Centred Weight Initialization in Neural Networks for Object Detection

Mengjie Zhang and Victor Ciesielski

Department of Computer Science,
Royal Melbourne Institute of Technology
GPO Box 2476V, Melbourne Victoria 3001, Australia
mengjie,vc@cs.rmit.edu.au

Abstract. We describe a new initialization method for pixel based neural networks for object detection problems in which the locations of relatively small objects in large pictures must be found. The networks have a square input field which is large enough to contain all objects of interest and are trained on examples which have been cut out from the large pictures. The trained networks are then applied, in moving window fashion, over the large pictures to locate the objects of interest. The new initialization method places the highest initial weights at the centre of the input field. Weights decrease uniformly to the perimeter. This is in contrast to the standard initialization method in which all weights are random. We have tested the method on three object detection problems of increasing difficulty. In all cases the method resulted in improved detection performance, as measured by precision and recall, and appears to be suited to object detection problems where the background is fairly uniform. Visualization of the weights in trained networks resulting from both initialization methods revealed that trained networks from both approaches contained feature detectors which 'made sense' for the domain, but learning in networks with centred initial weights was more focused on features which discriminated between the classes.

Keywords: Pixel based neural network; Network testing; Target recognition; Target detection.

1 Introduction

This paper addresses the problem of finding the locations of a number of different kinds small objects in a set of large pictures. For example, in the problem illustrated by middle picture of Figure 2 we want to find the centres of all of the 5 cent and 20 cent coins and whether the head or tail side is up. In the right hand picture of Figure 2, which is a picture of a human retina, we would like to find all of the micro aneurisms and haemorrhages. (Note: The picture is presented

[0] In Edwards J. (Ed.) *Proceedings of the 22nd Australasian Computer Science Conference* Springer, Singapore, p39-50, 1999.

at too coarse a level of resolution for this distinction to be evident). Examples of other problems of this kind include target detection problems [Gader et al., 1995; Waxman et al., 1995; Won et al., 1997] where the task is to find, say, all tanks, trucks and helicopters in a picture.

Two main approaches have been used in neural network systems for these kinds of problems – feature based and pixel based. In feature based approaches various features such as brightness, colour, size and perimeter are extracted from the sub-images of the objects of interest and used as inputs to the networks, as in [Casasent and Neiberg, 1995; Rogers et al., 1995; Roitblat et al., 1995; Winter et al., 1996]. In pixel based approaches [Ciesielski and Zhu, 1992; Jean and Wang, 1994; Shirvaikar and Trivedi, 1995] the pixel values are used directly as inputs. The approach described in this paper is pixel based. The paper suggests an improvement to a current approach and describes an investigation of its effect on pictures of increasing difficulty. More details of our approach can be found in [Zhang and Ciesielski, 1998]

It is important to note that finding objects in pictures with very cluttered backgrounds is an extremely difficult problem and that false detection rates of 200-2,000% (that is the detection system suggests that there are 200 times as many objects as there really are) are common [Roth, 1990; Shirvaikar and Trivedi, 1995].

1.1 The Object Detection System

A brief outline of our approach to object detection is as follows:

1. Assemble a data base of pictures in which the locations and classes of all of the objects of interest are manually determined. Reserve some of the pictures as 'unknowns' for measuring detection performance.
2. Determine an appropriate size (n) of a square which will cover all objects of interest and form the input field of the networks.
3. Build training and test sets by cutting out squares of size n. The $n \times n$ pixel values form the inputs of a training pattern and the classification is the output.
4. Choose a hidden layer size and train a three layer feed forward network by backward error propagation [Hinton, 1989].
5. Use the the trained network as a moving window template [Ballard and Brown, 1982] across the pictures from which the training data was extracted. Determine thresholds for each class.
6. Use the trained network as a moving window template on the pictures reserved with step 1. If the output for a class exceeds the threshold then report an object of that type at the current location.
7. Determine the recall, that is, the number of objects correctly reported as a proportion of the total number of objects, and the precision, that is, the number of objects correctly found as a proportion of the total number of objects reported.

