Using artificial neural networks as an image segmentation module of an OCR- system: a preliminary study

REGINA BERNHAUPT, SANDRA SCHÖNET

Department of Computer Science
University of Salzburg
Jakob Harringer Str.2
A-5020 Salzburg
AUSTRIA
{rber, schoen} @cosy.sbg.ac.at
http://www.cosy.sbg.ac.at/ {~rber, ~schoen}

Abstract: This preliminary study shows some impressive results on the image segmentation ability of a special kind of ANN. This ANN is the small segmentation module of a large (planned) document analysis and optical character recognition system. The segmentation module is built of an ANN using conductivity coded neurons and several complex network structures to implement edge and corner detection. Additionally color information is processed to make the ANN more flexible. We describe several image processing steps within this network. Our first studies can demonstrate how text can be segmented in an easy application - comics. To extract further information for the planned document analysis system, we show additionally the networks ability to learn the "hero" of the comic.

Key-Words: - optical character recognition, segmentation, document analysis, ANN, comic, color processing

1 Introduction

Today most information is either on paper or is present in the form of digitalized images of printed paper. The recognition accuracy of most OCR systems lies between 99.6 % and 99.9 % and this might make one believe that the problem of OCR is solved. But as soon as text not printed against clean backgrounds is used several problems arise. Recognition accuracy of text on structured backgrounds sinks and some non standard fonts are not recognized at all. Automatic reading and processing of documents is a task even more complex than OCR. It requires several processing steps and can reach from the task of document page segmentation to the task of content description in XML [5]. Today several document analysis methods are used to identify text regions. TextFinder [8] is an automatic system used to detect and recognize text in monochrome images. It uses several traditional image processing stages e.g. text segmentation with nonlinear transformation, edge finding algorithms, text cleanup with histogram based algorithms, and is thus rather slow (4.8 minutes / 1.512 pixel times 1.517 pixel, 92 percent of characters with a height greater than 10 pixels). Another major disadvantage of these systems, beside the time consuming computation, is the loss of information. Once the text is segmented, additional information in the images is not used any longer. In the area of image segmentation there are several kinds of methods used to solve the segmentation task. Traditional image processing algorithms are used in combination with artificial neural networks. [2] developed an ANN to segment pages with unstructured layout. This system needs a very large and properly chosen training set and uses modified back propagation algorithms.

In this short research note we describe the first module of a planned OCR system (with document classification and document (text) processing) - the segmentation module. This module is
able to segment unstructured pages using ANN without losing any information. These ANN use color information as well as edge detection mechanisms to segment documents. In our preliminary study we used comics to explore the segmentation abilities of the ANN. The ability for identification of text and non text regions shall be used to improve document classification in the final OCR and document analysis system. In the comic classification example we trained the network to recognize the hero and several other characters within the comic to use this information for classification. Module 2 of the OCR system is a traditional OCR tool which will be combined with a higher level error correction based on a special lexicon [3]. The ANN used in this preliminary "comic" study was first introduced in its original form by H. Geiger. These network models have been very successful used in industrial image quality control [1]. We will describe in detail the functionality of the CCM, the edge extraction and the special network structures to process color information as well as first segmentation results and further improvement possibilities.

2 Conductivity Coupled Neurons

We use the biologically inspired conductivity coupled neuron model (first introduced by H. Geiger). The neural activity is based on the membrane resting potential ($u_0 = -70 mV$) and on the input of excitatory (exc), inhibitory (inh), and normal or shunting (shu) synapses. We assume different ion types for each kind of synapse, thus we may use a typical resistance for each ion type ($r_i$). Equation 1 describes the membrane potential ($mp$) depending on resting potential ($u_0$), conductivity ($e_0$) and all other synaptic potentials ($u_i; i = exc, inh, shu$) with their typical conductivities ($r_i$) as well as the sum of all the input of the three synapse types ($net_i$):

$$mp = \frac{u_i r_i net_i + u_0 e_0}{net_i + e_0}.$$  \hspace{1cm} (1)

The membrane potential of a neuron thus depends nonlinearly on the synaptic input. The activity function depends linearly on the membrane potential. We use a monotonic increasing function between -65 mV (lower bound) representing zero activity and -25 mV (upper bound) representing maximal neural activity. The fixed activity may be viewed as coding a mean spike frequency of a fixed time interval. The direct representation of correlated neurons is lost in this model, since the exact membrane potential is not modeled (thus no firing time can be recorded). A possible solution is the multiplicative computation of correlation. As multiplication of numbers is not biologically plausible a CCM neuron with quadratic increasing activation function is used, as well as a special wiring to implement the correlation of two edge detection neurons. Several image processing steps (edge and corner detection, rotation and translation invariance, color) with a very unusual neural network architecture lead to excellent results in image segmentation and image processing.

3 The ANN’s Image Processing Steps

3.1 Edge Detection

In image processing we use network structures close to the biological equivalent of the vision system of cats [7]. We build a cube of several layers for edge detection (see figure 1). An edge detection neuron is sensitive to a special region on the retina (large receptive field). Neighboring neurons in all of the cube’s layers have neighboring receptive fields which do not overlap.

