

文章编号:1003-6180(2021)02-0023-07

## 基于 BP 神经网络结构的图像压缩技术研究

刘光宇<sup>1</sup>, 曹禹<sup>1\*</sup>, 黄懿<sup>1</sup>, 曾志勇<sup>1</sup>, 赵恩铭<sup>2</sup>, 邢传玺<sup>3</sup>

(1. 大理大学 工程学院, 云南 大理 671003; 2. 哈尔滨工程大学 物理与光电工程学院, 黑龙江 哈尔滨 150001; 3. 云南民族大学 电气信息工程学院, 云南 昆明 650504)

**摘要:**采用 BP (back propagation)神经网络方法进行图像数据压缩处理. 通过输入信号的正向传递和误差信号的反向传播算法, 直接为图像提供数据压缩的能力. 仿真实验表明, 通过合理调整 BP 神经网络模型隐含层神经元的个数, 可使图像压缩效果最佳.

**关键词:**BP 神经网络; 数据压缩; 归一化; 重构; 隐含层神经元

DOI:10.13815/j.cnki.jmtc(ns).2021.02.006  
[中图分类号]TP391 [文献标志码]A

## Research on Image Compression Technology Based on BP Neural Network structure

LIU Guangyu<sup>1</sup>, CAO Yu<sup>1\*</sup>, HUANG Yi<sup>1</sup>, ZENG Zhiyong<sup>1</sup>,  
ZHAO Enming<sup>2</sup>, XING Chuanxi<sup>3</sup>

(1. School of Engineering, Dali University, Dali 671003, China; 2. College of Physics and Optoelectronic Engineering, Harbin Engineering University, Harbin 150001, China; 3. School of Electrical and Information Technology, Yunnan Minzu University, Kunming 650504, China)

**Abstract:** BP (Back Propagation) neural network is used to compress image data. Through the forward transmission of the input signal and the back propagation algorithm of the error signal, the ability of data compression is directly provided for the image. The simulation results show that the optimal compression effect can be achieved by reasonably adjusting the number of hidden layer neurons in BP neural network model.

**Key words:** BP neural network; data compression; the normalized; reconsitution; hidden layer neurons

海量图像信息的无失真传输和快速存储的信息存储更是对高效压缩技术提出了迫切的要求.<sup>[1]</sup>为了减轻数字图像在信息存储和传输方面问题,使得多媒体技术面临巨大的压力,大容量

收稿日期:2020-12-14

基金项目:国家自然科学基金项目(61761048);云南省地方本科高校基础研究联合专项资金项目(202001BA070001-060);黑龙江省自然科学基金项目(LC2018026)

作者简介:刘光宇(1982-),男,黑龙江哈尔滨人,助理研究员,博士,硕士生导师,主要从事信号处理研究;曹禹(1994-),男,黑龙江牡丹江人,在读研究生,主要从事图像处理研究;黄懿(1998-),女,河北省邢台人,在读研究生,主要从事图像处理研究.

通信作者:曹禹

的压力,人们开始把人工神经网络的强大功能应用于图像压缩处理领域,并取得了良好的效果. BP (back propagation)神经网络模型鲁棒性高,容错性强,可以通过不断调整网络中的参数来实现数据压缩,达到理想的效果.<sup>[2]</sup>神经网络具有分布式处理、大规模并行、自学习、自组织等优点,是人工智能的主要技术手段.<sup>[3]</sup>在种类繁多的神经网络中,应用最为广泛的就是BP神经网络.<sup>[4]</sup>BP神经网络是典型的前馈型网络,它可以实现多维输入和多维输出的非线性映射.<sup>[5]</sup>BP神经网络有很强的自学习和自适应能力、泛化能力和容错能力<sup>[6]</sup>,其本质是以误差平方函数为目标函数,采用梯度下降法求目标函数的最小值.<sup>[7]</sup>另外,BP神经网络比传统的方法更为有效和方便,在非线形领域具有广阔的应用前景.<sup>[8]</sup>本文采用该方法进行图像数据压缩处理,通过仿真实验验证该方法的可行性和有效性.

