

Finding Templates for Cephalometric Landmarks using Pulse Coupled Neural Networks and Genetic Programming

Andrew Innes
Dept of Mechanical and
Manufacturing Engineering
RMIT University
Melbourne, Australia

Vic Ciesielski
School of Computer Science
and Information Technology
RMIT University
Melbourne, Australia

John Mamutil
Braces Pty Ltd
404 Windsor Road
NSW 2153, Australia

Sabu John
Dept of Mechanical and
Manufacturing Engineering
RMIT University
Melbourne, Australia

Abstract *We describe an approach to evolving detection programs to find cephalometric landmarks in digital X-rays. The approach involves the use of a moving window template which uses pixel statistics of different shaped regions under the moving window. Genetic programming is used to evolve detection programs based on the template. We show that analysis by pulse coupled neural networks can be used to determine region shapes. The approach has been tested on three landmark points. In all cases the pulse coupled analysis has improved detection performance over manually determined shapes, with a significant gain in improving detection rate of the lip from 85.1% to 96.3%.*

Keywords: Cephalometric analysis, Genetic Programming, Object detection, Pulse Coupled Neural Networks, PCNN, Computer Vision.

1 Introduction

Cephalometric analysis [1, 2, 3] is used by orthodontists to prepare treatment plans for patients with a range of orthodontic problems. Treatment planning requires accurate detection of a number of landmarks and measurement of the distances and angles between them. The landmarks used by Braces Pty Ltd are shown in figure 1. Our overall goal is to develop an automated system for performing a cephalometric analysis. In previous work we have established that accurate detection programs

for some of the landmarks can be obtained using genetic programming [4]. These detection programs use a manually constructed moving window template and features computed from the grey level image beneath the moving window. We have also previously established that analysis by pulse coupled neural networks would be a useful pre-processing step [5]. The aim of the work presented in this paper is to determine whether the outputs of the pulse coupled network can be used in the evolution of the landmark detection programs and whether this will increase the detection accuracy.

2 Methodology

The method developed here is to locate landmarks in regions categorised in terms of difficulty from relatively easy to hard (refer to Figure 2). The original image is scaled down to 20% of the original size. This reduces the number of genetic program evaluations during training and reduces the effect of the Gaussian noise on the image. The feature set is based on dividing the area surrounding a landmark into specific shapes (refer to Figure 4) individual to the landmark characteristics. The features consist of the mean and standard deviation of pixel intensities for each shape. The approach is based on [4] and [6] which used genetic programming to locate and classify objects such as heads/tails of different Australian coins in large images, and haemorrhages and micro-aneurisms in retina images.

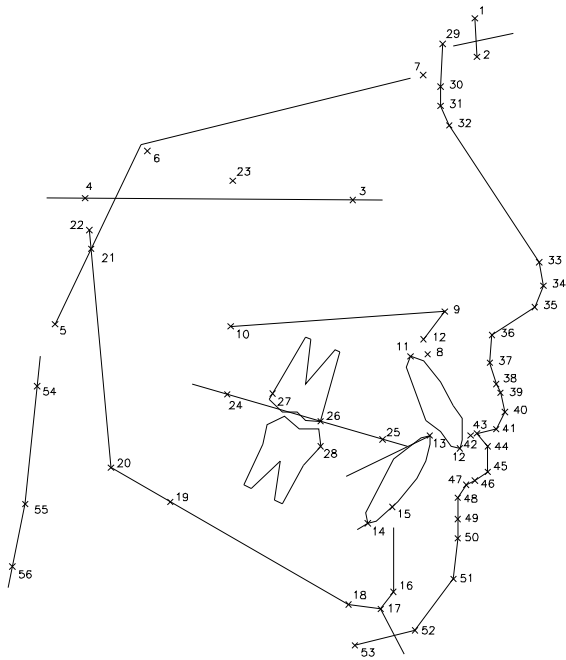


Figure 1: Line tracing of the fifty six landmarks required for a cephalometric analysis. Soft tissue landmarks are represented by points 29 -53.

