AUTOMATED STAR/GALAXY DISCRIMINATION WITH NEURAL NETWORKS

S.C. Odewahn, E.B. Stockwell, R.L. Pennington, R.M. Humphreys and W.A. Zumach

Department of Astronomy, University of Minnesota, 116 Church Street SE., Minneapolis MN 55455, U.S.A.

1 Introduction

Many of today's most relevant astrophysical problems concerning galactic structure and dynamics, environmental effects on galaxy formation and maintenance, and the large-scale distribution of matter in the Universe are approached in a statistical fashion using deep surveys of stars and galaxies over large areas of the sky. Some of the deepest and most complete galaxy catalogs such as those by Shane & Wirtanen (1967), Zwicky et al. (1961-1968), and Nilson (1973) have been compiled through visual inspection of photographic surveys. More recent efforts by Dickey et al. (1987), Heydon-Dumbleton et al. (1989), Slezak et al. (1988) and Rhee (1990) have used fast scanning machines and automated image detection and classification techniques to compile galaxy catalogs in specific areas of the sky.

A major contribution to both stellar and galaxy surveys will be the catalog of images contained on the 936 plate pairs of the first epoch Palomar Sky Survey currently being generated with the University of Minnesota Automated Plate Scanner (APS). Detailed discussions of this project are given by Humphreys & Pennington (1989). Machine classification is the only viable route in handling the massive data sets encountered in Schmidt plate surveys; but perhaps more importantly, the use of automated classification techniques will produce a more homogeneous sample selection. In this work we shall discuss the development of a novel approach to the problem of star/galaxy separation for the APS automated survey.

2 APS Image Data

We have experimented with a variety of methods for effectively parameterizing the images detected by the APS in threshold densitometry mode. Much of this work fol-

H. T. MacGillivray and E. B. Thomson (eds.), Digitised Optical Sky Surveys 215–224.

lows the methodology of Dickey et al. (1987), Rhee (1990), and Heydon-Dumbleton et al. (1989), but unlike the latter two works we are not using a density to intensity transformation in the image classification stage in order to decrease machine processing time. The image parameters adopted for use in this work are summarized in Table 1. The plate transmission, T, is used directly for the calculation of simple image moments and gradients. Each threshold contour is fit by an ellipse during the scanning process. In a post-processing phase, all pixels with transmission values above the scanning threshold (65% of the sky background) are used to reconstruct the image and compute additional parameters.

Table 1. Image parameters.

Diameter	dia
ellipticity	1 - b/a
average transmission	Tav
central transmission	Tc
(ellipse area) / (area from pixel count)	c1
log (area from pixel count)	c2
$(\Sigma T/r)/(\Sigma T)$	moment 1
rms error of ellipse fit to transit endpoints	fuzz
Y centroid error	jitter
$(T_4 - T_1)/(r_1 - r_4)$	gradient 14
$(T_3 - T_1)/(r_1 - r_3)$	gradient 13
$(T_2 - T_1)/(r_1 - r_2)$	gradient 12
$(T_3 - T_2)/(r_2 - r_3)$	gradient 23
$(T_4 - T_3)/(r_3 - r_4)$	gradient 34

The peak transmission, Tc, and average transmission, Tav, have proven to be very effective discriminators when combined with some measure of image size such as the diameter (see Fig. 1). Other useful parameters for image discrimination are:

1. $c1 = 2\pi a^2/A$ where a is the ellipse semimajor axis length, and A is the image area derived by summing the number of image pixels.
2. $c2 = \log A$.
3. $mom1 = (\sum(T(x,y)/r))/(\sum T(x,y))$ where r is the pixel radius measured from the image center (X_c, Y_c).
4. $Gij = (T_j - T_i)/(r_i - r_j)$, a simple image gradient in which T_i is the median transmission value in an elliptical annulus (having a shape which matches the threshold isophote) and semimajor axis length r_i.

Five distinct image gradients are formed using 4 image radii, to give a total of 14 classification parameters. These quantities comprise the input vector for the neural network classifiers to be discussed in section 3.