## 1 研究原理和方法

### 1.1 BP神经网络的图像压缩原理

BP神经网络对图像压缩的基本思想:让原始数据到达隐含层的网络瓶颈,使其在网络瓶颈处获得较为紧凑的数据表示,从而达到压缩的目的.即当隐含层的神经元个数较少时,意味着隐含层能用更少的数来表现输入模式,而这实际上就是压缩.网络模型和压缩原理流程见图 1. BP网络结构由输入层、隐含层和输出层三部分组成.<sup>[9]</sup>原始图像在输入端经过神经网络的处理后,通过隐含层得到输出数据,并由输出层来输出矢量数据,既为解压后重建的图像数据.一般对图像进行压缩处理时,选择三层的BP神经网络模型应用较多,原因是BP神经网络的整体收敛偏慢,当隐含层变多时,运行速度会变得更慢,其耗

时也会随之增加.

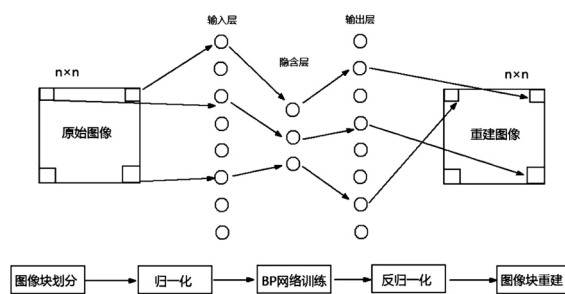


图 1 三层 BP 神经网络模型和压缩原理流程图

### 1.2 神经网络算法

信号在BP神经网络传播的过程中,包括一个正向学习过程和一个误差反向学习过程,这个过程是一个反复进行的网络训练学习过程.正向传播是由输入层到隐层的过程,反向传播是输出层到隐层的误差反传过程,这就是BP神经网络算法的基本原理.<sup>[10]</sup>正向传播时,每层神经元的状态只影响下一层神经元,且神经网络各层的阈值和权值在信号前向传播过程中是固定不会变的.<sup>[11]</sup>若在输出层得不到期望的输出,则转向误差信号的反向传播流程,随着输出端的误差逐渐减少,一直达到可接受的程度或设定的学习次数才会停止.<sup>[12]</sup>

#### • 正向传播算法:

设BP网络的输入层有  $n$  个节点,隐层有  $q$  个节点,输出层有  $m$  个节点,输入层与隐层之间的权值为  $v_{ki}$ ,隐层与输出层之间的权值为  $w_{jk}$ ,如图 2 所示.

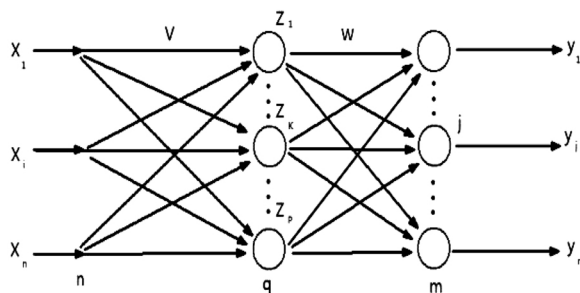


图 2 BP 神经网络拓扑图

图中,  $x_i$  是神经元  $i$  的输入,  $y_j$  为第  $j$  个神经元的输出, 第  $j$  个神经元的净输入值  $S_j$  为:

$$S_j = \sum_{i=0}^n w_{ji} x_i.$$

净输入  $S_j$  通过传递函数 (Transfer Function)  $f(\cdot)$  后, 便得到第  $j$  个神经元的输出  $y_j$ :

$$y_j = f(S_j) = f\left(\sum_{i=0}^n w_{ji} \cdot x_i\right).$$

设隐层的传递函数为  $f_1(\cdot)$ , 输出层的传递函数为  $f_2(\cdot)$ , 则隐层节点的输出为 (将阈值写入求和项中):

$$Z_k = f_1\left(\sum_{i=0}^n v_{ki} x_i\right), \quad k = 1, 2, \dots, q.$$

输出层节点的输出为:

$$y_j = f_2\left(\sum_{k=0}^q w_{jk} z_k\right), \quad j = 1, 2, \dots, m.$$

至此 BP 网络就完成了  $n$  维空间向量对  $m$  维空间的近似映射.