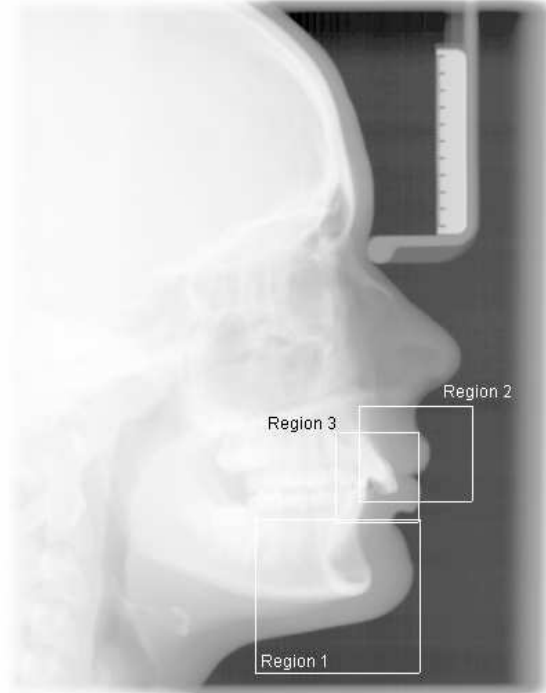


Figure 2: A digital cephalogram depicting three regions containing the menton, upper lip and incisal upper incisor landmarks.

A brief outline of the method used in this work is as follows:

1. Assemble a database of images with the known positions of landmarks to be located.
2. Use knowledge of facial geometry to extract regions in which the landmarks are expected to be. The left bottom end of the ruler (refer to Figure 2) can be reliably found. Each landmark will be in an area that can be defined by an angle and distance from this point. For example, the landmarks used in this paper - the menton point, the upper lip and the incisal upper incisor are expected to be in regions 1, 2 and 3 respectively.
3. Reserve some sub-images as ‘unknowns’ for measuring detection performance as the test set.
4. Determine SQUARE_SIZE, the size of a square centered on the landmark that will contain enough distinguishing information to permit the landmark to be identified. This is the size of the input window that will be used by the evolved programs. Manually determine the

appropriate shapes in the input window that are expected to generalise the training images and discriminate the landmark point from the background. These are shown in the first column of Figure 4.

5. Generate two shapes from the PCNN analysis as described in Section 2.1.
6. Invoke an evolutionary process to generate a program which can locate the landmark in its input field.
7. Apply the generated program as a moving template to the reserved test images from step 3, and obtain the position of the landmark. Calculate the detection rate and the false alarm rate on the test set as the measure of performance.

This method differs from the approach in [4] by incorporating the additional step outlined in item 5 that automatically extracts two shapes creating a further four features.

2.1 Pulse Coupled Neural Network

[5] proposed a method of using the PCNN to highlight key craniofacial features from digital-X-rays. The use of the PCNN for this problem was to automatically extract two shapes for each of the three landmarks as shown in the second column of Figure 4. The shapes for each landmark are extracted as follows:

1. A square cutout of size `SQUARE_SIZE` centred on the landmark is made for each image in the training set. The PCNN is applied to each cutout to segment the image into a binary format.
2. A template is created by averaging each pixel from the PCNN output images in the training set using the formula:

$$Template(i, j) = \frac{1}{n} \sum_{k=1}^n a_k(i, j),$$

where $a(i, j)$ is the segmented image from the PCNN, and n is the number of images in the training set. The templates for each landmark are shown in the second column in Figure 4.

3. Two shapes are then extracted. One shape (A) corresponds to the white pixels in the images in the second column of Figure 4. The other shape (B) consists of the black pixels. The grey pixels are not used.

Features F1-F4, shown in Figure 4, are the new features from the PCNN, and the remaining features are the same as those used in [4].

2.2 Genetic Programming

2.2.1 The Terminal Set

In the context of genetic programming in object detection problems, terminals correspond to image features. Because each landmark is distinct in shape, grayscale and contrast, a different set of features or shapes is required to define each landmark. The features presented in this paper correspond to different shapes with the resulting mean and standard deviation (refer to Figure 4). In addition to these features, a terminal that generates a random number in the range of [0,255] was included.

2.2.2 The Function Set

The function set $\{+, -, *, /\}$ consists of four arithmetic operations. The $+$, $-$ and $*$ have their usual meanings, while $/$ represents a protected division



Figure 3: The top, middle and bottom rows are indicative of the output from the PCNN for the menton, upper lip and incisal upper incisor respectively.

which is the usual division operator except that a divide by zero produces zero.

The output of the genetic program, *ProgOut*, is a floating point number which is interpreted as the likelihood that the evaluated position from the image is a landmark or background. During training the highest value of *ProgOut* from each image is used as the predicted position of the landmark. The predicted position given by the genetic program is then compared with the known true location and the example is classified as a true positive, a false alarm or unclassified (refer to section 2.2.3).

