



## 融合人类认知模式的自主发育网络及其在手势识别的应用

韦知辛, 方勇纯

引用本文:

韦知辛,方勇纯. 融合人类认知模式的自主发育网络及其在手势识别的应用[J]. 智能系统学报, 2023, 18(1): 144–152.

WEI Zhixin,FANG Yongchun. Autonomous developmental network incorporating human cognitive modes and its application in gesture recognition[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(1): 144–152.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202212002>

## 您可能感兴趣的其他文章

### 基于质心分水岭算法的静态手势分割算法模型

Static gesture segmentation algorithm model based on centroid watershed algorithm  
智能系统学报. 2019, 14(2): 346–354 <https://dx.doi.org/10.11992/tis.201804028>

### 基于快速密度聚类的RBF神经网络设计

Construction of RBF neural networks via fast density clustering  
智能系统学报. 2018, 13(3): 331–338 <https://dx.doi.org/10.11992/tis.201702014>

### 基于自组织模糊神经网络溶解氧控制方法研究

Dissolved oxygen concentration control method based on self-organizing fuzzy neural network  
智能系统学报. 2018, 13(6): 905–912 <https://dx.doi.org/10.11992/tis.201801019>

### 去冗余Top-k对比序列模式挖掘

Mining Top-k non-redundant distinguishing sequential patterns  
智能系统学报. 2018, 13(5): 680–686 <https://dx.doi.org/10.11992/tis.201702019>

### 基于PCNN的图像椒盐噪声滤除方法

Filtering image impulse noise by using a PCNN image noise reduction technique  
智能系统学报. 2017, 12(2): 272–278 <https://dx.doi.org/10.11992/tis.201605027>

DOI: 10.11992/tis.202212002

# 融合人类认知模式的自主发育网络 及其在手势识别的应用

韦知辛<sup>1</sup>, 方勇纯<sup>1,2</sup>

(1. 南开大学机器人与信息自动化研究所, 天津 300071; 2. 南开大学人工智能学院, 天津 300071)

**摘要:** 自主发育算法在智能机器人等领域有着广阔的应用前景, 考虑到现有方法的局限性, 本文提出了一种融合人类认知模式的自主发育神经网络, 并将该方法应用于手势识别任务中。通过动态改变神经元预响应值计算过程中的人为指导部分来模拟人类的学习方式, 提出了基于动态  $k$  值的 top-k 竞争机制, 并模拟人脑对知识的接收和记忆功能实现优胜神经元突触权重的更新, 最后基于神经元认知能力反馈进行突触重调整。对比实验结果表明, 与原有方法相比, 经该文改进后的自主发育网络在手势识别任务中具有更好的学习效果和更高的识别率。

**关键词:** 自主发育网络; 手势识别; 人类认知模式; top-k 竞争机制; 神经元认知能力反馈; 突触重调整; 增量分级回归树算法; 叶成分分析算法

**中图分类号:** TP183    **文献标志码:** A    **文章编号:** 1673-4785(2023)01-0144-09

**中文引用格式:** 韦知辛, 方勇纯. 融合人类认知模式的自主发育网络及其在手势识别的应用 [J]. 智能系统学报, 2023, 18(1): 144-152.

**英文引用格式:** WEI Zhixin, FANG Yongchun. Autonomous developmental network incorporating human cognitive modes and its application in gesture recognition[J]. CAAI transactions on intelligent systems, 2023, 18(1): 144-152.

## Autonomous developmental network incorporating human cognitive modes and its application in gesture recognition

WEI Zhixin<sup>1</sup>, FANG Yongchun<sup>1,2</sup>

(1. Institute of Robotics and Automatic Information System, Nankai University, Tianjin 300071, China; 2. College of Artificial Intelligence, Nankai University, Tianjin 300071, China)

**Abstract:** The autonomous development algorithm has broad application prospect in the intelligent robots and other fields. Considering the limitation of existing methods, this paper proposes an autonomous development neural network integrating human cognitive mode, and applies the method to gesture recognition tasks. Specifically, the proposed method dynamically changes the human-guided part for the computation of neuron pre-response values to simulate the way humans learn. In addition, we propose a top-k competition mechanism based on the dynamic  $k$  value. Besides, in the method, human brain's ability of receiving and memorizing knowledge is simulated to update the synaptic weights of the winning neurons, and neuron cognitive ability is utilized as the feedback for synaptic readjustment. Comparative experimental results show that, compared with the original method, the improved autonomic development network in this paper has better learning effect and higher recognition rate in gesture recognition task.

**Keywords:** autonomous developmental network; gesture recognition; human cognitive modes; top-k competition mechanism; neuron cognitive ability feedback; synaptic readjustment; incremental hierarchical discriminant regression; lobe component analysis

随着人工智能技术和机器人技术的迅猛发

展, 人机交互方式不断朝着智能化发展, 其中手势识别技术一直是人机交互领域的研究热点。手势交流具有表达方便、形象直观、信息丰富、无需接触等优点, 可以提供更自然、更高效的交互体验<sup>[1]</sup>。

收稿日期: 2022-12-02.

基金项目: 国家自然科学基金面上项目 (61873132); 先进计算与关键软件海河实验室项目 (22HHXCJC00003).

通信作者: 方勇纯. E-mail: fangyc@nankai.edu.cn.

