文章编号:1671-4598(2025)10-0174-09 DOI:10.16526/j.cnki.11-4762/tp.2025.10.023 中图分类号:TP183 文献标识码:A

基于动态扰动衰减神经网络和算法的图像识别

费春国, 赵扬帆
(中国民航大学 电子信息与自动化学院, 天津 300300)

摘要: 在处理大规模图像数据集时, 梯度下降法作为一种常用的训练方法容易在获得部分数据的最优解后收敛速度变慢, 导致无法获得整体数据的最优解; 针对梯度下降法在图像识别任务中遇到的问题, 基于现有神经网络架构, 结合生物神经网络中的大脑皮质-基底神经节回路和E/I扰动现象, 提出了一种解决方法: 动态扰动衰减网络和动态扰动衰减梯度下降算法; 此网络在现有网络的输入层上引入一层逐渐减小的扰动层, 随着迭代次数的增加, 扰动层对输入层施加的扰动逐渐趋近于零; 该方法不仅加快了图像识别任务中梯度下降法的收敛速度, 还在整个训练过程中避免了局部最优解和过拟合的问题, 从而提升了网络的性能; 通过在MNIST、CIFAR-10和CIFAR-100等图像识别数据集上, 使用不同的神经网络和算法进行实验, 验证了所提出的动态扰动衰减网络和算法的有效性; 相较于原始网络使用Adam和SGDM算法, 动态扰动衰减方法在测试准确度上分别取得了0.16%～1.4%和0.39%～1.38%的提升, 同时具备更快的收敛速度; 在进一步实验中, 引入数据增强, 并与WAdam算法和Adagrad算法对比, 仍显示出显著提升, 表明该方法具有较强的鲁棒性和广泛适用性。

关键词: 深度学习; 神经生物科学; 局部最优; 优化算法; 大脑皮质回路

Image Recognition Based on Dynamic Perturbation Attenuation Neural Networks and Algorithms

FEI Chunguo, ZHAO Yangfan
(College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: When dealing with large-scale image datasets, a gradient descent method, as a commonly used training method, is prone to slow down in convergence after finding out an optimal solution for some data, leading to the failure to obtain the optimal solution for entire data. To address the issues encountered by the gradient descent method in image recognitions, on the basis of which the existing neural network architecture is combined with the cortico-basal ganglia loops and E/I perturbation phenomena in biological neural networks, a solution of the dynamic perturbation attenuation network and dynamic perturbation attenuation gradient descent algorithm is proposed. In this network, a gradually diminishing perturbation layer is introduced into the input layer of the existing network. With an increase of iterations, the perturbation applied to the input layer gradually approaches zero. This method not only accelerates the convergence speed of the gradient descent method in image recognition tasks, but also avoids local optimum and overfitting during entire training process, thus enhancing the network's performance. By using different neural networks and algorithms on image recognition datasets such as MNIST, CIFAR-10, and CIFAR-100, experimental results verify the effectiveness of the proposed dynamic perturbation attenuation network and algorithm. Compared with the original network using the Adam and SGDM algorithms, the dynamic perturbation attenuation method improves the test accuracy by a range of 0.16% to 1.4% and 0.39% to 1.38%, respectively, while also exhibiting a faster convergence. In further experiments, with the introduction of data augmentation and comparisons with the WAdam and Adagrad algorithms, the proposed method significantly improves the accuracy, demonstrating a strong robustness and broad applicability.

Keywords: deep learning; neurobiology; local optima; optimization algorithms; cortical circuit

收稿日期:2024-07-26; **修回日期:**2024-10-21。

作者简介:费春国(1974-),男,博士,副教授。
通讯作者:赵扬帆(1997-),男,硕士研究生。

引用格式:费春国,赵扬帆. 基于动态扰动衰减神经网络和算法的图像识别[J]. 计算机测量与控制,2025,33(10):174-182,207.

0　引言

文献［1］指出深度学习（Deep Learning）是机器学习的一个分支，通常使用神经网络的方法进行训练。神经网络（NN，neural network）或称类神经网络，是一种用于机器学习和认知科学领域的数学或计算模型，其设计灵感源自生物神经网络的结构与功能。神经网络主要用于对函数进行估计或近似，能够通过模拟神经元之间的连接来处理复杂的计算问题。大脑的神经系统是通过多种多样的神经元来构成的，人工神经网络通过对其模拟建立起相应的网络结构。文献［2］指出随着神经网络的发展，智能故障诊断[3]，工业自动化[4]，目标检测[5]等方面的问题得到了新的解决方法，但是新方法的出现也带来了新的问题。

梯度下降法（Gradient Descent）是一种简单且常用的优化算法，常常用来解决许多可导凸优化问题，例如逻辑回归和线性回归问题，并且在非凸优化领域中也有重要的地位，因此它成为神经网络的优化算法之一。在深度学习中选用损失函数作为梯度下降法的目标函数，通过损失函数的最小化，使得神经网络可以更好地对训练数据集进行非线性表达，从而对测试集有更好的匹配度。为了使得网络表达更加接近真实值，在梯度下降法的基础上加入反向传播算法（BP，back propagation），该方法通过采用反梯度方向作为链式法则的传播方向来将输出层的误差传递到输入层，进而调整相应的权重参数使得更新后的输出逼近真实数据，从而实现收敛。但是随着数据集的扩大，训练过程中会面临存在许多局部优化解的情况，因此使用梯度下降法存在着梯度消失和陷入局部最优的问题[6]。在处理大规模数据集中避免局部最优一直是相关的研究热点。确定学习率和选择寻优方向是梯度下降算法研究的核 心。确定学习率和优化方向是梯度下降算法研究的关键问题。文献［7］指出近些年的研究，国内外在这一领域取得了大量成果，主要集中在以下两个方面：1）关于梯度下降算法学习率的研究显著提高了算法的收敛速度，并在一定程度上解决了非凸目标函数容易陷入局部次优解的难题；2）基于动量和方差缩减的随机梯度下降（SGD）相关研究则有效解决了 SGD 在优化过程中的不稳定性问题[8]。（SGDM，stochastic gradient descent with momentum）算法在传统 SGD 算法的基础上引入了动量项，该类算法能够有效减少参数更新时的震荡现象，同时加速了网络收敛到最优解的过程，代表算法有 SGDM，NAG，SVRG 等。文献［9］指出当学习率设置得过低时，算法将需要经过大量迭代才能收敛，导致训练时间过长；相反，若学习率过高，算法可能会陷入局部最优，无法找到全局最优解，甚至可能导致结果超出初始值，最终引发算法发散。因此，基于梯度的二阶矩估计来自适应的调整学习率是第二类算法的改进点，其中代表性算法包括 Adagrad、Adaela 和 Adam 等。在网络训练过程中，这些算法通过利用历史梯度信息，为每个参数的不同分量动态调整学习率。

