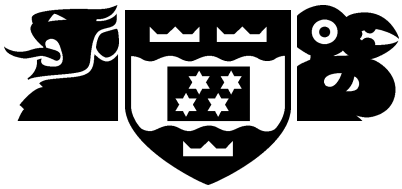
VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui



School of Mathematical and Computing Sciences
Computer Science

False Alarm Filters in Neural Networks for Multiclass Object Detection

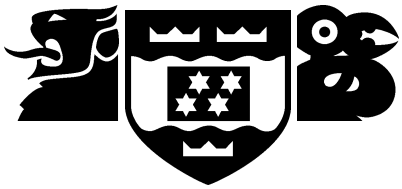
Mengjie Zhang, Bunna Ny

Technical Report CS-TR-04/5
March 2004

School of Mathematical and Computing Sciences
Victoria University
PO Box 600, Wellington
New Zealand

Tel: +64 4 463 5341
Fax: +64 4 463 5045
Email: Tech.Reports@mcs.vuw.ac.nz
http://www.mcs.vuw.ac.nz/research

VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui



School of Mathematical and Computing Sciences
Computer Science

PO Box 600
Wellington
New Zealand

Tel: +64 4 463 5341, Fax: +64 4 463 5045
Email: Tech.Reports@mcs.vuw.ac.nz
http://www.mcs.vuw.ac.nz/research

False Alarm Filters in Neural Networks for Multiclass Object Detection

Mengjie Zhang, Bunna Ny

Technical Report CS-TR-04/5
March 2004

Abstract

This paper describes a neural network approach to multiclass object detection problems in which both the classes and locations of relatively small objects in large images must be determined. Rather than using high level domain specific features or raw image pixels, this approach uses low level pixel statistics as inputs to neural networks. The networks are trained by the back propagation algorithm on examples which have been cut out from the large images. The trained networks are then applied, in a moving window fashion, over the large images to detect the objects of interest. To reduce the false alarm objects detected, a false alarm filter is developed. This approach is examined and compared with a basic neural network approach on three object detection problems of increasing difficulty. The results suggest that the new approach with the false alarm filter can perform very well on those object detection tasks and is more effective than the basic approach.

Keywords Neural networks, pixel statistics, false alarm filters, multiclass object detection, object recognition, computer vision.

Author Information

Mengjie Zhang is an academic staff member in computer science and Bunna Ny was a post-graduate student in computer science, School of Mathematical and Computing Sciences, Victoria University of Wellington, New Zealand.

False Alarm Filters in Neural Networks for Multiclass Object Detection

Mengjie Zhang and Bunna Ny

School of Mathematical and Computing Sciences
Victoria University of Wellington,
P. O. Box 600, Wellington, New Zealand,
Email: mengjie@mcs.vuw.ac.nz

Abstract. This paper describes a neural network approach to multiclass object detection problems in which both the classes and locations of relatively small objects in large images must be determined. Rather than using high level domain specific features or raw image pixels, this approach uses low level pixel statistics as inputs to neural networks. The networks are trained by the back propagation algorithm on examples which have been cut out from the large images. The trained networks are then applied, in a moving window fashion, over the large images to detect the objects of interest. To reduce the false alarm objects detected, a false alarm filter is developed. This approach is examined and compared with a basic neural network approach on three object detection problems of increasing difficulty. The results suggest that the new approach with the false alarm filter can perform very well on those object detection tasks and is more effective than the basic approach.

1 Introduction

Object detection tasks arise in a very wide range of applications, such as detecting faces from video images, finding tumors in a database of x-ray images, and detecting cyclones in a database of satellite images. In many cases, people (possibly highly trained experts) are able to perform the classification task well, but there is either a shortage of such experts, or the cost of people is too high. Given the amount of data that needs to be detected, automated object detection systems are highly desirable. However, creating such automated systems that have sufficient accuracy and reliability turns out to be very difficult.

The traditional approach to object recognition and detection task [1, 2] usually involves several stages of *preprocessing*, *segmentation*, and *feature extraction* before the final stage of *classification*. The goals of the earlier stages are to transform the image into a representation that makes the classification simpler. Feature extraction, in particular, transforms the very high dimensional pixel based representation of the original image into a much lower dimensional representation in which the objects are much more easily separated from each other and from the background. The design of the preprocessing and feature detection stages is a major component of the work involved in constructing an object

detection system using the traditional approach. Unfortunately, the design is usually dependent on the domain, and therefore may have to be repeated for each different task.

