

文章编号: 0253-2239(2001)01-0049-05

# 用于实时识别三维物体的级联神经网络模型的 优化方法\*

申金媛 刘 珺 陈 成 郭鹏毅 张延烁

(南开大学现代光学研究所, 国家教育部光学信息科学技术开放实验室, 天津 300071)

**摘要:** 利用级联神经网络模型对多个三维目标进行识别, 为提高其正确识别率, 提出多种优化方法, 它们可单独使用, 也可以联合使用。对不变性编码、算法、互连权重的二值化方法、样本优选等进行了研究和探讨。利用优化后的模型对三个飞机模型在视场内的任意位置、任意取向(面内旋转  $360^\circ$ , 面外旋转大于  $45^\circ$ ) 的投影进行识别。计算机模拟表明正确识别率达到 96% 以上。

**关键词:** 神经网络; 级联模型; 模式识别; 互连权重

中图分类号: TP183 文献标识码: A

## 1 引 言

实时三维目标识别在工业、军事及某些危险环境中有着广泛的应用, 在众多的识别方法中, 神经网络由于其并行性、容错性, 是一个用于识别多个三维目标的很好的方法。神经网络作为一种特殊的映射工具, 它可以将不同的模式映射到特征空间中的不同区域, 从而达到分类的目的。单层神经网络只能解决线性可分问题, 而不变性识别是一个线性不可分问题, 所以无法用单层网络实现。拓扑理论证明, 若要用神经网络若要实现任意的空间映射(异联想), 网络结构除了输入层和输出层以外, 至少包含一个中间层<sup>[1]</sup>。多层前馈神经网络被证明具有任意复杂的模式分类能力和良好的多维函数的映射能力, 原则上可以实现多目标位移、旋转及尺度不变性识别<sup>[2,3]</sup>, 但它存在许多缺点, 如相邻层间的互连权重由误差背传算法(BP)获得, 其学习训练过程繁杂, 收敛速度慢, 特别是学习样本数目较多、网络规模较大、中间层数目较多时, 该算法很难收敛, 且难以避免局域最小问题。此外多层前馈网络的多个隐藏层之间没有明确分工, 每一隐藏层不具有独立而明确的物理意义, 所以隐藏层神经元个数的确定具有一定的盲目性, 往往是根据输出结果不断尝试进行调整, 在实际应用中存在许多困难。

级联神经网络是一种前馈型神经网络<sup>[4,5]</sup>。它与一般多层前馈网络所不同的是, 每一层在结构和功能上被赋予了明确的意义, 且相对独立。这一点具有生物学意义。视觉生理和心理的研究表明: 人类的视觉对外界物体的识别过程是分层次的。一个完整的视觉过程, 包括对物体形状、尺寸、颜色、灰度、方位、空间取向以及运动状态等不同因素的识别, 对于其中的每一个因素而言, 由于其本身的复杂性, 又可分成许多子因素, 视觉对所有这些因素确定显然不可能一次解决, 而是分成不同层次逐步完成。

一个三维物体可以由其不同方向的投影图像得到, 所以对它的识别可以利用其投影图像的识别来实现。在实现实时的三维目标识别中, 真实三维物体可有无数二维投影图, 且这些投影图是通过实时采集任意放置、任意旋转(面内、面外)的真实物体得到的, 因此得到的投影图不仅具有旋转变化, 而且具有平移和尺度变化, 可以说每个三维目标的样本是千变万化的。用原来的级联模型<sup>[5]</sup>对真实目标进行识别, 正确识别率并不高, 这主要是由于算法选择的不够好导致其容错能力、存储容量都不够大, 且学习样本选择得也不够好, 不能包含目标的所有特征, 所以用它来对无穷变化的投影进行识别不能得到理想的识别效果。为此, 我们对级联模型进行了研究探讨, 从不变性编码的选择、算法的选择、互连权重的二值化方法、学习样本的优选及网络结构的选择等方面提出了改进方法。这些方法根据实际情况可以单独使用, 也可以联合使用。