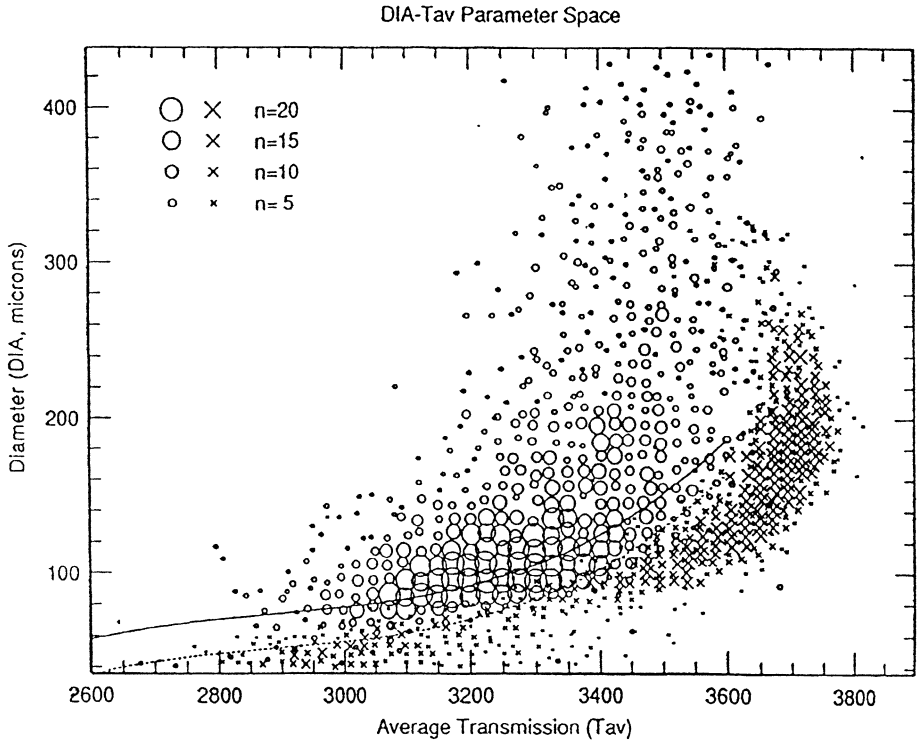

Fig. 1. The distribution of galaxies (open circles) and stars (crosses) in the diameter (DIA) vs. average transmission (Tav) parameter plane. The point size scales linearly with the number of objects occupying that position in parameter space. Notice that galaxies generally possess a larger diameter than stars for a given average transmission value. The solid curve may be used to select galaxies contaminating a star sample. All objects lying above this line have a very high probability of being a nonstellar source. The dashed line is used in a similar fashion to select contaminating stellar objects in a nonstellar sample.

Initial samples of galaxy and star images were collected from two regions of equal area on P323 (the POSS plate containing the Coma cluster of galaxies). As a final check, a smaller area, which extends from the center of the plate to the edge, was also surveyed and used purely as a source of test data for the image classification software. Region 1 was centered on the core of the Coma cluster. Galaxies were collected in this portion by cross-matching APS image coordinates with positions of Coma galaxies in the deep catalog of Godwin et al. (1983), hereafter referred to as GMP83. In this initial study we have restricted our sample to a minimum APS image diameter of D = 40μm (corresponding to B = 21 for a stellar image) and galaxies brighter than $B_{26.5} \leq 20.0$, where $B_{26.5}$ is the isophotal B band magnitude from GMP83 which measures the flux integrated within the μ_B = 26.5 B-mag arcsec^{-2} isophote. A sample of stellar objects was selected in this region by locating objects not matched to the GMP83 catalog and obeying the diameter restriction. This provided an initial list of approximately 2400 galaxies (B $\leq$ 20) and 750 stars

$(11.75 \leq B \leq 21)$.

An area having the same size as Region 1, but located in the opposite plate quadrant, provided the Region 2 sample. Incomplete samples of objects in progressively smaller image diameter intervals (down to $D = 40\mu m$) were classified as galaxies or stars by visual inspection. This produced a list of 290 galaxies ($B \leq 20$) and 1350 stars ($11 \leq B \leq 21$). The last area surveyed, which is used as test data only, is referred to as Region 3. In Region 3, we visually classified all images in a $94' \times 205'$ rectangular box extending roughly from the plate center, with diameters greater than $73\mu m$ (corresponding to approximate B band magnitudes of 19.5 for stars and 20.1 for galaxies). In this case, a total of 4135 detected objects were classified into 5 major categories: galaxies, stars, merged stellar images, plate defects and uncertain types. For this last survey, all classifications were done on the glass copy plate using a 8× loop magnifier for images with diameters larger than $130\mu m$, and a variable magnification binocular microscope for all smaller images. The resultant sample was comprised of 2380 stars and 936 galaxies to be used exclusively for testing the automated image classifiers developed in this work.