Neural networks have been applied in object detection by a number of researchers. Some have taken a traditional feature based approach [3, 4] in which a set of domain specific, hand designed, feature detectors are used to transform the image into a low dimensional vector. This feature vector is then used as input to the classification network.

Other researchers [5–7] have explored a domain independent approach by using the raw pixel values of the image as inputs to the neural network classifier. The approach we have used in previous work [8, 9] was such an approach (referred to as *the basic approach*), where the neural network was trained on the raw image pixels of the object examples cut out from the large images. The trained network was then applied in a moving window fashion over the large images to detection the objects of interest. A centre-finding algorithm was used to find the centres of object detected for evaluation. The major advantage here is that the cost and domain specificity of the preprocessing and feature extraction and hand-crafting of programs for feature extraction were successfully avoided. Past work has demonstrated the effectiveness of this approach on detecting small, well presented objects in large images with a relatively uniform background.

However, the basic approach has some problems. One problem is that the input vector to the neural network is typically very large. The effect of this is to require a large number of training examples and long training time. Although long training time is not necessarily a critical problem, generating a large number of correctly classified training examples may be infeasible in many applications. A second problem is that this approach often produces a large number of false alarms even though the objects of interest could usually be successfully detected. The problem is even serious when the objects are unclear or irregular, or the background is cluttered. This is mainly because the centre-finding algorithm was not sufficiently effective for reducing false alarms.

The goal of this approach is to investigate an alternative approach using domain independent, statistically based features (referred to as *pixel statistics*) and a new false alarm filter for object detection problems. Like domain specific features, these pixel statistics transform an image to a lower dimensional vector. However, the features are based on standard statistical measures, such as means and standard deviations, of parts of the image, and do not depend on any particular properties of the image domain. They are therefore equally relevant for any kind of image and class of object. The new false alarm filter aims to reduce false positive objects classified by the neural networks. A further characteristic of our approach is that, unlike most of the current work in the object detection area, where the task is to detect only objects of one class [10, 11], our objective is to detect objects from a number of classes.

The rest of paper is organised as follows. Section 2 describes the approach. Section 3 gives the three image data sets and section 4 presents the experimental results. Section 5 draws the conclusions and gives future directions.

2 The Approach

2.1 Overview of the Approach

This section describes the new approach to the use of pixel statistics and false alarm filter in neural networks for domain independent, multiclass object detection problems. An overview of the approach is shown in Figure 1.

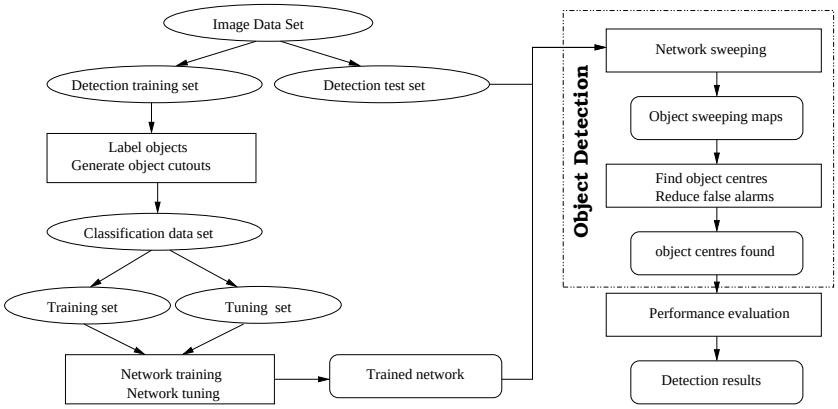


Fig. 1. Overview of the approach

First, the locations of all objects of interest in an image database are manually determined and their classifications are labelled and recorded along with their locations. Then this image database is split into two data sets: one is the *detection training set*, which is used for training the network, and the other is the *detection test set* reserved for evaluating the object detection performance of the network.

The neural network is never directly applied to an entire image, but only to sub-images of a fixed size. We refer to these sub-images as *input images*. For the network training stage, the input images are created by cutting out samples from images in the detection training set. For the object detection stage, the system sweeps a window (or *input field*) of the appropriate size over the entire image, creating an input image for each possible position. In both stages, a set of pixel statistics (low level image features) are calculated from the input image and used as inputs to the neural network.

