A NEURAL NETWORK APPROACH TO A CLASSIC IMAGE RECOGNITION PROBLEM

Wilson Harron, Calvin Pettinger and Robert Dony

University of Guelph
School of Engineering
Guelph ON

ABSTRACT

Scenes from "Where's Waldo?" are used as data sets to explore a method of pattern recognition using local histograms and neural networks. The method consists of splitting the image into a number of sub images, selecting random training images, training the neural networks and then passing the whole set of puzzles through the image to find the character Waldo. Results are shown which demonstrates that this method is effective for finding the pattern found in Waldo's shirt.

Index Terms— object recognition, pattern recognition, neural networks, Where's Waldo

1. INTRODUCTION

The "Where's Waldo" books are a well known activity from the early 1990s. The books contain a number of scenes, or puzzles. Each of these puzzles contains the main character (Waldo) hidden among a large number of other incidental characters[1]. These characters are both humourous and numerous and their sole purpose is to obscure Waldo from sight. Some of the puzzles contain similar looking characters to Waldo, such as the puzzle 'Waldoworld"[2]. The scenes presented are a relevant data set for exploring pattern recognition techniques, as the puzzles are rich in patterns and colours. The scenes also mirror real world situations where a user would need to find a person in a crowd, such as facial recognition in airport security videos or where an object is hidden among similar looking objects like when drug enforcement officials use aerial surveillance to determine whether illicit drugs are being grown in a crop.

Ennesser and Medioni[3] approached this problem without using a neural network approach. To find Waldo a histogram of the segmented image was compared to the histogram of the known Waldo image. While [3] did not present any statistics on the success rate of the system produced, the results did show that the local histogram method of finding Waldo is successful. This paper presents a similar method to the local histogram analysis. A number of different neural networks are used to find Waldo using the local histogram of sub images in the puzzles as the inputs.

Previous tests of this method have shown that a neural network solution is workable. These tests all consisted of using simple neural networks and local histogram analysis to find Waldo among a number of images. This paper presents the latest results from methods that have been improved over these previous tests.

One of the distinctive qualities of 'Waldo' is the colour of his clothing, especially the shirt and hat which are a pattern of alternating red and white stripes. This feature makes colour analysis an excellent area to focus on to find Waldo hidden among the scenes. The problem is well suited for a neural network approach because there is a large input set and the set of puzzles can be used as an effective method of training.

2. METHOD

As mentioned, Waldo's shirt is a very distinctive pattern of alternating red and white stripes. This distinctive colouring pattern was used as the basis for the pattern recognition methods explored. To obtain a representation of the colouring a histogram was taken of all of the colours composed in a number of sub images. These sub images were images that were selected from the main puzzle and were 60×80 pixels in size for the first data set and 30×40 pixels in size for the second data set.

The first task was to go through all of the puzzles and determine the centre point for Waldo's location in the puzzle. This task was done manually and the centre points were chosen as a point centred in Waldo's chest. Typically this was done so that the point corresponded with the second or third stripe in the shirt. However, due to inconsistencies in the drawing of the Waldo images such as orientation and scaling, the points sometimes did not directly correspond to the second or third stripe. Furthermore, if Waldo's shirt was obscured by some object in the image where the centre point would typically be chosen then the centre point was moved slightly so that it would be resting on the Waldo image.

While this method came up with the centre points for Waldo, it was quickly realized that the sub imaging process

978-1-4244-1643-1/08/$25 ©

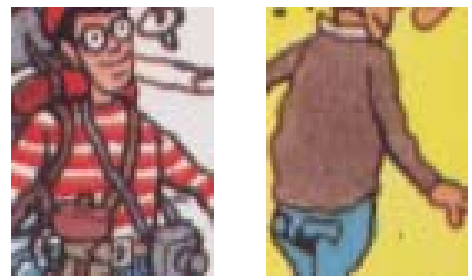
could possibly split Waldo into sections if not done very precisely. Therefore the sub imaging process would split the image into a number of overlapping elements. However, this raised the problem of which images would include Waldo. There was questions as to whether the good images were to be the images containing Waldo or only the image that had a centre point closest to Waldo's centre point. The solution was a compromise. It was determined through inspection that Waldo was approximately 50 pixels across for most of the images, therefore the images that were chosen to be 'Waldo-positive' all had centre points that were within 25 pixels of the determined centre point.