\* 国家自然科学基金(69877005)、国家科委 863 高科技项目及天津市高教局资助课题。

收稿日期: 1999-07-02; 收到修改稿日期: 1999-09-29

## 2 级联神经网络模型及其优化方法

级联神经网络是将几个相互独立的级网络(或子网络)依次连接而成,前一层的输出即是后一层的输入。更重要的是,每一级在结构和功能上被赋予明确的意义,且相对独立,这不但简化了学习程序,提高了学习效率,而且这些不同类型的网络在性能上还可以相互补充。此外,级联神经网络的网络结构是开放的,可以根据需要加入具有特定功能的中间隐藏层,增强系统的综合性能,而不必改变原有的训练结果。因此,级联神经网络继承了多层前馈网络强大的识别性能同时有效地解决了前馈神经网络存在的困难。但若想使一个级联网络具有最佳性能,还需要从以下几方面着手对网络进行选择优化。

### 2.1 网络结构和算法的选择

级联模型网络结构的选择主要根据实际需要决定。本文在计算机模拟时采用的级联神经网络模型由两级四层网络组成,结构如图 1 所示。第一级网络是一个多对一的异联想网络(Hetero-Associative),实现对输入目标的不变性编码;第二级网络为胜者全取网络(WTA)。第一级网络包括输

入层  $L_1$  和输出层  $L_2$ (第一隐藏层)两层。本文采用具有极大存储容量和容错能力的胜者全取网络<sup>[6]</sup>作为第二级网络,其输入层即为  $L_2$ ,此外还包括隐藏层  $L_3$  和输出层  $L_4$ ,它基于相似度原理进行识别,并利用异联想输出直观的结果。

用光电混合系统实现上述的级联模型时,存在很多的硬件误差,主要有光源的空间分布不均匀、时间分布不稳,空间光调制器的对比度不高及衍射误差、对准误差、光电检测器的光敏不均匀等等。在多个三维物体的实时识别中物体的旋转、平移、尺度等变化非常大,所以模型的容错性只能消除很小一部分硬件误差。且由于目前光电器件及技术的限制,想要提高硬件系统的正确识别率,最容易的方法是修正模型使它有更大的存储容量和容错能力。为此改变网络结构如图 2 所示,将两级的第一级多对一异联想网络改变成三层网络,即在  $L_1$  和  $L_2$  间插入隐藏层  $L'_1$ ,训练权重  $W'_{ij}$  以消除硬件误差;第二级胜者全取网络不变。