3 A Neural Network Approach

Neural networks are a family of artificial intelligence techniques that are capable of performing difficult pattern classification tasks. Their design and development have been inspired by biological neural networks, but the algorithms we have used do not accurately model real biological systems. Many introductory references on neural networks are available: Lippmann 1989 and Rumelhart & McClelland 1988.

We have use the perceptron and backpropagation neural network algorithms to create accurate classifiers for separating star and galaxy images and to inspect their parameter spaces. Backpropagation and perceptron networks are supervised learning techniques. The networks start from a random initial state. A set of training patterns is used to 'teach' the network to perform the desired classification function. The training set must contain a representative sample of patterns for each class. Starting from a random initial configuration, the network is used to classify each pattern in the training set. Each time a pattern in the training set is misclassified, an error term is computed and used to adjust the network's configuration. The patterns in the training set are repeatedly presented to the classifier in this manner until the entire training set is correctly classified or until the network is unable to learn any more patterns.

A perceptron is a simple classifier that forms a hyperplane in parameter space to separate the two classes. It is trained using a simple gradient descent procedure to minimize an error criteria function. The *Perceptron Convergence Theorem* (Duda & Hart 1973) guarantees that training will converge to a solution vector if the classes are linearly separable.

Backpropagation networks are capable of learning much more complex functions and are not restricted to linearly separable classification problems. A backpropagation network consists of one or more layers of nodes ('neurons'). Each node computes

the following function *:

$$output = \frac{1}{1 + \exp -(\vec{i} \cdot \vec{w} + b)} \tag{1}$$

where $\vec{i}$ is the node's input. This is either the network's input vector or the output from the previous layer. The vector $\vec{w}$ and b are the weight vector and bias term and are unique for each node. Initially they contain random values. The weight and bias values are modified by the training process according to the generalized delta rule (Rumelhart & McClelland 1988).

Nodes are arranged in layers. The number of nodes and layers in a network determines the complexity of the function that it computes, as well as the amount of information that it contains. We have experimented with networks consisting of two and three layers of nodes. Figure 2 illustrates a network with an input vector of length five: four nodes in the first layer, two nodes in the second layer, and three in the output layer. As a shorthand, such a network can be written as $\{5 : 4, 2, 3\}$. To use a backpropagation network as a classifier, each class is assigned to an output node. The node with the greatest output value determines the classification of the input pattern. While a simple perceptron classifier is restricted to two classes, a backpropagation network can be used for an arbitrary number of classes.

Several different networks were generated. We have trained a perceptron for a small diameter regime ($73\mu m \leq D \leq 137\mu m$), referred to as SP; and one for a large diameter regime ($146\mu m \leq D \leq 330\mu m$), referred to as LP. Using the same diameter regimes, we have trained $\{14 : 14, 13, 2\}$ backpropagation networks which are referred to as S1 for the small regime, and L1 for the large regime. These are the primary networks discussed in this paper. Additional networks were also trained to investigate specific features. Networks SR (small reduced) and LR (large reduced) were trained with smaller training sets to explore the number of prototypes necessary to produce an accurate classifier. A backpropagation network which was allowed fewer training passes and was used to experiment with using the network's output values as a measure of the confidence of image classification is referred to as SC. The training sets and network configuration for the classifiers are summarized in Table 2.

Considering the network trained with the full set of training data (SP, LP, S1, L1), the networks had greater difficulty learning to separate the star and galaxy classes for small diameter images than for large diameter images. This is indicated by the higher rate of misclassified training patterns for both the small diameter backpropagation and perceptron networks. Also, S1 required many more training passes than L1 before it reached its final state.

Our results suggest that the parameter space for diameters greater than $137\mu m$ separates well and is nearly linearly separable. Smaller images have a more complex parameter space and a linear classifier is not adequate. This also demonstrates the superior learning capacity of the backpropagation network in comparison to the perceptron algorithm. One possible drawback is that if errors exist in the training

* This function is most commonly used. Other functions with a similar shape such as $x/(c + |x|)$ may also be used (Stockwell 1991).