©《智能系统学报》编辑部版权所有

在人机手势交流过程中, 人们更希望机器人能够像人一样具有自主学习的能力, 可以通过自主发育实现对陌生手势的感知和理解, 从而适应人与机器人复杂多变的手势交互。

## 1 相关研究

大多数早期的手势识别研究都是基于复杂的人工特征提取和手势分割<sup>[2]</sup>。Huang 等<sup>[3]</sup>使用 Gabor 特征提取器和支持向量机 (support vector machine, SVM) 分类器进行手势识别任务。Tarvekar 等<sup>[4]</sup>提出了一种皮肤阈值分割方法, 通过分割图像中的手部区域并使用 SVM 分类器从中提取颜色和边缘特征来识别和分类手势。基于机器学习的手势识别方法一般需要在分类前提取特征, 不仅操作复杂, 而且效率低、泛化能力差<sup>[5]</sup>。随着深度学习技术特别是卷积神经网络 (convolutional neural network, CNN)<sup>[6]</sup> 的蓬勃发展, 基于深度学习的手势识别已成为主流方向<sup>[7]</sup>。Simonyan 等<sup>[8]</sup>使用双流卷积神经网络实现视频中的手势识别。Tran 等<sup>[9]</sup>提出了一种 3D CNN 模型 C3D (convolutional 3D), 解决了基于视频的手势识别需要同时处理时间和空间特征的问题。Molchanov 等<sup>[10]</sup>将收集到的图像和雷达信息融合到卷积神经网络中进行手势分类。Oyodotum 等<sup>[11]</sup>首先对图像进行二值化预处理操作, 然后设置阈值来定位手势, 最后使用卷积神经网络识别手势。ElBadawy 等<sup>[12]</sup>将归一化图像送入 3D 卷积神经网络进行手势识别。Carreira 等<sup>[13]</sup>提出了一种新的用于手势识别的二维卷积神经网络模型 (inflated 3D convnet, I3D)。Fang 等<sup>[14]</sup>提出了一种基于 CNN 的手势识别系统, 使用一种深度卷积生成对抗网络 (deep convolution generative adversarial networks, DCGAN) 来解决过拟合问题。Koller 等<sup>[15]</sup>使用 GoogLeNet<sup>[16]</sup>的 Inception-v1 从手语视频中提取数百万个弱标记数据, 并将其输入隐马尔可夫模型 (hidden Markov model, HMM) 以获得手势识别结果。为了提高深度神经网络的识别准确率, 研究人员不断加深模型深度, 设计出越来越复杂的网络模型。然而, 复杂的模型深度和庞大的模型参数对计算能力提出了很高的要求, 不利于实际应用。此外, 将神经网络应用到机器人上实现人机交互时, 机器人原有的知识和技能无法适应新任务、新目标, 对于不同的手势识别任务则需要人为重新设计方案并且提供大量新的训练样本, 其学习性能和泛化性能比较有限。

自主发育神经网络 (developmental network,

DN) 是一种新兴的、自适应更新和受大脑启发的学习系统<sup>[17]</sup>。自主发育模型的思想最早由 Weng 等<sup>[18]</sup>提出, 并提出了增量分级回归树算法 (incremental hierarchical discriminant regression, IHDR)<sup>[19]</sup>, 通过模仿大脑联想皮质的自动发育计算模型, 给高维的回归或决策空间增量式地构建一棵决策树或者回归树。黄敏等<sup>[20]</sup>利用自主发育模型, 在 IHDR 算法和反向传播 (back propagation, BP) 神经网络复合框架下, 实现机器人对智能空间情景下用户所需服务的准确认知; Lu 等<sup>[21]</sup>从智能家居数据中的离散部分构造一个 IHDR 树, 用于存储机器人的历史经验, 并结合强化学习方法实现家庭服务机器人以类人的方式自主开发服务认知技能。为了使机器人充分利用人类的心智发育模式, 建立一个通用的智脑模型, 钱夔等<sup>[22]</sup>利用 DN 实现机器人室内场景识别任务; 王东署等<sup>[23]</sup>利用 DN 对不同背景光线图像中人脸的朝向进行识别; Castro-Garcia 等<sup>[24]</sup>利用 DN 实现多种语言的识别任务; 针对 DN 的内部表示缺乏层次结构的问题, Zheng 等<sup>[25]</sup>在发育网络中提出了新的机制, 通过动态初始化具有不同连接范围的神经元来扩展复杂的层次结构, 进而完成视觉导航任务; 余慧瑾等<sup>[26]</sup>针对机器人场景识别任务, 在自主发育网络的基础上, 加入了负向学习机制, 利用连续型样本的特点加强学习, 并提出了将自主发育网络与多传感器融合技术相结合的场景识别方法<sup>[27]</sup>, 在一定程度上提高了机器人场景识别的准确率。

自主发育神经网络虽然对不同的任务在理想条件下都能取得不错的效果, 但针对同类内差别大的情况仍然存在神经元对输入信息学习不够彻底、学习效率欠佳、神经元利用率较低、识别效果较差等问题。考虑到传统神经网络方法的局限性, 本文提出了一种融合人类认知模式的自主发育网络: 在神经元预响应值计算过程中, 随着神经元的学习逐渐减小人为指导部分的比例, 从而更好地模拟人类的学习方式, 促进神经元对输入信息的学习; 采用动态值的 top-k 竞争机制, 机动性选择优胜神经元以提高神经元的利用率, 并改进优胜神经元的突触权重更新方程, 进一步提高神经元的学习效率; 最后通过神经元的认知反馈信息修正神经元的突触权重, 对神经元的学习进行优化, 从而提高网络的识别效果。利用改进后的自主发育神经网络实现手势识别任务, 实验结果表明, 改进后的方法比原有方法可以更好地应对类内变化大的问题, 并且具有更好的学习效果