#### • 反向传播算法:

定义误差函数: 输入  $P$  个学习样本, 用  $x^1, x^2, \dots, x^p$  表示. 第  $p$  个样本输入到网络后得到输出  $y_j^p (j = 1, 2, \dots, m)$ . 采用平方型误差函数, 于是得到第  $p$  个样本的误差  $E_p$ :

$$E_p = \frac{1}{2} \sum_{j=1}^m (t_j^p - y_j^p)^2.$$

式中,  $t_j^p$  为期望输出. 对于  $P$  个样本, 全局误差为:

$$E = \frac{1}{2} \sum_{p=1}^P \sum_{j=1}^m (t_j^p - y_j^p)^2 = \sum_{p=1}^P E_p.$$

输出层权值的变化: 采用累计误差 BP 算法调整  $w_{jk}$ , 使全局误差  $E$  变小, 即:

$$\begin{aligned} \Delta w_{jk} &= -\eta \frac{\partial E}{\partial w_{jk}} = -\eta \frac{\partial}{\partial w_{jk}} \left( \sum_{p=1}^P E_p \right) \\ &= \sum_{p=1}^P \left( -\eta \frac{\partial E_p}{\partial w_{jk}} \right). \end{aligned}$$

式中,  $\eta$  为学习率.

定义误差信号为:

$$\delta_{yj} = -\frac{\partial E_p}{\partial S_j} = -\frac{\partial E_p}{\partial y_j} \cdot \frac{\partial y_j}{\partial S_j}.$$

其中, 第一项:

$$\begin{aligned} \frac{\partial E_p}{\partial y_j} &= \frac{\partial}{\partial y_j} \left[ \frac{1}{2} \sum_{j=1}^m (t_j^p - y_j^p)^2 \right] \\ &= -\sum_{j=1}^m (t_j^p - y_j^p). \end{aligned}$$

第二项为输出层传递函数的偏微分:

$$\frac{\partial y_j}{\partial S_j} = f'_2(S_j).$$

于是, 误差信号可变为:

$$\delta_{yj} = \sum_{j=1}^m (t_j^p - y_j^p) f'_2(S_j).$$

由链式定理得:

$$\begin{aligned} \frac{\partial E_p}{\partial w_{jk}} &= \frac{\partial E_p}{\partial S_j} \cdot \frac{\partial S_j}{\partial w_{jk}} = -\delta_{yj} z_k \\ &= -\sum_{j=1}^m (t_j^p - y_j^p) f'_2(S_j) \cdot z_k. \end{aligned}$$

所以, 输出层各神经元的权值调整公式为:

$$\Delta w_{jk} = \sum_{p=1}^P \sum_{j=1}^m \eta (t_j^p - y_j^p) f'_2(S_j) z_k.$$

隐含层权值的变化:

$$\begin{aligned} \Delta v_{jk} &= -\eta \frac{\partial E}{\partial v_{ki}} = -\eta \frac{\partial}{\partial v_{ki}} \left( \sum_{p=1}^P E_p \right) \\ &= \sum_{p=1}^P \left( -\eta \frac{\partial E_p}{\partial v_{ki}} \right). \end{aligned}$$

误差信号变为:

$$\delta_{zk} = -\frac{\partial E_p}{\partial S_k} = -\frac{\partial E_p}{\partial z_k} \cdot \frac{\partial z_k}{\partial S_k}.$$

其中, 第一项:

$$\frac{\partial E_p}{\partial z_k} = \frac{\partial}{\partial z_k} \left[ \frac{1}{2} \sum_{j=1}^m (t_j^p - y_j^p)^2 \right]$$

$$= - \sum_{j=1}^m (t_j^p - y_j^p) \frac{\partial y_j}{\partial z_k}.$$

依链式定理有:

$$\frac{\partial y_j}{\partial z_k} = \frac{\partial y_j}{\partial S_j} \cdot \frac{\partial S_j}{\partial z_k} = f'_2(S_j) w_{jk}.$$

第二项是隐含层传递函数的偏微分:

$$\frac{\partial z_k}{\partial S_k} = f'_1(S_k).$$

于是,误差信号可变为:

$$\delta_{zk} = \sum_{j=1}^m (t_j^p - y_j^p) f'_2(S_j) w_{jk} f'_1(S_k).$$

由链式定理得:

$$\begin{aligned} \frac{\partial E_p}{\partial v_{ki}} &= \frac{\partial E_p}{\partial S_k} \cdot \frac{\partial S_k}{\partial v_{ki}} = -\delta_{zk} x_i \\ &= - \sum_{j=1}^m (t_j^p - y_j^p) f'_2(S_j) w_{jk} f'_1(S_k) x_i. \end{aligned}$$