2.2.3 The Fitness Function

The fitness of a program during training is calculated by using detection rate and false alarm rate. The fitness is calculated as follows:

1. The program is applied as a moving window to each training image and the program output (*ProgOut*) evaluated at each pixel location. The predicted position of the detected landmark is recorded as the location corresponding to the highest value of *ProgOut* for each image.
2. If another position of the image has a *ProgOut* value within 5% of highest evaluation (providing the position is not within `SQUARE_SIZE/2` of the recorded position), the landmark for that image is recorded as unclassified.
3. If the image is not unclassified, then a comparison is made between the recorded position

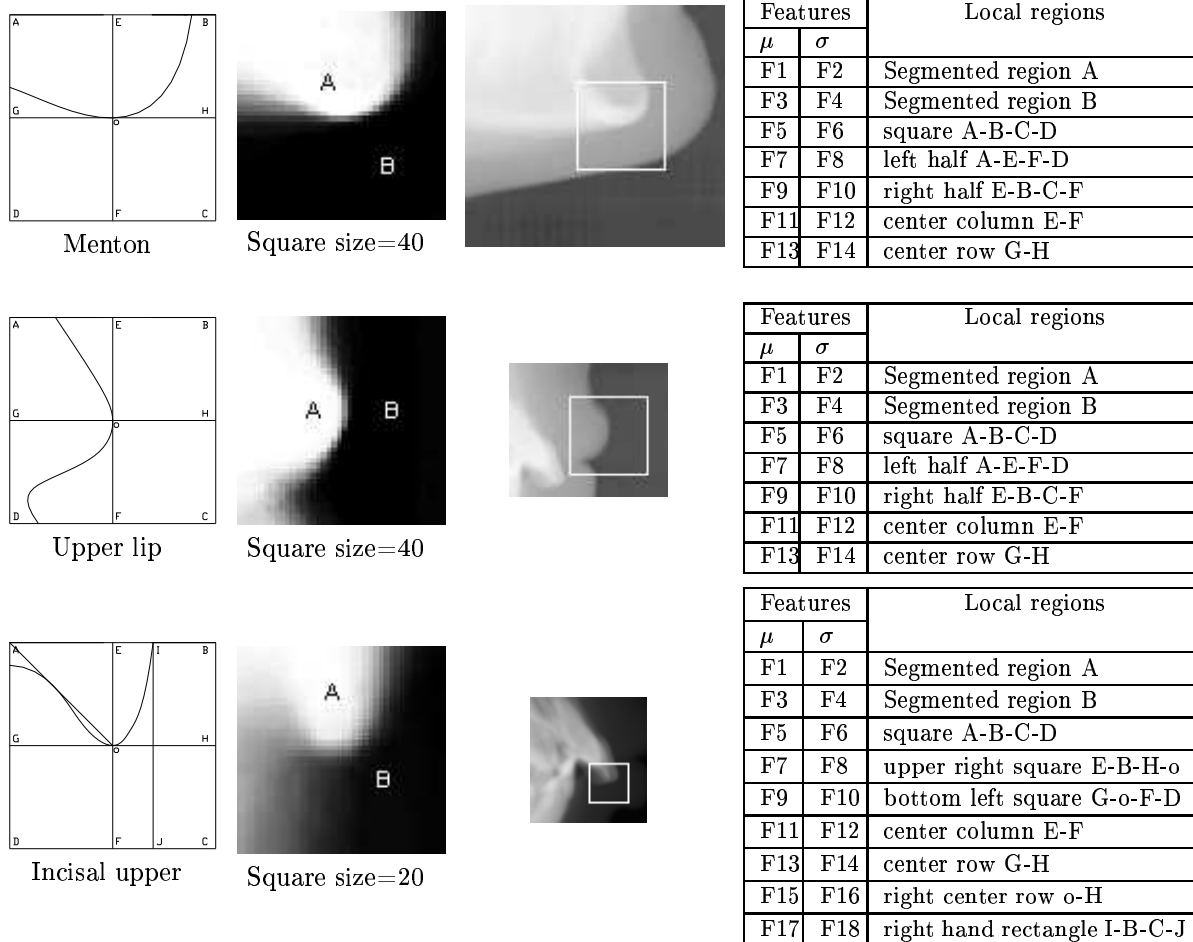


Figure 4: The diagrams in the left column show the shapes that have been manually determined. The second column shows the shapes that were constructed using the PCNN. The third column shows the size of the template (shown as the white square) relative to the sub-image. The fourth column shows the features in their terminal set.

and the known true location of the landmark. A match (true positive) occurs when the comparison is within a set TOLERANCE of 5 pixels or 2mm. If the comparison is not within the set TOLERANCE then the landmark for the respective image is recorded as a false alarm.

