

神经网络后向传播算法的 Dammann 光栅设计*

廖江红 桑涛 卢振武 翁志成

(中国科学院长春光机所 应用光学国家重点实验室, 长春 130021)

提要 提出了一种基于神经网络后向传播算法的二元光学设计新方法, 该法的突出特点是可以获得较快的收敛速度和较好的均匀性。最后给出 6 个 Dammann 光栅的设计结果。

关键词 Dammann 光栅, 二元光学, 神经网络

1 引言

Dammann 光栅^[1]是一种周期性的纯位相结构。在单色平面波照明下可产生人们所期望的各种光强分布图案。近些年的研究表明, 这种光栅在一物成多象^[1]、激光束相干叠加^[2]、光计算^[3]和光纤通信等领域有广泛用途, 因而成为二元光学研究的热点之一。目前人们已提出多种设计方法, 从本质上看可以分为两大类: 第一类包括误差递减下山法、梯度下降法、输入输出算法和投影算法^[4]等, 这类算法强调减少计算量, 不具备跳出局部极值的能力。尽管它们在形式上不同, 但由于其目的在于寻找同时满足物域和频域要求的函数, 实质上都构成了一个正反傅里叶变换的闭环, 故本质上都相同。第二类主要包括模拟退火法和遗传算法^[5], 这一类与前一类的差别在于它以一定概率接受优化过程中出现的坏结果, 从而使算法有了从问题的局部极值跳出的可能, 具备了寻求全局最小值的能力。但由于接受了坏结果, 使运算量变得很大。遗传算法具有并行处理能力, 这对于减少计算量十分有意义。值得一提的是, 这类算法在跳出局部极值性能方面并非十分理想, 如果算法逐步减少接受坏结果概率的历程与算法当前所处问题区域本身特性不相符合, 结果仍然可能停留在局部优化解上。由此可见, 二元光学设计总是处在提高收敛速度和跳出局部极小值这两个相互矛盾的问题之中。

本文提出一种加快收敛速度的神经网络后向传播的算法, 据我们所知, 这是首次将神经网络概念用于二元光学设计, 尽管该法依然存在上述局部极值问题, 但我们对 $(0, \pi)$ 型 Dammann 光栅的设计结果表明, 在选取合适的初始结构下, 该算法可以快速地收敛到极值, 均匀性可以明显地得到改善。

* 本工作曾获得应用光学国家重点实验室基金、博士后基金、中国科学院基础局基金的部分支持。

收稿日期: 1995年1月2日; 收到修改稿日期: 1995年3月27日

2 理论模型

神经网络后向传播算法是一种 supervised 型训练神经网络的方法^[6],设计思路表述如下:把自二元光学元件出射的光波复振幅赋以神经网络权重的意义,将光波从二元光学元件

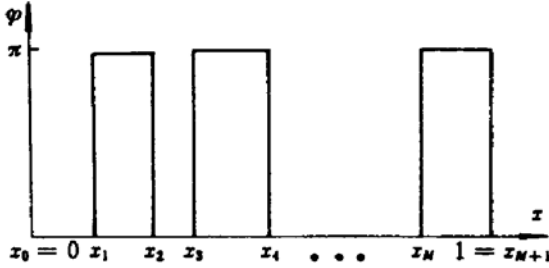


Fig. 1 The phase structure of the grating

到谱面传播的傅里叶变换关系类比为信号从输入神经元到输出神经元的传播关系,输出神经元的输出就是光波传播到谱面的复振幅,将之取模平方即为谱面的光强。该光强与我们所需的光强相比较获得误差信号,并依据神经网络后向传播的原理将该误差逆着原来的传播方向传播至输入层,并利用有关规则修改该层的权重,使得

下次传播的输出误差减小。每一次正向传播和一次逆向传播合称一次循环,经过若干次循环后输出结果与所需结果相比小于某一阈值或不再变化时便完成对神经网络的训练,由此时的神经元权重便可获得二元光学元件的优化位相结构。

为了进一步阐述上述思想,现藉设计输出为 $(2N + 1)$ 个光点的一维 $(0, \pi)$ 型 Dammann 光栅来加以说明。假设位相光栅单位周期内有 M 个可变坐标点 $[0, X_1, \dots, X_M, 1]$,位相的空间结构如图 1 所示(图中 M 为奇数),则谱面上各亮点振幅 U_k 是上述位相光栅傅里叶变换的线性叠加^[7]

$$U_k = \begin{cases} 1 + 2 \sum_{m=1}^{M+1} (-1)^{m+1} X_m & k = 0 \\ \frac{1}{i\pi k} \cdot \sum_{m=1}^{M+1} (-1)^m \exp(-i2\pi k X_m) & k \neq 0 \end{cases} \quad (1)$$