所以,隐层各神经元的权值调整公式为:

$$\Delta v_{ki} = \sum_{p=1}^P \sum_{j=1}^m \eta (t_j^p - y_j^p) f'_2(S_j) w_{jk} f'_1(S_k) x_i.$$

BP网络实质上是对任意非线性映射关系的一种逼近,由于采用的是全局逼近的方法,因而BP网络具有较好的泛化能力.

### 1.3 BP 神经网络设计

BP神经网络使用 feedforwardnet 函数创建,用LM训练法.因为LM算法的BP神经网络是基于误差不断减小这一原则,采用逐层向前移动的方法,使输出值与期望值之间的误差平方在阈值范围内.<sup>[13]</sup>

网络训练参数的设置:训练目标最小误差为 0.001,最大训练步数为 1 000,学习率设置为 0.05.确定目标误差和最大训练步数以及学习率后,需要新建一个初始化的网络,网络的初始化包括输入层和映射层之间权值的初始化.<sup>[14]</sup>启动一个计时器记录训练所用的时间,调用train函数

进行训练,最后利用sim函数实现神经网络对数据非线性映射后的编码仿真结果.

## 2 仿真实验和结果分析

### 2.1 样本原图的处理

实验原图选用云南省大理市崇圣寺三塔照片,见图3(a).该照片为png格式的灰度图像,尺寸为1080×1440,宽度1080像素,高度1440像素,位深度为8.对该样本原图做以下处理:首先将原图修改为312×312的像素尺寸,然后将此图像分割成若干不重复的4×4像素块,利用函数reshape将像素块重构成16×1的列向量,即将样本原图转化成了16×6084的像素矩阵,对此像素矩阵除以255来进行归一化处理,保证矩阵中的每一个值在[0,1]之间,提高其在BP训练网络中的收敛性.

### 2.2 图像重构和反归一化

将BP神经网络对图像压缩后的输出数据进行重建.由于BP神经网络输入与输出采用的是同一个矩阵,即将之前构建的16×6084像素矩阵作为目标输出矩阵,再将每个列向量恢复成图像块,把所有子图像块合成一幅完整图像,完成图像的重构工作.最后将矩阵元素乘255并取整,进行反归一化操作,目的是将各像素值从[0,1]恢复到[0,255]区间内重建图像.

### 2.3 仿真结果分析

由于隐层神经元数量对BP网络的性能影响很大,若神经元数量过小,则计算精度较低;若神经元数量过大,则会出现过拟合现象.<sup>[15]</sup>因此,对BP神经网络的隐含层神经元个数分别设置为2,4,6,8,16,然后分别进行实验.