用光电实现系统进行实时识别时,图 1 所示的模型正确识别率低于 45%,图 2 所示的模型正确识别率高于 85%。

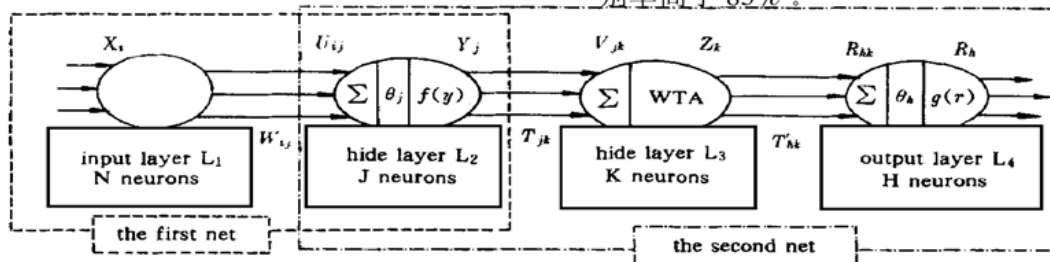


Fig. 1 The constructing scheme of 4-layer cascaded neuron

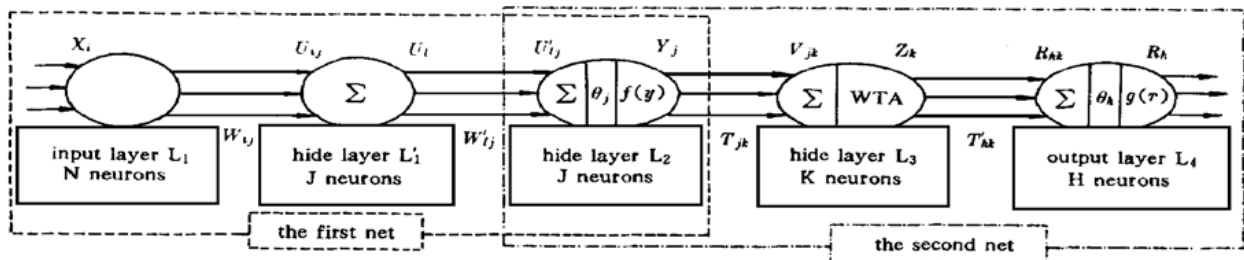


Fig. 2 The constructing scheme of 5-layer cascaded neuron

有很多方法可以对三维多目标进行识别,主要体现在子网络模型的选择上,模型不同得到互连权重不同的算法不同,求第一级网络的互连权重有很多的方法:如基于 Hebb 规则的相关学习算法,瓦块递增算法<sup>[7]</sup>,径向基函数(RBF)网络<sup>[8]</sup>,胜者全取神经网络模型<sup>[6]</sup>等,主要依据算法的复杂程度、学习训练时间的长短、光电实现的容易度等因素进行选择。

另外待识别目标对算法的选择也有较大的影响,有时不同的识别目标对于不同的算法有很大的差别,如非常相近的目标用胜者全取神经网络模型时正确识别率比相差较大的目标要差得多。本文对三个非常相似的飞机模型进行识别时,采用基于 Hebb 规则的相关学习算法得到第一级网络的互连矩阵,对每个目标都事先指定其对应于一组 6 位编码

$C_k (j = 6), k = 1, 2, 3$  对应于 3 个识别目标。显然对于任意一个训练样本, 据其所属目标都对应于某个编码。对于 4 层网络的级联模型, 如果设定  $L_1$  层与  $L_2$  层间神经元之间的互连权重为  $W_{ij}$ , 第  $k$  个目标的第  $m$  个学习样本 (简称为  $km$  个学习样本) 的第  $i$  个神经元的激发状态为  $S_i^{km}$ , 与之对应的编码为  $\{C_{jk}\}$ , 则互连权重值  $W_{ij}$  为:

$$W_{ij} = W_{ij}^0 + \lambda \sum_{k=1}^K \sum_{m=1}^M S_i^{km} C_{jk}, \quad (1)$$

$W_{ij}^0$  为互连权重的初始值, 一般随机选取,  $M$  为每个目标的学习样本数目,  $\lambda$  为学习训练步长。设输入模式为  $X$ , 其神经元数目为  $N$ , 设  $L_1$  层第  $i$  个神经元状态为  $X_i$ ,  $L_2$  层第  $j$  个神经元状态为  $Y_j$ ,  $L_2$  层第  $j$  个神经元的输入为:

$$U_{ij} = W_{ij} X_i = W_{ij}^0 X_i + \lambda \sum_{k=1}^K \sum_{m=1}^M S_i^{km} C_{jk} X_i = W_{ij}^0 X_i + \lambda \sum_{k=1}^K C_{jk} \sum_{m=1}^M S_i^{km} X_i. \quad (2)$$

令  $U_j = \sum_{i=1}^N U_{ij}$ , 根据神经元激发函数, 输出层神经元的实际输出为

$$Y_j = f(U_j - \theta_j) = f\left[\sum_{i=1}^N W_{ij}^0 X_i + \lambda \sum_{k=1}^K C_{jk} \sum_{m=1}^M A_{km} - \theta_j\right], \quad (3)$$

其中  $A_{km} = \sum_{i=1}^N S_i^{km} X_i$  为输入模式与第  $km$  个学习样本间的内积,  $\theta_j$  为该神经元的阈值,  $f(x)$  为硬取阈函数。

为便于光电技术的实现, 将多灰阶的互连权重二值化, 为避免简单的截取所造成的信息丢失而影响系统的识别率, 我们采用了适应截值法进行灰阶压缩<sup>[9]</sup>, 最大限度地减少了由灰阶压缩所引起的正确识别率下降。

