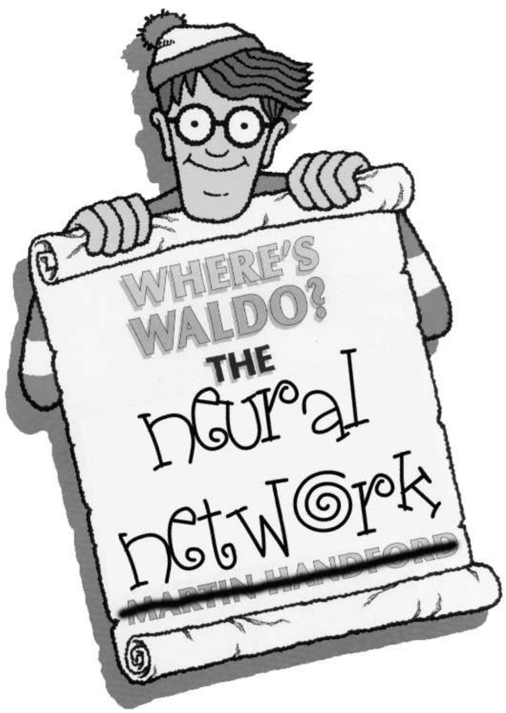
Specific Face Recognition Using a Backpropagation Neural Network

by

Samuel Inverso



Neural Networks
Dr. Gaborski
0603-755-70
February 21, 2000

Abstract

Many papers discuss using neural networks to find generic faces in images. In this paper neural networks were experimented with to find a specific face in an image. A binary and bipolar backpropagation neural network system was used to examine small windows of an image, and decide whether each image contains the specific face. An architecture of the network, explanation of how the training data was gathered, and an analysis of the each network's results are included.

Introduction

In this paper, I report on an experiment to recognize a specific face in a scene using a binary and a bipolar backpropagation neural network. For the specific face I used Waldo from the Where's Waldo books. I chose Waldo because images of him were readily available in many different scenes, and the Where's Waldo series was created to find the specific face of Waldo.

Throughout this report an image refers to a single page from a Waldo book. The naming convention of these images is an 'L' or 'R' for the left or right position of the image, the scene number of the image, and the book abbreviation the scene is in. For example 'R2Now' would refer to the right page, of the second scene, in the Where's Waldo Now book. Another example is 'L12Orig' which refers to the left page, of the twelfth scene, in the original Where's Waldo book. Table 1 shows the abbreviation of the book names matched with the actual book name.

Table 1: Image file abbreviations and corresponding book names

Abbreviation	Book name
Fan	Where's Waldo The Fantastic Journey
Now	Where's Waldo Now
Orig	Where's Waldo
Won	Where's Waldo The Wonder Book

Description of the System

The network was built in MATLAB because it offered the best environment for image processing and neural network creation. Also, through testing I found MATALB's neural network tool kit is significantly faster then my Java coded backpropagation neural network. A benefit of using MATLAB is it has more learning algorithms than I have coded in my Java net to choose from.

I used the resilient backpropagation learning algorithm to train my network because it is very fast, and does not have a big memory overhead. I tried using the quasi-Newton backpropagation and scaled conjugate gradient backpropagation learning algorithms but they required too much memory to train the net. I also tried the Levenberg-Marquardt algorithm but it reached the minimum gradient decent and halted before the net was reasonably trained.

I trained the binary network to a mean squared error (MSE) of 1e-5 because I wanted the net to be accurate but to not over train. After looking at the results for the binary net I decided the network should have trained to a smaller MSE, so I trained the bipolar net to an MSE of 1.65e-08. However, the new MSE did not prove to help much.

Both network architectures consist of 500 inputs, 50 hidden units, and 1 output unit. The 500 input units come from a 20 x 25 pixel region of the image. I refer to this 20 x 25 pixel region as a window because it is a 'window' the net looks through to see a portion of the 850 x 1150 pixel image that is a page from a Waldo book. The size of the window was chosen because the largest Waldo head in the data domain fits inside it. There are 50 hidden units because the network has to be able to recognize Waldo and recognize all the images that are not Waldo. One output unit is used because either Waldo is in the window or not in the window.