$k = 0, \pm 1, \pm 2, \dots, \pm N;$

相应的强度记为

$$I_k = I_{-k} = |U_k|^2 \quad (2)$$

图 2 为与此对应的双层神经网络模型,对于 Dammann 光栅设计,入射光复振幅设为 1,于是只需一个输入神经元,取值为 1,用 $Out_0 = 1$ 表示。如图 2 所示,从输入层到隐层代表光波透过光栅的传播过程,且其间连接的权重设为 W_{1m}

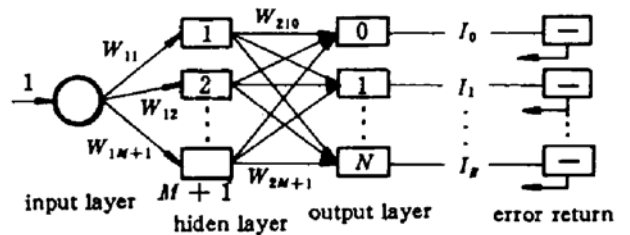


Fig. 2 The neural network model

$$W_{1m} = X_m \quad m = 1, 2, \dots, M + 1 \quad (3)$$

若当前开始第 n 次训练神经元,则用上角标表示

$$W_{1m}^{[n]} = X_m^{[n]} \quad m = 1, 2, \dots, M + 1 \quad (4)$$

其隐层输出用 $Out_{1m}^{[n]}$ 表示

$$Out_{1m}^{[n]} = Out_0 \cdot W_{1m}^{[n]} \quad (5)$$

隐层各神经元向输出层的传播代表了光波从光栅传播到谱面的夫朗和费衍射过程,这两层神经元连接的权重设为 $W_{2mk}^{[n]}$,而隐层各神经元代表谱面上各亮点的复振幅 $Out_{2k}^{[n]}$,于是由(1)和(4)两式可知

$$W_{2mk}^{[s]} = \begin{cases} 2 \cdot (-1)^{m+1} + \frac{1}{(1+M)X_m^{[s]}} & k = 0 \\ \frac{(-1)^m \exp(-i2\pi k X_m^{[s]})}{ik\pi X_m^{[s]}} & k \neq 0 \end{cases} \quad (6)$$

$$k = 0, \pm 1, \pm 2, \dots, \pm N; \quad m = 1, 2, \dots, M+1$$

并且输出是

$$\text{Out}_{2k}^{[s]} = \sum_{m=1}^{M+1} W_{2mk}^{[s]} \text{Out}_{1m}^{[s]} \quad (7)$$

将 $\text{Out}_{2k}^{[s]}$ 取模平方后便是输出神经元的输出: $I_k^{[s]} = |\text{Out}_{2k}^{[s]}|^2$ 亦即为谱面上各亮点光强。

为了评价每一次训练神经网络的误差变化情况,引入消费函数(costfunction)

$$\text{cost}^{[s]} = \frac{1}{2} \left[C_1 \sum_{K=0}^N (\hat{I} - I_k^{[s]})^2 + C_2 \sum_{K=0}^N (\bar{I} - I_k^{[s]})^2 \right] \quad (8)$$

其中 $C_1 + C_2 = 1$; $\hat{I} = \eta / (2N + 1)$, η 是希望的 Dammann 光栅衍射效率

$$\bar{I} = \sum_{k=0}^N I_k^{[s]} / (2N + 1)$$

是输出的平均光强,(8)式右边的第一项表征了输出光强相对于希望设计的光强的大小;第二项表征了输出光强均匀性的大小,训练神经网络的目的就是使消费函数减少到某一个希望的极小值以内。

对于每一个输出神经元,引入综合性误差

$$\delta_{3k}^{[s]} = - [C_1 (I_k^{[s]} - \hat{I}) + C_2 (\bar{I} - I_k^{[s]})] \quad (9)$$

式中负号的作用是当它逆向传播到输入层时它引起权重的变化是使每一个神经元偏离理想时的综合误差减小。

依据神经网络后向传播原理,该偏差逆着原来的传播方向最终传播到输入层引起权重的变化并引起下一次正向传播时的误差减小。在输出层引起其输出的变化

$$\delta_{2k}^{[s]} = \frac{\partial I_k^{[s]}}{\partial \text{Out}_{2k}^{[s]}} \delta_{3k}^{[s]} \quad (10)$$

