Research Proposal: Where’s Waldo?

Introduction

For my final project, I’m going to build a system that takes in an image and finds Waldo in it. The system will use neural networks for texture classification (red and white stripes for his shirt and hat) and face recognition to distinguish Waldo from other red and white stripy areas of the image. This builds on the work done with texture and face recognition using neural networks, and also on the image processing work discussed in the paper I presented on Traffic Sign Recognition. Ultimately, I hope to have a system that can take in an image and highlight the section containing Waldo.

Problem Statement

For this project, I intend to build a system that can isolate a figure in an image and return the location of the figure. The figure will be specified by its texture and its face, which will allow the system to distinguish between Waldo and some other stripy area. The system will first search the image for sections that match a red and white striped texture, and then search for a face in each of those sections. It will find a confidence value for each section of the image, and overlay the original image with a greyscale mask based on those confidence values, such that the areas that definitely do not contain Waldo are blacked out, and areas that definitely do contain Waldo are not masked at all.

System Requirements and Program Flow

1. The system will acquire an image from the user.
2. The system will segment that image into small (between 10 and 30 pixels) square chunks.
3. The system will analyze the image chunks, converting them into a format that the texture recognition network will accept.
4. The image chunks will be run through a neural network that recognizes textures, in order to isolate those chunks that contain red and white stripy sections. The neural network will output a confidence score, indicating the amount of matching texture in the image chunk.
5. Any image chunks with a texture score greater than some predetermined threshold will be further processed, in order to convert them into a format that the face recognition network will accept.
6. The image chunks will be run through a neural network that recognizes Waldo’s face, in order to find the chunk(s) that contain Waldo. The network will output a confidence score indicating the probability that Waldo is in that image chunk.

7. For each image chunk, the scores from the texture and face networks will be added (if the chunk was not searched for a face, it’s face score is 0) and a mask will be generated based on those scores. The transparency of each mask chunk will be based on its total confidence score; a score of 0 means that the mask will be solid black at that point, 50 means the mask will be 50% opaque, and 100 means that the mask will be transparent at that point.

8. The system will overlay the original image and the mask and display them to the screen.

Risks

The greatest risk in this project is the scope; it requires two fully trained neural networks, which will take time and many human hours to create training and testing data for those networks. The risks are reasonable however, and I feel that they will be minimal if I manage time efficiently.

In addition, the networks are simple compared to typical networks used for texture detection and face recognition, as the texture and the face to recognize are not at all complex. I will need to produce a set of training data made of images of Waldo, which should not be difficult considering the number of images from the Where’s Waldo? books available online.