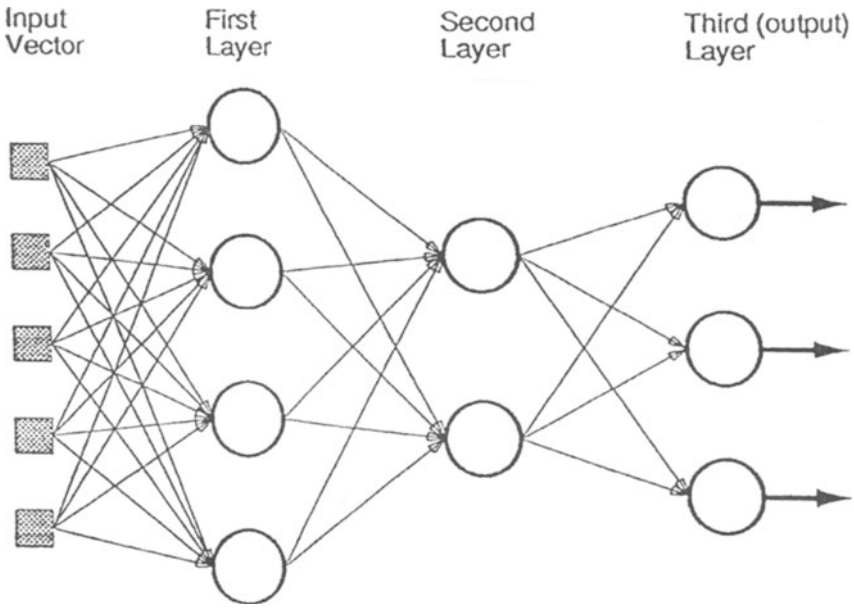

Fig. 2. A schematic illustration of a network with an input vector of length five: four nodes in the first layer, two nodes in the second layer, and three in the output layer. As a shorthand, such a network can be written as: {5:4,2,3}.

Table 2. Neural network training parameters.

Network	Diameter Range (microns)	Patterns in Training Set			Training Passes
		Galaxies	Stars	Total	
L1	$d > 137.3$	1116	1050	2166	925
LP	$d > 137.3$	1116	1050	2166	50000
LR	$d > 137.3$	300	300	600	834
S1	$146.5 > d > 73.2$	1584	719	2303	6540
SP	$146.5 > d > 73.2$	1584	719	2303	50000
SR	$146.5 > d > 73.2$	300	300	600	3873
SC	$146.5 > d > 73.2$	1584	719	2303	218

set, the backpropagation network is more capable of learning these as well. The perceptron, by virtue of its limitation to simple division of parameter space, is required to learn a more general function.

For the classification results listed so far, the image classification for the backpropagation networks is based on the node with the highest output value. Since the network is trained with idealized target outputs of zero and one, a pattern that produces an output between 0.1 and 0.9 must deviate from the patterns that the classifier learned when it was trained. Although many images are correctly classified

with output values that fall in this intermediate range, the frequency of misclassification is higher. Although it cannot be used directly as a probability of class membership, it is clear that the network's output value can provide some measure of classification confidence and a means of producing data sets that are relatively free from contamination.

4 Analysis of the Classifier Success Rate

In order to test the quality of the final perceptron and backpropagation networks, which were trained using galaxy and star images collected in P323 Regions 1 and 2, a third independent set of test images was established. As described previously, over 4500 images with diameters larger than 73μm were classified in order to provide an adequately large test sample. The image parameter sets for these images were measured and normalised in the same manner as the training data. Depending on the diameter of a test image an appropriate network was then used to classify the test image. We judge the success of the automated network classifiers by comparing their results to those classifications established by visual inspection. It should be stated explicitly that all visual classifications were made prior to image analysis by the network.

In Fig. 3 we plot the success rate, S=(number of successes/total number) $\times$ 100, as a function of the O plate B band magnitude. The curves were established by binning the data in equal number diameter bins containing roughly 30 objects each in the case of galaxies, and 80 objects each in the case of stars. The point symbols represent mean points established in binning intervals of 0.5 magnitudes. The success rate of the backpropagation network classifier remains above 90% until B $\approx$ 20.0 for galaxies. The success rate for stars fluctuates in the range between 95% and 100% out to B $\approx$ 19.5. Very similar results were obtained with the perceptron classifiers.