神经生物科学中，决策这一行为通过对外界各种证据的感知，再对证据进行积累后做出判断而产生，整个过程由大脑中的众多区域和各种回路合作完成[10]。大脑皮质－基底神经节回路是大脑皮质内部的神经元与基底神经节相互连接形成的网络，负责执行高级的认知功能，包括感知、思维、决策等。皮质－基底神经节回路控制活动的基本原理是控制兴奋（Excite）和抑制（Inhibition）之间的突触平衡（E/I 平衡），破坏这一平衡会导致认知缺陷如决策受损等精神疾病[11]。在皮质－基底神经节回路中，纹状体作为输入的门户具有 D1 和 D2 两种受体，两种受体分别对纹状体起到兴奋和抑制的作用。正常情况下，皮质－基底神经节回路通过不同的神经递质调节来达到 E/I 平衡的状态，但是当负责接受输入信息的纹状体的两种受体受到抑制后就会导致 E/I 平衡受到扰动。在受到不同的扰动后，皮质－基底神经节回路对于证据积累的过程便会受到不同程度影响，最终造成决策改变。根据生物神经网络系统动力学[12]，漂移扩散模型（Drift Diffusion Model）建立相应的决策预测模型后可以发现在 E/I 比率升高的情况下，决策过程受早期证据的影响较大，进而做出冲动的决策。而在 E/I 比率降低的情况下，决策过程会变得犹豫不决，证据整合速率会减缓[13]。

基于神经生物科学中的这一现象，本文从调整网络结构以模拟 E/I 扰动对整个神经网络的优化角度出发，提出了一种名为动态扰动衰减网络（DPDN，dynamic perturbation decay network）的网络结构和相应的动态扰动衰减下降算法（DPAA，dynamic perturbation attenuation algorithm）。该方法综合了 SGD 算法的优点[14]，通过改进深层网络的结构，使网络性能得到提升。本文将动态扰动衰减网络嵌入－几种网络模型中，并使用不同的数据集对改进后的网络模型和相应的算法进行对比实验，测试后的结果表明改进后的动态扰动衰减网络和动态扰动衰减算法有更好的效果。

1　网络模型和算法介绍

神经网络通过模拟神经生物网络的方式来完成机器学习和人工智能的任务。神经生物网络是生物体内神经元相互连接形成的网络，其功能和结构受到生物学的制约，存在着许多未知的复杂信息传递方式和信息处理系统[15]。神经网络在结构上模仿了生物神经网络的基本组成，包括人工神经元和连接方式，但是在对生物神经

网络中的复杂调节机制和化学信号扰动等方面的模拟上有所欠缺。

生物神经网络中，大脑皮质层作为最终的控制中心，通过将大量兴奋性神经递质谷氨酸投射给基底神经节，在经过一系列的信息加工后经过丘脑返回皮质。基底神经节与大脑皮质之间构成三条基底神经节回路[16]：

1）直接回路（皮质－含有 D1 类受体的纹状体－苍白球内侧/黑质网状部－丘脑－皮质）；

2）间接回路（皮质－含有 D2 类受体的纹状体－苍白球外侧－底丘脑－苍白球内侧/黑质网状部－丘脑－皮质）；

3）超直接回路（皮质－底丘脑－苍白球－皮质）。

大脑皮质－基底神经节回路的整体结构如图 1 所示。

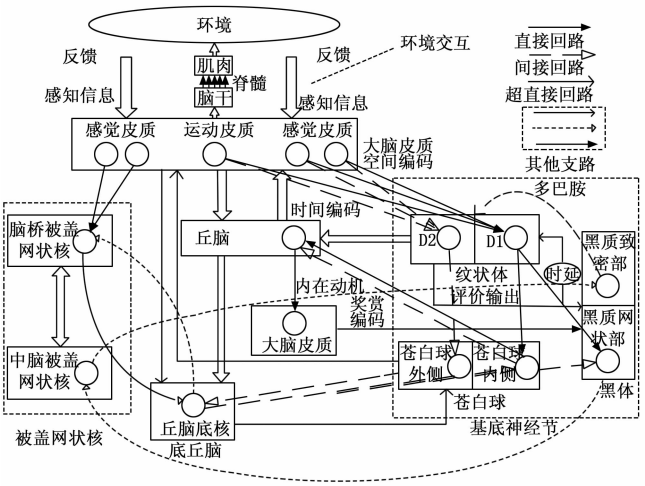

图 1 皮质－基底神经节回路

在大脑皮质层与基底神经节构成的回路中，纹状体是信息传入的关键节点，接受来自大脑皮质层的兴奋性信号以及来自黑质致密部（SNC）的抑制输入。苍白球与纹状体协同作用，将运动信号进行汇聚整合，将空间分布的神经元的并行输入转化为不同的控制输出信号，传递给丘脑[17]。丘脑扮演信息输出的重要通道，它将经过基底神经节整合的神经冲动传递给大脑皮质。三条回路中，直接回路对皮质层起兴奋性作用，间接回路对皮质层起抑制性作用[18]，超直接回路的作用目前尚不明确[19]。黑质致密部释放的多巴胺对直接回路和间接回路起着重要的调节作用，其表现为对直接回路有着兴奋性作用，对间接回路则有着抑制性作用[20]。因此皮质－基底神经节回路通过调节纹状体，丘脑等结构所分泌的神经递质的量来调整回路的输出，进而控制皮质层做出各种动作和决策。

人工神经网络以全连接网络为例，整个网络分为输入层、隐藏层和输出层三层。其中输入层与大脑皮质回路中的纹状体的功能相似，负责信息的输入；隐藏层则与苍白球及其神经投射构成的“黑盒”网络功能相似，负责信息的处理；输出层则与丘脑相似，负责输出最终的结果，对应结构如图 2 所示。

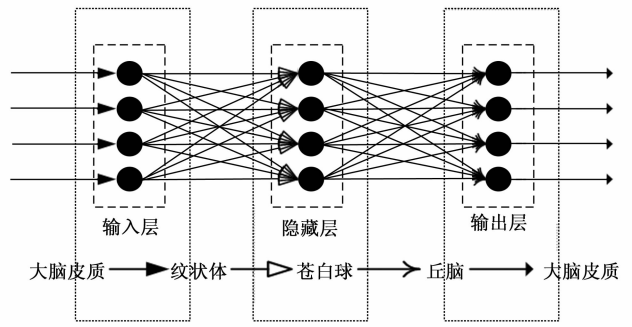

图 2 全连接网络与皮质－基底神经节回路网络对应结构

在网络结构上，以 E/I 扰动现象为基础，增加一扰动层，扰动层中的扰动元数值在（0，1）中获取并随着迭代次数的增加不断衰减。扰动层生成扰动信号后将其引入到输入层的每个神经元上，并设置相应的衰减系数来模拟大脑皮质－基底神经节回路中纹状体的受体功能衰退所造成的 E/I 扰动现象，从而构造出一种动态扰动衰减网络结构（DPDN）。扰动层的引入增加了网络的深度，有效减少了网络对部分数据的过拟合，提高了模型的非线性表示能力，使其在大数据集训练中表现更为优越。

