grating based on 10.6 micron laser is proposed. Since the wavelength of the light source evolved from the millimeter wave to micrometer wave, the characteristic scale of the neuron are reduced to 20 micrometers. Compared with the existing optical computing neural network, the feature size of the deep learning neural network is reduced by 80 times, which laid the foundation for further large-scale integration of photonic computing chips.

Key words: Photonic chip, diffraction grating, deep learning, neural network

PACS: 42. 50. Vx, 42. 40. Lx, 42. 25. Fx

收稿日期: 2019-04-14, 修回日期: 2019-11-15

Received date: 2019-04-14, Revised date: 2019-11-15

基金项目: 高等学校学科创新引智计划(D17021), 国家自然科学基金重点项目(51535002), 促进高校内涵发展项目(5211910905)

Foundation items: Project supported by Program of Introducing Talents of Discipline to Universities (D17021), the Key Projects of the National Natural Science Foundation of China. (51535002), Promoting the connotation development project of colleges (5211910905)

作者简介(Biography):牛海莎(1984-), 女, 河北石家庄人, 博士后, 北京信息科技大学讲师, 2008年于武汉大学获得工学学士学位, 2017年于北京航空航天大学获工学博士学位, 主要从事激光器自混合干涉效应的精密测量、光学器件设计研究工作。

*通讯作者(Corresponding author): Email: zhulianqing@sina.com

引言

随着时代的发展,更加高速、携带信息更多的计算技术是当今科技工作者关注的热点问题^[1-2]。光子计算^[3]由于其天然优势被认为是最有希望的未来技术:光的并行、高速决定了光计算的并行处理能力很强、且具有超高的运算速度;光子不带电荷,因此传输光路之间不存在电磁场相互作用;光子没有质量,既可以在真空中传播又可以在介质中传播,传播速度为电子在导线中的1 000倍;此外,光计算不仅能够实现超大规模的信息存储容量,且具有能量散耗小、发热量低等优点。

目前,光子计算的研究已经取得了系列进展:金贤敏团队^[4]通过“飞秒激光直写”技术制备出节点数达49×49、世界上最大规模的光子计算芯片;同年,Qiang^[5]等使用光子编码量子比特,通过对光子的量子操控及测量实现量子计算,利用硅基光波导集成技术开发出一种通用光量子芯片执行不同的量子信息处理任务,在推动光量子计算机大规模实用化上迈出重要一步;在人工智能领域,构建全光神经网络、对网络参数进行训练为图像识别、语言翻译、安防监控、AI金融等提供高性能的硬件支持,也已取得了突出进展:2017年,Shen^[6]等利用马赫-曾德干涉单元构建神经网络,通过前向传输步骤计算网络输出、反向传播过程优化参数训练了3类分类问题,正确率达到92%;2018年,Lin^[7]等提出了一种衍射深度神经网络架构及全光机器学习^[8]的物理机制,以3D打印^[9]的方式制作了太赫兹波段的深度学习衍射光栅,实现了手写数字及物品的识别,开启了光计算在人工智能领域的新里程碑。

要进一步将光栅机器学习神经网络计算芯片推向应用,小型化与集成化是这一技术革新的关键步骤。受限于3D打印的精度,Lin等提出的0.75 mm波段全光衍射光栅尺度为8 cm×8 cm,神经元尺寸为数十微米,离大规模集成尚有距离。紫外光刻技术^[10-12]直接决定了大规模集成电路的特征尺寸,是大规模集成电路制造的关键工艺。将光刻工艺运用于衍射光栅的制作,将大幅度降低神经元及神经网络的特征尺度。为此,本文提出一种基于10.6 μm激光的全光衍射深度神经网络光栅设计及实现方法,与文献[7]的最新报道相比,由于光源波长的由毫米波向微米波进化,光栅器件的尺寸缩小至1 mm,神经元的尺度缩小至微米量级,将深度学习神经网络的尺寸缩小80倍,为进一步实现大规模集成

奠定了基础。

1 全光深度神经网络设计

在训练神经网络模型时,数据被馈送到输入层并通过前向传播步骤计算输出,然后通过反向传播过程优化参数。根据Rayleigh-Sommerfeld原理^[13],每一个光栅衍射神经网络上的单元都可以视为下一层神经网络的入射源,可表达为

$$w_i^l(x,y,z) = \frac{z - z_i}{r^2} \left(\frac{1}{2\pi r} + \frac{1}{j\lambda} \right) \exp\left(\frac{j2\pi r}{\lambda}\right), \quad (1)$$