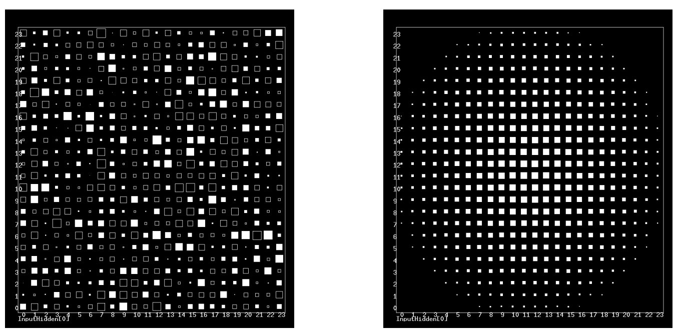
1.2 Goals

In the standard backward error propagation algorithm the network weights are initialized to small random values. In our approach we expect that the central pixels in a square are more important than those on the perimeter and initialize network weights accordingly. We investigate whether the centred initial weights:

- Will decrease the number of epochs to convergence;
- Will improve network accuracy on test data;
- Will result in better object detection, that is, higher recall and higher precision; and
- Will be effective for problems of increasing difficulty.

2 Centred Weight Initialization

In the standard backward error propagation algorithm the weights are initialized with small random weights, we used numbers between -0.5 and +0.5. This would result in a pattern of weights between the input layer and a hidden unit as shown in Figure 1a. In this figure the full squares represent positive weights and the outline squares represent negative weights, while the size of the square is proportional to the magnitude of the weight. To facilitate visualization the weights are shown as a parallel array to the input field. Thus $weight(i,j)$ in Figure 1a corresponds to $pixel(i,j)$ in the input field. Figure 1b shows centred initial weights.



(a) Random Initialization (b) Centred Initialization

Fig. 1. Initial Input-Hidden Weights

The initialization algorithm requires an input parameter, max_weight, the size of the central weight. The four central pixels are set to this weight. We then calculate

$$weight\_gap = 2 * max\_weight / size\_square \quad (1)$$

The weights corresponding to the neighbours of the central pixels are then found by subtracting the weight gap from the central weights and similarly the weights of their neighbours. The process continues until all weights to the perimeter have been set. Finally a small random amount, based on a Gaussian distribution with mean 0 and variance $weight\_gap/30$ is added to each weight. This ensures that the weights are largest at the centre and decrease uniformly to the perimeter as shown in Figure 1b. While the weights are not truly random they provide a satisfactory starting point for the back propagation algorithm. Results for different choices of the central weight are shown below in section 4.

3 The Image Data Bases

We used three different data bases in the experiments. Example pictures and key characteristics are given in Figure 2. The pictures were selected to provide problems of increasing difficulty. Data base 1 (Easy) was generated to give well defined objects against a uniform background. The pixels of the objects were generated using a Gaussian generator with different means and variances for each class. The coin pictures were intended to be somewhat harder and were taken with a CCD camera over a number of days with relatively similar illumination. In these pictures the background varies slightly in different areas of the image and between images and the objects to be detected are more complex, but still regular. The retina pictures were taken by a professional photographer with special apparatus at a clinic and contain irregular objects on a fairly cluttered background. Note that in each of the data bases the background counts as a class.

4 Results

4.1 Training and Test Results

Training and test results for the easy pictures are shown in Table 1. Line 1 of the figure shows that 15 runs with random initial weights terminated at an average of 199.40 epochs with an average standard deviation of 18.09 and the mean squared error on the test set was 0.0509 with a standard deviation of 0.03. The same data for centred initialization with different values of max_weight are given in the other lines of the table. It is evident that for $0.07 \leq max\_weight \leq 0.28$ the centred weight method is superior.

A similar series of experiments was performed for the other two data bases. In both cases there was a range of values for max_weight which gave improved training times and lower test mean squared error (MSE). These results are summarized in Table 2. On the coins data base, for example, a 3 layer network will show improved performance for $0.024 \leq max\_weight \leq 0.90$. There is an average decrease in training epochs of 27.15% and and 19.23% in test MSE. There does not appear to be a relationship between problem difficulty and improvement.