The data domain for the net consists of the four Waldo books: Where's Waldo, Where's Waldo Now, Where's Waldo The Fantastic Journey, and Where's Waldo The Wonder Book. I randomly took 5 Waldo images and 20 non-Waldo images from each book for training data. The 20 Waldo images used to train the net are in Appendix A, and random samplings of the non-Waldo images are in Appendix B. In the binary net the 20 total Waldos were trained with a target of 1 and the 200 total non-Waldos were trained with a target of 0. In the bipolar net the Waldos were trained with a target of 1 and the non-Waldos with a target of –1. Figure 1 shows training over time for the binary net and figure 2 shows training over time for the bipolar net.

The binary and bipolar systems operate in three stages: parsing an image into windows, detecting a Waldo in a window, and reconstructing the image with boxes around windows the net thinks are Waldos.

Figure 1: Training of binary network. Mean square error vs. epoch.

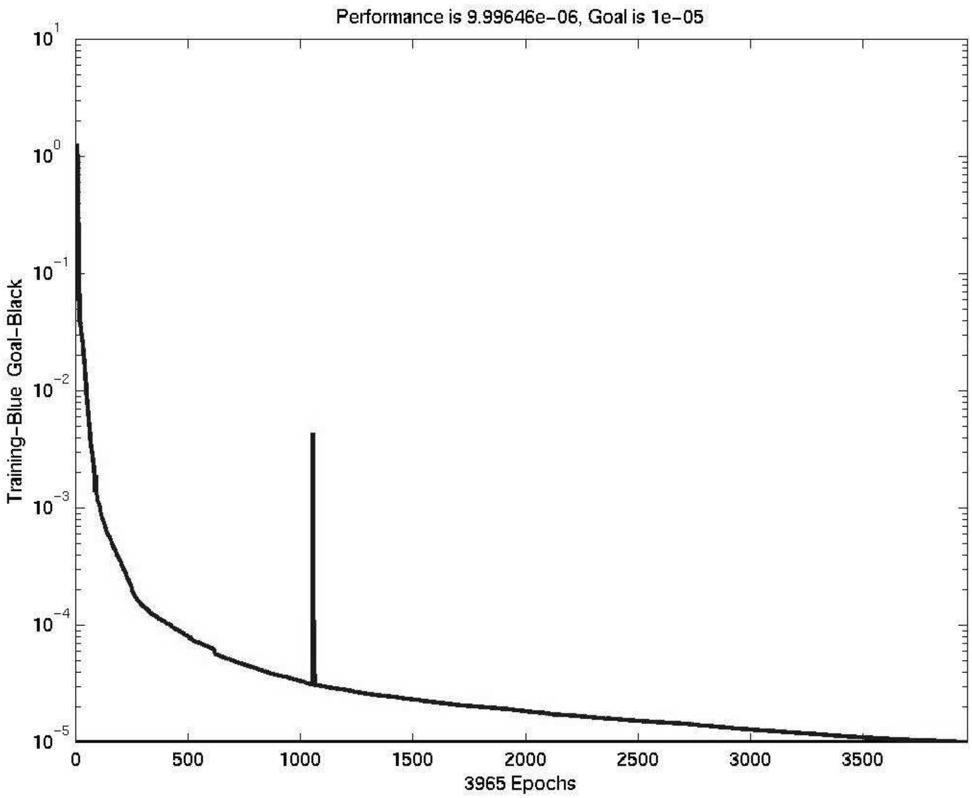
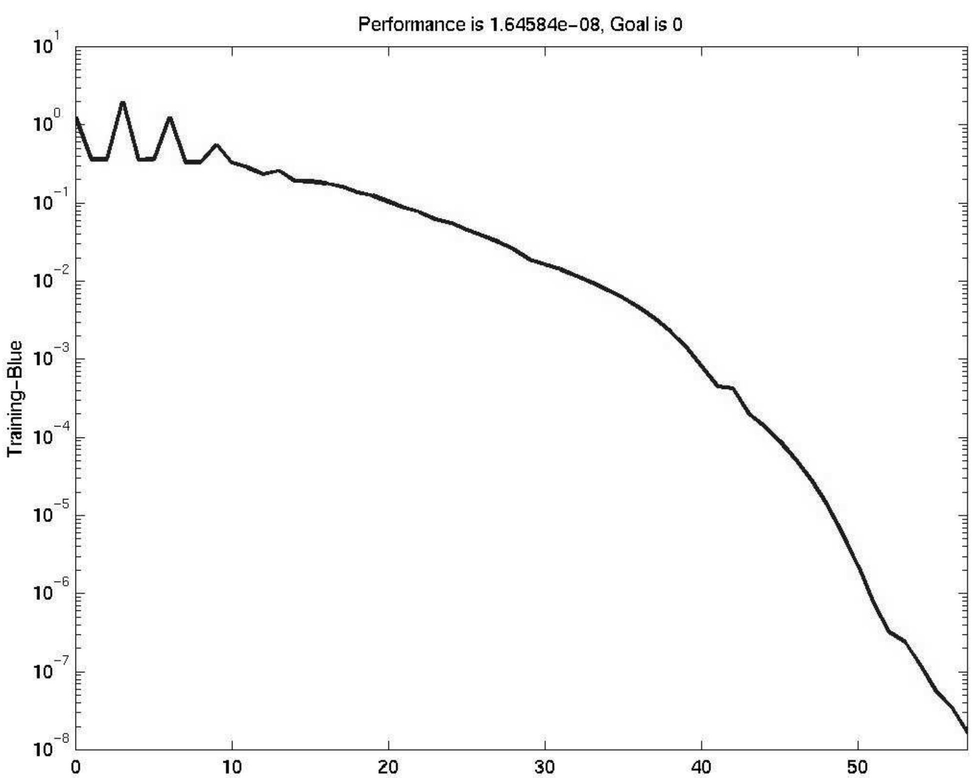