The size of the input images is determined by the size of the object of interest in each of the image databases. Following the heuristic in [8], the input field is chosen to be sufficiently large to characterise the background but small enough to contain only a single object of interest. The largest object in each of the image databases can be used to determine this size.

The set of input images constructed from the detection training set is called the *classification data set.* This data set consists of input images centred on each of the objects in the images in the detection training set, labeled with the object class, plus a collection of input images that are not centred on the objects, and are labeled class *other*. The classification data set is randomly partitioned into a *classification training set* and a *classification tuning set*.

2.2 Neural Network Architecture

We use three layer, feed forward neural networks with one input layer, one output layer and one hidden layer. The number of input nodes is the number of pixel statistics computed from an input image. The number of output nodes depends on the number of object classes in the image database. The number of hidden nodes is empirically determined through experiments, using the classification tuning set to choose the optimal number of hidden nodes.

2.3 Pixel Statistics

Pixel statistics are low-level, domain independent image features, computed from the pixel intensities in a region of the input image. In this approach, we used two sets of pixel statistics, as shown in figure 2. The first set was constructed by the means and standard deviations computed from a series of concentric square regions centred in the input image window. The second set consisted of the means and standard deviations computed from a series of concentric circular regions.

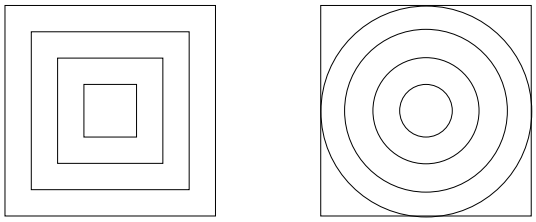


Fig. 2. Local square and circular features

2.4 Network Training and Tuning

We use the backward error propagation algorithm [12] with the variations of online learning and the fan-in factor [13, 14] to train the networks. In the *online learning* procedure, also called the *stochastic gradient procedure*, weight changes are applied to the network after each training pattern. The fan-in factor normalises the weights by dividing them by the number of inputs of the node to which the connection belongs. The normalisation is applied before network training and also to the weight changes during network training.

The training is terminated when the classification accuracy in the classification training set reaches a pre-defined percentage. When training is terminated, the trained network weights and biases are saved for the use in network tuning or subsequent resumption of training.

The trained network is then applied to the classification tuning set. If the performance is reasonable, then the trained network is ready to be used for object detection. Otherwise, the network architecture and/or the learning parameters need to be changed and the network re-trained, either from the beginning or from a previously saved, partially trained network. The classification tuning set

is also used as a "validation set" for monitoring network training process in order to obtain good parameters of the learned network for object detection.

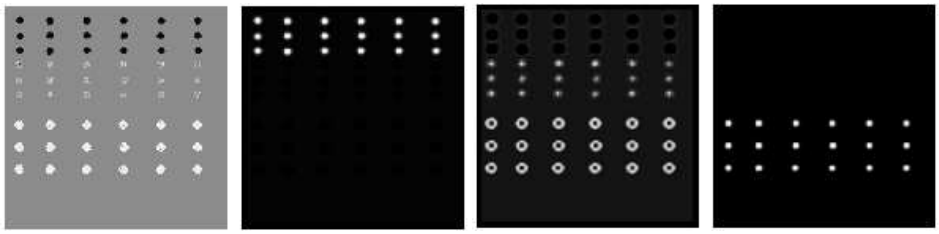
During network training and tuning, the classification is regarded as correct if the output node with the largest activation value corresponds to the desired class of a pattern.

2.5 Object Detection

In object detection, the trained network is applied to square input fields of the large images in the detection test set to detect the objects of interest. This consists of *network sweeping* and *false alarm reduction*.

Network Sweeping. During *network sweeping*, the successfully trained neural network is used as a template matcher, and is applied, in a moving window fashion, over the large pictures to detect the objects of interest. The template is swept across and down these large pictures, pixel by pixel in every possible location. The sweeping window selects a portion of a full image as the input image and passes the raw pixel values of the input image to the pixel statistics computation described earlier. The computed pixel statistics are then provided as inputs to the trained neural network and the network classifies the input image.