Unfortunately there does not appear to be a reliable way of choosing the best max_weight and some experimentation is needed. However, as suggested

Easy (Circles and Squares)	Medium Difficulty (Coins)	Very Difficult (Retinas)
Number of Images: 7	Number of Images: 20	Number of Images: 15
Number of Classes: 4	Number of Classes: 5	Number of Classes: 5
Max size of Object: 12 pixels	Max size of Object: 22 pixels	Max size of Object: 16 pixels
Number of Objects: 240	Number of Objects: 600	Number of Objects: 64
Picture Size 700x700	Picture Size 640x480	Picture Size 1024x1024

Fig. 2. Object Detection Problems of Increasing Difficulty

earlier, the major problem in these kinds of object detection problems is a very high number of false positives. If the centred weight method can lower this number significantly then a short search for a good *max_weight* is a small price to pay. The next section compares the detection performance of the two weight initialization methods.

4.2 Object Detection Results

This section describes a series of comparisons between the object detection performance of networks initialized with the standard random method and networks initialized with the centred weight method. In all cases the procedure described in section 1.1 is repeated 10 times and the averages are presented.

Table 3 shows the comparison for the easy pictures. In each run it was always possible to find a threshold for the network output for class 1, black circles, and class 3, white circles, which resulted in finding all objects of these classes (recall 100%). Furthermore there were no false positives (precision 100%). However this was not the case for class2 (grey squares). As the detection threshold of the network is lowered the recall increases but so do the false positives. This is a common occurrence in detection systems of this kind and systems are compared by looking at the precision at various levels of recall. At a recall of 90% the precision is 88.47 for the random method and 98.93 for the centred method. In fact the precision of the centred method is better at each level of recall. Note that

Expt No.	Initial Weights format	Max-Wei	Wei-Gap	Epochs ($\mu \pm \sigma$)	Training MSE ($\mu \pm \sigma$) ($\times 10^{-2}$)	Test MSE ($\mu \pm \sigma$) ($\times 10^{-2}$)
1	random			199.40±18.09	5.09±0.30	5.17±0.27
2	centred	0.420	0.060	430.80±29.78	4.81±0.10	4.85±0.11
3	centred	0.350	0.050	263.00±24.56	5.01±0.07	5.03±0.05
4	centred	0.280	0.040	219.47±14.20	4.47±0.12	4.55±0.11
5	centred	0.210	0.030	143.80±23.07	4.63±0.17	4.70±0.15
6	centred	0.140	0.020	109.87±10.64	5.04±0.11	5.08±0.12
7	centred	0.070	0.010	128.87±9.42	5.29±0.10	5.35±0.07
8	centred	0.035	0.005	139.33±9.03	5.37±0.12	5.45±0.11
9	centred	0.014	0.002	153.40±11.76	5.11±0.06	5.19±0.04

Table 1. Network training epochs and test performance on the "easy" pictures. (Network architecture: 196-4-4; Learning rate = 0.5; Momentum = 0; Stop when test performance is 100%; Number of repetitions = 15)

Database	Network Arch.	Range of Centred Initial Parameter (max_wei/wei_gap)	Improvement of Training Speed (μ/σ)	Improvement of Test Performance (μ/σ)
Circles and Squares (Easy)	196-4-4	> 0.07/0.01 and < 0.28/0.04	44.90%/41.18%	9.09%/44.44%
	196-5-4	> 0.07/0.01 and < 0.28/0.04	40.59%/35.48%	8.35%/30.00%
Coins (Medium Difficulty)	576-3-5	>0.024/0.002 and <0.090/0.0075	27.15%/78.59%	19.23%/45.65%
	576-5-5	>0.030/0.0025 and <0.120/0.0100	22.95%/5.70%	21.77%/2.78%
Retina (Very Difficult)	256-4-5	> 0.16/0.02 and < 0.48/0.06	19.5%/17.32%	2.74%/27.78%
	256-5-5	> 0.28/0.035 and < 0.80/0.10	25.57%/78.64%	3.48%/82.28%

Table 2. Summary of the improvement in training time and test performance of the centred weight initialization method.

the precision is directly related to false positive rate – a low precision indicates a high false positive rate while a high precision indicates a low false positive rate.