Figure 2: Training of bipolar network. Mean square error vs. epoch.



I decided to move the window through the image at increments of 5 pixels. I chose this incre ent because it is large enough to make certain the images are processed in a reasonable amount f time, and it is small enough to ensure a Waldo will be positioned in a window in such a way that the net has a reasonable chance of r ecogni zing him. The images were cropped to the 2 x 25 pixel window size using

In the detecting stage the window is converted to a 500 x 1 matrix and passed to the MATLAB *sim* function, which simulates the network on the window. After doing some preliminary tests I decided a window had Waldo in it if the network's output for that window is greater than 0.95. I was going to do winner takes all, where the window with the highest network output would be the window the net thinks Waldo is in, but the system would never find Waldo. When a window is found with an output greater than 0.95 system moves to the boxing stage where a box is placed around that 20 x 25 pixel window in the image. If the output is less than or equal to 0.95 the window is ignored.

Results

The binary network was tested on nine images randomly sampled from the four books. This test set consisted of five images the network had trained on and four images it had never seen before. Each image contains 40,670 windows. Out of the nine images the binary network tested it produced an output greater than 0.95 for two if them; these two images were trained on by the network. The network came close to finding Waldo in another two images with outputs of 0.9316 and 0.8675. However the network performed poorly on two images containing Waldo with outputs of 9.905e-4 and –0.0144. These results are summarized in figure 3. Next to each bar in figure 3 is the window image the network saw. Figure 3 also contains the windows with the highest network output for each image file. As can be seen in table 2, the binary network was 22% correct in finding Waldo and its rate of false detection is 1 in 210 images.

Table 2: Results of the binary and bipolar networks on test data.

System	Missed Waldos	Detect Rate	False Detects	False detect rate
Binary Net	7	22.22%	1741	1/210
Bipolar Net	3	25%	246	1/661

The bipolar network was tested on four images randomly sampled from the test set of binary network. The test set consisted of two images the network had trained on and four images it had never seen before. Each image contains 40,670 windows. Of the four images the network tested it produced an output greater than 0.95 for one of them. The other three images had outputs close to –1.