With the Waldo-positive images determined the problem of splitting the images was solved. To ensure that the images would not miss Waldo the sub images were overlapping. While this created a large data set, it was found to be more accurate than a splitting algorithm that did not have the final sub images to be overlapping. To split the images two vectors of X and Y locations were created. Each had a set starting and ending point and were both stepped by a constant. This caused there to be more images in the Y direction than in the X direction. After the two vectors of X and Y centre points were created, the sub images were created by selecting the pixels surrounding (X_i, Y_i). The number of pixels was determined by the desired sub image size. If the sub image's desired boundaries were larger than the picture would allow the boundaries were chosen as the edge of the picture. For the purpose of finding Waldo the sub images were selected at every 10 pixels, with a starting centre point of (10,10). The original image was 1275×1750 pixels, resulting in 22225 sub images for each image. Each image in the set was one half of each puzzle.

To capture Waldo's distinctive colour pattern a histogram was constructed from each sub image. This histogram was then normalized so that each histogram value was a percentage of the total number of pixels in the image. In earlier trials the histograms of each of the red, green and blue values were used as inputs. This lead to some issues where many of the false positives returned by the system had a large amount of brown. To alleviate this problem the histogram was constructed from a combined colour value. The calculation for the colour values is shown in Equation 1. After the colour value was calculated, the histogram was calculated using the algorithms described in [4].

$$C = R + 8 * G + 64 * B \tag{1}$$

Two networks were created to find Waldo. The first network was a feed forward back propagation network (FFBP) with 512 input neurons, two hidden layers with five and three neurons and a single output neuron. This network was chosen from previous tests of the system that showed FFBP networks were able to solve the problem. The training algorithm was the Levenberg-Marquardt algorithm, with a goal for the mean squared error to be zero and a maximum of 20 training epochs. The transfer function for all of the neurons was selected as the sigmoid transfer function. The second network was a perceptron based network. This network was a single perceptron using the perceptron learning algorithm, a goal for the MSE to be zero and a maximum of 50 epochs. Both of the networks were trained using a randomly selected set of Waldo-negative sub images from the puzzles and the six sub images that had the previously determined Waldo centre points as the sub image centre points. An example of both a positive and a negative training image is shown in Figure 2.



(a) Positive (b) Negative

Fig. 1. Sample Training Images

There were two different data sets used to test each network. The first data set consisted of the images from the first six puzzles. This data set contained the training images for both networks and was expected to be more accurate than the second data set. This data set was split into sub image sizes of 60×80 pixels. The second data set contained the images from the last six puzzles in [1].

3. RESULTS

The results for the feed forward back propagation network and the perceptron network for both data sets are found in Tables 1 and 2.

Table 1. Results for FFBP Network

Data Set	Accuracy	TP	TN
1	94.59%	88.70%	94.59%
2	87.30%	54.24%	87.31%

Table 2. Results for Perceptron Network

Data Set	Accuracy	TP	TN
1	76.90%	95.65%	76.88%
2	66.88%	82.20%	66.87%

The true positive (TP) percentage was determined as the percentage of Waldo-positive sub images in the image that were identified as Waldo-positive by the network. The true negative (TN) percentage was determined as the percentage of the Waldo-negative sub images in the image that were identified as Waldo-negative by the network. The accuracy was determined from the number of correct (TP and TN) results in the whole data set.

4. ANALYSIS

As expected the results from the first data set showed that Waldo was found consistently and accurately by the system. The feed forward back propagation network tended to reject more images than the perceptron based network as shown by both the TP percentage and the TN percentage. The lower values of the TP percentage indicates that more Waldo-positive sub images were rejected by the FFBP than the perceptron based network. Similarly, the higher TN percentage shows that more Waldo-negative sub images were rejected by the FFBP network than the perceptron based network. This leads to a higher accuracy rating over the perceptron system. The number of Waldo-positive sub images in an image is approximately 20 and the number of sub images is 22225. Therefore a higher rejection rate for Waldo-negative sub images (or TN perecentage) will influence the accuracy rating much more than the TP percentage.