如图 3 所示，以 3 层全连接网络为例，动态扰动衰减下降算法（DPAA）的参数更新表达式如下所示。由于扰动层的加入，整体的网络结构也有所改变，设输入层神经元个数为 m，隐藏层个数为 n，输出层神经元个数为 l。输入数据为 x_i，扰动信号输入为 δ，扰动的衰减权重为 ω_i'。输入层与隐藏层之间的连接权重设为 ω_{ij}，其中（$i=1,2,3,\cdots,m$；$j=1,2,3,\cdots n$）。隐藏

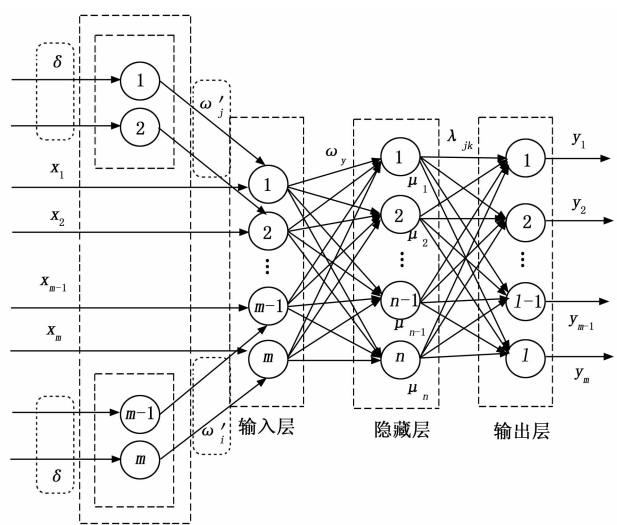

图 3 3 层 DPDN 结构

层与输出层之间的连接权重设为 λ_{jk}，其中（$j=1,2,3,\cdots,n$；$k=1,2,3,\cdots,l$）。设隐藏层的输出为 μ_j，输出层的输出设为 y_k。激活函数设置为 f，隐藏层神经元的偏置设置为 b_j，输出层神经元的偏置设置为 φ_k，则隐藏层的输出可以表示为：

$$\mu_j = \sum_{i=1}^{m} f(\omega_{ij}(x_i+\omega_i'\delta)+b_j) \tag{1}$$

可以推出输出层的输出为：

$$y_k = \sum_{j=1}^{n} f(\lambda_{jk}\mu_j+\varphi_k) \tag{2}$$

因此，整个 DPDN 网络可以表达为函数：

$$y_k = \sum_{j=1}^{n} f(\lambda_{jk}\sum_{i=1}^{m} f(\omega_{ij}(x_i+\omega_i'\delta)+b_j)+\varphi_k) \tag{3}$$

误差函数 $F(\omega)$ 设置为交叉熵函数，步长设为 η，根据反向传播算法，DPDN 的参数更新公式（以不同层的连接权重 ω 为例）：

$$\Delta\omega = -\eta\times\frac{\partial F}{\partial \omega_{t-1}} \tag{4}$$

$$\frac{\partial F}{\partial \omega_{t-1}} = \frac{\partial F}{\partial y}\times\frac{\partial y}{\partial \omega_{t-1}} = \frac{\partial(-y_k'\log y_k)}{\partial y_k}\times\frac{\partial y_k}{\partial \omega_{t-1}} \tag{5}$$

$$\frac{\partial F}{\partial \omega_{t-1}} = -\frac{y'_{t-1}}{y_{t-1}}\times f'\lambda_{t-1}f'\sum_{i=1}^{n}(x_i+\omega'_{t-1}\delta) \tag{6}$$

$$\omega_t = \omega_{t-1}-\eta\times\frac{y'_{t-1}}{y_{t-1}}\times f'\lambda_{t-1}f'\sum_{i=1}^{n}(x_i+\omega'_{t-1}\delta) \tag{7}$$

扰动衰减权重 ω'_t 在每次的迭代后进行一次衰减，其衰减系数为 α，其衰减表达式为：

$$\omega'_t = (1-\alpha)\omega'_{t-1} \tag{8}$$

算法 1：动态扰动衰减下降法（DPAA）

Parameters：全局学习率 η，扰动信号 δ，扰动衰减权重 ω'，扰动衰减系数 α，最大迭代次数 T，初始参数 ω，ω'，时间步长 t

Require：学习率 $\eta=0.001$，$\alpha=0.02$，$\delta=0.09$

Require：初始化参数 ω，ω'

$\omega\leftarrow\omega\sim N(0,1^2)$，$\omega'\leftarrow\omega'\sim N(0,1^2)$

Require：定义参数的损失函数 $F(\omega)$

While：$F(\omega)$ 不收敛 do

For all i in T do

$t\leftarrow t+1$

$\Delta\omega_{t+1}\leftarrow\nabla F_{t+1}(\omega_t)$：计算 t 时刻损失函数 $F(\omega)$ 的梯度

$\omega_{t+1}\leftarrow\omega_t+\Delta\omega_{t+1}$：更新参数 ω

$\omega'_{t+1}=(1-\alpha)\omega'_t$：更新扰动衰减权重值 ω'

End for

Return ω，ω'

End while

2　收敛性分析

在本节中，在证明 DPAA 算法的收敛性前引入相应的数学理论。

理论 1：多元泰勒展开（Multivariate Taylor Expansion）

当一个二阶函数 $f(\omega):\mathbb{R}^n\to\mathbb{R}$ 可导时，那么它的泰勒展开式就可以表示为：

$$f(\omega) = f(\mu)+\nabla f(\mu)^T(\omega-\mu)+\frac{1}{2}(\omega-\mu)^T\nabla^2 f(\gamma)(\omega-\mu) \tag{9}$$

其中：ω 和 μ 是函数 $f(\omega)$ 定义域中的任意两点，γ 是 ω 和 μ 众多的凸组合中的一种，其表达式为 $\gamma=n\omega+(1-n)\mu$[21]。