和识别准确度。

## 2 自主发育算法

最简单的 DN 由 3 层自组织网络组成: 输入层  $X$ 、大脑神经层  $Y$ 、效应层  $Z$ , 其架构如图 1 所示。

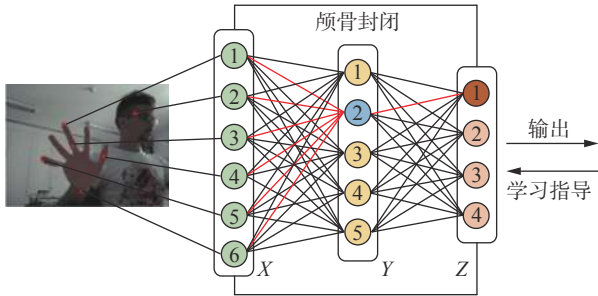


图 1 DN 的架构

Fig. 1 Architecture of DN

DN 算法在训练阶段, 通过通路  $X \rightarrow Y \leftarrow Z$  引入自下而上和自上而下的信息流, 并通过 top-k 竞争机制和叶成分分析算法 (lobe component analysis, LCA) 对相同模式重复激活从而加强映射, 训练过程中主要包含以下步骤:

1) 从  $X$  层获取外部环境信息输入向量  $\mathbf{x}_b$ , 从  $Z$  层获取对应的人为指导输入向量  $\mathbf{z}_t$ , 根据输入向量对  $Y$  层的每个神经元  $i$  计算预响应值  $r_i$ :

$$r_i = (1 - \alpha) \frac{\mathbf{V}_{ybi} \cdot \mathbf{x}_b}{\|\mathbf{V}_{ybi}\| \cdot \|\mathbf{x}_b\|} + \alpha \frac{\mathbf{V}_{yti} \cdot \mathbf{z}_t}{\|\mathbf{V}_{yti}\| \cdot \|\mathbf{z}_t\|} \quad (1)$$

式中:  $\mathbf{V}_{ybi}$  为该神经元对应的自底向上权重向量,  $\mathbf{V}_{yti}$  则为自顶向下权重向量, 人为指导系数  $\alpha$  ( $0 < \alpha < 1$ ) 为人为指导输入对于预响应值计算的影响系数,  $\alpha$  越大意味着神经元学习所需的人为指导越多。

2) 根据 top-k 竞争机制, 选出预响应值最大的前  $k$  个神经元作为  $Y$  层优胜神经元。根据 Hebbian 学习机制更新  $Y$  层优胜神经元  $i$  的突触权重  $\mathbf{V}_{ybi}$  和  $\mathbf{V}_{yti}$ , 以及神经元年龄  $g_{yi}$ :

$$\mathbf{V}_{ybi} \leftarrow \omega_1(g_{yi})\mathbf{V}_{ybi} + \omega_2(g_{yi})r_i\mathbf{x}_b \quad (2)$$

$$\mathbf{V}_{yti} \leftarrow \omega_1(g_{yi})\mathbf{V}_{yti} + \omega_2(g_{yi})r_i\mathbf{z}_t \quad (3)$$

$$g_{yi} \leftarrow g_{yi} + 1 \quad (4)$$

式中:  $\omega_1(g_{yi})$  和  $\omega_2(g_{yi})$  为控制神经元突触权重更新率的参数, 通过遗忘平均算法 (amnesic mean) 计算得到。

3) 选择环境信息输入对应类别的  $Z$  层神经元  $j$  作为  $Z$  层的优胜神经元, 并更新其突触权重  $\mathbf{V}_{zbj}$  以及神经元年龄  $g_{zj}$ :

$$\mathbf{V}_{zbj} \leftarrow \omega_1(g_{zj})\mathbf{V}_{zbj} + \omega_2(g_{zj})\mathbf{R} \quad (5)$$

$$g_{zj} \leftarrow g_{zj} + 1 \quad (6)$$

式中:  $\mathbf{R}$  为  $Y$  层所有神经元的预响应值组成的预响应向量。

重复以上步骤, 直到迭代学习结束, 神经元将学习到的知识存储在突触权重中。

## 3 融合人类认知模式的自主发育算法

DN 算法虽然具有较好的自主发育能力, 能够模仿人类的思考方式根据已有知识做出决策, 但是 DN 在发育学习过程中的自适应性较差, 对于类内变化大的情况往往只能学习到比较模糊的知识表示, 导致识别正确率比较低。实际上, 人类在知识学习和认知发展过程中, 会根据具体情况不断调整其学习方式。因此, 人类认知模式的启发, 本文在原始 DN 算法的基础上融合了人类的认知模式, 进行了以下改进。

首先, 针对 1) 中神经元预响应值的计算, 本文根据神经元的学习情况和实际的表现, 动态改变人为指导部分的比例, 从而使神经元对输入信息做出准确有效的反应。其次, 针对 2), 本文提出了基于神经元整体预响应值水平的动态  $k$  值的 top-k 竞争机制, 灵活选择优胜神经元, 并在优胜神经元突触权重的更新过程中, 通过优胜神经元预响应值大小决定其学习率的大小, 实现优胜神经元突触权重的有差别更新, 以保证整体的学习效果。最后, 通过向神经元反馈其对于输入信息的认知能力来重新调整神经元的突触权重, 利用有效的反馈信息来修正神经元的认知偏差, 从而帮助神经元更好地学习, 进一步提高神经元的学习效率。

### 3.1 模拟人类学习方式的预响应值计算

在神经网络的指导学习阶段,  $Y$  层将同时接收到  $X$  层的外部环境信息输入和  $Z$  层的人为指导输入,  $Y$  层神经元在这两个输入的共同作用下产生最终的预响应, 其计算方式如式 (1) 所示, 其中人为指导系数  $\alpha$  的值在网络学习开始前就根据实际任务设置为一个定值, 在网络的学习过程中, 人为指导部分所占的比例总是固定的。其实, 预响应值衡量的是神经元对输入信息反应强度, 反应越强烈说明神经元对输入信息越熟悉。在学习过程中, 神经元对输入信息的熟悉程度不断提高, 需要的人为指导逐渐减弱。此时, 若人为指导部分在神经元的學習过程中固定不变, 则容易对神经元造成误导, 降低了学习效率。

事实上, 当个体或组织在一项任务中学习到更多经验, 其效率就会随之提高。根据学习曲线 (learning curve) 效应<sup>[28]</sup>, 越是经常地执行一项任务, 每次所需的成本就减少。类似的, 在神经元