其中 l 代表第 l 层光栅神经网络, i 为坐标位置 (x_i, y_i, z_i) 的第 i 个神经元, λ 为入射波长,且

$$r = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}, \text{且 } j = \sqrt{-1}, \quad (2)$$

因此, l 层光栅的第 i 个神经元的输出光可表达为

$$\begin{aligned} n_i^l(x,y,z) &= w_i^l(x,y,z) \cdot t_i^l(x_i,y_i,z_i) \cdot \sum_k n_k^{l-1}(x_i,y_i,z_i) \\ &= w_i^l(x,y,z) \cdot |A| \cdot e^{j\Delta\theta} \end{aligned}, \quad (3)$$

其中,第 l 层的第 i 个神经元处的输入为前一层的所有输出项叠加,如图1所示。 A 为第 i 个神经元输出的幅度因子, $\Delta\theta$ 为输入光波在 i 处叠加产生的相位延迟因子。光场在各层之间衍射并在神经元上的叠加干涉数据、包括幅度和相位因子流入神经元,喂给训练过程。光的传输系数可以表达为

$$t_i^l(x_i,y_i,z_i) = a_i^l(x_i,y_i,z_i) \exp(j\phi_i^l(x_i,y_i,z_i)), \quad (4)$$

对于纯相位衍射深度神经网络光栅来说,忽略光在传播过程中的损失,将幅度因子设为理想值1。

为简化前向模型的架构,式(3)可以简写为

$$\begin{aligned} n_{i,p}^l &= w_{i,p}^l \cdot t_i^l \cdot m_i^l \\ m_i^l &= \sum_k n_{k,i}^{l-1} \\ t_i^l &= a_i^l \exp(j\phi_i^l) \end{aligned}. \quad (5)$$

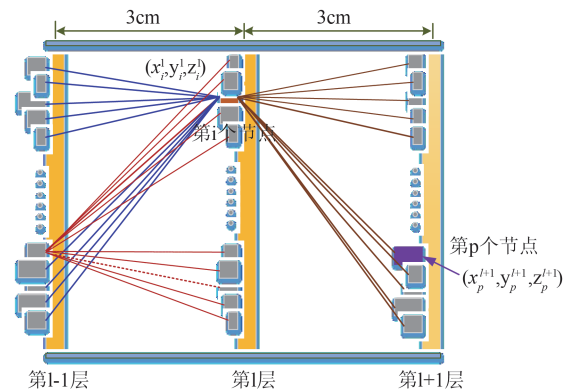


图1 全光接连神经网络结构示意图

Fig. 1 Schematic diagram of the structure of all-optical depth neural network

如图1所示, i 表示第 l 层的第 i 个神经元, p 表示第 $l+1$ 层第 p 个神经元,各层之间的神经元通过衍射效应实现全连接。输入层的输入因子 \mathbf{h} 是一个复数,其幅度和相位信息中携带了输入图像的信息,输入因子在光栅之间的衍射传递可表示为

$$n_{k,p}^0 = w_{k,p}^l \cdot h_k^0 \quad . \quad (6)$$

对于一个具有 N 层结构的光栅神经网络来说,输出光场的光强可以表示为

$$s_i^{N+1} = |m_i^{N+1}|^2 \quad . \quad (7)$$

将光栅神经网络的输出结果与输入进行比较,并将误差反向传递以迭代优化光栅参数,根据目标特性定义损失函数来评估衍射神经网络的输出,并采用随机梯度下降优化算法迭代优化光栅的相位参数,定义损失函数为输出平面光强和目标之间的误差均方差

$$E(\phi_i^l) = \frac{1}{K} \sum_k (s_k^{N+1} - g_k^{N+1})^2 \quad , \quad (8)$$

K 表示输出平面的测量点数,光栅相位优化算法可表示为

$$\min_{\phi_i^l} E(\phi_i^l), s.t., 0 \leq \phi_i^l < 2\pi \quad . \quad (9)$$

误差函数对相位因子求偏导计算相对误差梯度,表达式

$$\frac{\partial E(\phi_i^l)}{\partial \phi_i^l} = \frac{4}{K} \sum_k (s_k^{N+1} - g_k^{N+1}) \cdot \text{Real} \left\{ (m_k^{N+1})^* \frac{\partial m_k^{N+1}}{\partial \phi_i^l} \right\} \quad , \quad (10)$$