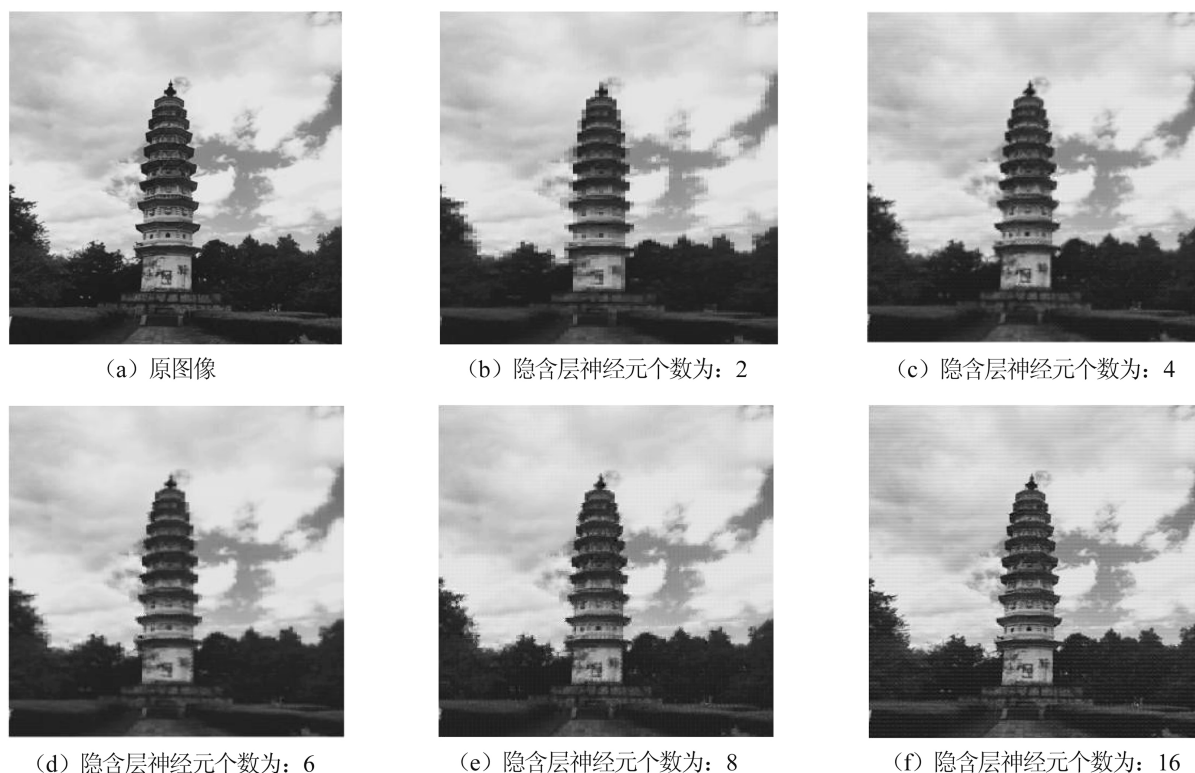


图 3 原图像与不同神经元个数的图像对比

为了更直观的对实验压缩重构后的图像质量进行评价,将原始图像与不同神经元个数的重构图像做对比,如图 3 所示.对比发现:隐含层的不同神经元个数对重构后的图像质量会有很大的影响.当隐含层神经元个数越多时,图像会在某种程度上出现块效应现象,这是由于对图像处理时简单进行分成像素块造成的,但是观察图像压缩重构后的效果可以发现,图像的清晰度有明显的提升,故可以通过增加隐含层的神经元个数提高图像的质量.

采用压缩所用的时间、重构后的图像信噪

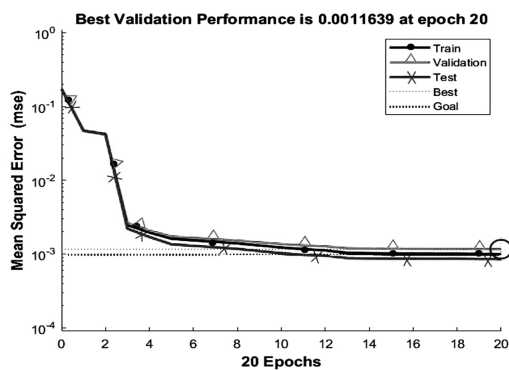
比、峰值信噪比和压缩率等性能指标对实验效果进行评价,实验性能指标结果见表 1.

结果显示,改变 BP 神经网络的隐含层神经元个数会对实验的性能指标产生较大的影响.当对 BP 神经网络的隐含层神经元个数选取为 4 时,其性能指标要好于对其他神经元个数的选取,且隐含层神经元个数为 4 时,压缩所用时间仅在 1 秒左右,拥有的压缩率也显示出可以大幅减少占用存储空间.

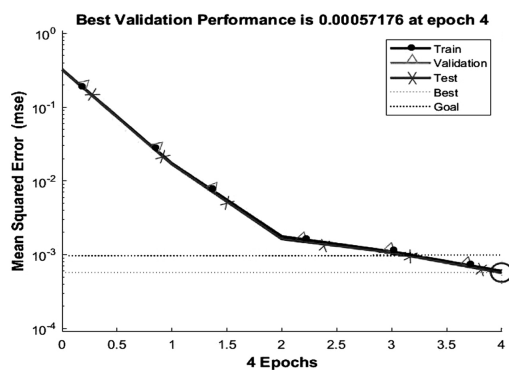
通过用户界面可以综合分析各参数运行的状况.当隐含层神经元个数为 2 时的图像数据压