学习过程中也应遵循同样的规律。因此, 受学习曲线启发, 考虑将人为指导比例系数 $\alpha$ 设置为随着神经元学习时间 $t$ 变化的幂函数 $\alpha_t$ , 从而更好地模拟人类的学习方式, 其计算方式为

$$\alpha_t = \alpha_0 \cdot t^n$$

$$n = \frac{\lg b}{\lg 2} \quad (7)$$

式中:  $\alpha_0$ 为初始学习所需要的人为指导比例 (在手势识别任务中设置为 0.6),  $t$ 为神经元对于同一类型知识学习的时间,  $b$ 为神经元的学习率 (设置为 0.8)。经过改进后的神经元预响应值计算方式为

$$r_i = (1 - \alpha_t) \frac{V_{ybi}}{\|V_{ybi}\|} \cdot \frac{x_b}{\|x_b\|} + \alpha_t \frac{V_{yti}}{\|V_{yti}\|} \cdot \frac{z_i}{\|z_i\|}$$

### 3.2 动态 $k$ 值的 top-k 竞争机制

在当前的自主发育神经网络中, 通常采用固定  $k$  值的 top-k 竞争机制, 在训练前就预先根据任务特点确定  $k$  值的大小, 神经网络对于不同知识内容的学习或者在不同的学习进程中, 每次都只激活预响应值最大的前  $k$  个神经元。这种做法导致神经元的利用率太低, 并且无法对神经元整体的学习水平有一个较好的评估, 无论神经元学习情况的好坏, 总是选择相同数量的神经元进行突触权重的更新, 在一定程度上降低了学习性能。

考虑到以上弊端, 本文提出了动态  $k$  值的 top-k 竞争机制, 在神经元的学习过程中, 每次均根据神经元突触与输入信息匹配的完美程度对  $k$  值进行动态调整, 即根据神经元预响应值整体分布情况动态选择  $k$  值, 综合考虑所有神经元的预响应值, 利用统计学的思想, 选择响应值较大且与其他预响应值相比差异较大的神经元作为优胜神经元, 在所有预响应值中选出较大的离群值, 这些离群值的个数即为  $k$  值。对于不同的识别任务, 神经元的响应值分布不同, 选择离群值的方法也有所差异, 例如, 对于手势识别的任务, 神经元响应值的分布近似为正态分布, 因此根据响应值分布的特点, 选择使用标准差法选择离群值,  $k$  值的计算方式为

$$r_q > \mu + 2\sigma$$

$$k = \max q$$

式中:  $r_q$ 为将预响应值由大到小顺序排列后的第  $q$  个预响应值,  $\mu$ 为所有神经元预响应值的均值,  $\sigma$ 为标准差。

经过上述对 top-k 竞争机制的改良, 实现了对优良神经元资源的合理利用, 并且提高了网络的学习效果, 在一定程度上提高了发育网络的识别正确率。

### 3.3 模拟人脑接收知识方式的突触权重更新

自主发育神经网络中, 神经元学习知识的体现就是对突触权重的不断更新, 神经元将学习到的知识转化为“记忆”存储在突触权重中。对于  $Y$  层神经元而言, 不仅有自底向上的突触权重, 还有自顶向下的突触权重, 其更新方式分别如式 (2)、(3) 所示。对于  $Z$  层神经元而言, 只有自底向上的突触权重, 其更新方式如式 (5) 所示。优胜神经元  $i$  的突触权重更新方程可以表示为

$$V_i \leftarrow \omega_1(g_i)V_i + \omega_2(g_i)r_iP \quad (8)$$

式中:  $r_i$ 为神经元响应值,  $P$ 为神经元的输入向量。

在优胜神经元的突触权重向量更新过程中, 不仅要学习新的知识技能, 还要保留已经学习到的知识经验。从式 (8) 的更新方程可以看出, 所有优胜神经元对于新旧知识的学习力度均通过学习因子  $\omega_1(g_i)$  和  $\omega_2(g_i)$  控制。

事实上, 神经元的响应值在一定程度上体现了神经元的学习能力, 神经元的响应值越大, 其对于新知识的学习能力则越强。因此, 响应值最大的神经元对新知识的学习能力最强, 随着响应值逐渐减小, 逐步降低新知识学习比例, 而增加旧知识的学习比例, 根据神经元响应值大小不断调整权重更新方程, 从而模拟人脑对知识的接收及记忆功能。优胜神经元之间对于新旧知识的学习比例差值与响应值的相对大小有关。

综合上述考虑, 引入一个新知识影响因子来衡量新知识对于神经元突触权重更新的影响。将预响应值按照由大到小的顺序排列后, 假设  $r_i$  为第  $i$  个优胜神经元的预响应值, 则新知识影响因子  $\beta$  的计算方式为

$$\beta_i = \frac{r_i}{r_1}$$

对于响应值最大的优胜神经元, 将其新知识影响因子最大为 1, 其余优胜神经元的响应值越大则新知识影响因子越大。在新知识影响因子和原有的学习因子的综合之下, 形成了神经元最终的突触权重更新速率, 更改后优胜神经元  $i$  的突触权重更新方程为

$$V_i \leftarrow \delta_1 V_i + \delta_2 r_i P$$

$$\delta_1 = 1 - \beta_i + \beta_i \omega_1(g_i)$$

$$\delta_2 = \beta_i \omega_2(g_i)$$

式中:  $\delta_1$  和  $\delta_2$  为最终控制神经元突触权重更新速率的参数,  $\omega_1(g_i)$  和  $\omega_2(g_i)$  则为神经元原有的学习因子, 且两者的和恒为 1, 因此  $\delta_1$  与  $\delta_2$  的和也恒为 1。