4. At the conclusion of evaluating the program for each image in the training set the detection rate (Dr) and false alarm rate (Fr) are calculated.
5. The fitness is computed as per equation 1.

$$fitness = A \times Fr + B(1 - Dr) \quad (1)$$

where Fr is the false alarm rate, Dr is the detection rate, and A and B are constants that reflect the relative importance of the false

alarm rate and detection rate.

The fitness function defined in equation 1 is constructed so that as better programs are evolved the fitness value approaches zero. It is also designed to evolve programs so an evaluation will be significantly higher at the predicted position of the landmark than at any other position.

2.2.4 Genetic Programming Parameters

The values used in the genetic programming training are defined in Table 1. Population size is the number of individuals in the population, Elitism (%) is the percentage of best individuals in the current population copied to the next generation, Cross rate (%) is the percentage of individuals to be produced by cross over, Mutation rate (%) is

the percentage of individuals to be created by mutation, Cross chance term (%) is the probability that in a crossover operation two terminals will be swapped, Cross chance func (%) is the probability that in crossover operation random sub-trees will be swapped, Maximum depth is the maximum depth allowed for programs, Maximum generations is the termination of the evolutionary process. A, B, Tolerance and Square size have already been defined. As shown in the table, the same parameter values have been used for all 3 points, except for square size.

Parameters	Region 1, 2, 3
Population size	100
Elitism (%)	10
Cross rate (%)	70
Mutation rate (%)	20
Cross chance term (%)	15
Cross chance func (%)	85
Maximum depth	6
Maximum generations	100
A	200
B	1000
Tolerance (pixels)	5 (2mm)
Square size (pixels)	40, 40, 20

Table 1: Parameters used during the genetic search for the three regions containing the menton, upper lip and incisal upper incisor landmark.

3 Results

To determine whether the proposed approach can be used to locate landmarks ranging from easy to hard, three landmarks were selected. The regions containing the landmarks from easy to hard contain the menton, upper lip and incisal upper incisor. Landmarks are classed as easy if the features surrounding the landmark are subject to minimal biological variation such as the menton landmark, whereas difficult landmarks such as the incisal upper incisor are subject to greater biological diversity. A landmark is classified as found if the genetic program locates the landmark position within 2mm (5pixels) of the actual position as found by an orthodontist. If the landmark position is not within 2mm of the actual position the landmark is recorded as a false alarm. When a landmark has not been found by the genetic program it is unclassified (neither found or false alarm). Training and test images are independently and randomly chosen from a database of images. Accuracy estimation was done by three fold cross-validation. For all

three points, 30 independent runs were performed for each fold and the results given below are from the fittest program of these runs. The results are summarised in Table 2.

Figures 6, 8 and 10 are examples of the three landmark points. In general, the pictures in the top row are examples that were quite straight forward and the landmarks were accurately found. The pictures in the bottom row are difficult and exhibit some ambiguity indicative of the errors found. The pictures in the middle row are of intermediate difficulty.

3.1 Region 1 (Menton)

As can be seen from Table 2 the menton point is detected very accurately. Images in Figure 6 are indicative of the type of images from the test set containing the menton landmark. The black dot corresponds to the location of the menton as found by the genetic program. It is worth noting the fittest program from each fold incorrectly found the landmark in the same image (refer to the bottom center image in Figure 6).

```
(+ 147.011 (/ (+ (* (+ f4 (- f13 (* f9 f4))) (/ f3 f13)) (* (* f14 f10) (/ 44.2753 f4))) (+ (+ (/ f11 f8) (/ (/ 44.2753 f4) f13)) (+ (/ (/ 44.2753 f4) f13) (/ f10 f11))))))
```

Figure 5: A generated program for the menton landmark.