第二级胜者全取网络本身又是一个两级三层网络, 第一级得到输入模式与存储模式间的相似度, 第二级是一个异联想网络, 直观地输出识别结果。本文取输入模式与存储模式的内积为相似度, 所以  $T_{jk} = C_{jk}$ ,  $L_3$  层神经元的输入为  $V_{jk} = Y_j T_{jk}$ , 通过自适应取阈得到相应的输出  $Z_k (k = 1, 2, \dots, K)$  为:

$$Z_k = \begin{cases} 1, & V_k = \max(V_p), \\ 0, & V_k \neq \max(V_p), \end{cases} \quad p = 1, 2, \dots, K,$$

其中  $V_k = \sum_{j=1}^J V_{jk}$ ;  $L_4$  层与  $L_3$  层间的互连权重为

$T'_{kh}$ , 由操作者根据需要随意选择, 本文选择的  $T'_{kh}$  使得网络的输出是文字 “The input target is plane  $k$ ”,  $k$  为飞机模型的序号  $1, 2, \dots, K$ , 这样网络若能正确识别, 输入第二个目标的某个投影面时, 输出就为 “The input target is plane 2”。

## 2.2 不变性编码的选择

在网络结构选定以后, 首先要选择不变性编码  $(C_1, C_2, C_3, \dots, C_J)$ , 其中  $C_j = (C_{j1}, C_{j2}, \dots, C_{jK})$ ,  $K$  为待识别的目标数; 原则上最短的码长  $J$  为  $\log_2 K$ , 码的位数越长, 系统的容错性越大, 码的冗余性可提高系统的容错性; 对于一定的算法, 当掉换  $C_j$  和  $C_1$  的位置时, 相应的互连子矩阵掉换, 不改变系统的识别率, 所以最长的码长  $J$  为  $2^K$ ; 当码长选择  $2^K$  时, 其中有两位码对所有的目标都是 1 或 -1, 识别时, 总可以通过改变阈值使得输入所有的目标都输出 1 或 -1, 因而它们对识别结果没有意义, 所以实际可用的最长码为  $2^K - 2$  位。

有两种方法使码长大于最长码  $2^K - 2$ , 一种方法是把待识别目标以外的目标作为一个目标进行编码学习, 这样待识别的目标增加了一个, 而可用的最长码变为  $2^{K+1} - 2$ , 几乎增加一倍, 从而提高系统的容错性。另一种方法是采用多种算法求互连子矩阵, 设希望得到的码长为  $J$ , 当  $J > 2^K - 2$  位时, 第一个  $2^K - 2$  个互连子矩阵利用算法 1 得到, 第二个  $2^K - 2$  个互连子矩阵利用算法 2 得到, 其他的利用算法 3 得到; 以此类推, 可以根据需要任意增加码长。

此外, 为充分利用神经网络的容错性, 根据第二级网络胜者全取网络的特点, 选择的不变性编码应满足不同目标的码两两正交或汉明 (Hamming) 距离最大。

## 2.3 训练样本的选取

学习的目的在于应用和推广。在网络模型及规模选定以后, 训练样本集的选择是提高神经网络推广性的关键环节。在以往的三维目标旋转不变识别中, 不论选取何种模型都以物体的旋转角度作为样本间的距离, 认为两个样本间的取向角度越大, 两个样本间的差别越大, 选择训练样本时以均匀距离进行选择。显然, 这样选择的样本并不一定合理。样本的选择应与所采用的模型结合, 与识别目标时所利用的目标特征结合才可能选出好的训练样本。

由 (3) 式可见, 级联模型是利用样本间的内积相似度作为主要特征进行识别的, 针对此特点, 我们利用样本间的汉明距离 (样本  $X$  与  $S^{km}$  之间的汉明距

离:  $d_H(X, S^{km}) = N - A_{km}$ ,  $A_{km}$  为  $X$  与  $S^{km}$  之间的内积) 选择训练样本集。

显然训练样本选得过密和过疏都会影响识别效果, 好的训练样本应具有相容性、遍历性和致密性及尽可能少的训练样本数。

相容性是指训练样本中不能存在矛盾样本, 即在系统的识别精度范围内无法区分的样本。我们所进行的三维多目标的识别是基于三维物体可以由无数的二维投影图像合成而识别的, 即将一个三维物体的所有的二维投影图像看成是一个目标进行编码学习及识别。不同的三维物体在某一个方向上的二维投影图像可能非常接近甚至完全相同, 当识别这样的二维图像时, 它既可以归为甲类也可以归为乙类, 这样的样本就为矛盾样本, 若将它们作为训练样本, 则训练程序进入死循环或不能使训练样本 100% 地正确识别并导致模型识别混乱, 所以它们不能作为训练样本。