经过这样的更改, 网络可更为合理地利用神经元, 降低了网络在学习过程中对知识的遗忘, 提高了神经元的学习效率, 增强了神经网络的容错

性,并在一定程度上提高了网络的识别正确率。

### 3.4 基于神经元认知能力反馈的突触重调整

在神经元学习过程中,对于  $Y$  层神经元,随着神经元学习到的知识越来越多,神经元对  $X$  层输入信息的认知能力越来越大。但是,也存在神经元对输入信息的认知能力大量减小的情况,这是由于新输入的信息与神经元已学习的模式相差较大,此时,若仍按照先前的学习方式进行突触权重的更新,则会导致神经元在已学习模式和当前输入模式这两种差别较大的模式之间反复振荡,使得神经元对先前已学习的模式产生遗忘,并且无法对新模式产生较好的学习。

综合上述考虑,本文提出基于神经元认知能力反馈的突触重调整策略,向神经元反馈其对于当前输入信息的认知能力,根据反馈信息重新对神经元进行指导与突触权重的调整,避免神经元由于认知偏差带来的知识缺陷。优胜神经元  $i$  对于当前输入信息  $x_i$  的认知能力  $A_{i,t}$  计算式为

$$A_{i,t} = \frac{\mathbf{V}_{i,t}}{\|\mathbf{V}_{i,t}\|} \cdot \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}$$

式中  $\mathbf{V}_{i,t}$  是优胜神经元  $i$  在当前时刻的突触权重,即神经元已经学习的模式。

通过神经元当前认知能力  $A_{i,t}$  与其最大认知能力  $A_{i,\max}$  的比较,重新调整神经元的学习策略,讨论如下:

#### 1) $A_{i,t} \geq A_{i,\max}$ :

优胜神经元  $i$  的认知能力提升,意味着神经元对当前输入信息模式有着较强的学习能力,此时,对神经元的指导与突触的重新调整规则取决于认知能力  $A_{i,t}$  的大小,通过神经元认知能力阈值  $\tau$  判断,阈值  $\tau$  根据实际任务设置 (如在手势识别的任务中设置为 0.96)。

#### ① $A_{i,t} \geq \tau$ :

如果  $A_{i,t}$  大于阈值,则说明神经元已经学习到较多的知识,神经元对于当前的输入可以完美地响应,神经元对当前任务已经具备较强的能力,则可以保留原来的学习方式进行突触权重的更新。

#### ② $A_{i,t} < \tau$ :

如果  $A_{i,t}$  小于阈值,说明神经元对当前任务的能力比较欠缺,则适当增加人为指导部分,即适当增大式 (7) 中的学习率  $b$ ,并提高神经元突触权重更新时的学习率,即将新知识影响因子  $\beta_i$  设置为 1,加强神经元对输入知识的学习。

#### 2) $A_{i,t} < A_{i,\max}$ :

优胜神经元  $i$  的认知能力降低,说明神经元对当前输入信息较为陌生,此时的调整规则取决于

神经元认知能力变化的大小,通过神经元认知能力变化阈值  $\varepsilon$  来判断,该值根据实际任务设置 (如在手势识别的任务中设置为 0.02)。

#### ① $A_{i,\max} - A_{i,t} \leq \varepsilon$ :

若神经元对当前输入的认知能力只降低了较少的一部分,说明当前输入模式与神经元已经学习到的模式相差不大,神经元需要从输入中学习一些新的模式,则保留原来的学习方式。

#### ② $A_{i,\max} - A_{i,t} > \varepsilon$ :

若神经元对当前输入的认知能力降低比较多,说明当前输入模式与神经元已经学习到的模式相差较大,则降低神经元的学习率,即减小新知识影响因子  $\beta_i$ ,增强神经元对已有知识的记忆能力。本文针对手势识别任务,将新知识影响因子  $\beta_i$  设置为

$$\beta_i = \frac{r_i}{2r_1} e^{-\sqrt{A_{i,\max} - A_{i,t}}}$$

## 4 实验结果与分析

利用本文改进后的自主发育神经网络实现静态手势识别的任务,实验所用的数据集为帕多瓦大学多媒体技术与电信实验室提供的 Creative Senz3D 数据集<sup>[29-30]</sup>,该数据集通过 Creative Senz3D 相机采集得到,其中包含 4 个不同的人执行的 11 种不同的手势,并且每个手势重复执行 30 次。即每种手势共 120 个样本。数据集中 11 种手势的部分样例如图 2 所示。



图 2 Creative Senz3D 数据集中 11 种手势的部分样例  
Fig. 2 Samples of the 11 gestures in the Creative Senz3D dataset

考虑到 DN 网络针对的是输入图像的全局信息,为了减少背景像素的干扰,首先对数据进行手势提取。本文利用数据集中的深度图像,通过阈值分割、形态学操作等步骤获得手部区域掩码,通过掩码获取 RGB 图像的手势部分并进行分割,最后将得到的手势图像转化为  $100 \times 100$  像素的灰度图像,作为网络的输入数据。预处理过程如图 3 所示。



