Document Resizing Using a Multi-Layer Neural Network

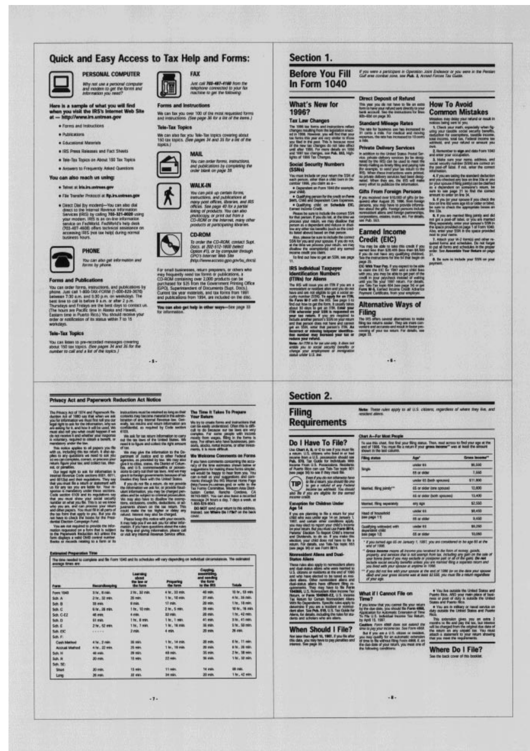
Mohamed N. Ahmed, Brian E. Cooper, and Shaun T. Love
Lexmark International, Inc.
Lexington, KY, USA

Abstract

In this paper we present a resizing neural network for edge and detail preserving image interpolation. The multilayer neural network is trained by using pairs of high resolution and low resolution imagery. The high resolution is an 8-bit image scanned at 600 dpi. The low resolution image (300 dpi) is either a processed version of the high resolution image, or it is scanned independently. pixels are extracted from the low (high) resolution image and are used as inputs to the neural networks. The interpolated pixels obtained as output are compared with the high (low) resolution pixels after enhancement and the error is used to train the neural network.

1. Introduction

Image interpolation is used for several purposes such as picture and document resizing for display and printing, image reconstruction, and geometrical distortions correction. In this paper, we will use image interpolation to resize a digital copy. Digital copies are obtained using a scanning device and then printed. This process involves a variety of inherent factors that compromise image quality. Ordered halftone patterns in the original document interact with the periodic sampling of the scanner, producing objectionable moiré patterns. These are exacerbated when the copy is reprinted with an ordered halftone pattern. In addition, limited scan resolution blurs edges, degrading the appearance of detail such as text. Fine detail also suffers from flare, caused by the reflection and scattering of light from the scanner's illumination source. Flare blends together nearby colors, blurring the high-frequency content of the document. A typical examples of a scanned document is shown in Figure 1.


Quick and Easy Access to Tax Help and Forms:

Section 1.

Before You Fill In Form 1040

Privacy Act and Paperwork Reduction Act Notice

Section 2.

Filing Requirements

Figure 1: *Examples of a scanned document at 300 dpi.*

2. Data Interpolation

The interpolation problem, in its strict sense, may be stated as follows [1]: Given a set of N different points: $\{x_i \in \Re^m \mid i = 1, 2, .., N\}$ and a corresponding set of N real numbers: $\{d_i \in \Re^n \mid i = 1, 2, .., N\}$ find a function F:

$$F : \Re^m \rightarrow \Re^n \mid F(x_i) = d_i,\ i = 1,\ 2,\ ...,\ N\ , \quad (1)$$

where m and n are integers. The interpolation surface is constrained to pass through all the data points. The interpolation function can take the form:

$$F(x) = \sum_{i=1}^{N} w_i\, \phi\,(x,\ x_i)\ , \quad (2)$$

where $\{\phi\ (x,\ x_i) \mid i = 1,\ 2,\ ...,\ N\}$ is a set of N arbitrary functions known as the radial basis functions.