其中, $\frac{\partial m_k^{N+1}}{\partial \phi_i^l}$ 为输出层复数光场的梯度, $(m_k^{N+1} = \sum_{k_1} n_{k_1,k}^M)$ 为前一层光栅网络神经元的相位因子,对于任意一层神经网络,复数光场的梯度可按如下公式计算

$$\begin{aligned} \frac{\partial m_k^{N+1}}{\partial \phi_i^{l=N}} &= j \cdot t_i^N \cdot m_i^N \cdot w_{i,k}^N, \\ \frac{\partial m_k^{N+1}}{\partial \phi_i^{l=N-1}} &= j \cdot t_i^{N-1} \cdot m_i^{N-1} \cdot \sum_{k_1} w_{k_1,k}^N \cdot t_{k_1}^N \cdot w_{i,k_1}^{N-1}, \\ \frac{\partial m_k^{N+1}}{\partial \phi_i^{l=N-2}} &= j \cdot t_i^{N-1} \cdot m_i^{N-2} \cdot \sum_{k_1} w_{k_1,k}^N \cdot t_{k_1}^N \cdot w_{i,k_1}^{N-1} \cdot w_{i,k_2}^{N-2} \\ &\dots \\ \frac{\partial m_k^{N+1}}{\partial \phi_i^{l=N-L}} &= j \cdot t_i^{N-L} \cdot m_i^{N-L} \cdot \sum_{k_1} w_{k_1,k}^N \cdot t_{k_1}^N \dots \\ &\quad \sum_{k_L} w_{k_L,k_{L-1}}^{N-L+1} \cdot t_{k_L}^{N-L+1} \cdot w_{i,k_L}^{N-L} \quad , \quad (11) \end{aligned}$$

其中, $3 \leq L \leq N-1$ 。每一次误差背向传输迭代,都将一批训练过的数据反馈到衍射神经网络进行每一层的相位梯度计算,并根据计算结果更新优化衍射神

神经网络系统的相位分布。为将光栅网络训练成数字分类器,采用了图2所示MNIST的手写数字数据库作为输入层。

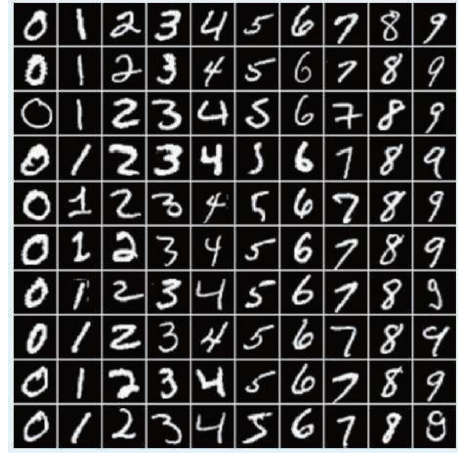


图2 MNIST手写数字数据库(部分输入层)

Fig. 2 MNIST handwritten digital database (partial input layer)

将光栅神经网络的间距设置为1 mm,将MNIST的55 000对数据分成550批,每次训练使用一批,共训练两次,得到5层光栅神经网络的相位分布分别如图3(a)~(e)所示。其中每个光栅神经元的特征尺寸为 $5 \mu\text{m}$,光栅特征尺寸为 $1 \text{ mm} \times 1 \text{ mm}$ 。

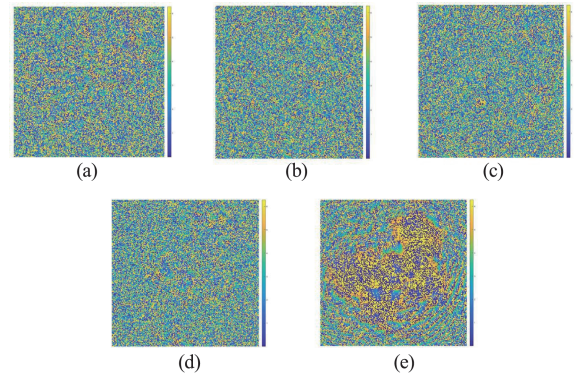


图3 训练的光栅神经网络相位分布

Fig. 3 Trained grating neural network phase distribution

2 光栅神经网络的刻写

根据

$$\Delta h = \lambda \phi / 2\pi (n_m - n_a) \quad , \quad (12)$$

其中, n_m 为材料的折射率, n_a 为空气折射率, ϕ 为训练的光栅相位, λ 为入射光源的波长,可计算出光栅的深度 Δh 。选取的光源为 $10.6 \mu\text{m}$ 的二氧化碳激光器,光栅的刻写材料为锗, $10.6 \mu\text{m}$ 下锗的折射率