For each object class, the sweeping process will generate an *object sweeping map* — a bitmap representing the output value of the neural network for the given class at each position in the original image. Sample object sweeping maps for the three classes in an original image with an easy detection problem is shown in Figure 3. A black pixel in a sweeping map indicates a neural network output of 0 ("object not present") at that position; a white pixel indicates an output of 1 ("object definitely present"). Grey pixels indicate a partial match to the class at that position.



Original picture (a) | Class1-sweeping-map (b) | Class2-sweeping-map (c) | Class3-sweeping-map (d)

Fig. 3. Sample object sweeping maps in object detection.

False Alarm Reduction. Typically, a well trained neural network will detect an object not only when the sweeping window is centred over an object, but also when it is within a few pixels of the centre of an object. A simplistic interpretation of the neural network output would therefore suggest that there are multiple

objects present in the image within a few pixels of each other, which is obviously false. The 18 filled white areas in figure 3 (b), (c) and (d) show examples of this kind. Clearly, there is only one object of interest in each of these areas, but more pixels in these areas were considered potential objects. Also notice that there were also 18 white "ring" in figure 3 (c) with a black area in the centre of each ring. Clearly, those areas are false alarms (caused by partial objects of the white circles and pieces of the background), which need to be eliminated or at least reduced to one false alarm for each ring rather than several false alarms.

We developed a false alarm filter to deal with the above two situations. The algorithm is described as follows.

S1 For each class, choose a threshold.
S2 For each pixel in the object sweeping map, if the pixel value is less than the threshold, set the pixel to "off", indicating that this pixel would not be considered as the centre of any object by the network; otherwise, the pixel is "on", indicating that this pixel is a potential object centre.
S3 Scan the object sweeping map, pixel by pixel from the top left corner. If the current pixel is "off", ignore this position and continue the scan; otherwise (the current pixel is "on"), this pixel would be either a part of an object, or a part of a false alarm, and do the following:
   - if the pixel immediately above the current pixel is also "on", then the current pixel and the pixel above belong to the same object and the current pixel is referenced to the same object of the previous pixel;
   - if the pixel above is "off" but the left pixel is "on", reference the current pixel to the same object of the left pixel;
   - if both the above and the left pixels are "on" but they are referenced to different objects, then *merge* the two objects into one — they belong to the same object;
   - if both the above and the left pixels are "off", then mark the current pixel as a new object centre.
S4 Each of these marked objects (connected areas with pixels marked as "on") is either an detected object or a detected false alarm. Calculate the centre of them by taking the average of the minimum and the maximum coordinators of the pixels marked as "on" in each of these marked objects.
S5 Determine whether those areas are detected objects or false alarms. If the calculated centre is "on", then this area is a detected object; otherwise, it is a detected false alarm, which will not be reported by the neural networks.

It is possible that "object centres" for two (or more) different classes may be found at the same position. Only the class with the highest activation level at the position will be selected. We expect that this algorithm would greatly reduce false alarms in various object detection problems and accordingly improve the performance.

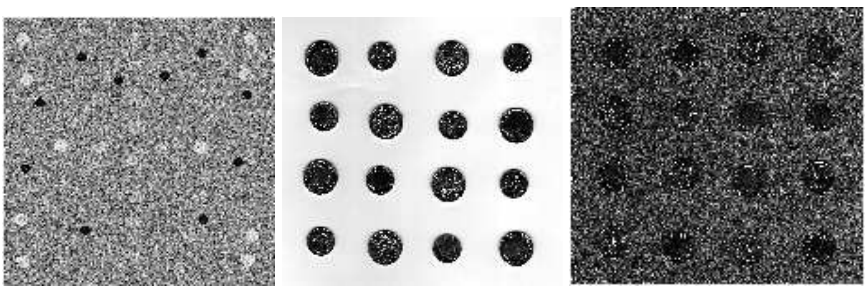
2.6 Object Detection Performance Measurement

The set of objects detected is reported by the trained neural network (more accurately the false alarm filter). To measure the performance of the network,

the system compares the reported object centres and classes against the desired known object centres and reports the number of objects correctly detected. We allow a tolerance of 4 pixels in the x and y directions.

We use the *detection rate* and *false alarm rate* to measure the object detection performance. The detection rate is the number of objects correctly reported as a percentage of the total number of real objects and the false alarm rate is the number of objects incorrectly reported as a percentage of the total number of real objects. For example, a detection system looking for grey squares in Figure 3 (a) may report that there are 25. If 15 of these are correct the detection rate will be $(15/18) = 83.3\%$. The false alarm rate will be $(10/18) = 55.6\%$. It is important to note that detecting objects in pictures with very cluttered backgrounds is an extremely difficult problem and that false detection rates of 200%–2,000% are common [11, 6].