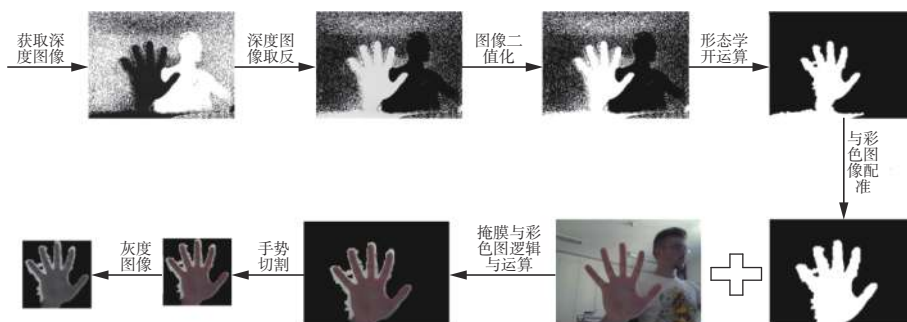


图 3 输入图像的预处理过程

Fig. 3 Preprocessing process of input image

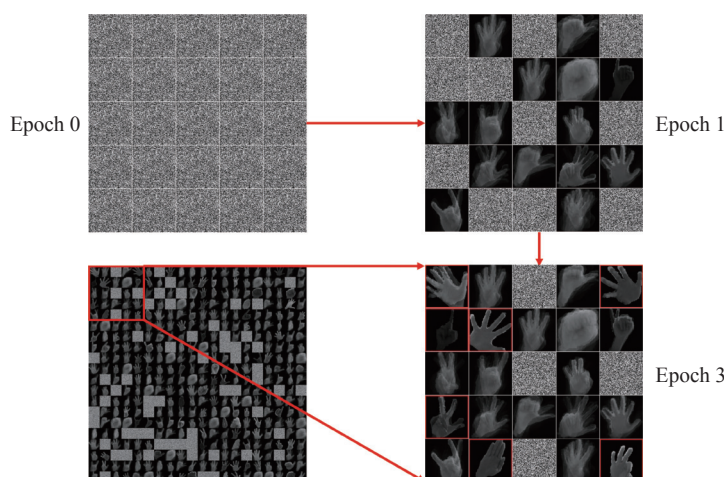
实验过程中, 对于同一种手势, 每个人分别选取 5 张图像作为训练图像, 其余作为测试图像, 训练与测试的数据划分比例为 1:5。根据任务要求设置  $Y$  层神经元的个数为 400。 $Z$  层神经元的个数设置为 12, 每个神经元代表不同的手势, 其中一个代表背景。训练开始前, 设置神经网络迭代学习的次数为 3。为了验证网络改进后的性能, 对原始 DN 和改进后的 DN 以相同的训练图像、相同的模型规模进行训练, 训练开始前, 对于网络中神经元的初始突触权重进行同样设置, 并且以相同的测试图像进行识别和验证。其中原始 DN 中 top-k 竞争机制的  $k$  值大小设置为 3。

#### 4.1 训练过程 $Y$ 层神经元突触权重的可视化

神经元突触权重的可视化可以展示神经元的學習情况, 图 4 给出了改进后训练过程中  $Y$  层神经元

在训练过程中自下而上的内部权重。每个框代表 1 个  $Y$  神经元, 矩形格呈现出图像表明对应神经元学习到了相应的手势特征, 随机图像则意味着神经元还没有被激活。图中具有红框的矩形格代表神经元从第 1~3 次迭代学习过程中新学习到的内容。

从图 4 可以发现, 改进后的 DN 方法中, 大部分神经元可以学习到手势特征, 从第 1~3 次迭代学习过程中不断有新的神经元学习到新的手势特征。由于在训练过程中对神经元人为指导比例的动态改变以及对突触权重的不断调整, 神经元学习到的手势特征都比较清晰, 对于同一手势类别不相同的输入图片, 神经元也能达到一个比较好的学习效果, 这在很大程度上提高了网络的学习效果和识别正确率, 增强了网络的容错性, 这也体现了改进后 DN 方法的优势。

图 4  $Y$  层神经元的权重可视化Fig. 4 Internal weights of  $Y$  neurons

#### 4.2 神经元利用率比较与分析

$Y$  层神经元的利用率在一定程度上体现了网络的学习效果, 图 5 记录了算法改进前后  $Y$  层神经元的利用率对比情况。可以发现, 原始的 DN 方法中神经元的利用率在整个迭代学习中均一样, 说明在学习过程中没有新的神经元学习到

知识, 并且神经元利用率非常低, 只有 8.25%, 这在一定程度上造成了内存资源的浪费, 并影响系统的识别性能。而改进后的 DN 方法在迭代学习过程中, 神经元利用率不断增大, 在第 3 次迭代学习之后神经元的利用率达到了 84.75%, 合理利用神经元的同时提高了网络的识别效果。

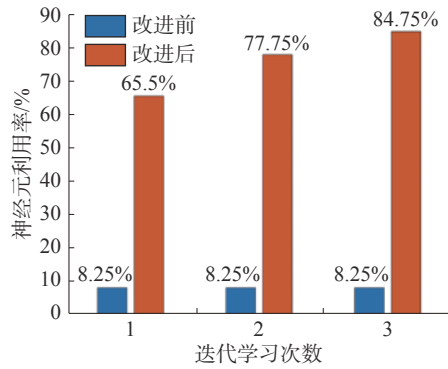


图 5 算法改进前后 Y 层神经元利用率对比

Fig. 5 Comparison of the utilization of Y neurons before and after algorithm improvement

#### 4.3 手势识别结果与分析

为了评估所提方法的性能,用数据集上测试的结果生成混淆矩阵。在目前的工作中,给定数

据集中总共 11 类静态手势,计算正确和错误分类比例的混淆矩阵将以大小为  $11 \times 11$  的方表形式布局,其中对角线部分记录正确分类的手势的比例,而非对角线部分记录错误分类的比例,得到的混淆矩阵如表 1 所示。

由表 1 可以发现,所有手势的识别精度均超过 90%,而大多数手势的准确率都在 95% 以上,且手势 G4、G5 的准确率高达 100%,11 种静态手势的平均识别率达到了 95.27%,对于大多数手势都以相当高的正确率被成功识别,而少数相似手势之间不可避免地混淆。

针对 Creative Senz3D 数据集上的静态手势识别方法,将本文方法与几种现有的主流手势识别技术<sup>[30-34]</sup>以及原始的 DN 方法进行了比较,对比结果如表 2 所示。