Inserting the interpolation conditions, we obtain the following set of simultaneous linear equations for the unknown coefficients (weights) of the expansion w_i:

$$\begin{pmatrix} \phi_{11} & \phi_{12} & \cdot & \cdot & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdot & \cdot & \phi_{2N} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{N1} & \phi_{N2} & \cdot & \cdot & \phi_{NN} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ w_N \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ \cdot \\ \cdot \\ d_N \end{pmatrix} \tag{3}$$

where

$$\phi_{ji} = \phi(x, x_i), \quad j, i = 1, 2, ..N . \tag{4}$$

Let the $N \times 1$ vectors $\mathbf{d}$ and $\mathbf{W}$ be defined as

$$\mathbf{d} = [\mathbf{d_1}, \mathbf{d_2}, ..., \mathbf{d_N}]^{\mathbf{t}} \tag{5}$$

$$\mathbf{W} = [\mathbf{w_1}, \mathbf{w_2}, ..., \mathbf{w_N}]^{\mathbf{t}} \tag{6}$$

which represent the desired response vector and the linear weight vector, respectively. Let

$$\Phi = \{\phi_{i,j} , i, j \in [1, N]\} \tag{7}$$

denote the $N \times N$ interpolation matrix. Hence, equation (3) can be rewritten as

$$\Phi \mathbf{W} = \mathbf{d} . \tag{8}$$

Provided that the data points are all distinct, the interpolation matrix Φ is positive definite [1]. Therefore, the weight vector W can be obtained by

$$\mathbf{W} = \Phi^{\mathbf{-1}} \mathbf{d} , \tag{9}$$

where Φ^{-1} is the inverse of the interpolation matrix Φ. Theoretically speaking, a solution to the system in (9) always exits. Practically, however, the matrix Φ can be singular. In such cases, regularization theory can be used. Here, the matrix Φ is perturbed to $\Phi + \lambda \mathbf{I}$ to assure positive definiteness [1].

Based on the interpolation matrix Φ different interpolation techniques are available [2]-[7]. Some of these techniques will be reviewed in next section.

3. Classical Interpolation Techniques

3.1. Spline Interpolation

Shepard formulated an explicit function for interpolating scattered data [2]. It is composed of a sum of radial basis functions. The basis functions are radially symmetric about the points at which the interpolating function is evaluated. Conceptually, the method is simple to understand in terms of a thin, deformable plate passing through the data points collected off the surface of the object. The thin plate spline radial basis functions are obtained from the solution of minimizing the energy of the thin plate constrained to pass through loads positioned at the cloud data set. The modeling surface is constructed from the radial basis functions $\beta_i(x, y)$ by expanding them in a series of $(n + 3)$ terms with c_i coefficients:

$$S(x, y) = \sum_{i=1}^{n} c_i \beta_i(x, y), \tag{10}$$

where the basis functions are given by

$$\beta_i(x, y) = r_i^2 ln(r_i). \tag{11}$$

The modeling surface function $S(x, y)$ has the form

$$S(x, y) = a_0 + a_1 x + a_2 y + \sum_{i=1}^{n} c_i r_i^2 ln(r_i). \tag{12}$$

The coefficients are determined by substituting the discrete data set into and solving the resulting set of linear equations:

$$\sum_{i\,=\,1}^{n} c_i = 0, \tag{13}$$

$$\sum_{i\,=\,1}^{n} x_i c_i = 0, \tag{14}$$

$$\sum_{i\,=\,1}^{n} y_i c_i = 0, \tag{15}$$

$$f(x_i, y_i) = a_0 + a_1 x + a_2 y + \sum_{i\,=\,1}^{n} c_i r_i^2 ln(r_i). \tag{16}$$

Bilinear and bicubic interpolation belong to the spline general class of interpolation functions.

3.2. Interpolation artifacts

Since the result of the interpolation process is only an approximation, one should expect artifacts. Those has been classified into four categories called ringing, aliasing, blocking, and blurring.