然后,逆向传播到隐层引起该层各神经元的变化

$$\delta_{1m}^{[s]} = \sum_{k=0}^N W_{2mk}^{[s]} \delta_{2k}^{[s]} \quad m = 1, 2, \dots, M \quad (11)$$

最终它引起输入层的权重变化^[6]

$$\Delta W_{1m}^{[s+1]} = \text{Out}_0 \cdot \text{Out}_{1m}^{[s]} \cdot \delta_{1m}^{[s]} \quad (12)$$

因此训练第 $(n+1)$ 项时的新权重是

$$W_{1m}^{[s+1]} = W_{1m}^{[s]} + \alpha \Delta W_{1m}^{[s+1]} + \beta \Delta W_{1m}^{[s]} \quad (13)$$

式中 α 和 β 是小于 1 的影响收敛速度的实验参数,其大小依据计算机模拟过程来决定。

值得注意的是,公式(10)存在复数域中求导数的情况,该导数幅角是不确定的(见附录),然而由于权重 $\Delta W_{1m}^{[s+1]}$ 的变化是一个实数,因此由(10),(11),(12)式可以获得对该幅角的一个约束条件

$$\sin(2\theta_k^{[s]}) \text{Im} \left\{ \sum_{k=0}^N W_{2mk}^{[s]} \text{Out}_{2k}^{[s]} \delta_{3k}^{[s]} \right\} - \cos(2\theta_k^{[s]}) \text{Re} \left\{ \sum_{k=0}^N W_{2mk}^{[s]} \text{Out}_{2k}^{[s]} \delta_{3k}^{[s]} \right\} = \text{Im} \left\{ \sum_{k=0}^N W_{2mk}^{[s]} \text{Out}_{2k}^{[s]} \delta_{3k}^{[s]} \right\} \quad (14)$$

这充分表明,对于光强的微小变化,只有当复振幅的变化沿满足(14)式约束条件的方向变化时,才能保证各坐标点 $[0, X_1, X_2, \dots, X_M, 1]$ 的变化是最佳的,即使误差以最快的速率递减,这就是该算法比其他算法可以较快的收敛的重要依据。

3 设计结果及分析

依据上述的神经网络模型,用 C 语言编制了设计一维 $(0, \pi)$ 型 Dammann 光栅的程序,以此设计了若干个 Dammann 光栅,发现该模型有较快的收敛速度,尤其是优化均匀性效果明显,在选取适当的初始结构条件下取 $C_1 = C_2 = 0.5$ 时,设计 1×15 点的 Dammann 光栅的优化结果如表 1 所示,衍射效率达 81.7%,均匀性为 3.339×10^{-2} ;在此基础上取 $C_1 = 0, C_2 = 1$,以优化均匀性为目的,在 IBM386 兼容机上运行约 15 min 就可获得表 2 的结果,其衍射效率为 81.6%,均匀性为 3.7×10^{-5} 。

Table 1 The optimum result of a 15×15 fan-out Dammann grating, when $C_1 = 1.0$, and $C_2 = 1.0$

	0	1	2	3	4	5	6	7	8
Coordinate: X_m	0.0	0.14272	0.26720	0.61389	0.70410	0.78132	0.86935	0.91856	1.0
Intensity: $I_i \times 10^{-1}$	0.53723	0.55202	0.53387	0.54836	0.54714	0.54093	0.54745	0.54423	
Average intensity: $\bar{I} = 0.05443$	Diffraction efficiency: 0.8165					Uniformity: 0.03			

Note: Uniformity = $[\max(I_i) - \min(I_i)]/\bar{I}$, the same bellow

Table 2 The optimum result of a 15×15 fan-out Dammann grating, when $C_1 = 0$, and $C_2 = 1.0$

	0	1	2	3	4	5	6	7	8
Coordinate: X_m	0.0	0.14370	0.26763	0.61413	0.70511	0.78188	0.87009	0.91972	1.0
Intensity: $I_i \times 10^{-1}$	0.54376	0.54376	0.54376	0.54376	0.54375	0.54376	0.54377	0.54376	
Average intensity: $\bar{I} = 0.054376$	Diffraction efficiency: 0.8156					Uniformity: 0.00004			