遍历性是指学习样本应分布于整个样本空间。

致密性则是指训练样本之间的间距不能太大, 否则会影响正确识别率。显然训练样本的致密性和尽可能少的训练样本是互相矛盾的, 训练样本之间的间距太小, 则不仅会增长学习训练时间, 而且可能导致更高的互连权重灰度阶, 因而互连权重二值化更会影响整个系统的识别率; 而训练样本之间的间距太大, 则不能包含整个样本空间的目标特征。要完美地满足这两个相互矛盾的条件是很困难的, 只能根据实际情况决定。本文采用的方法如下:

首先在想要识别的空间内以旋转角度为距离, 较密地基本上是等间隔地采集样本, 然后利用汉明距离从中选择训练样本, 首先将同一目标中相互间汉明距离大于某个值  $a$  的样本作为训练样本, 然后去掉其中的矛盾样本, 得到初始的训练样本集。为使训练样本数尽可能少, 选择较大的  $a$  (计算机模拟时  $a = 600$ ), 这样初始的训练样本比较少, 仅用这些样本无法完全表示出模式的特征, 所以需要插入一些新的样本, 使样本集达到一定的致密性。初始训练样本集利用(1)式得互连矩阵, 然后用其他样本进行测试, 记录误判样本(错误识别的样本)得误判样本集, 然后从中选择新的训练样本加入到初始的训练样本集中, 为减少选择次数, 这一次  $a$  值应比较小(模拟时  $a = 250$ )。若使网络得到最佳识别效果, 还应不断地采集新的样本进行测试识别, 找出新的误差样本集和新的训练样本集增加到原有的训练样本集中。如此反复进行, 训练样本集即可扩充到整个样

本空间, 且不会导致过多的训练样本数, 同时使得网络得到最佳的识别效果。

### 3 计算机模拟及结论

利用上述方法我们对 3 个非常相似的飞机模型进行了实时识别, 选 4 层级神经网络模型作为识别模型, 不变性编码为 6 位码, 如表 1 所示, 模拟识别系统如图 3 所示, 利用直径为 1 m 的云台放置飞机模型, 云台可进行面内 360° 旋转, 面外 45° 倾斜, 因此可以利用 CCD 实时采集任意位置、任意旋转和倾斜的飞机投影进行识别。

Table 1 The invariance code of flyer models

	$C_{1k}$	$C_{1k}$	$C_{1k}$	$C_{1k}$	$C_{1k}$	$C_{1k}$
flyer1	1	1	1	-1	-1	-1
flyer2	1	-1	-1	1	-1	1
flyer3	-1	1	-1	1	1	-1

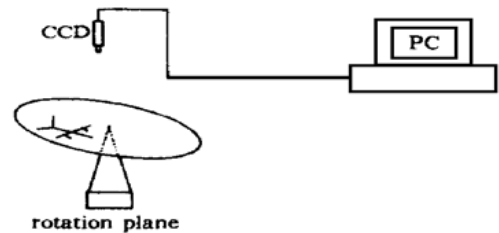


Fig. 3 Real time simulation system

系统在能正确识别之前要进行有导学习训练, 图 4 表示计算机实时学习和识别过程, 虚线表示学习训练过程, 实线表示实时识别过程。

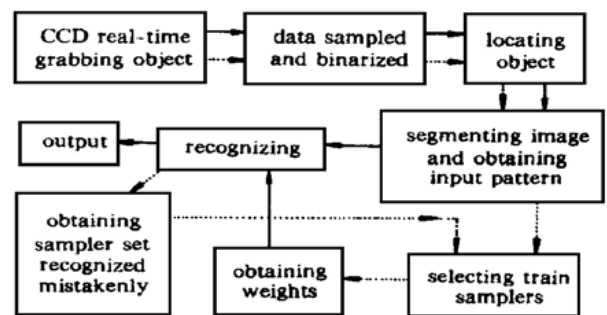


Fig. 4 The diagram of the real-time learning recognizing system of the computer simulation