3 Image Data Sets



Object size:18×18 (Shape) | Object size:90× 90 (Coin1) | Object size:70×70 (Coin2)

Fig. 4. Object Detection Problems

We used three data sets in the experiments. Example images are given in figure 4. These data sets provide object detection problems of increasing difficulty. Data set 1 (Shape) was generated to give well defined objects against a noisy background. The pixels of the objects were generated using a Gaussian generator with different means and variances for different classes. There are three classes of small objects of interest in this database: black circles, grey squares and white squares. Note that the grey squares have very similar intensities to the background. Data set 2 (Coin1) consists of scanned images of New Zealand coins. There are four object classes of interest: head side and tail side of the 5 cent coins and 10 cent coins (*head005, tail005, tail010, tail010*). In the approach, neural networks need to determine where those coins are, whether those coins are 5 cents or 10 cents, and distinguish either tail side up or head side up. The objects in each class have a similar size but are located at arbitrary positions and with different rotations. Compared with the shape data set, the detection problems here are more difficult. Data set 3 (Coin2) contains two object classes of interest (the head side and the tail side of New Zealand 5 cents) with various

orientations, but the background is highly cluttered, which makes the detection problem much harder. Given the low resolution of the images in the two coin data sets, the detection tasks are very difficult — even human eyes could not distinguish those classes perfectly.

In the experiments, we used ten images in both the detection training set and the detection test set for each of the three data sets.

4 Results

This approach was examined on the above three image data sets and compared with the basic approach with the same sets of pixel statistics. For all cases, the experiments were repeated 10 times and the average results on the detection test set were presented.

4.1 *Shape* Data Set

The results of the new approach and the basic approach for the shape data set are shown in table 1. Detecting black circles and white circles was relatively straightforward and both the basic approach and the new approach produced perfect detection results. For detecting the grey circles, however, the two approaches gave different results. In the basic approach, the square region features produced ideal performance, but the circular features resulted in a false alarm rate of 17.0% on average. The new approach with the false alarm filter achieved ideal performance, suggesting that the new approach outperformed the basic approach for this data set.

Table 1. Object detection results on the shape data set.

Image Data Set			black circle	grey square	white circle
Best Detection Rate(%)			100	100	100
Average False Alarm Rate (%)	square regions	Basic approach	0	0	0
		The new approach	0	0	0
	circular regions	Basic approach	0	17.0	0
		The new approach	0	0	0

It is also note that square region features achieved better results than the circular features. This is probably because the square region features contain more heuristics than the circular region features for the grey square objects. Also notice that these results are quite good since these grey square objects are not easy to detect — even human eyes can not detect those objects easily due to the fact that the grey square objects are too similar to the cluttered background.

4.2 *Coin1* Data Set

The results in the *Coin1* data set (table 2) shows a similar pattern to the *shape* data set in that the new approach with the false alarm filter greatly reduced the false alarms and achieved a much better performance than the basic approach for both sets of features.

Table 2. Object detection results on the *coin1* data set.

Image Data Set			5 cent coins		10 cent coins	
			tail	head	tail	head
Best Detection Rate(%)			100	100	100	100
Average False Alarm Rate (%)	square regions	Basic approach	66.3	14.3	0	0
		The new approach	21.3	4.4	0	0
	circular regions	Basic approach	25	55	0	0
		The new approach	0	33.1	0	0

4.3 *Coin2* Data Set

The results in the *Coin2* data set shows exactly the same pattern as in *Coin1*. While it is not concluded which feature set is better for this data set, the new approach with the false alarm filter is certainly better than the basic approach.

Table 3. Object detection results on the *coin2* data set.

Image Data Set			tail	head
Best Detection Rate(%)			100	100
Average False Alarm Rate (%)	square regions	Basic approach	163	5.6
		The new approach	0	0
	circular regions	Basic approach	134	9.4
		The new approach	0	0

The basic approach produced a large number of false alarms with both sets of features. However, all the false alarms were eliminated by the false alarm filter in the new approach.

It is important to note that both feature sets achieved similar results on the two coin data sets and it is hard to conclude one set is better than the other. This is different from our early hypothesis that the circular feature set would be better than the square features.