理论 2：利普西茨连续（Lipschitz continuity）

给定两个度量空间 (M,d_M)，(N,d_N)，$U\subseteq M$，其中 d_M 是集合 M 的测度，d_N 是集合 N 的测度。若对于函数 $f:U\to N$，存在常数 L 使得 $d_N[f(a),f(b)]\leqslant Ld_M(a,b)\quad\forall a,b\in U$，则说明函数 f 符合利普西茨连续[22]。

因此若函数 $f:\mathbb{R}^n\to\mathbb{R}$ 利普西茨连续，则对于任意两点 ω 和 μ 都满足：

$$\|f(\omega)-f(\mu)\|\leqslant L\|\omega-\mu\| \tag{10}$$

理论 3：下降引理（The Descent Lemma）

若一个二阶函数 $f(\omega):\mathbb{R}^n\to\mathbb{R}$ 可导，其导数 $\nabla f(\omega)$ 满足利普西茨连续条件，则它的海森矩阵（Hessian Matrix）满足 $\nabla^2 f(\omega)\leqslant LI$，其中 L 为利普西茨常数。

其中导数 $\nabla f(\omega)$ 变化最快的方向为其海森矩阵 $\nabla^2 f(\omega)$ 绝对值最大特征值所对应的特征向量。由于海森矩阵 $\nabla f(\omega)$ 为对称矩阵（symmetric matrix），则根据瑞利商（Rayleigh quotient）可得：对于任意的 σ 都满足：

$$\sigma^T\nabla^2 f(\omega)\sigma\leqslant\lambda_{\max}\|\sigma\|^2 \tag{11}$$

对于函数 $f(\omega)$ 而言，它任意点的海森矩阵都满足 $\lambda_{\max}\leqslant L$。

因此根据理论 1 和理论 2，对于一个二阶可导，且导数 $\nabla f(\omega)$ 满足利普西茨连续的函数 $f(\omega)$ 进行多元泰勒展开可得[23]：

$$f(\omega) = f(\mu)+\nabla f(\mu)^T(\omega-\mu)+\frac{1}{2}(\omega-\mu)^T\nabla^2 f(\gamma)(\omega-\mu)\leqslant f(\mu)+\nabla f(\mu)^T(\omega-\mu)+\frac{L}{2}\|\omega-\mu\|^2 \tag{12}$$

根据以上理论，为了证明 DPAA 算法的收敛性，首先设损失函数 $F(\omega)$ 是一个连续的凸函数，学习率为 η。由理论 2 可得，对于所有的 $x,y\in R^d$ 都成立如下不等式：

$$\|\nabla F(\omega_{t+1})-\nabla F(\omega_t)\|\leqslant L\|\omega_{t-1}-\omega_t\| \tag{13}$$

其中：$\|x\|$ 代表了向量的模长，L 即为利普西茨常数，$L\leqslant\frac{1}{\eta}$。通过对损失函数 $F(\omega)$ 进行泰勒二次展开，

则对于任意 $x,y\in R^d$，都满足如下不等式：

$$|F(\omega_{t+1})-F(\omega_t)-\nabla F(\omega_t)^T(\omega_{t+1}-\omega_t)|\leqslant \frac{L}{2}\|\omega_{t+1}-\omega_t\|^2 \tag{14}$$

将 $\omega_{t+1}=\omega_t-\eta\nabla F(\omega_t)$ 代入式（14）中，得到：

$$F(\omega_{t+1})\leqslant F(\omega_t)+\nabla F(\omega_t)^T(\omega_{t+1}-\omega_t)+\frac{L}{2}\|\omega_{t+1}-\omega_t\|^2$$

$$F(\omega_{t+1})=F(\omega_t)-\nabla F(\omega_t)^T\eta\nabla F(\omega_t)+\frac{L}{2}\|\eta\nabla F(\omega_t)\|^2$$

$$F(\omega_{t+1})=F(\omega_t)-\eta(1-\frac{L\eta}{2})\|\nabla F(\omega_t)\|^2$$

$$\because L\leqslant\frac{1}{\eta},\text{并且 }\eta,L>0$$

$$\therefore 1-\frac{L\eta}{2}\leqslant\frac{1}{2}$$

得到：

$$F(\omega_{t+1})\leqslant F(\omega_t)-\frac{1}{2}\eta\|\nabla F(\omega_t)\|^2 \tag{15}$$

当 DPAA 算法收敛时，算法 1 中的目标函数 $F(\omega)\geqslant F(\omega^*)$，即存在一个精确值 ε，使得任意参数 ω_t 满足[24]：

$$\sum_{t=1}^{T}E[\nabla F(\omega_t)-\nabla F(\omega^*)]\leqslant\varepsilon \tag{16}$$

设 $\omega_{t+1}=\omega_t-\eta\nabla F(\omega_t)$，将 $\omega^*=\omega$ 代入式（15）中，得到：

$$\sum_{t=1}^{T}E[F(\omega_t)-F(\omega^*)]\leqslant\frac{1}{2T}\eta\left\|\lambda_t\sum_{i=1}^{m}(x_i+\omega'_t\delta)\right\| \tag{17}$$

证明：设 $\omega^*=\operatorname{argmin}\sum_{t=1}^{T}F(\omega_t)$，根据理论 3 和式（14），得到：

$$F(\omega_{t+1})\leqslant F(\omega^*)+\nabla F(\omega_t)(\omega_t-\omega^*)-\frac{1}{2}\eta\|\nabla F(\omega_t)\|^2$$

$$F(\omega_{t+1})-F(\omega^*)\leqslant\frac{\omega_t-\omega_{t+1}}{\eta}(\omega_t-\omega^*)-\frac{1}{2\eta}\|\omega_t-\omega_{t+1}\|^2\leqslant$$

$$\frac{1}{2\eta}\|\omega_t-\omega^*\|^2-\frac{1}{2\eta}(\|\omega_t-\omega^*\|^2-$$

$$2\eta\nabla F(\omega_t)(\omega_t-\omega^*)+\|\eta\nabla F(\omega_t)\|^2)$$

$$F(\omega_{t+1})-F(\omega^*)\leqslant\frac{1}{2\eta}(\|\omega_t-\omega^*\|^2-\|\omega_{t+1}-\omega^*\|^2)$$

$$\sum_{t=1}^{T}F(\omega_t)-TF(\omega^*)\leqslant\frac{1}{2\eta}(\|\omega_0-\omega^*\|^2-\|\omega_T-\omega^*\|^2)$$

$$\sum_{t=1}^{T}E[F(\omega_t)-TF(\omega^*)]\leqslant\frac{1}{2T}\eta\|\nabla F(\omega^*)\|^2=$$

$$\frac{1}{2T}\eta\left\|\lambda\sum_{i=1}^{m}(x_i+\omega'_t\delta)\right\|^2=O(\frac{1}{T})$$