The second data set showed a similar trend. The perceptron based network found a significantly higher number of Waldo-positive images than the FFBP network, and the performance did not degrade as much when comparing TP percentage. The FFBP network rejected a large number of Waldo-negative images, along the same lines as the first data set. From the results, the TN percentage was 87.31%, a 7% decrease from the first data set. For the perceptron based network the TN percentage decreased over 10% from the first data set. However the number of waldo positive images found by the neural network also decreased a significant amount in the FFBP network. Only 54.24% of the Waldo-positive images were found by the FFBP, while the perceptron based system found 82.20%. Therefore a significant number of Waldo-positive images were not found by the FFBP network. In this aspect, the FFBP network did not perform as well as the perceptron based network in either the first or the second data set.

One result that is important to note is that both networks found at least one Waldo-positive image in the every puzzle. This means that while there is technically a high number of false negatives in the network results, the networks will always find at least a part of the Waldo images. For this application it is debatable as to whether these should be considered as false negative images. If the false negative is defined by this new standard, then neither system has any false negatives and the FFBP is considered more accurate than the perceptron network.

The results from the second data set were less accurate than the results from the first data set. This is expected because the networks were trained with examples from the first data set. However the accuracy seemed to be more related to the image contents than whether or not it was contained in the training set. The amount of false positives in certain puzzles in the first data set was actually higher than the amount of false positives in some of the puzzles in the second set. This is due to the fact that as the books progress, the complexity of the puzzles increases. This was found when the puzzles were being solved manually. The later puzzles in the book contain more of the pattern found of Waldo's shirt. For example, in some puzzles there are towels and shirts that have similar patterns. While these are considered false positives by the analysis, the patterns are matching Waldo's shirt. Examples of patterns that the networks identified as true are shown in Figure 2. This problem exists in the first data set, but it is less prevalent than in the second data set. The puzzle that has the largest number of false positives in the first data set is puzzle 5. This puzzle contains large areas of red colours and large areas of white colours. Due to the colour matching method, the networks would identify the sub images that contain the edges between the red and white would match it to the pattern because the count of red and white pixels would be similar.

(a) Found by Perceptron Network

(b) Found by FFBP Network

Fig. 2. Example of a pattern closely matching Waldo's Shirt

Another factor to consider in the accuracy is the assumption that was made with respect to Waldo. It was assumed earlier on in the investigation that Waldo would be the same size and oriented in the same way throughout the puzzles. This is why the distance of 25 pixels was used as the maximum distance from the center point of the manually found Waldo. This assumption proved to be incorrect as Waldo is significantly smaller in puzzle 10 than he is throughout the other puzzles. Therefore the manual system of measuring whether Waldo is in an image may contain a number of false positives. It is unlikely that the manual method contains any false negatives. This factor could influence the number of false negatives in the system because the networks may rightly identify sub images as not having Waldo, but the manual method marked them incorrectly as being Waldo-positive. Examples of these types of false positives are demonstrated in Figure 3.


(a) Perceptron

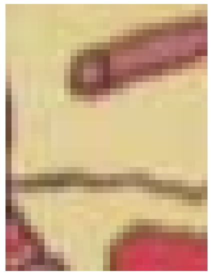

(b) FFBP

Fig. 3. False Negative examples

5. CONCLUDING REMARKS

Using neural networks to find Waldo has been shown to be an effective method, especially a feed forward back propagation network. It has also been shown that for the data set contain in “Where’s Waldo,” the FFBP network is more accurate than the perceptron based network but it finds fewer Waldo-positive images. It has also been shown that the method of colour matching is effective but not a perfect method for matching Waldo due to similarities between the pattern in Waldo’s shirt and other patterns contained throughout the puzzles.

As shown in Figure 2, the neural networks are very effective at finding the pattern of Waldo’s shirt. This method of finding Waldo can be employed alongside another method of identifying Waldo, such as identifying the specific shapes of Waldo. These shapes can be any number of things, from the shape of the glasses to the shape of Waldo’s glasses. This would eliminate the false positive shown in Figure 2(b). Matching the shape of Waldo’s head would eliminate the false positive shown in Figure 2(a).

6. REFERENCES

[1] Martin Hanford, *Where’s Waldo?*, Candlewick Press, 1997.

[2] Martin Hanford, *Where’s Waldo?: The Fantastic Journey*, Candlewick Press, 1997.

[3] F. Ennesser and G. Medioni, “Finding waldo, or focus of attention using color information,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, 1995.

[4] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, Prentice Hall, 2002.