5 Conclusions

The goal of this paper was to investigate the effectiveness of the new neural network approach with two sets of pixel statistics and the false alarm filter for object detection problems. The approach was tested and compared with the basic approach on three object detection problems of increasing difficulty.

For all the three object detection problems investigated here, the new approach achieved very good results. The results also showed that the new approach always outperformed the basic approach using the same set of features. Compared with the basic approach, the new approach appears to be more effective in reducing false alarm rates.

While the local concentric region features were more effective than the circular features for the shape data set, it did not appear to conclude one feature set is better than the other for the two coin data sets. This contradicts our original hypothesis that the circular region features would be better than the square region features for the two data sets. Further investigation needs to be carried out in the future.

References

1. Olivier Faugeras. *Three-Dimensional Computer Vision – A Geometric Viewpoint*. The MIT Press, 1993. ISBN 0-262-06158-9.
2. A. Yli-Jaaski and F. Ade. Grouping symmetrical structures for object segmentation and description. *Computer Vision and Image Understanding*, 63(3):399–417, May 1996.
3. David P. Casasent and Leonard M. Neiberg. Classifier and shift-invariant automatic target recognition neural networks. *Neural Networks*, 8(7/8):1117–1129, 1995.
4. Philip Winter, Shahab Sokhansanj, Hugh C.Wood, and William Crerar. Quality assessment and grading of lentils using machine vision. In *Agricultural Institute of Canada Annual Conference*, Saskatoon, SK S7N 5A9, Canada, July 1996. Canadian Society of Agricultural Engineering. CASE paper No. 96-310.
5. V. Ciesielski and J. Zhu. A very reliable method for detecting bacterial growths using neural networks. In *Proceedings of the International Joint Conference on Neural Networks*, pages 62–67, Beijing, November 1992.
6. Mukul V. Shirvaikar and Mohan M. Trivedi. A network filter to detect small targets in high clutter backgrounds. *IEEE Transactions on Neural Networks*, 6(1):252–257, Jan 1995.
7. Mengjie Zhang and Victor Ciesielski. Using back propagation algorithm and genetic algorithm to train and refine neural networks for object detection. In Trevor Bench-Capon, Giovanni Soda, and A Min Tjoa, editors, *Proceedings of the 10th International Conference on Database and Expert Systems Applications (DEXA'99)*, pages 626–635, Florence, Italy, August 1999. Springer-Verlag. Lecture Notes in Computer Science, (LNCS Volume 1677).
8. Mengjie Zhang. *A Domain Independent Approach to 2D Object Detection Based on the Neural and Genetic Paradigms*. PhD thesis, Department of Computer Science, RMIT University, Melbourne, Australia, August 2000.
9. Mengjie Zhang and Victor Ciesielski. Neural networks and genetic algorithms for domain independent multiclass object detection. *International Journal on Computational Intelligence and Applications*, 4(1), March 2004.
10. Paul D. Gader, Joseph R. Miramonti, Yonggwan Won, and Patrick Coffield. Segmentation free shared weight neural networks for automatic vehicle detection. *Neural Networks*, 8(9):1457–1473, 1995.
11. Michael W. Roth. Survey of neural network technology for automatic target recognition. *IEEE Transactions on neural networks*, 1(1):28–43, March 1990.
12. D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, and the PDP research group, editors, *Parallel distributed Processing, Explorations in the Microstructure of Cognition, Volume 1: Foundations*, chapter 8. The MIT Press, Cambridge, Massachusetts, London, England, 1986.
13. Dick de Ridder. Shared weights neural networks in image analysis. Master's thesis, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands, Feb 1996.
14. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1989.

A Multiclass Object Detection

Object detection refers to the detection of small objects in large pictures. It consists of both object classification, which involves determining the classes of objects of interest, and object localisation, which identifies the positions of all the objects in the large pictures.

In most automatic object detection systems, all the objects of interest are considered to be a *single class*. Such systems address the single class object detection task, which is the problem of distinguishing *objects* or *targets* from *non-targets* or the *background*. In contrast, *multiclass object detection* refers to the case where there is more than one class of objects and both the classes and locations of all the objects must be determined. In general, multiclass object detection problems are harder than single class detection problems, and multiclass detection using a single trained program (such as a neural network) is an even more difficult problem.