Experiments with the coin images gave similar results. These are shown in Table 4. Detecting heads and tails of the 5 cent coin and tails of the 20 cent coin turned out to be relatively straight forward, while detecting the heads of the 20 cent coins was a difficult problem, as shown in Table 4b. However the precision of the centred weights method at all levels of recall was superior to the random method.

		object classes															
		class1	class3	class2													
recall (%)		100	100	100	96.67	93.33	90.00	86.67	83.33	80.00	76.67	73.33	70.00	66.67	63.33	60.00	<= 56.67
best precision (%)	random	100	100	52.31	74.97	86.72	88.47	89.17	91.81	93.15	96.06	96.26	96.48	96.80	98.10	99.47	100
	centred	100	100	68.29	98.69	98.97	98.93	98.89	98.85	99.60	99.58	100	100	100	100	100	100

Table 3. Comparison of object detection in easy pictures using random and centred initial weights

The results for the retina pictures showed similar trends, but the precision was very low.

The results are summarized in Figure 3 where precision is plotted against recall for the difficult classes. In all cases the centred weight results are superior – the precision at all levels of recall is always higher, in some cases very much so.

For the easy and coins pictures network thresholds can be chosen which give good precision and recall (Figure 3a and 3b). This is not the case for the more difficult retina pictures (Figure 3c and 3d).

4.3 Analysis of Weights

Figure 4 shows weights from a trained 576-3-5 network which was able to successfully detect all of the classes in the coins pictures. Part (a) shows the weights from the input units to the first hidden unit, part (b) from inputs to second hidden unit and part (c) the inputs to the third hidden unit. The weights are shown in a 24x24 square to facilitate visualization. Figure 4d shows the weights from the hidden layer to the 5 output units. The 5 rows in this matrix, in top to bottom order, correspond to the classes *head005, tail005, head020, tail020* and *other*.

Inspection of the first column of Figure 4d reveals that weight matrix (a) has a positive influence on *head005* and, a very strong positive influence on *tail*005. This same matrix has a negative effect on the other classes. Inspection of the second column reveals that matrix (b) has a strong influence in differentiating between 5 cent and 20 cent coins. If we regard the nodes of the hidden layers as representing feature detectors learnt by the network then Figures 4a-4c are a visual representation of these features. Visually these features 'make sense' as there are regions corresponding to the 5 cent coin, the annulus remaining when a 5 cent coin is 'removed' from the centre of a 20 cent coin and to the background.

Figure 5 shows a similar set of weight diagrams for a 576-3-5 network which has been initialized with the centred weights method. The same features are

		object classes				
		head005	tail005	tail020		
recall (%)		100	100	100	93.75	<= 87.5
best precision (%)	random	100	100	98.421	99.375	100
	centred	100	100	100	100	100

(a)

		object classes															
		head020															
recall (%)		100	93.75	87.50	81.25	75.00	68.75	62.50	56.25	50.00	43.75	37.50	31.25	25.00	18.75	12.50	6.25
best precision (%)	random	35.45	37.07	38.48	41.58	44.73	50.47	54.85	61.43	70.06	80.91	81.59	80.94	79.74	95.00	96.67	100
	centred	70.74	72.18	75.86	77.61	81.89	90.58	93.37	98.00	100	100	100	100	100	100	100	100

(b)

Table 4. Result of regular object detection in coin pictures by using random/centred initial weights

evident as in Figure 4 but in this case they are much better defined. It appears that the centred weights initialization method has resulted in learning which is focused on features necessary to discriminate the classes.

5 Conclusions

The goal of the work described in this paper was to investigate a new method of setting initial weights in pixel based neural networks for image detection problems. Our results show that, for the three detection problems investigated, the new method, centred weight initialization, is superior to standard random initialization.

The methods were compared on three detection problems of increasing difficulty. On the easy (circles and squares), medium difficulty (coins) and very difficult (retinas) problems it was possible to find centred initial weights which resulted in fewer training epochs and lower test mean squared error. More importantly however, the centred weight method produced networks which were much better, in terms of recall and precision, at the task of finding the locations of objects of interest in large pictures. The amount of improvement did not appear to be related to the difficulty of the problem. Overall, precision and recall for the