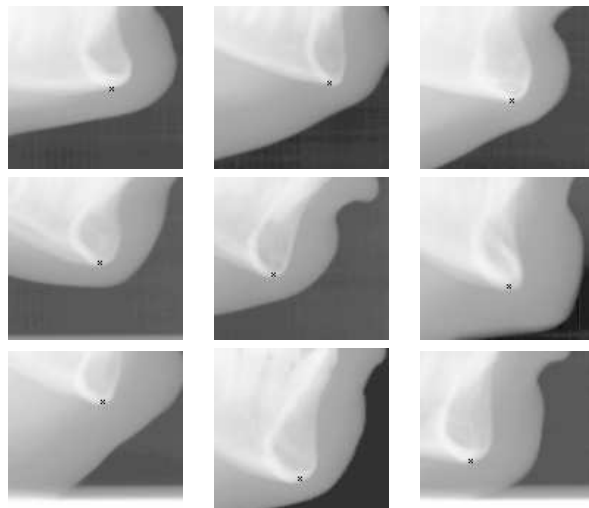


Figure 6: The 9 images are indicative of the variations in shape surrounding the menton landmark. The black point indicates the position found using the program shown in Figure 5.

	Menton		Upper lip		Incisor	
	Training	Testing	Training	Testing	Training	Testing
No. of objects	74	36	74	36	74	36
No. of objects found	73	35.7	72	34.7	71	34.3
No. of false alarms	1	0.3	2	1.3	2.3	1.7
No. unclassified	0	0	0	0	0.7	0
Detection rate (%)	98.6	99.1	97.3	96.3	95.9	95.3
False alarm rate (%)	1.4	0.9	2.7	3.7	3.2	3.7
Positional error (mm)	NA	0.8	NA	0.7	NA	0.8

Table 2: Results are for three landmarks ranging in terms of difficulty from easy (menton) to hard (incisal upper incisor). The results are based on averages from three fold cross-validation. Positional error is calculated using the average difference between the detected landmark and the position found by the orthodontist.

3.2 Region 2 (Upper lip mid)

The program in Figure 7 is a result of the evolutionary process stopping after 100 generations. Test results produced a slightly lower detection rate than the menton and nose landmark because of the ambiguity with the lower lip.

```
(* (/ (* (/ f12 f7) (+ (+ f1 129.901) f12)) (+ (/ f5
f6) (* f9 f4))) (/ (* (+ (/ (+ f12 (- (/ f12 f3) f2)) f1)
f14) (+ (* f12 f8) f1)) (* (* f9 f8) (+ f1 129.901))))
```

Figure 7: A generated program for the upper lip landmark.

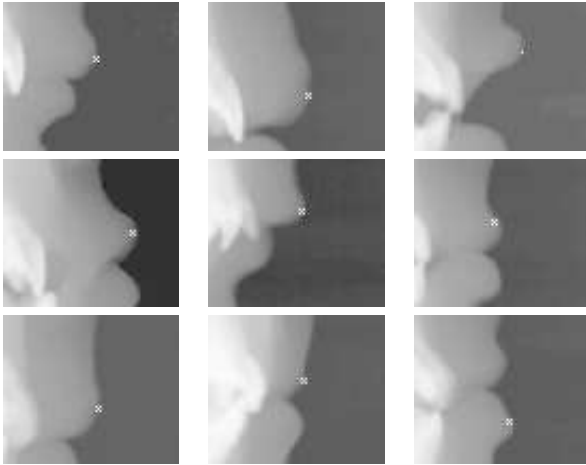


Figure 8: The 9 images are indicative of the diversity surrounding the mid upper lip. The white point in the images indicates the position found using the program in Figure 7.

3.3 Region 3 (Incisal upper incisor)

Landmark detection generally performed better on images subject to overbite as opposed to images

containing under-bite.

```
(+ (/ f15 f17) (- (+ (/ (- (* f11 (/ f5 f13)) (* (/ f3
f7) (- 234.02 f10))) f17) (- 8.98289 (+ (/ f17 (* f3
(/ 70.8492 f13))) (/ f17 (* f17 (/ 70.8492 f13))))))
(+ (/ f17 (* (* f17 (/ 70.8492 f1)) (/ 70.8492 f13)))
(/ f17 f17))))
```

Figure 9: A generated program for the incisal upper incisor.