表 1 仿真实验性能指标结果

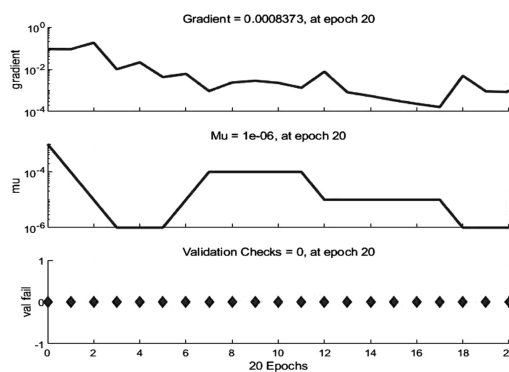
神经元个数	所用时间/秒	信噪比 SNR	峰值信噪比 PSNR	压缩率/%
2	2.076 476	17.928 835	36.976 6	6.38
4	1.000 146	20.165 010	37.517 9	10.29
6	1.593 623	19.119 698	37.239 0	11.72
8	2.309 079	19.378 313	37.217 8	14.04
16	6.259 990	20.223 274	35.468 0	30.14



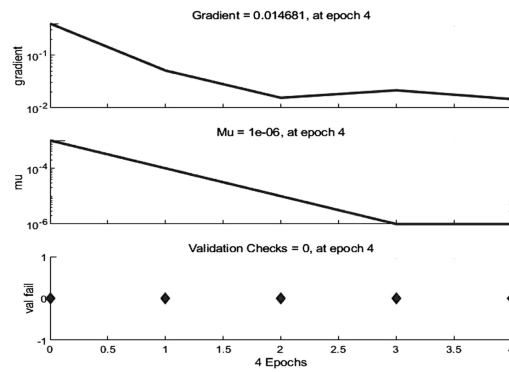
隐含层神经元个数为 2 的误差性能曲线



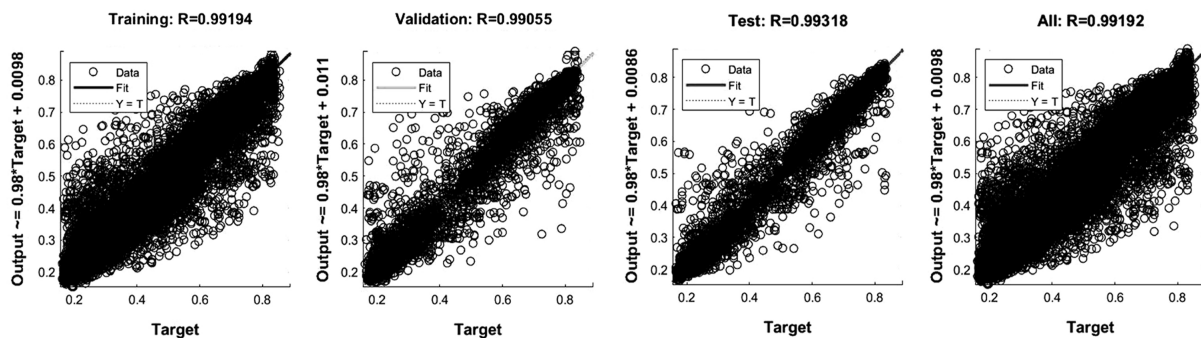
隐含层神经元个数为 16 的误差性能曲线



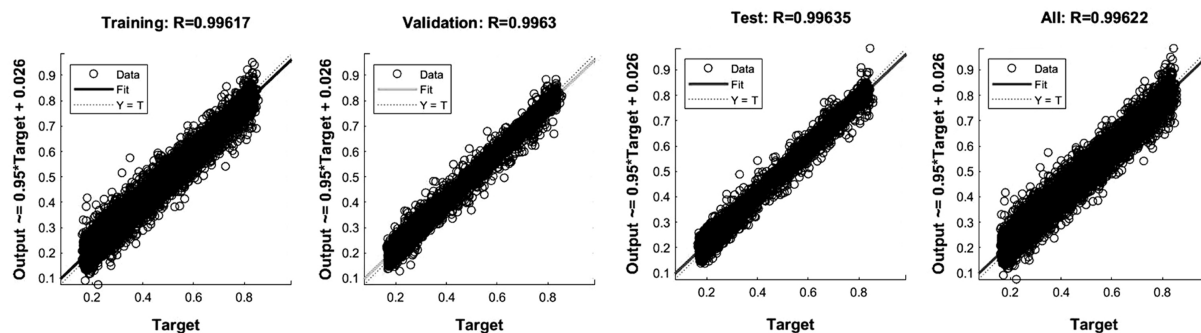
隐含层神经元个数为 2 的训练状态



隐含层神经元个数为 16 的训练状态



隐含层神经元个数为 2 的回归图



隐含层神经元个数为 16 的回归图

图 4 用户界面各参数运行状况对比

缩率最好,但图像质量较差;隐含层神经元个数为 16 时的图像数据压缩率较差,但图像质量最好.因此,可以综合分析隐含层神经元个数为 2 和 16 时的用户界面各参数运行的状况,如图 4 所示.

实验选用的隐含层神经元个数可使训练目标的最小误差在 0.001 内,隐含层神经元个数越多,训练的步数会变得更少,回归效果也会拟合的更好,但相应的数据压缩率会变差以及所花费的时间会更多.

### 3 总结

本文采用 BP (back propagation) 神经网络方法进行图像数据压缩处理.通过输入信号的正向传递和误差信号的反向传播算法,直接为图像提供数据压缩的能力.仿真实验证明,合理调整 BP 神经网络的隐含层神经元的个数,可以实现最大化节约图像传输所用的时间,减少数据压缩后的大小,满足人们对重构后的图像质量的要求.

### 参考文献

- [1]李崇. 基于神经网络的实时性图像压缩算法研究[D]. 北京:北京邮电大学. 2019.
- [2]诸葛木子,李文,张兴会. 基于 BP 神经网络的数字图像压缩方法的研究[A]. 中国高科技产业化研究会智能信息处理产业化分会. 第十届全国信号和智能信息处理与应用学术会议专刊[C]. 中国高科技产业化研究会智能信息处理产业化分会:中国高科技产业化研究会, 2016:5.