In the edge detection cube each layer is sensitive for a special edge orientation. Edge detection is realized implementing a kind of filter in the network structure. The neuron of one layer is connected with its receptive field using excitatory and inhibitory connections. If we want to detect a vertical edge, neurons "left" of the vertical line are connected excitatory, neurons "right" of the vertical line are connected inhibitory. The edge we want to detect, runs along the line between left and right part. An edge-detection-neuron will thus get activated if an edge of preferred orientation (a multiple of $\frac{\pi}{n}$ where n is the number of edge detection layers) is present.
To improve the selectivity of our edge detection neurons a special wiring between neurons is used. All neurons of the receptive field lying on a line between the separating line of excitatory and inhibitory neurons have synapses of equal strength. We use a special computation to maximize the influence of these connections to the firing of the edge detector. We further influence the activation function by adding a parameter $\sigma$ representing a gauss function to manipulate the selectivity of the edge detection neurons. Higher gradients in the activation function lead to a more selective edge detection neuron.

3.2 Corner Detection

Results of the edge detection cube are used to detect corners. A corner is present if two edges are activated in the same "time span". To compute the simultaneous activity of two neurons we simply have to multiply the two neuronal activities. Biologically this is not plausible, thus a special neuron type, which has a nearly quadratic course of the activation function, is used. For an overview of the ANN see figure 1.
3.3 Translation- and Rotation Invariance

To achieve translation and rotation invariance, neurons of edge detection and corner correlation cube are summarized and normalized. Invariant object recognition is not possible using the neurons of the cubes, as the neurons are sensitive to special regions. We need to remove the information of edges on certain locations and replace this information by a certain frequency of edges within the retina. Thus a simple summarizing layer enables us to use the location invariant information. To achieve that edges of differing length are normalized too, a norming layer is introduced. Now objects which do only differ in size can be recognized as the same object. The rotation and translation invariant information is processed further by coding the network activity topologically to ease learning.

3.4 Color Processing

In a first step the retinal input is duplicated. Then a color difference image is computed. Based on the color difference image several histograms are computed (red-green difference, blue-green difference, black-white difference, and several other image filters)[6]. The results of these histograms are another part of the information which is used to train the net.

3.5 Learning

Learning takes place between topological coded summary layers and norming layers [edge-processing-cube and corner-processing cube] and the output layer as well as between histogram layers and output layer. We used several types of learning but \( \delta \)-learning seemed to be fastest in the OCR-task. All results here use \( \delta \)-learning.

4 Preliminary Study

4.1 Experiment 1: ”Text/NonText”

In a preliminary study we tested the ability of the ANN to segment the page of a comic (Asterix and Obelix). Taking 2 parts of a comic to train the text and 3 parts of a comic to train the non text regions (training time less than 3 seconds) we could achieve the result shown in figure one (left). The major problem with this test image is the wrong classification of white areas and white areas with black lines. The ANN did only use color information to learn the difference between text and non text, but did not use any information from the edge and corner detection cubes. In a second step we enlarged the training set, adding image areas with text on colored background and image areas with white (colored) text on black background. This leads to better recognition of text areas. We used as a first sample 5 comic pages (selection by chance) recognizing all of the text areas in balloons.
Learning the 7 image sample (training set) takes about 4 sec (Pentium II, 133). The classification of the image depends on the used grid. The grid shown in image 2 is 15 times 15, and the neural network makes a decision (text / non text) for every (non overlapping) region within the grid separately.

To improve the results we used an overlapping grid step, that means we made a decision for image part $I_0$ (upper left corner) moved half of the grid step to the right, made a decision for the half step, move to $I_1$, and so on. The performance time is about 0.1 sec for the non overlapping grid. Experiment 1 showed that the ANN is fairly easy able to segment text and non text regions, if the text is within balloons. If text is within images, we had to use further training examples.

More information forces the network to distribute its knowledge and thus results are more stable. We have to point out, that we could not solve some specialized text areas (see for example figure 4: Idefix is saying GRRR).

### 4.2 Experiment 2: "Asterix and Obelix"

To make further information available to the OCR-modules for document analysis we tried to train the net on hero of the comics to gain further information for document analysis. The idea was to learn, who is going to act on the page of the comic, to give further information to a document analysis tool or database. One might think of the ability to have a database query "Idefix rescues trees" or "Idefix falls in love".

Figure 3 demonstrate the ability of the network to classify a main character (Obelix) in two test images. Figure 3 left, shows the ability of the network to simply give a classification if a hero is somewhere within the whole image.

The network was trained to recognize 4 different characters. The three main characters Asterix, Obelix and Falbala and the dog Idefix. The characters Obelix, Asterix and Falbala work quite well using the one image grid size. Unfortunately the character Idefix seems to be too small to be recognize properly. It can never "win" against any other hero present in the image. Using a more specific grid (figure 3 right) we can show, that the network classifies some regions for Obelix (crossed regions). The lower left regions are classified as Idefix (figure 4).
5 Summary/Future Work

Our preliminary study shows promising results in image segmentation. We could achieve with a very small amount of training examples very good results. An image (shown in figure 2) takes about 0.1 sec to be classified in this resolution. This fast procession time is mainly achieved by the used mathematical model [4].

The segmentation module using the introduced ANN, will be part of a large document analysis and OCR tool. The module thus will provide additional information from the images to other modules to ease the identification of the document or parts of the document. The aim is to build a system able to convert a scanned image into an XML structure.

To improve the system and combine the two experiments the introduction of time coded neurons might help to solve several hypotheses in time, or to represent information at different times (like text at time t_1 and images at time t_2). This improvement includes the idea to solve the segmentation of a scene by backward projection [1].

References