首先通过 CCD 实时采集目标, 计算机通过图像采集卡获取样本, 对数据进行抽样以减小样本的规模并二值化, 然后利用图像形态学方法对目标进行定位, 分割图像得到输入样本, 开始我们得到 125 个飞机 1、135 个飞机 2 和 126 个飞机 3 的投影样本。根据第二部分叙述的优选样本的方法从中选取 25 + 25 + 30 个样本作为训练样本 ( $a = 600$ ), 再用(1)式求得互连权重  $W_{ij}$ , 对所有 125 + 135 + 126 个样本

进行测试, 并对三个飞机模型各新采集的 100 个样本进行实时测试, 记录误差样本集, 从中再次优选训练样本 ( $a = 250$ ), 最后共得 57+ 60+ 45 个训练样本, 再由(1) 式修改  $W_{ij}$ , 并利用适应截值法对  $W_{ij}$  进行二值化, 然后测试所有已采集的样本, 正确识别率为 100%。

用实时采集的样本进行识别, 检验整个网络的识别性能。对每个飞机在面内面外任意位置任意旋转和倾斜的 200 个投影模式进行实时识别, 正确识别率达 96% 以上。如果对正确识别率不满意, 还可以继续将实时识别时误差的样本记录下来, 累计到一定的数目后, 再次从优选训练样本以训练网络权重, 提高网络的识别性能。

通过对 3 个飞机模型的实时采集识别, 不难发现级联神经网络模型有很多的优点, 可以推断通过对其网络结构、算法、样本等的选择, 它不仅可以用来实时识别真实三维物体, 还可以用于其他的模式识别中, 如手写字的认识、汽车标牌的识别、人面孔识别等。

## 参 考 文 献

- [1] Hornik K, Combe M S, White H. Multilayer feedforward neural networks are universal approximators. *Neural Networks*, 1989, **2**(3): 359~ 366
- [2] 傅京孙 主编. 模式识别及其应用. 北京: 北京大学出版社, 1990
- [3] 焦李成 著. 神经网络系统理论. 西安: 西安电子科技大学出版社, 1990
- [4] Zhang Y X, Huang W Q, Chen T L *et al.* A cascaded model of neural network for pattern recognition. *Opt. Memories. & Neural Networks*, 1992, **1**(3): 199~ 205
- [5] 杨行峻, 郑君里. 人工神经网络. 北京: 高等教育出版社, 1992
- [6] Shen J Y, Zhang Y X, Mu G G. Optical pattern recognition system based on a Winner-Take-All model of a neuron network. *Opt. Engng.*, 1993, **32**(5): 1053 ~ 1058
- [7] 李豫华, 孙 颖, 申金媛 等. 神经网络模式识别系统互连权重二值化的研究. 光学学报, 1996, **16**(10): 1497 ~ 1500
- [8] Poggio T. A network that learns to recognize 3-D object. *Lett. Nat.*, 1990, **343**(18): 2637
- [9] Cheng S *et al.* Adaptive clipped model and its optical implementation. *Optik*, 1998, **107**(4): 135~ 140

## The Option Methods Used in Cascaded Neural Network Model for Recognizing 3-D Targets

Shen Jinyuan Liu Yue Cheng Shu Guo Pengyi Zhang Yanxin

(*Institute of Modern Optics, Nankai University, Optical Information Technical Science Lab, Education Ministry of China, Tianjin 300071*)

(Received 2 July 1999; revised 27 September 1999)

**Abstract:** The cascaded neuron network model is used to recognize 3-D targets. In order to improve the recognizing rate, several methods of choosing the construct and algorithm are proposed. They can be used together or solely according to the requirement. The invariance encoding, the algorithm, the binarizing of the interconnection weights and the selecting methods of the training samples, and so on are studied. The computer simulation of recognizing three plane models is completed based on this cascaded neuron network model. The recognizing rate of the model is over 96% as the three planes arbitrarily positioned and directed.

**Key words:** neural network; invariance encoding; real-recognition, interconnection weights