It should be noted that while these results are very satisfactory, implying that our POSS galaxy catalog will be as deep as the often used Lick survey, it is unclear whether the fall-off in galaxy classification success rate at the faint end is due to lack of information in the small image parameter sets, inadequacies of the neural network method when applied to noisy data, or lack of precision by the human classifier in the case of small, faint images. Efforts are underway, using deeper, higher resolution imagery to differentiate these effects. Even if the success rate function implied in Fig. 3 is dictated by the deepness and resolution of the POSS copy plate material, the fundamental advantage of the automated classification approach is evident when we consider that the human classifications were made over several weeks of tedious work, while the plate digitization, image parameterization and network classification were carried out in a matter of hours.

One disadvantage of neural network classifiers is that they give no direct information regarding why an object is assigned a particular classification. In an attempt to determine the significance of the various parameters in our 14 element input vector, we have used an empirical approach in which each component of the input vector is distorted in such a way that its information input is nullified. In

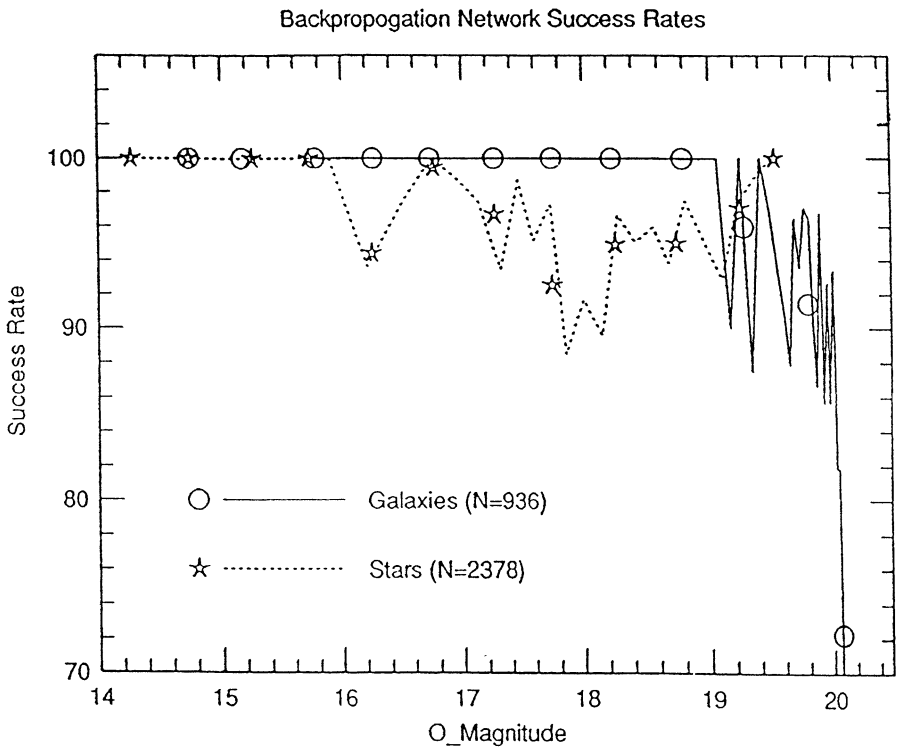

Fig. 3. The classification success rates of the backpropagation network developed in this work. The curves are derived by binning equal numbers of objects into diameter bins. The symbols denote mean success rates in binning intervals of 0.5 magnitudes. Note that both stars and galaxies are classified with success rates above 90% down to the relatively faint O magnitude of 19.5.

our case, the network is run on 14 separate test data sets, each set having a percentage level of random noise added to one of its components. A parameter, X, is adjusted using the relation $X_a = X \pm \beta X$, alternating the sign of the incremental adjustment for every other image. If a parameter carries high weight in the network, then the rate of classification success should be decreased when that information is effectively removed from the calculation. This method is simplistic in its approach in that the network might rely upon interdependencies between several sets of parameters, and removing just one component at a time may not severely cripple the network's ability to make accurate classifications. This is one of the positive features of a neural network which produces a very robust classifier in the presence of observational scatter. This simple empirical method should allow a rough determination of which input parameters are most important to our present networks.

We have run eight sets of calculations, varying the input β value, for each diameter regime. So as not to bias the results of this experiment, we have used approximately equal numbers of stars and galaxies in each diameter regime. Clas-

sification success rates decrease with increasing β value. The four most significant image parameters for each diameter regimes (those parameters which produce the fastest success rate degradation) were found to be:

1. Tav, g34, g14 and g13 for the large diameter set,
2. Tav, Tc, c2 and g14 for the small diameter set.

The results of this simple test are encouraging in that they confirmed our intuitive feeling that image contrast parameters such as the gradient and moment terms would be most useful in the large images. For small images, the calculation of these parameters is complicated by poor resolution, and image classification must be based on simpler global parameters. This exercise also serves to demonstrate the robust nature of the neural network classifiers. A large level of noise ($\beta = 0.5$ and greater) was required to significantly degrade the classifier success rate.