The bipolar network results are summarized in figure 4. Next to each bar in figure four is the window image the network saw. Figure 4 also contains the windows with the highest network output for each image file. As can be seen in Table 2, the bipolar network was 25% correct in finding Waldo and its rate of false detection was 1 in 661.

Analysis of Results

Both networks did poorly finding the Waldos in the images. If the cut off point for a window to have a Waldo in it were dropped from 0.95 to 0.85 the binary network would have a detection rate of 44%. This is still very low to be acceptable and by lowering the cut-off the false detection rate will increase a lot. While the binary system can not seem to find Waldo it does seem to do a good job of figuring out what images are not Waldo. Out of the 366,030 images it looked at the net had 1741 false positives, which is a false detection percentage of 0.4%.

Appendix C and D contain selected images with boxes around the regions the binary network thought was Waldo. As can be seen from the images, the network is making the same kind of mistake over and over again. Taking this into account the network is actually doing a very good job of limiting false detections. If a few of the false detected windows were sent through the network to train against the false detection rate would significantly drop.

The bipolar system also performed poorly in detecting Waldos. Because the network outputs such a low number for the Waldo images it does not match, it seems that the network was trained too long so that it could not generalize to other Waldo images. It is interesting that the bipolar network was trained on a Waldo image from R2Now, but it outputs a significantly low number for the Waldo in that image. Figure 5 shows a side-by-side comparison of the R2Now Waldo the network was trained on and the R2Now Waldo the network tested on.

Figure 5: Comparison of R2Now tested and R2Now trained with the bipolar network.

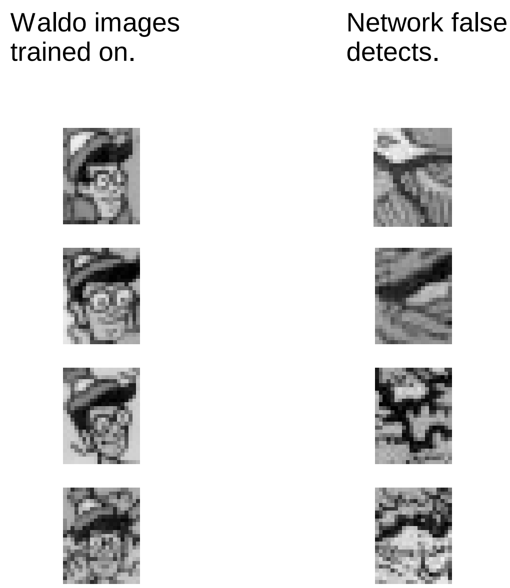
R2Now Trained | R2Now Tested

As can be seen in figure 5, the two images of Waldo are not the same. The trained image left of center in the window with part of Waldo's hat missing, while the tested image is a little right of center with most of Waldo's hat in view. The trained Waldo was tested on the bipolar network and it yielded an output of 0.9994. This comparison supports the theory that the bipolar network was over trained and cannot generalize a Waldo face because even though the trained Waldo does not match the test Waldo other trained Waldo images are similar in location and size.

Like the binary system the bipolar system does a good job of limiting the false detections. Out of the 162,680 images it looked at it falsely detected 246 Waldos, this is a false detection percentage of 0.1%. Appendix E and F contain selected images with boxes around the regions the bipolar network thought was Waldo. Looking at these images we see that there is a reduction in the amount of repeated errors the binary system had a problem with.