因此当迭代次数达到最大迭代次数时 $\lim_{t\to T}\omega'_t=0$，根据式（16），可以找到一个精确值 $\varepsilon=O\left(\frac{1}{T}\right)$，使得 DPAA 算法收敛，收敛率为 $O\left(\frac{1}{T}\right)$。

3 数值实验分析

为了验证 DPDN 网络的有效性，实验中使用了不同的网络结构来对属于图像识别领域的 MNIST[25]、CIFAR-10 和 CIFAR-100 数据集进行测试。其中 MNIST 数据集是 NSIT 数据集的一个子集，包含 70 000 张手写数字图像。CIFAR-10 数据集是一个广泛用于计算机视觉领域的小型图像分类数据集，包括飞机、汽车、鸟类、猫、鹿、狗、青蛙、马、船和卡车。CIFAR-100 是 CIFAR 数据集的一个扩展，包含 100 个类别的图像，每个类别有 600 个图像。对 MNIST 数据集，使用了包含 3 层全连接层的网络和 LeNet-5 卷积神经网络[26]；对 CIFAR-10 数据集的训练采用 ResNet18 网络[27]；使用 ResNet50 网络对 CIFAR-100 数据集进行训练。为了进一步优化网络模型，在采用的 3 种网络模型中引入 DPAA 算法，通过多次数值实验获取可靠结果来对比 DPAA 算法与另外两种优化算法（Adam 和 SGDM）在相同数据集上的性能。为确保实验的公正性，在保持其他共同参数一致的情况下，本文对每种算法进行了 5 次独立实验取平均值，在 SGDM 算法中，设定动量项 $\beta=0.9$；在 Adam 算法中，分别设置 $\beta_1=0.9$ 和 $\beta_2=0.999$，记录并计算每种方法在训练集和测试集上的准确度以及损失值。所有实验均基于 Python 语言实现。

3.1 全连接网络测试 MNIST 数据集

在 MNIST 数据集上使用 3 层全连接网络进行测试。通过引入 DPDN 方法，在输入层前加入一层扰动层，网络的参数设置如下：网络的损失函数设置为交叉熵函数，学习率 $\eta=0.001$，Batch-size 设置为 100，动态扰动 $\delta=0.09$，衰减系数 $\alpha=2\times10^{-4}$。输入层包含 900 个神经元，隐藏层设置为 600 个神经元，输入层和隐藏层均采用 ReLU 激活函数，输出层采用 SoftMax 作为激活函数。

图 4 和图 5 显示了 DPDN 网络上采用 DPAA 算法与全连接网络采用另外两种算法的训练和测试时的准确度和损失值过程曲线。

表 1 展示了在 MNIST 数据集上使用全连接网络和 DPDN 网络采用 3 种不同的优化算法：DPAA、Adam 和 SGDM 进行实验的结果。

表 1 MNIST 数据集在全连接网络上

算法	训练损失值	训练准确度/%	测试损失值	测试准确度/%
DPAA	0.087	98.52	0.083	98.51
Adam	0.106	98.40	0.103	98.35
SGDM	0.101	98.06	0.096	98.12

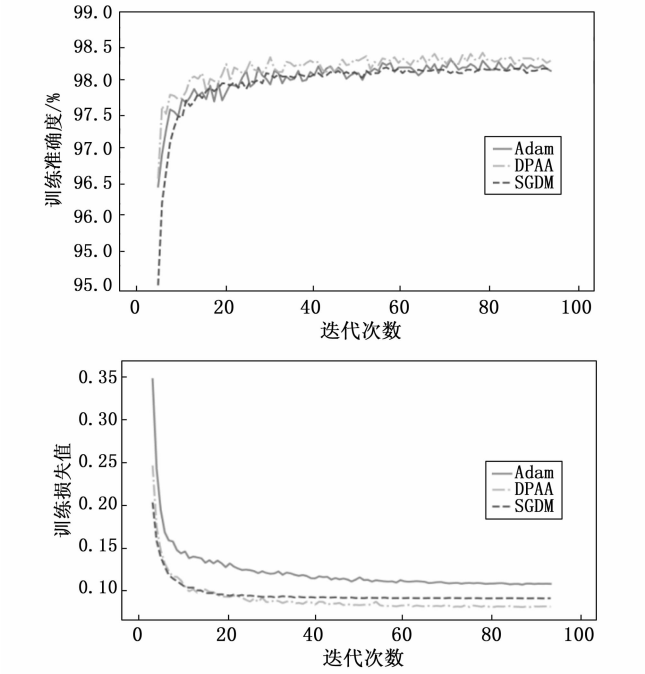

图 4　训练的准确度和损失值

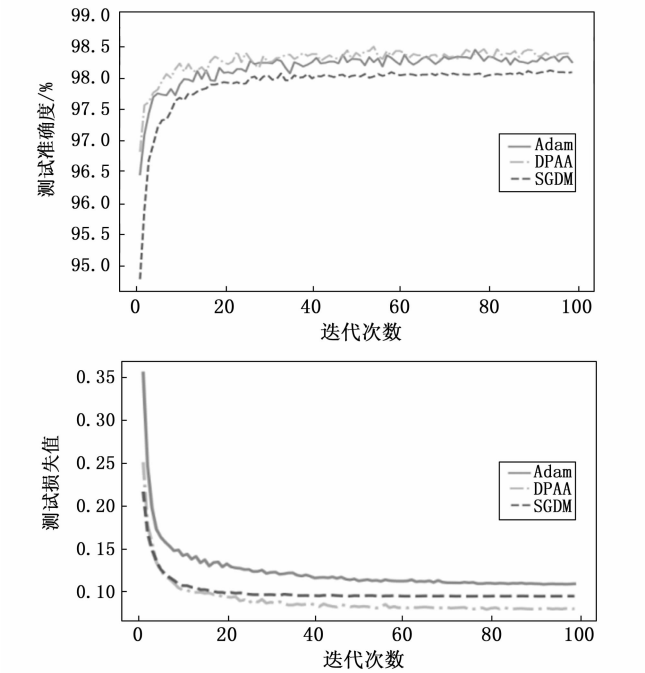

图 5　测试的准确度和损失值

通过表 1 的对比可以观察到，在 MNIST 数据集上，DPAA 算法在训练和测试数据上均表现最佳，这表明 DPAA 算法在优化能力和提高分类精度方面表现更为突出。多次实验结果表明，DPAA 相较于 Adam 在训练损失值上降低 0.019，测试损失值降低 0.02，训练准确度提升了 0.12%，测试准确度提升了 0.16%。与 SGDM 相比，DPAA 在训练损失值上降低了 0.014，测试损失值降低了 0.013，训练准确度提升了 0.46%，测试准确度提升了 0.39%。这些结果显示出 DPAA 在分类准确度和损失值的优化上具有显著优势，相比 Adam 和 SGDM 有明显的提升。由图 4 和图 5 可以看出，在网络的收敛速度上，DPAA 算法明显快于其他两种方法。综合上述分析来看，DPAA 算法在 MINIST 数据集上使用全连接网络结构训练比 Adam 算法和 SGDM 算法有着更好的准确度和收敛速度。