- [3]俞立平,阮先鹏,陈一涛,等. 基于因子-BP 人工神经网络的期刊评价方法选择研究[J]. 现代情报, 2020, 40(11):128-135+153.
- [4]张尧,胡骏,宋美荣,等. 神经网络在彩色图像压缩中的应用[J]. 计算机技术与发展, 2012, 22(11):9-12+17.
- [5]李敬德,康维新. 基于信息熵和 BP 神经网络的信号奇异点智能检测[J]. 牡丹江师范学院学报:自然科学版, 2017(4):1-5+43.
- [6]王虹,徐佑宇,谭冲,等. 基于改进粒子群的 BP 神经网络 WSN 数据融合算法[J]. 中国科学院大学学报, 2020, 37(5):673-680.
- [7]潘程艳. 带动置的 BP 神经网络在线梯度学习算法的收敛性[D]. 上海:华东理工大学. 2019.
- [8]Yan-Wen Li, Ke Cao. Establishment and application of intelligent city building information model based on BP neural network model [J]. Computer Communications, 2020, 153:382-389.
- [9]兰欣. 基于 BP 神经网络的乙型病毒性肝炎辅助诊断的应用研究[J]. 中国数字医学, 2016, 11(12):19-21.
- [10]张文胜,郝孜奇,朱冀军,等. 基于改进灰狼算法优化 BP 神经网络的短时交通流预测模型[J]. 交通运输系统工程与信息, 2020, 20(2):196-203.
- [11]常志萍. 基于 BP 神经网络的投标报价博弈研究[D]. 西安:西安理工大学. 2020.
- [12]赵晨. 神经网络预测模型在 A 股市场预测中的应用[J]. 时代经济, 2012, (18):185-185.
- [13]Hongzhi Xue, Hongwei Cui. Research on image restoration algorithms based on BP neural network[J]. J. Vis. Commun. Image R. 2019, 59:204-209.
- [14]别威,田丽. 基于 SOM 神经网络的电能质量综合评估[J]. 牡丹江师范学院学报:自然科学版, 2019(4):34-37.
- [15]陈乐瑞,曹建福,胡河宇,等. 一种基于 Volterra 频域核的非线性频谱智能表征方法[J]. 中南大学学报:自然科学版, 2020, 51(10):2867-2875.

编辑:琳莉