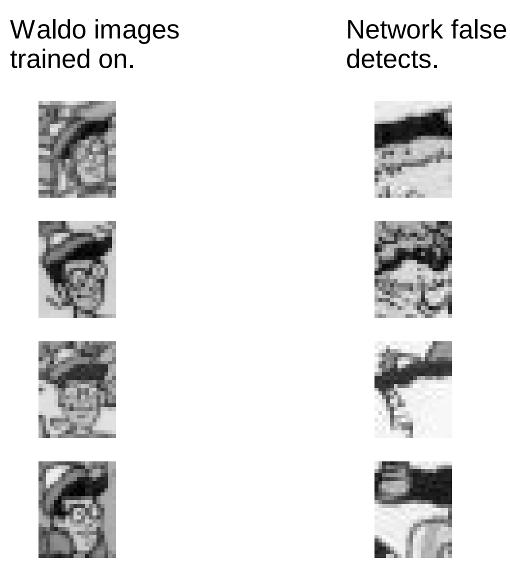
Looking closely at the false detected windows for both the binary and bipolar networks we see that one of the errors the system is making occurs when the window contains a black line around the center of the box. Figure 6 shows a side-by-side comparison of some of the trained Waldo images and some falsely detected windows. The false detect windows contain black lines that the network is mistaking for the hair coming out from under Waldo's hat. This error accounts for a significant portion of the errors the nets are making.

Figure 6: Comparison of Waldo images trained on and network false detects.



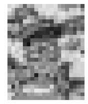
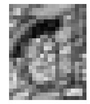
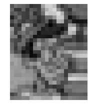
Looking at more false detection windows we see that another error the networks make occur in images where there is a color intensity transition from dark to light. Figure 7 shows some trained Waldo images against false detect windows that have the color intensity transition. We see that the Waldo images have dark to light transitions from Waldo's hair to Waldo's face. In looking at the training data I noticed that there are only a few non-Waldo pictures that have

Figure 7: Comparison of Waldo images and network false detects for color intensity transitions.



this feature, therefore the networks used the color intensity transition as a Waldo classifier given the non-Waldo images did not have this color intensity transition.   Those explanations account for most of the false detections in the system, but they do not explain why the detection rate for Waldo is low. Table 3 shows the Waldo images the networks did not detect.

Table 3: Waldo images the network had trouble detecting did not detect and their outputs.

			
File	R7Fan	R2Now	R12Orig
Binary net	0.0074	9.9052e-4	0.9316
Bipolar net	-1	-0.1	-.9999

We notice that in R7Fan Waldo’s head is very small and you can see much of his chest. There are no Waldos in the training set that have this configuration. In R2Now Waldo’s head was not placed in the center of the window. And in R12Orig the right side of Waldo’s face blends with the background so much that it is hard to tell where his face ends and the background begins.

Conclusion and Future Research

The results of this experiment were good and bad. On the good side both networks did a good job of limiting false detections. On the bad side they both had a poor rate of finding Waldo. One reason why the networks had trouble-finding Waldo is that Waldo is not always drawn the same way. Bye looking at the training data we see Waldo ranges from having a pointed chin to having a wide chin and sometimes he has a long face or a round face. I think if specific face recognition were done with a person the network would have a much easier time with detection because human features do not vary as widely as Waldo faces.

One of the biggest problems with this system was the quality of training data. By doing a random sampling some key non-Waldo images were missed. If I were to do this experiment again I would increase the set size of the non-Waldo images to 1,932. This would be done because an 850 x 1150 pixel image can be broken into 1,932 20 x 25 pixel regions creating a grid over the image. This is a better training set because it covers more of the data domain. I would also increase the number of hidden units to 200 because of the larger training set.

For future research it would be interesting to see how well a neural network can recognize a human face from different angles, especially the same face from the frontal and side profiles. Depending on how well the human face network performs it could be used to authenticate people for security.