3.2　卷积网络测试 MNIST 数据集

LeNet-5 是经典的卷积神经网络（CNN，convolutional neural networks）架构，专为手写数字识别任务（如 MNIST）设计。它由多个卷积层、池化层和全连接层组成，本文将其中的全连接层替换为 DPDN 网络来测试 DPAA 算法的效果。整体网络的具体设置如下：最大迭代次数为 100 次，学习率 $\eta = 0.001$，批次大小设置为 128，扰动层扰动 $\delta = 0.08$，加入正则化方法来防止过拟合的发生，其中权重衰减系数的设置为 $\alpha = 2 \times 10^{-4}$，引入 Dropout 层的 rate 设置为 0.5。图 6 和图 7 展示了在卷积神经网络以及经过改进的 DPDN 网络上使用 3 种不同算法的整体训练和测试过程中的损失值曲线和准确度曲线。

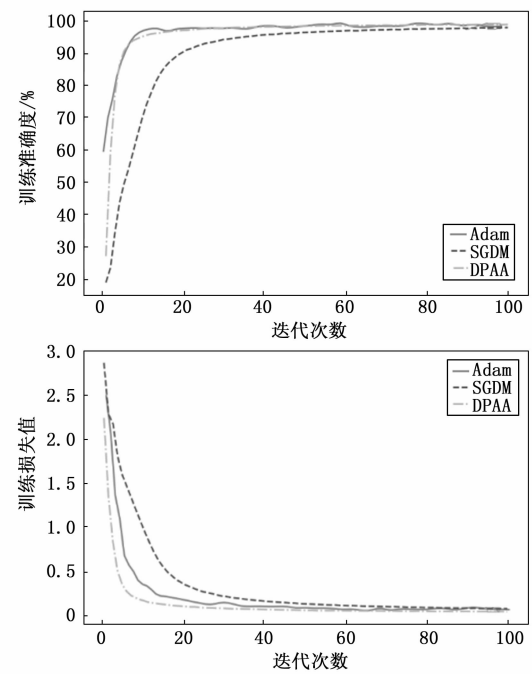

图 6　训练的准确度和损失值

表 2 展示了在 MNIST 数据集上使用 CNN 网络和 DPDN 网络采用 3 种不同的优化算法：DPAA、Adam 和 SGDM 进行实验的结果。

表 2　MNIST 数据集在卷积网络上

算法	训练损失值	训练准确度/%	测试损失值	测试准确度/%
DPAA	0.052	98.49	0.031	99.12
Adam	0.082	97.67	0.049	98.61
SGDM	0.091	97.46	0.053	98.49

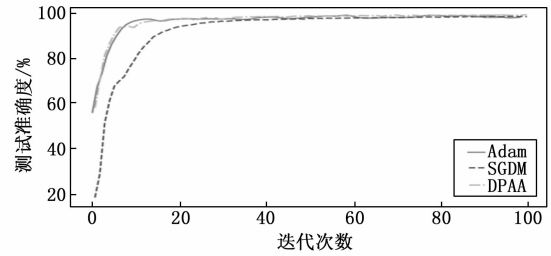

图 7 测试的准确度和损失值

由表 2 可以发现 DPAA 算法在训练和测试阶段的表现均优于其他两种算法。具体而言，DPAA 算法的训练损失为 0.052，测试损失为 0.031，相比 Adam 算法的训练损失 0.082 和测试损失 0.049，分别减少了 0.030 和 0.018；与 SGDM 算法的训练损失 0.091 和测试损失 0.053 相比，DPAA 算法的训练损失减少了 0.039，测试损失减少了 0.022。此外，DPAA 的训练准确度为 98.49%，测试准确度为 99.12%，相比 Adam 算法分别提高了约 1.82%和 0.51%；与 SGDM 算法相比，DPAA 算法在训练准确度和测试准确度上分别提高了约 1.03%和 0.63%。通过对图 6 和图 7 的观察，可以看出 DPAA 算法在收敛速度上与 Adam 算法相当，两者要快于 SGDM 算法；在准确度方面 DPAA 算法略优于 Adam 算法和 SGDM 算法。尤其在损失值方面，DPAA 算法的降低幅度更为显著。

3.3 残差网络测试 CIFAR-10

对于 CIFAR-10 数据集，采用 18 层的残差网络（ResNet18）进行测试验证，其中将全连接网络层替换为 DPDN 网络，并进行 DPAA 算法、Adam 算法和 SGDM 算法的比较。为了进一步验证动态扰动网络和 DPAA 算法的鲁棒性和泛化性，对数据集进行数据增强，并增加了 Wadam 和 Adagrad 算法来进行对比。网络参数设置如下：损失函数选用交叉熵函数，激活函数使用 ReLU 函数，并设定最大迭代次数为 40 次。学习率 $\eta = 0.001$，Batch-size 设置为 256，动态扰动 $\delta = 0.12$，衰减系数 $\alpha = 2.0 \times 10^{-4}$。

表 3 显示了 3 种方法在训练和测试时的最佳准确度和损失值。

由表 3 可以发现 DPAA 算法在训练集上的准确度上略低于 Adam，较高于 SGDM 算法，三者在损失值上相差较少。但是在测试集上，DPAA 算法的损失值相比 Adam 和 SGDM 算法有所降低，准确度上相比 Adam 算法提升 0.83%，相比 SGDM 算法提升 1.38%。在引入 WAdam 和 Adagrad 算法后，可以发现 DPAA 算法相较于另外两种算法有较大的优势。图 8 和图 9 展示了 CIFAR-10 数据集在 Res-Net 18 网络上的实验结果。可以发现 DPAA 算法的整体收敛速度是要快于 Adam 算法和 SGDM 算法的。因此，DPAA 算法在 Resnet-18 网络结构上比其他算法更具优势。

表 3 CIFAR－10 数据集在 Res－Net 18 上

算法	训练损失值	训练准确度/%	测试损失值	测试准确度/%
DPAA	0.20	93.4	0.47	86.04
Adam	0.15	94.74	0.54	85.21
SGDM	0.21	92.76	0.53	84.66
Aadgrad	0.33	89.05	0.69	78.02
WAdam	0.27	91.36	0.55	82.76