Table 3 Some optimum results of Dammann gratings

Fan-our number	5	7	13	15	19	23
Variable coordinates number	3	3	7	7	9	11
Diffraction efficiency: (%)	77.5	78.7	79.6	81.6	82.1	83.6
Uniformity	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-5}	10^{-4}

由于仍不是求解全局最优解的方法,我们采用多次优化从中选择最佳值的方法,表 3 是设计的 $(0, \pi)$ 型 Dammann 光栅的部分设计结果,均匀性和衍射效率均比我们已知的文献报道的高。

关于解决全局最优问题,我们将采用模拟退火的方法,有关研究结果将另文叙述。

参 考 文 献

- 1 H. Dammann, K. Gortler. High-efficiency in-line mutiple imaging by mutiple phase holograms. *Opt. Commun.*, 1971, 3:313
- 2 Wilfrid B. Veldkamp, James R. Leger, Gary J. Swanson. Coherent summation of laser beams using binary phase gratings. *Opt. Lett.*, 1986, 11(5):303

- 3 J. Fagerholm, J. Turunen, A. Vasara *et al.*. Computer-generated Fourier-transform holograms of optical computing. in *Holographic Optics I: Principles and Applications*, G. M. Morris ed. *Proc. Soc. Photo-opt. Instrum. Eng.*, 1989, 1136:253~260
- 4 H. Stark, W. C. Catino, J. Locicero. Design of phase gratings by generalized projections. *J. Opt. Soc. Am.*, 1991, A8:566~571
- 5 Uri Mahlab, Joseph Shamir, H. John Caulfield. Genetic algorithm for optical pattern recognition. *Opt. Lett.*, 1991, 16(9):648~650
- 6 P. D. Wasserman. *Neural Computing Theory and Practice*. Van Nostrand Reinhold, 1989. 43~60
- 7 R. L. Morrison. Symmetries that simplify the design of spot array phase grating. *J. Opt. Soc. Am. A*, 1992, 9:9494

Design of Dammann Gratings by Neural Network Backpropagation Algorithm

Liao Jianghong Sang Tao Lu Zhengwu Weng Zhicheng

(State Key Lab. of Applied Optics, Changchun Institute of Optics and Fine Mechanics,
Academia Sinica, Changchun 130022)

Abstract This paper proposed a novel method to design binary optical elements based on neural network backpropagation, by which we have gotten a faster convergence speed and the better uniformity. At last, we give six Dammann gratings' designed results.

Key words Damman grating, binary optical elements, neural network

附录

因

$$I_k = \text{Out}_{2k} \cdot \text{Out}_{2k}^*$$

则

$$\frac{\partial I_k}{\partial \text{Out}_{2k}} = \left(\text{Out}_{2k}^* + \text{Out}_{2k} \frac{\partial \text{Out}_{2k}^*}{\partial \text{Out}_{2k}} \right) \quad (\text{A}-1)$$

如图 3 所示, 设 $\Delta \text{Out}_{2k} = r_k \exp(i\theta_k)$, θ_k 是增量的幅角, 为不确定因子。

则

$$\begin{aligned} \frac{\partial \text{Out}_{2k}^*}{\partial \text{Out}_{2k}} &= \lim_{r_k \rightarrow 0} \frac{(\text{Out}_{2k} + r_k \exp(i\theta_k))^* - \text{Out}_{2k}^*}{r_k \exp(i\theta_k)} \\ &= \exp(-i2\theta_k) \end{aligned}$$

因此

$$\frac{\partial I_k}{\partial \text{Out}_{2k}} = \text{Out}_{2k}^* + \text{Out}_{2k} \cdot \exp(-i2\theta_k) \quad (\text{A}-2)$$

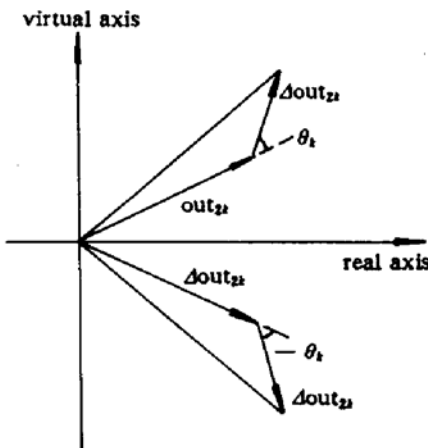


Fig. 3 The demonstration of the variation of Out_{2k}