Ringing arises because most synthesis functions are oscillating. Aliasing is related to the discrete nature of the data. When it is desired to represent a coarser version of the image using fewer samples, the optimal procedure is first to create a piecewise representation of the coarse image using every available sample, and then downsample the coarse representation. Typical visual signatures of aliasing are moiré effects and the loss of texture. Blocking arises when the support of the interpolant is finite. In this case, the influence of any given pixel is limited to its surroundings. Nearest-neighbor methods exacerbate this effect. Blurring arises from the averaging process and the inclusion of edge pixels in the interpolation procedure.

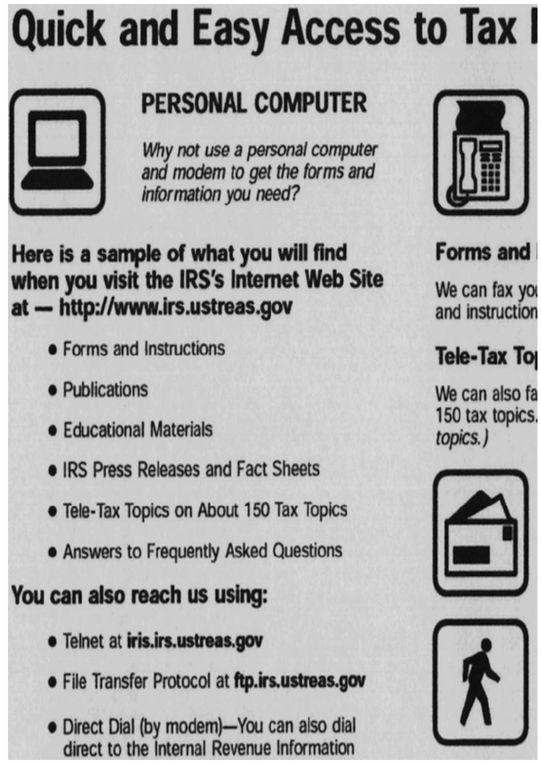
Bicubic interpolation [4] provides a reasonably simple but effective method for enlarging many images. Transitions between expanded pixels remain smooth, and edge content is preserved better than with bilinear interpolation. Nevertheless, detail in document images remains overly blurry, even with bicubic interpolation. As shown in Figure 2, the resulting document image generally suffers from objectionable edge blurring and aliasing.

Rather than trying increasingly complex interpolation algorithms, we propose to use a multi-layer neural network as an image interpolator with enhancement capabilities.

4. Neural Network as Universal Approximator

Although classification is a very important form of neural computation, neural networks can also be used to find an approximation of a multivariable function $F(x)$ [8]. This may be approached through a supervised training of an input-output mapping from a data set. The learning proceeds as a sequence of iterative weight adjustments until a weight vector is found that satisfies certain criteria.

In a more formal approach, multilayer networks can be used to map $\Re^n$ into $\Re$ by using P examples of the function $F(x)$ to be approximated by performing nonlinear mapping with continuous neurons in the first layer then computing the linear combination by the single node of the output layer as follows:



Quick and Easy Access to Tax I

PERSONAL COMPUTER

Why not use a personal computer and modem to get the forms and information you need?

Here is a sample of what you will find when you visit the IRS's Internet Web Site at — http://www.irs.ustreas.gov

- Forms and Instructions
- Publications
- Educational Materials
- IRS Press Releases and Fact Sheets
- Tele-Tax Topics on About 150 Tax Topics
- Answers to Frequently Asked Questions

You can also reach us using:

- Telnet at iris.irs.ustreas.gov
- File Transfer Protocol at ftp.irs.ustreas.gov
- Direct Dial (by modem)—You can also dial direct to the Internal Revenue Information

Forms and

We can fax yo and instruction

Tele-Tax To

We can also fa 150 tax topics. topics.)