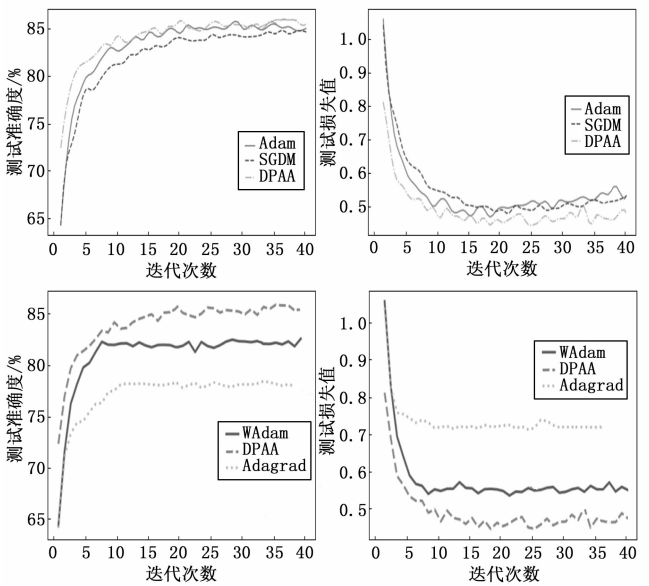

图 8 测试的准确度和损失值

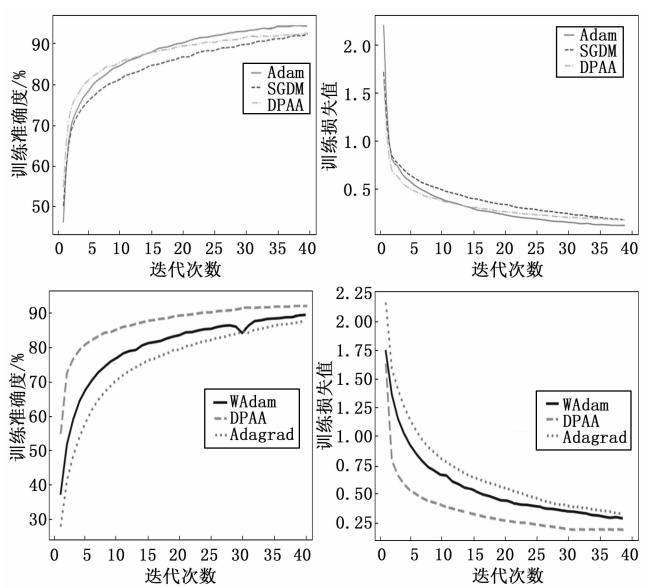

图 9 训练的准确度和损失值

3.4　残差网络测试 CIFAR-100

本文使用 50 层的残差网络 ResNet50 在 CIFAR-100 数据集上进行实验验证，将初始网络加入扰动层搭建为 DPDN 网络，加入数据增强来对动态扰动衰减网络和 DPAA 算法的鲁棒性和广泛性进行验证。网络参数设置如下：设置最大迭代次数为 40 次，学习率 $\eta = 0.001$，批次 Batch-size 设置为 256，动态扰动 $\delta = 0.08$，衰减系数 $\alpha = 1.5 \times 10^{-4}$。

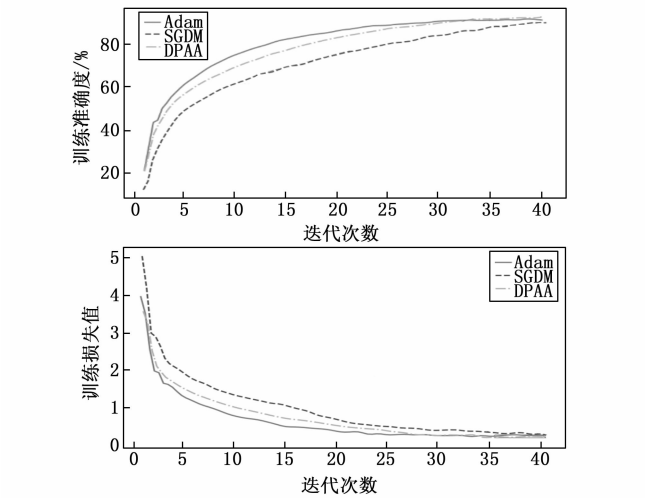

图 10　训练的准确度和损失值

表 4 展示了在 CIFAR-100 数据集上使用 Resnet50 网络和 DPDN 网络采用 3 种不同的优化算法：DPAA、Adam 和 SGDM 进行实验的结果。

表 4　CIFAR-100 数据集在 ResNet50 网络上

算法	训练损失值	训练准确度/%	测试损失值	测试准确度/%
DPAA	0.209	94.05	1.503	63.63
Adam	0.244	92.76	1.595	62.23
SGDM	0.288	91.14	1.513	62.56

由表 4 可以发现在训练集上 DPAA 算法和 Adam 算法要优于 SGDM 方法。DPAA 算法在训练和测试阶段均优于其他两种算法，表现出较低的训练损失值和较高的训练准确度。对于 Adam 算法来说，使用 DPAA 算法可以使平均准确度提高 1.4%，损失值降低了 0.092；对于 SGDM 算法而言，使用 DPAA 算法可以提高 1.07%的准确度，降低 0.01 的损失值。图 10 和图 11 显示了 DPDN 网络上采用 DPAA 算法与 Resnet50 网络采用另外两种算法的训练和测试时的准确度和损失值过程曲线。可以看出，DPAA 算法在训练集上的收敛速度与 Adam 算法相似，但最终收敛效果优于 Adam 算法，并且整体表现优于 SGDM 算法。在测试集中，DPAA 算法的曲线较为稳定，在准确度和损失值方面相较于其他两种算法具有一定的优势。

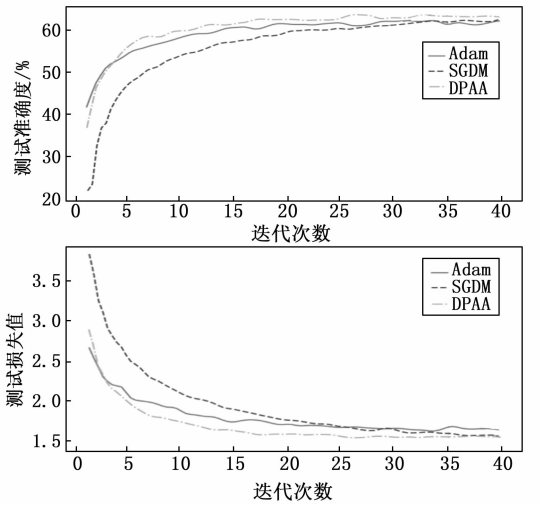

图 11　测试的准确度和损失值

3.5　参数量和时间分析

为了进一步的验证动态扰动衰减网络在添加扰动层后对原本网络模型的性能影响，在不同的数据集上采用不同的优化算法对模型参数量以及平均推理时间进行评价，验证其模型参数量和平均推理时间的改变。本文采用的硬件设备为 RTX4080，CUDA 版本为 11.8。