5 Conclusion

A key step in the compilation of an automated survey such as that being conducted with the APS is the discrimination between stellar and nonstellar images. The development of a machine automated technique for performing this task ensures that the tremendous flow of data generated by such a project may be processed in a reasonable period of time. Perhaps more importantly, such a technique produces a very homogeneous catalog whose constituents are selected in an objective and consistent fashion. In this work we have presented a rather novel automated image classification technique employing a neural network which shows great promise. Classifications into stellar and nonstellar categories are based upon a 14-element image parameter set. We have shown that a number of parameter spaces formed with these vector elements are effective in separating a sample of images into the two basic populations of stellar and nonstellar objects. The application of a neural network to this problem allows a large number of image parameters to be used simultaneously in distinguishing a classification.

We have experimented with a simple linear neural network classifier known as a perceptron, as well as with a more sophisticated backpropagation neural network with the result that we are able to attain classification success rates of 99% for galaxy images with $B \leq 18.5$ and above 95% for the magnitude range $18.5 \leq B \leq 19.5$. Based on an analysis of 3601 galaxy images and 4460 stellar images on the POSS field containing the Coma cluster, we have determined the success rate of these classifiers as a function of image diameter and integrated magnitude. Simple numerical experiments illustrate the robust nature of this method and identify the most significant image parameters used by the networks in distinguishing image class. The results from these experiments are extremely promising and indicate that the APS survey of the POSS copy plates will reach a limiting magnitude fainter than that of the Lick integrated counts (Shane & Wirtanen 1967).

References

Dickey, J.M., Keller, D.T., Pennington, R., Salpeter, E.E., 1987. *Astron. J.*, **93**, 788.

Duda, R.O., Hart, P.E., 1973. *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York.

Godwin, J.G., Metcalfe, N., Peach, J.V., 1983. *Mon. Not. R. Astron. Soc.*, **202**, 1123.

Heydon-Dumbleton, N.H., Collins, C.A., MacGillivray, H.T., 1989. *Mon. Not. R. Astron. Soc.*, **283**, 379.

Humphreys, R.M., Pennington, R.L., 1989. *Workshop on Digitized Optical Sky Surveys*, p. 1, eds. C. Jaschek, H.T. MacGillivray.

Lippman, R.P., 1989. *IEEE Communications Magazine*, **27**, 11.

Nilson, P., 1973. *Uppsala General Catalog of Galaxies*, Uppsala Offset Center.

Rhee, G., 1990. *The Structure of Rich Clusters of Galaxies: Clues to Formation and Origin*, PhD thesis, University of Leiden.

Rumelhart, D.E., McClelland, J.L., 1988. *Parallel Distributed Processing*, Vol 1, MIT Press, Cambridge, Mass.

Shane, C.D., Wirtanen, C.A., 1967. *Publications of Lick Observatory*, Technical Report 22, Lick Observatory.

Slezak, E., Bijaoui, S., Mars, G., 1988. *Astron. Astrophys.*, **201**, 9.

Stockwell, E.B., 1991. *IEEE Transactions on Neural Networks*. Submitted.

Zwicky, F., Herzog, E., Kowal, C.T., Wild, P., Karpowicz, M., 1961-1968. *Catalogue of Galaxies and Clusters of Galaxies*, Pasadena: California Institute of Technology.

Discussion

Deul :

How much time does the training of the network take, and is your final network fully connected (no bypasses, no feedbacks)?

Odewahn :

For training our {14:14,13,2} backpropagation networks on a sample of approximately 5,000 objects has required three to four (human) hours on an unloaded Sun4. We have not yet experimented with disconnecting various nodes in the network.

Parker :

Are you going to have to retrain the neural network for each new plate that you wish to scan, as of course plates come in a variety of different qualities, etc.?

Odewahn :

We are investigating the plate to plate variations present in POSS copies in general. We believe that a set of systematic parameter space corrections will be applied before treatment by the neural network classifier. Hence, only one extensively trained network (in theory!) will be required. Certainly this is an area which will require much more work.