表 1 光栅相位与高度

Table 1 Grating phase and height

光栅相位(rad)	0	1	2	3	4	5	6	2π
光栅深度(μm)	0	0.563	1.125	1.688	2.250	2.815	3.378	3.54

为 $4.003^{[15]}$, 空气的折射率近似取 1, 则不同相位下光栅的高度分布如表 1 所示。

由表 1 可见, 相位分布在 $0\sim 2\pi$ 范围内的光栅沟槽深度为 $0\sim 3.54\ \mu\text{m}$, 单位弧度的光栅的刻蚀深度为 $563\ \text{nm}$ 。3D 打印的方式难以达到百纳米的精度, 因此, 光栅的制作采用光刻技术实现。电感耦合等离子体刻蚀(ICP)技术作为干法刻蚀技术的一种, 具有刻速快、选择比高、各向异性高、刻蚀损伤小、大面积均匀性好、刻蚀断面轮廓可控性高等优点。理论上训练优化出的光栅深度为连续值, 为提高台阶的精细度, 采用先进的五次套刻的方式制作光栅。流程如图 4 所示: 锗片上涂覆负光刻胶, 掩膜曝光后显影, 通过 ICP 刻蚀(第一次刻蚀深度设置为 $1.7667\ \mu\text{m}$)、去胶, 再换掩膜版 2 重复上述步骤进行套刻。

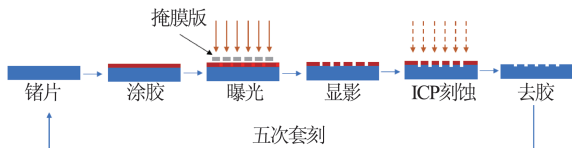


图 4 光刻及套刻示意图

Fig. 4 Photolithography and engraving

按照图 4 所示的步骤循环 5 次, 且每次的刻蚀深度为上一次的二分之一, 得到五次套刻的台阶理论分辨率为 $3.54\ \mu\text{m}/32=110.6\ \text{nm}$, 每次刻蚀的掩膜版计算分别如图 5(a-e) 所示。

采用原子力显微镜(Bruker, Dimension Icon)对

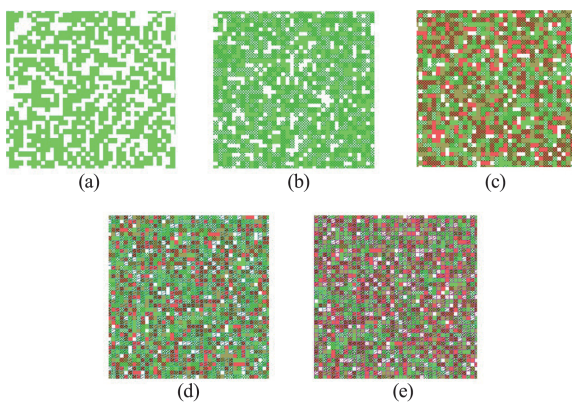


图 5 单片光栅的五次套刻掩膜版

Fig. 5 Five times engraved mask for monolithic grating

五次套刻的光栅神经元进行成像, 其平面及高度分布分别如图 6(a)、(b) 所示。

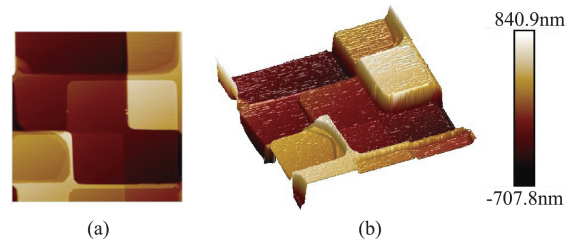


图 6 衍射光栅神经元

Fig. 6 Diffraction grating neurons

3 光栅神经网络正确率测试

如图 7 所示, 在根据 MNIST 手写数字数据库训练的数字分类器神经网络中输入手写数字 8, 按照训练的输出, 系统输出光强应集中在“8”的区域, 输出结果如图 8(a) 所示。