Figure 2: Result of applying a smoothing function to remove noise, followed by bicubic interpolation. The resulting document image suffers from edge blurring and aliasing.

$$y = \Gamma[\,\mathbf{VX}\,] \tag{17}$$

$$O = \mathbf{W^t} y \tag{18}$$

where $\mathbf{V}$ and $\mathbf{W}$ are the weight matrices for hidden and output layer respectively, and $\Gamma[\cdot]$ is a diagonal operator matrix consisting of nonlinear squashing functions $\phi(\cdot)$

$$\Gamma = \begin{pmatrix} \phi(\cdot) & 0 & 0 & \cdot & 0 \\ 0 & \phi(\cdot) & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \phi(\cdot) \end{pmatrix} \tag{19}$$

A function $\phi(\cdot) : \Re \rightarrow [0, 1]$ is a squashing function if:

1. It is nondecreasing,
2. $lim_{\lambda \rightarrow \infty}\, \phi(\lambda) = 1$,
3. $lim_{\lambda \rightarrow -\infty}\, \phi(\lambda) = 0$.

Here we have used a bipolar squashing function of the form

$$\phi(x) = \frac{2}{1+e^{-\lambda x}} - 1. \qquad (20)$$

The studies of Funanashi [10], Hornik, and Stinchcombe [9] prove that multilayer feedforward networks perform as a class of universal approximators. Although the concept of nonlinear mapping, followed by linear mapping, pervasively demonstrates the approximating potential of neural networks, the majority of the reported studies have dealt with the second layer also providing the nonlinear mapping [8]-[9]. The general network architecture performing the nested nonlinear scheme consists of a single hidden layer and a single output O such that

$$O = \Gamma(\mathbf{W}\Gamma[\mathbf{VX}]). \qquad (21)$$

This standard class of neural networks architecture can approximate virtually any multivariable function of interest provided that a sufficient number of hidden neurons is available.

4.1. Approximation Using Multilayer Networks

A two-layer network was used for surface approximation. The x and y coordinates of the data points were the input to the network, while the function value $F(x, y)$ was the desired response d.

The learning algorithm applied was error back propagation. This technique calculates an error signal at the output layer and uses this signal to adjust network weights in the direction of the negative gradient descent of the network error E so that, for a network with I neurons in the input layer, J neurons in the hidden layer,and K neurons the output layer, the weight adjustment is as follows:

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}}, \;\; k = 1,2,...,K \;\; j = 1,2,..J \qquad (22)$$

$$\Delta v_{ji} = -\eta \frac{\partial E}{\partial v_{ji}}, \;\; j = 1,2,...,J \;\; i = 1,2,..I \qquad (23)$$

where

$$E = \frac{1}{2} \sum_{k=1}^{K} (d_k - O_k)^2. \qquad (24)$$

The size J of the hidden layer is one of the most important considerations when solving actual problems using multilayer feedforward networks. The problem of the size choice is under intensive study with no conclusive answers available thus far for most tasks. The exact analysis of the issue is rather difficult because of the complexity of the network mapping and due to the nondeterministic nature of many successfully completed training procedures [8]. Here, we tested the network using different numbers of hidden neurons. The degree of accuracy reflected by the mean square error was chosen to be 0.05. Results are provided later in the paper.

5. Results and Conclusions

We tested our algorithm with several images scanned at 300 and 600 dpi. The multilayer neural network is trained by using pairs of high resolution and low resolution imagery. The high resolution is an 8-bit image scanned at 600 dpi. The low resolution image (300 dpi) is either a processed version of the high resolution image, or it is scanned independently. pixels are extracted from the low (high) resolution image and are used as inputs to the neural networks. The interpolated pixels obtained as output are compared with the high (low) resolution pixels after enhancement and the error is used to train the neural network. Results are shown in Figure 3. The neural network was successfully trained to perform the role of image interpolation and enhancement in the same step.

References

[1] S. Haykin. *Neural Networks: a comprehensive foundation*. Macmillan College Co., 1994.