表 1 11 种手势的混淆矩阵

Table 1 Confusion matrix for the 11-class hand gestures

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
G1	<b>0.91</b>	0.00	0.05	0.01	0.00	0.00	0.00	0.03	0.00	0.00	0.00
G2	0.00	<b>0.96</b>	0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
G3	0.00	0.02	<b>0.97</b>	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
G4	0.00	0.00	0.00	<b>1.00</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G5	0.00	0.00	0.00	0.00	<b>1.00</b>	0.00	0.00	0.00	0.00	0.00	0.00
G6	0.00	0.00	0.01	0.02	0.00	<b>0.93</b>	0.00	0.01	0.00	0.00	0.03
G7	0.00	0.01	0.00	0.01	0.00	0.02	<b>0.95</b>	0.00	0.00	0.00	0.01
G8	0.01	0.03	0.00	0.03	0.00	0.00	0.00	<b>0.93</b>	0.00	0.00	0.00
G9	0.00	0.00	0.01	0.03	0.00	0.03	0.00	0.00	<b>0.93</b>	0.00	0.00
G10	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.03	0.00	<b>0.95</b>	0.00
G11	0.00	0.00	0.01	0.02	0.00	0.02	0.00	0.00	0.00	0.00	<b>0.95</b>

表 2 Creative Senz3D 数据集上不同方法的准确率

Table 2 Gesture recognition accuracy of different methods %

方法	准确率
3D 特征向量+ SVM 分类 <sup>[30]</sup>	90.0
Gabor 特征 <sup>[31]</sup>	90.6
形状特征描述符+ SVM 分类器 <sup>[32]</sup>	93.3
CNN 模型 <sup>[33]</sup>	94.61
MLP 分类器 <sup>[34]</sup>	95.98
原始 DN 方法	82.13
本文方法	95.27

从表 2 可以看出,相比原始 DN 方法 82.18% 的准确率,本文改进后的 DN 方法准确率提升了

13%,并且本文方法优于绝大多数文献中提出的手势识别方法。值得注意的是,本文方法仅仅利用很小一部分训练数据就达到较高的识别精度,而目前主流的手势识别方法中绝大部分都需要大量的训练数据才能获得较好的识别精度,在缺少训练数据的情况下,本文方法具有较大的优势。此外,本文方法不需要复杂的特征提取,其计算需求非常有限,可以快速得到训练模型,且可以实现快速实时识别。

## 5 结束语

针对自主发育神经网络对输入信息学习不彻底、学习效率低下、神经元利用率低、识别效果差等问题,本文提出了一种融合人类认知模式的自



主发育神经网络。在神经元预响应值计算过程动态改变人为指导部分的比例; 提出了基于神经元整体预响应值水平的动态  $k$  值的 top-k 竞争机制; 对优胜神经元突触权重进行有差别更新; 并通过神经元的认知反馈来重新调整神经元的突触权重。通过静态手势识别任务的实验结果表明, 本文方法在使用比较少训练数据的情况下实现了较好的识别精度, 与原有的算法相比, 本文提出的方法具有更好的学习效率, 并且在识别正确率方面有明显的提高。

在未来的研究工作中主要集中于以下两点:

1) 考虑增加新的注意力机制, 使神经元专注于输入图片的有效信息而忽略杂乱的背景, 在不使用手势分割的基础上实现手势识别, 从而在更加复杂和不受限制的环境中提高手势识别任务上的性能; 2) 提高网络的智能发育水平和特征表达能力, 实现在动态环境下的实时手势识别任务, 并将发育网络应用到机器人上, 利用手势识别在复杂环境下实现人机交互。

## 参考文献:

- [1] FENG Zhao, WU Jinlong, NI Taile. Research and application of multifeature gesture recognition in human-computer interaction based on virtual reality technology[J]. *Wireless communications and mobile computing*, 2021, 10: 1–5.
- [2] LI Jiayao, RAY S, RAJANNA V, et al. Evaluating the performance of machine learning algorithms in gaze gesture recognition systems[J]. *IEEE access*, 2021, 10: 1020–1035.
- [3] HUANG Dengyuan, HU W C, CHANG S H. Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination[J]. *Expert systems with applications*, 2011, 38(5): 6031–6042.
- [4] TARVEKAR M P. Hand gesture recognition system for touch-less car interface using multiclass support vector machine[C]//2018 Second International Conference on Intelligent Computing and Control Systems. Madurai: IEEE, 2019: 1929–1932.
- [5] RAHIM M A, MIAH A S M, SAYEED A, et al. Hand gesture recognition based on optimal segmentation in human-computer interaction[C]//2020 3rd IEEE International Conference on Knowledge Innovation and Invention. Kaohsiung: IEEE, 2021: 163–166.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the acm*, 2017, 60(6): 84–90.
- [7] GUO Xing, XU Wu, TANG Wen quan, et al. Research on optimization of static gesture recognition based on convolution neural network[C]//2019 4th International Conference on Mechanical, Control and Computer Engineering. Hohhot: IEEE, 2020: 398–3982.
- [8] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems–Volume 1. Montreal: ACM, 2014: 568–576.
- [9] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks [C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2016: 4489–4497.
- [10] MOLCHANOV P, GUPTA S, KIM K, et al. Multi-sensor system for driver's hand-gesture recognition[C]//2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Ljubljana: IEEE, 2015: 1–8.
- [11] OYEDOTUN O K, KHASHMAN A. Deep learning in vision-based static hand gesture recognition[J]. *Neural computing and applications*, 2017, 28(12): 3941–3951.
- [12] ELBADAWY M, ELONS A S, SHEDEED H A, et al. Arabic sign language recognition with 3D convolutional neural networks[C]//2017 Eighth International Conference on Intelligent Computing and Information Systems. Cairo: IEEE, 2018: 66–71.
- [13] CARREIRA J, ZISSERMAN A. Quo vadis, action recognition? A new model and the kinetics dataset[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 4724–4733.
- [14] FANG Wei, DING Yewen, ZHANG Feihong, et al. Gesture recognition based on CNN and DCGAN for calculation and text output[J]. *IEEE access*, 2019, 7: 28230–28237.
- [15] KOLLER O, NEY H, BOWDEN R. Deep hand: how to train a CNN on 1 million hand images when your data is continuous and weakly labelled[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 3793–3802.
- [16] JAHANDAD, SAM S M, KAMARDIN K, et al. Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3[J]. *Procedia computer science*, 2019, 161: 475–483.
- [17] WENG Juyang, LUCIW M. Dually optimal neuronal layers: lobe component analysis[J]. *IEEE transactions on autonomous mental development*, 2009, 1(1): 68–85.
- [18] WENG Juyang, MCCLELLAND J, PENTLAND A, et al.