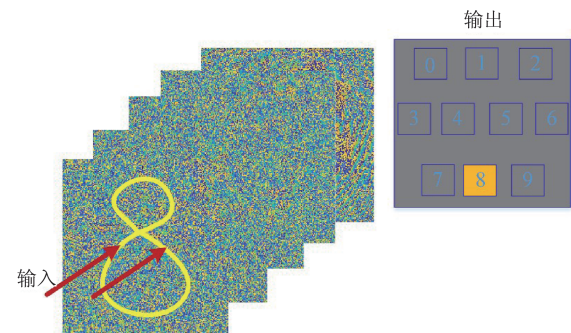


图 7 深度学习光栅神经网络测试示意图

Fig. 7 Schematic diagram of deep learning grating neural network

如图 8(b) 所示, 训练得到的五层光栅神经网络数字分类器对 0~9 的识别正确率分布在 75%~97% 之间。光栅的套刻加工会产生百纳米量级的误差, 因此, 对训练的光栅神经网络单个像元分别引入误差并重新对手写数字进行识别, 得到的识别准确率如图 9 所示。

结果表明, 引入加工误差后, 数字 0, 1, 4, 6, 8 的识别率下降、数字 2, 9 的识别率增加、数字 3, 7 的识别率持平。对 4 和 8 的识别率影响最大, 准确率下降了 6%, 其余数字影响均在 3% 以内, 且整体识别

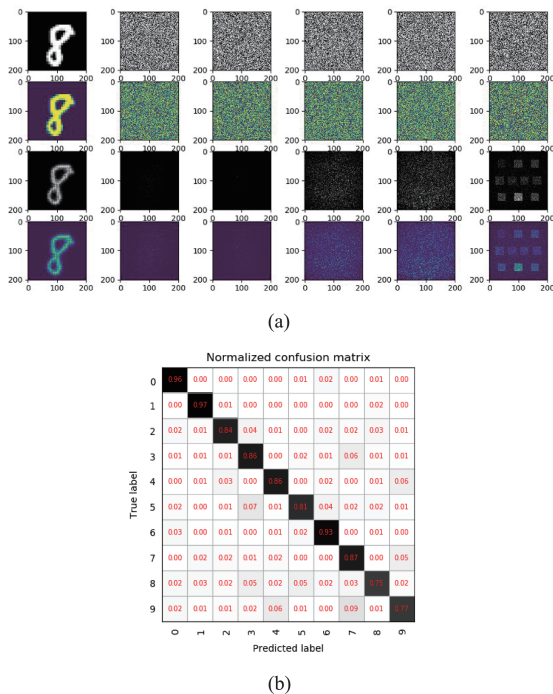


图8 (a)数字分类光栅神经网络输出结果,(b)数字分类器正确率

Fig.8 (a) Digital classification grating neural network output results,(b) Digital classifier recognition accuracy

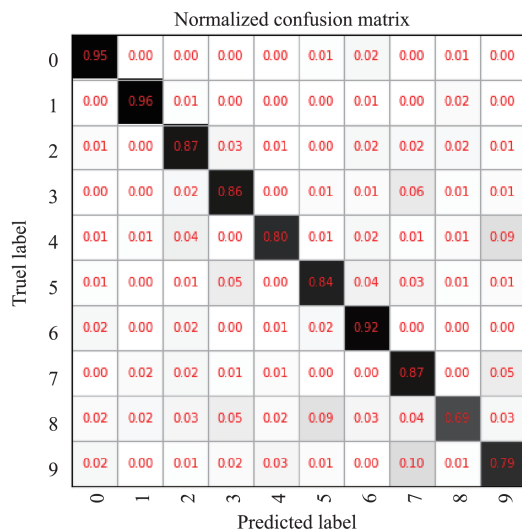


图9 引入加工误差后的光栅神经网络识别准确率

Fig.9 Accuracy of grating neural network recognition after introducing errors

率趋势不受光栅高度误差影响,说明光栅神经网络整体容错性能较好。目前的全连接光栅神经网络训练模型中尚未引入非线性计算,提取数字特征的能力尚有提升空间,对于具有相似特征的数字识别率较低,例如将8识别为3和5的概率均为5%,而将9识别为7的概率高达10%,因此10个手写数字中,

对8和9的识别率均低于80%。

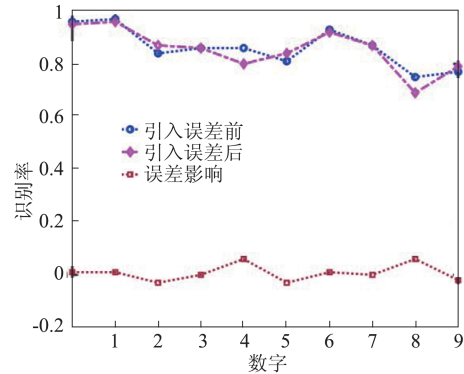


图10 引入误差前、后识别准确率比较

Fig.10 Comparison of recognition accuracy before and after the introduction of error