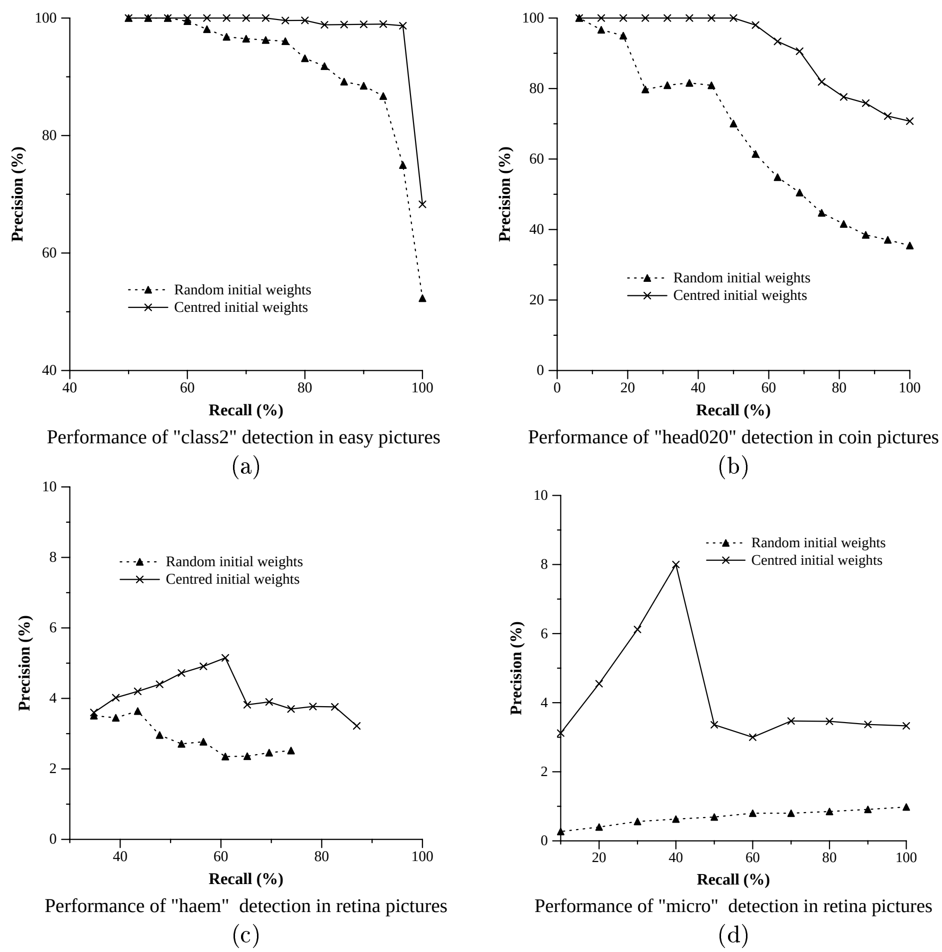


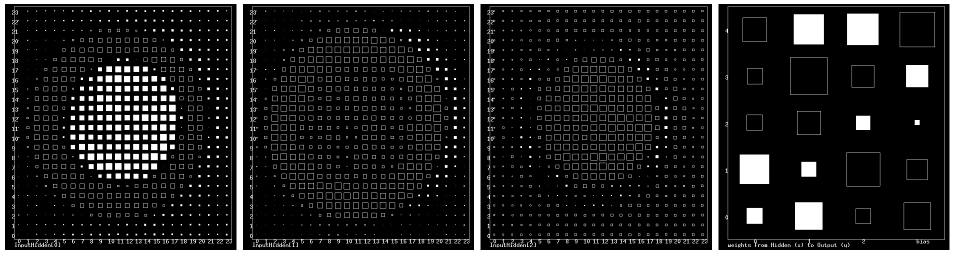
Fig. 3. Some typical results for object detection in the three databases

easy and medium difficulty problems was very good, but precision and recall for the difficult retina problem were much too low. However this is consistent with the performance of other methods on similar difficult problems.

The central weight initialization has the disadvantage of requiring an empirical search for a good initialization parameter but this is more than offset by the increase in detection accuracy.

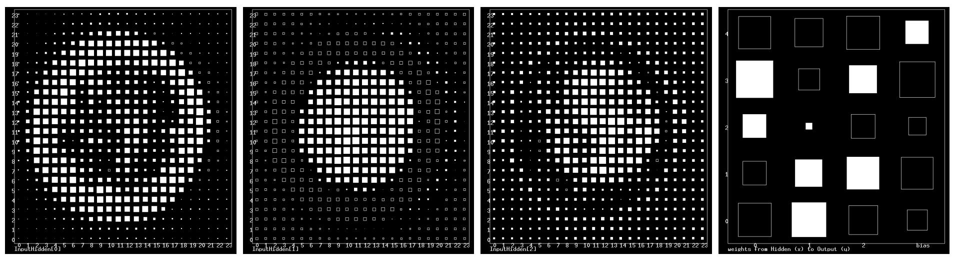
Visualization of the weights in trained networks resulting from both initialization methods revealed that trained networks from both approaches contained feature detectors which 'made sense' for the domain, but learning in networks with centred initial weights was more focused on features which discriminated between the classes.