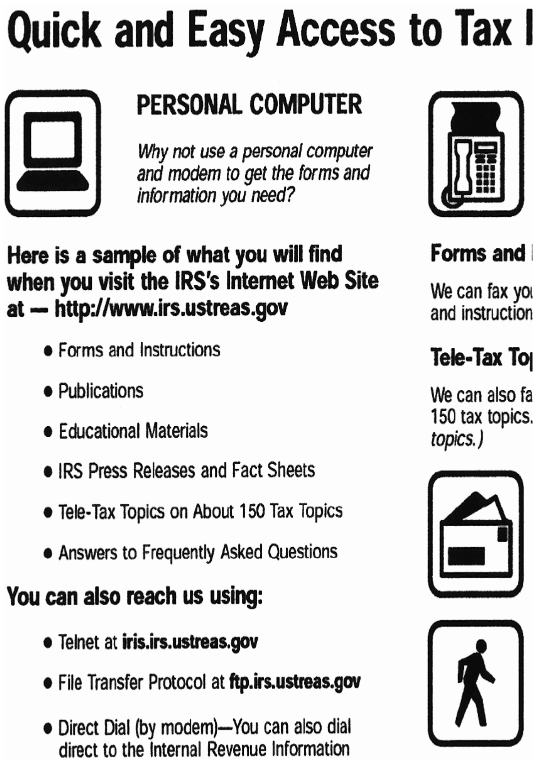
[2] D. Shepard. A two-dimensional interpolation function for irregularly spaced data. *Proc. ACM National. Conf.*, pp. 517–524, 1964.

[3] M. Ahmed, B. E. Cooper, and S. Love, "Document Image Segmentation Using a Two-stage Neural Network," *Proc. of SPIE* , Vol. 3962, pp. 25-33, January 2000.

[4] M. Unser, A. Aldroubi, and M. Eden, "Enlargement or Reduction of Digital Images with Minimum Loss of Information," *IEEE Trans. on Image Processing*, Vol. 4, No. 3, pp. 247-258, March 1995.

[5] C. Bradley and G. Vickers, Free-form Surface Reconstruction for Machine Vision Rapid Prototyping, *Optical Engineering*, Vol. 32, No. 9, pp. 2191-2200, September 1993.

[6] B. Bhanu and C. C. Ho, CAD-based 3-D object representation for robot vision, *IEEE Comput.*, Vol. 20, No. 8, pp. 19-36, Aug. 1987.


Quick and Easy Access to Tax I

PERSONAL COMPUTER

Why not use a personal computer and modem to get the forms and information you need?

Here is a sample of what you will find when you visit the IRS's Internet Web Site at — http://www.irs.ustreas.gov

- Forms and Instructions
- Publications
- Educational Materials
- IRS Press Releases and Fact Sheets
- Tele-Tax Topics on About 150 Tax Topics
- Answers to Frequently Asked Questions

You can also reach us using:

- Telnet at **iris.irs.ustreas.gov**
- File Transfer Protocol at **ftp.irs.ustreas.gov**
- Direct Dial (by modem)—You can also dial direct to the Internal Revenue Information

Forms and

We can fax yo
and instruction

Tele-Tax To

We can also fa
150 tax topics.
topics.)

Figure 3: Enlarged and enhanced document using multi-layer neural network

[7] R. L. Hardy. Multiquadratic equations of topography and other irregular surfaces, interpolation using surface splines. *J. Geophys*, 76, 1971.

[8] J. Zurada. *Introduction to Artificial Neural Systems*. West Publishing, 1992.

[9] K. Hornik and M. Stinchombe. Multilayer feedforward networks as universal approximators. *Neural Networks*, pp. 359–366, 1989.

[10] K. I. Funanashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, pp. 183–192, 1989.

[11] Mohamed N. Ahmed and Aly A. Farag. Free form surface reconstruction using neural networks. *Proc. of ANNIE'94*, Vol. 1, pp.51-56, 1994.

[12] P. H. Pao. *Adaptive Pattern Recognition and Neural Networks*. Addison Wesley Publishing, 1989.

[13] T. Poggio and F.Girosi. Networks for approximation and learning. *Proc. IEEE*, pp. 1481-1497, 1990.