表 5 展示了 MINIST 数据集下两种网络模型使用不同优化算法的模型参数量和训练时间。

表 5　MNIST 数据集的模型评估

网络	算法	模型参数量/百万	平均推理时间/(秒/次)
全连接	DPAA	0.24	0.001
	Adam	0.24	0.002
	SGDM	0.24	0.001
Le-net5	DPAA	0.06	0.002
	Adam	0.06	0.005
	SGDM	0.06	0.003

可以发现在 MINSI 数据集上在使用全连接网络模型和卷积网络模型的情况下，动态扰动衰减网络和 DPAA 算法的模型参数量没有太大的改变，而平均推理时间有所降低。

表 6 展示了 CIFAR-10 数据集下使用 resnet-18 的模型参数量和训练时间。

表 6　CIFAR－10 数据集的模型评估

网络	算法	模型参数量/百万	平均推理时间/(毫秒/次)
Resnet18	DPAA	11.18	0.006
	Adam	11.18	0.009
	SGDM	11.18	0.007
	Adagrad	11.18	0.009
	WAdam	11.18	0.006

可以发现在使用 Resnet18 模型的情况下，对 CIFAR-10 数据集进行训练，动态扰动衰减网络和 DPAA

算法的模型参数量没有太大的改变，而平均推理时间有所降低。

表 7 展示了 CIFAR-100 数据集下使用 Resnet-50 的模型参数量和训练时间。

表 7 CIFAR-100 数据集的模型评估

网络	算法	模型参数量/百万	平均推理时间/(毫秒/次)
Resnet50	DPAA	23.71	0.13
	Adam	23.71	0.19
	SGDM	23.71	0.14

通过对 3 种数据集的实验对比，可以发现使用动态扰动衰减网络和 DPAA 算法在模型参数量上变化不大，但是在平均推理时间上相较于其他的算法有一定的优势。

4 结束语

本文基于神经生物科学原理以及 E/I 扰动现象，针对神经网络在大数据集情况下训练时，通过模拟大脑皮质-基底神经节回路的 E/I 扰动现象，在全连接层上增加一层扰动层，通过加入相应的衰减权重构建了一种动态扰动衰减网络，并提出了相应的动态扰动衰减下降法来优化梯度下降法和反向传播算法易陷入局部最优的问题，提高了网络获取全局最优解的概率，加快了收敛速度。通过数学证明了该算法的收敛性和稳定性，表明该网络可以更好地表示输入与输出之间的非线性关系。4 组实验结果表明，在 MNIST、CIFAR-10 和 CIFAR-100 数据集上，将 DPDN 网络替换原始网络后，DPAA 算法相比 Adam 和 SGDM 算法展现出更快的收敛速度和更高的测试精度。与 Adam 算法相比，DPAA 算法的测试准确度提升了 0.16%～1.4%；相较于 SGDM 算法，测试准确度的提升范围在 0.39%～1.38%之间。此外，通过准确度和损失值曲线以及对模型进行参数量和平均推理时间的评估表明，DPAA 算法能够更迅速地收敛到全局最优解。另外在数据增强的情况下，加入 WAdam 和 Adagrad 算法进行对比实验验证后发现，DPAA 算法仍然具有一定的优势，进而表明 DPAA 算法具有一定的鲁棒性和泛化性。综上所述，本文提出的方法对于大数据集下神经网络的训练易陷入局部最优解和提高收敛速度等问题具有一定意义。

参考文献：

[1] 邱锡鹏. 神经网络与深度学习 [M]. 北京：机械工业出版社，2020.

[2] SCHMIDHUBER，JüRGEN. Deep learning in neural networks：an overview [J]. Neural Netw，2015，61：85－117.

[3] 宫文峰，陈 辉，张美玲，等. 基于深度学习的电机轴承微小故障智能诊断方法 [J]. 仪器仪表学报，2020，41 (1)：195－205.

[4] 江 兵，杨 春，杨雨亭，等. 基于 aco 优化 BP 神经网络的变压器热点温度预测 [J]. 电子测量与仪器学报，2022 (10)：235－242.

[5] WANG A，LU J，CAI J. Large-margin multi-modal deep learning for rgb-d object recognition [J]. IEEE Transactions on Multimedia，2015，17 (11)：1887－1898.

[6] GLOROT X，BENGIO Y. Understanding the difficulty of training deep feedforward neural networks [J]. Journal of Machine Learning Research，2010，9：249－256.

[7] 王 昕. 梯度下降及优化算法研究综述 [J]. 电脑知识与技术，2022，18 (8)：71－73.

[8] 谢 涛，张春炯，徐永健. 基于历史梯度平均方差缩减的协同参数更新方法 [J]. 电子与信息学报，2021，43 (4)：956－964.

[9] 李兴怡，岳 洋. 梯度下降算法研究综述 [J]. 软件工程，2020，23 (2)：1－4.

[10] LARISSA A，GUSTAVO D. The encoding of alternatives in multiple-choice decision making [J]. Proceedings of the National Academy of Sciences of the United States of America，2009，106 (25)：10308－10313.

[11] MARIA C，ARVID C. Schizophrenia：A subcortical neurotransmitter imbalance syndrome? [J]. Schizophr Bull，1990 (3)：425－432.

[12] 陆启韶，刘深泉，刘 锋，等. 生物神经网络系统动力学与功能研究 [J]. 力学进展，2008 (6)：766－793.

[13] LAM N H，THIAGO B，JAIME H，et al. Effects of altered excitation-inhibition balance on decision making in a cortical circuit model [J]. The Journal of Neuroscience：the Official Journal of the Society for Neuroscience，2021，42 (6)：1－7.

[14] RUDER S. An overview of gradient descent optimization algorithms [J]. CoRR，2016，abs/1609.04747.

[15] IVANCEVIC V G，IVANCEVIC T T. Macro- and microscopic self-similarity in neuro and psycho-dynamics [J]. International Journal of Biomathematics，2009，2 (3)：243－251.

[16] ALEXANDER G E，CRUTCHER M D. Functional architecture of basal ganglia circuits：Neural substrates of parallel processing [J]. Trends in Neurosciences，1989，13 (7)：266－271.

[17] FRANK M J. Dynamic dopamine modulation in the basal ganglia：a neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism [J]. Journal of Cognitive Neuroscience，2005，17 (1)：51－72.

(下转第 207 页)