- Autonomous mental development by robots and animals[J]. *Science*, 2001, 291(5504): 599–600.
- [19] WENG Juyang, HWANG W S. Incremental hierarchical discriminant regression[J]. *IEEE transactions on neural networks*, 2007, 18(2): 397–415.
- [20] 黄敏, 路飞, 李晓磊, 等. 基于 IHDR 算法和 BP 神经网络复合框架的机器人服务自主认知和发育系统[J]. *机器人*, 2019, 41(5): 609–619.
- HUANG Min, LU Fei, LI Xiaolei, et al. Autonomous cognition and development system of robot service based on a composite framework combining IHDR algorithm with BP neural network[J]. *Robot*, 2019, 41(5): 609–619.
- [21] LU Fei, HUANG Min, LI Xiaolei, et al. Learning and development of home service robots' service cognition based on a learning mechanism[J]. *Applied sciences*, 2020, 10(2): 464.
- [22] 钱夔, 宋爱国, 章华涛, 等. 基于自主发育神经网络的机器人室内场景识别[J]. *机器人*, 2013, 35(6): 703–708, 743.
- QIAN Kui, SONG Aiguo, ZHANG Huatao, et al. Robot indoor scenes recognition based on autonomous developmental neural network[J]. *Robot*, 2013, 35(6): 703–708, 743.
- [23] 王东署, 谭达佩, 韦晓琴. 基于发育网络的人脸朝向识别研究[J]. *郑州大学学报(工学版)*, 2017, 38(5): 23–27.
- WANG Dongshu, TAN Dapei, WEI Xiaoqin. Study of face orientation recognition based on development network[J]. *Journal of Zhengzhou university (engineering science edition)*, 2017, 38(5): 23–27.
- [24] CASTRO-GARCIA J L, WENG Juyang. Emergent multi-lingual language acquisition using developmental networks[C]//2019 International Joint Conference on Neural Networks. Budapest: IEEE, 2019: 1–8.
- [25] ZHENG Zejia, WU Xiang, WENG Juyang. Emergent neural Turing machine and its visual navigation[J]. *Neural networks*, 2019, 110: 116–130.
- [26] 余慧瑾, 方勇纯. 基于改进型自主发育网络的机器人场景识别方法[J]. *自动化学报*, 2021, 47(7): 1530–1538.
- YU Huijin, FANG Yongchun. A robot scene recognition method based on improved autonomous developmental network[J]. *Acta automatica sinica*, 2021, 47(7): 1530–1538.
- [27] 余慧瑾, 方勇纯, 韦知辛. 基于多传感融合的自主发育网络场景识别方法[J]. *机器人*, 2021, 43(6): 706–714.
- YU Huijin, FANG Yongchun, WEI Zhixin. A scene recognition method of autonomous developmental network based on multi-sensor fusion[J]. *Robot*, 2021, 43(6): 706–714.
- [28] 崔丽, 曾凤章. 基于“学习曲线”效应的长期质量成本模型[J]. *商业研究*, 2004(24): 46–48.
- CUI Li, ZENG Fengzhang. A model of long-term quality cost based on learning curve effect[J]. *Commercial research*, 2004(24): 46–48.
- [29] MEMO A, MINTO L, ZANUTTIGH P. Exploiting silhouette descriptors and synthetic data for hand gesture recognition[C]//Smart Tools and Applications in Graphics - Italian Chapter Conference 2015. Verona: The Eurographics Association, 2015.
- [30] MEMO A, ZANUTTIGH P. Head-mounted gesture controlled interface for human-computer interaction[J]. *Multimedia tools and applications*, 2018, 77(1): 27–53.
- [31] BAKHEET S, AL-HAMADI A. Hand gesture recognition using optimized local Gabor features[J]. *Journal of computational and theoretical nanoscience*, 2017, 14(3): 1380–1389.
- [32] BAKHEET S, AL-HAMADI A. Robust hand gesture recognition using multiple shape-oriented visual cues[J]. *EURASIP journal on image and video processing*, 2021, 2021(1): 26.
- [33] TASMER D, AHMED B, DAS S R. Real time hand gesture recognition in depth image using CNN[J]. *International journal of computer applications*, 2021, 174(16): 28–32.
- [34] SAHANA T, BASU S, NASIPURI M, et al. MRCS: multi-radii circular signature based feature descriptor for hand gesture recognition[J]. *Multimedia tools and applications*, 2022, 81(6): 8539–8560.

#### 作者简介:



韦知辛, 硕士研究生, 主要研究方向为机器视觉及发育神经网络。



方勇纯, 教授, 博士生导师, 南开大学人工智能学院院长, 主要研究方向为机器人视觉控制、欠驱动吊运系统控制、仿生机器人运动控制和微纳米操作。主持国家重点研发计划项目、国家自然科学基金重点项目、“十二五”国家技术支撑计划课题、国家自然科学基金仪器专项等国家级项目 10 多项。获吴文俊人工智能自然科学奖一等奖、天津市专利奖金奖、天津市自然科学一等奖、高等教育教学成果一等奖等多项奖励, 发表学术论文 100 余篇。