The centred weight initialization method produces pixel based networks that



(a) (b) (c) (d)

Fig. 4. Weights for coin recognition network for initial random weights.



(a) (b) (c) (d)

Fig. 5. Weights for coin recognition network for initial centred weights.

work well on objects on a relatively uniform background. However more work is needed for objects on non uniform and cluttered backgrounds such as our retina pictures.

Acknowledgements

We would like to thank Dr. James Thom for useful discussions on image retrieval, Steven Hayes whose data preparation and neural network simulation programs were used by the project and Chris Kamusinski who provided the retina pictures.

References

Ballard, Dana H. and Brown, Christopher M. (1982). *Computer Vision*. Englewood Cliffs, N.J: Prentice-Hall, Inc.

Casasent, David P. and Neiberg, Leonard M. (1995). Classifier and shift-invariant automatic target recognition neural networks. *Neural Networks*, 8(7/8):1117–1129.

Ciesielski, V. and Zhu, J. (1992). 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.

Gader, Paul D., Miramonti, Joseph R., Won, Yonggwan, and Coffield, Patrick (1995). Segmentation free shared weight neural networks for automatic vehicle detection. *Neural Networks*, 8(9):1457–1473.

Hinton, Geoffrey E. (1989). Connectionist learning procedures. *Artificial Intelligence*, 40(1-3):185–234. Reprinted in J. Carbonell, editor, "Machine Learning: Paradigms and Methods", MIT Press, 1990. Also appears as Technical Report CMU-CS-87-115 (version 2), Carnegie Mellon University, Pittsburgh, PA, December 1987.

Jean, Jack S. N. and Wang, Jin (1994). Weight smoothing to improve network generalization. *IEEE Transactions on neural networks*, 5(5):752–763.

LeCun, Y., B.Boser, J.S.Denker, D.Henderson, W.Hubbard, R.E.Howard, and L.D.Jackel (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, (1):541–551.

Rogers, Steven K., Colombi, John M., Martin, Curtis E., Gainey, James C., Fielding, Ken H., Burns, Tom J., Ruck, Dennis W., Kabrisky, Matthew, and Ocley, Mark (1995). Neural networks for automatic target recognition. *Neural Networks*, 8(7/8):1153–1184.

Roitblat, H. L., Au, W. W. L., Nachtigall, P. E., Shizumura, R., and Moons, G. (1995). Sonar recognition of targets embedded in sediment. *Neural Networks*, 8(7/8):1263–1273.

Roth, Michael W. (1990). Survey of neural network technology for automatic target recognition. *IEEE Transactions on neural networks*, 1(1):28–43.

Shirvaikar, M. and Trivedi, M. (1995). A network filter to detect small targets in high clutter backgrounds. *IEEE Transactions on Neural Networks*, 6(1):252–257.

Waxman, Allen M., Seibert, Michael C., Gove, Alan, Fay, David A., Bernandon, Ann Marie, Lazott, Carol, Steele, William R., and Cunningham, Robert K. (1995). Neural processing of targets in visible, multispectral ir and sar imagery. *Neural Networks*, 8(7/8):1029–1051.

Winter, Philip, Sokhansanj, Shahab, C.Wood, Hugh, and Crerar, William (1996). Quality assessment and grading of lentils using machine vision. In *Agricultural Institute of Canada Annual Conference*, Saskatoon, SK S7N 5A9, Canada. Canadian Society of Agricultural Engineering. CASE paper No. 96-310.

Won, Yonggwan, Gader, Paul D., and Coffield, Patrick C. (1997). Morphological shared-weight networks with applications to automatic target recognition. *IEEE Transactions on Neural Networks*, 8(5):1195–1203. ISSN 1045-9227.

Zhang, Mengjie and Ciesielski, Victor (1998). Centred weight initialization to improve the performance of network training speed and the performance of object detection. Technical Report TR 98-10, Department of Computer Science, RMIT University, Melbourne, Australia.