Sources

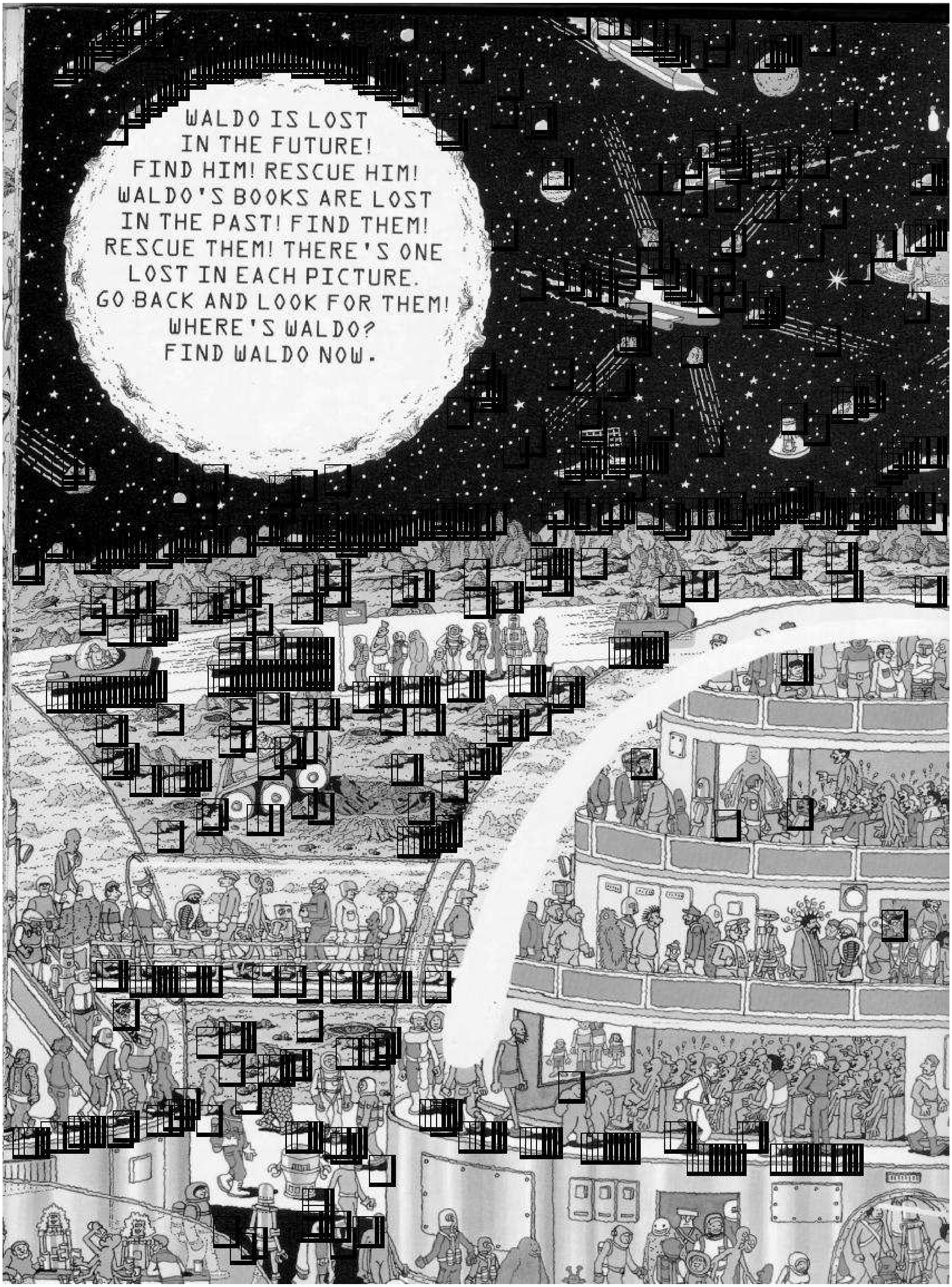
The following books were used as source material for the data domain of the neural networks:

Handford, M. (1997). Where's Waldo. Cambridge, MA: Candlewick Press.
Handford, M. (1988). Where's Waldo Now. Cambridge, MA: Candlewick Press.
Handford, M. (1989). Where's Waldo The Fantastic Journey. Cambridge, MA: Candlewick Press.
Handford, M. (1997). Where's Waldo The Wonder Book. MA: Candlewick Press.