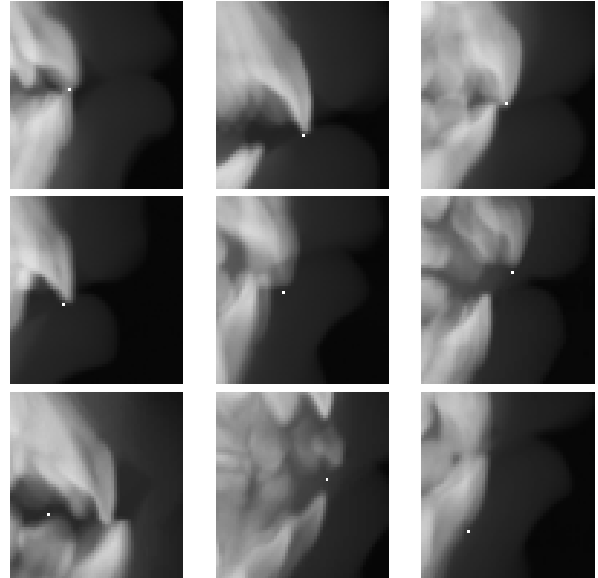


Figure 10: The 9 images are indicative of the variations in shape surrounding the incisal upper incisor. The white point in the images indicates the position found using the program in Figure 9. The top two rows are images where the program found the landmark and the bottom row are false alarms.

Table 3 shows the difference in detection performance due to the use of the PCNN features for

	Without PCNN features	
	Detection rate (%)	False alarm rate (%)
Nose	97.2	0
Lip	85.1	11.1
Incisor	88.4	11.6
	With PCNN features	
Nose	100	0
Lip	96.3	3.7
Incisor	95.3	3.7

Table 3: Detection performance using original features in [4] compared to using the additional PCNN features.

the three landmarks used in [4] (nose, lip and incisor). As can be seen from the table, the accuracy for these landmarks is significantly higher if the PCNN features are used. We expect even more improvement for landmarks with more clutter and less contrast.

4 Conclusion

The goal of the work presented in this paper was to investigate whether a PCNN analysis could be used to develop templates for detecting cephalographic landmarks and whether the use of these templates would give greater detection accuracy than the manually constructed ones. We have shown that additional features for use in the template can be constructed by applying an averaging process to the PCNN outputs and that detection accuracy on two landmarks is improved by using these new features. The images containing these two landmarks contained a considerable variety of human shapes and sizes so the accuracy achieved is not trivial. In further work we intend to explore the approach on the other landmarks and to fully automate the process of mask construction.

Acknowledgments

This work is funded by the Australian Research Council (ARC) SPIRT grant scheme in partnership with Braces Pty Ltd from grant no. C00107119 and supported by grant EPPNRM054 from the Victorian Partnership for Advanced Computing.

References

- [1] J. Cardillo and M.A. Sid-Ahmed, “An image processing system for locating craniofacial landmarks,” *IEEE transaction on Medical Imaging June 1994*, vol. 13, no. 2, pp. 275–289, 1994.
- [2] T.J. Hutton, S. Cunningham, and P. Hammond, “An evaluation of active shape models for the automatic identification of cephalometric landmarks,” *European Journal of Orthodontic*, vol. 22, no. 5, pp. 499–508, October 2000.
- [3] Y. Chen, K. Cheng, and J. Liu, “Improving cephalogram analysis through feature subimage extraction,” *IEEE engineering in Medicine and Biology*, pp. 25–31, 1999.
- [4] Andrew Innes, Vic Ciesielski, John Mamutil, Sabu John, and Alan Harvey, “Landmark detection for cephalometric radiology images using genetic programming,” in *Proceedings of the 6th Australia-Japan Joint Workshop on Intellignet and Evolutionary Systems*, Ruhul Sarker, Bob McKay, Mitsuo Gen, and Akira Namatame, Eds., Canberra, Nov. 2002, pp. 125–132, <http://www.cs.rmit.edu.au/~vc/papers/aust-jap-ec02.pdf>.
- [5] Andrew Innes, Vic Ciesielski, John Mamutil, and Sabu John, “Landmark detection for cephalometric radiology images using pulse coupled neural networks,” in *Proceedings of the 2002 International Conference on Artificial Intelligence and Workshop on Computational Models of Scientific Reasoning and Applications*, 2002, pp. 390–395, <http://www.cs.rmit.edu.au/~vc/papers/isai02.doc>.
- [6] Mengjie Zhang and Victor Ciesielski, “Genetic programming for multiple class object detection,” in *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence*, Norman Foo, Ed. Dec 1999, vol. 1747, Lecture Notes in Artificial Intelligence, pp. 180–191, Springer, Heidelberg, <http://www.cs.rmit.edu.au/~vc/papers/ai99.pdf>.