4 总结

基于全光衍射光栅机器学习神经网络在人工智能领域具有广阔的应用前景。本文通过前向传输步骤计算网络输出、反向传播过程优化参数训练了10.6微米全光衍射光栅深度神经网络数字分类器,正确率达到75%~97%。采用紫外光刻技术对设计的衍射光栅进行了5次套刻,套刻精度为110.62 nm,首次实现了特征尺寸为5 μm 量级的全光机器学习神经元制作。在后续工作中将进一步优化光栅训练算法提高识别正确率,并对光栅刻写误差对神经网络的影响进行量化分析及改进,为进一步实现光子计算芯片大规模集成奠定基础。

References

- [1] Hermans M, Burm M, Van Vaerenbergh T, *et al.* Trainable hardware for dynamical computing using error back-propagation through physical media[J]. *Nature Communications*, 2015, **6**:6729.
- [2] Sicre R, Jurie F. Discriminative part model for visual recognition ☆ [J]. *Computer Vision & Image Understanding*, 2015, **141**(C):28–37.
- [3] Husko C, Vo T D, Corcoran B, *et al.* Ultracompact all-optical XOR logic in a slow-light silicon photonic crystal waveguide[J]. *Optics Express*, 2011, **19**(21):20681–90.
- [4] Tang H, Lin X F, Feng Z, *et al.* Experimental two-dimensional quantum walk on a photonic chip[J]. *Science Advances*, 2018, **4**(5):eaat3174.
- [5] Xiaogang Q, Xiaoqi Z, Jianwei W, *et al.* Large-scale silicon quantum photonics implementing arbitrary two-qubit processing[J]. *Nature Photonics*, 2018.
- [6] Shen Y, Harris N C, Skirlo S, *et al.* [IEEE 2017 IEEE Photonics Society Summer Topical Meeting Series (SUM) – San Juan, PR, USA (2017.7.10–2017.7.12) 2017 IEEE Photonics Society Summer Topical Meeting Series (SUM) –

- Deep learning with coherent nanophotonic circuits[C]. 2017 IEEE Photonics Society Summer Topical Meeting Series (SUM). IEEE, 2017:189-190.
- [7] Xing L, Yair R, Yardimci N T, et al. All-optical machine learning using diffractive deep neural networks [J]. *Science*, 2018:eaat8084.
- [8] Lecun Y, Bengio Y, Hinton G. Deep learning.[J]. *Nature*, 2015, **521**(7553):436.
- [9] Guo C, Jin L I, Shang X B, et al. Novel microwave/millimeter-wave passive waveguide devices based on 3-D printing techniques [J]. *J. Infrared Millim. Waves*, (郭诚, 李津, 尚小邦, 等. 基于3-D打印技术的新型微波/毫米波无源波导器件. *红外与毫米波学报*).2017, **36**(1): 81-91.
- [10] Cao Y Y, Zhang Y G, Yao-Yao L I, et al. Fabrication of column shape two dimensional photonic crystals: double developments in holographic lithography process [J]. *J. Infrared Millim. Waves*, (曹远迎, 张永刚, 李耀耀, 等. 全息光刻和二次显影法制备柱形二维光子晶体. *红外与毫米波学报*)2014, **33**(1):45-49.
- [11] Bolk J, Ambrosius H, Stabile R, et al. Deep UV Lithography Process in Generic InP Integration for Arrayed Waveguide Gratings [J]. *IEEE Photonics Technology Letters*, 2018, PP(99):1-1.
- [12] JIA Jia. Improving the aspect ratio of ion beam etched trenches in HgCdTe, [J]. *J. Infrared Millim. Waves*, (离子束刻蚀碲镉汞的沟槽深宽比改进. *红外与毫米波学报*) 2015, **34**(3):282-285.
- [13] Shen F, Wang A. Fast-Fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula.[J]. *Appl Opt*, 2006, **45**(6):1102-1110.
- [14] Sheppard C J, Lin J, Kou S S. Rayleigh-Sommerfeld diffraction formula in k space.[J]. *Journal of the Optical Society of America A Optics Image Science & Vision*, 2013, **30**(6):1180-1183.
- [15] He J, Chen L, Wang Q. Measurements of Infrared Materials Refractiveindex Using Infrared Interferometer, [J]. *Acta Photonica Sinica*(贺俊, 陈磊, 王青. 使用红外干涉仪测量红外材料折射率. *光子学报*), 2010, **39**(6): 1125-1128.