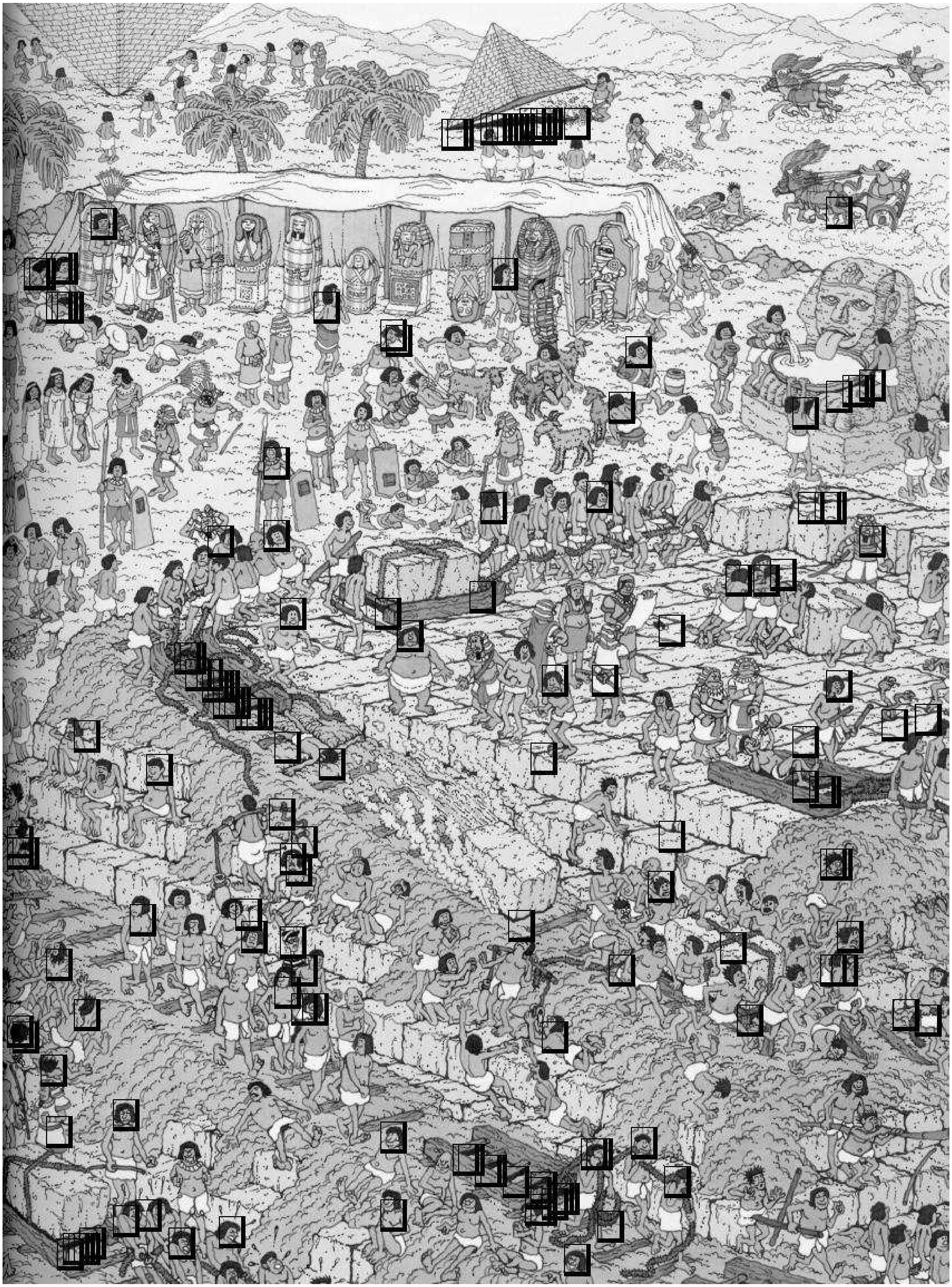
Appendix A: Waldo images trained on.

Appendix B: Select non-Waldo images trained on.

Appendix C: L12Now Binary Boxed



Appendix D: R2Now Binary Boxed



Appendix E: R2Now Bipolar Boxed

Appendix F: R2Orig Bipolar Boxed