University of Wollongong
Research Online

Faculty of Informatics - Papers (Archive)

Faculty of Engineering and Information Sciences

2008

A Biologically Inspired Visual Pedestrian Detection System

F. H. Tivive
University of Wollongong, fok-hing-chi_tivive@uow.edu.au

A. Bouzerdoum
University of Wollongong, a_bouzerdoum@uow.edu.au

Publication Details

This conference paper was originally published as Tivive, FHC & Bouzerdoum, A, A Biologically Inspired Visual Pedestrian Detection System, 2008 International Joint Conference on Neural Networks (IJCNN 2008), Hong Kong, 1-6 June 2008, 703-709. Copyright Institute of Electrical and Electronics Engineers 2008. Original conference paper available here

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

A Biologically Inspired Visual Pedestrian Detection System

Abstract

In this paper, we present a biologically inspired method for detecting pedestrians in images. The method is based on a convolutional neural network architecture, which combines feature extraction and classification. The proposed network architecture is much simpler and easier to train than earlier versions. It differs from its predecessors in that the first processing layer consists of a set of pre-defined nonlinear derivative filters for computing gradient information. The subsequent processing layer has trainable shunting inhibitory feature detectors, which are used as inputs to a pattern classifier. The proposed pedestrian detection system is evaluated on the DaimlerChrysler pedestrian classification benchmark database and its performance is compared to the performance of support vector machines and Adaboost classifiers.

Disciplines

Computer Sciences | Physical Sciences and Mathematics

Publication Details

This conference paper was originally published as Tivive, FHC & Bouzerdoum, A, A Biologically Inspired Visual Pedestrian Detection System, 2008 International Joint Conference on Neural Networks (IJCNN 2008), Hong Kong, 1-6 June 2008, 703-709. Copyright Institute of Electrical and Electronics Engineers 2008.
Original conference paper available here

This conference paper is available at Research Online: http://ro.uow.edu.au/infopapers/690

A Biologically Inspired Visual Pedestrian Detection System

Fok Hing Chi Tivive, *Member, IEEE*, and Abdesselam Bouzerdoum, *Senior Member, IEEE*

***Abstract*—In this paper, we present a biologically inspired method for detecting pedestrians in images. The method is based on a convolutional neural network architecture, which combines feature extraction and classification. The proposed network architecture is much simpler and easier to train than earlier versions. It differs from its predecessors in that the first processing layer consists of a set of pre-defined nonlinear derivative filters for computing gradient information. The subsequent processing layer has trainable shunting inhibitory feature detectors, which are used as inputs to a pattern classifier. The proposed pedestrian detection system is evaluated on the DaimlerChrysler pedestrian classification benchmark database and its performance is compared to the performance of support vector machines and Adaboost classifiers.**

I. Introduction

In recent years, there has been much attention focused on the problem of pedestrian detection in real-world images due to its potential applications, for example, in video surveillance and driver-assistance systems. The vision-based pedestrian detection problem is very challenging from a machine vision perspective because pedestrians can appear in highly cluttered environments and have a wide range of appearances due to different pose, body size, clothing, and outdoor lighting conditions. The general idea of solving this type of visual pattern recognition problem is to extract salient features from the region of interest (ROI), features that can differentiate the full human figure from another entity in the scene. In a human face, facial features such as the nose, eyes and mouth form vital cues for classifying faces from non-face objects; however, for pedestrian classification more abstract features are required.

In this paper, we propose a pedestrian detector inspired by the human visual system. It comprises a hierarchical architecture that preserves the topographical mapping between the input image and the feature-extraction layers; the last layer is reserved for classification of the input window into pedestrian or non-pedestrian patterns. Here shunting neurons are used as feature detectors, and the visual information processing is segregated into two parallel pathways: the "on" and "off" channels. Furthermore, the feature extraction and classification stages are designed within the same framework and optimized simultaneously with respect to one another.

The organization of the paper is as follows. Section II presents a brief review of previous work on pedestrian detection. Section III describes the architectural concepts of two well-known convolutional neural networks (CoNNs), followed by the description of the proposed pedestrian detection system and its training method in Section IV. The training and testing procedures are described in Section V. Finally, concluding remarks are presented in Section VI

Fok Hing Chi Tivive and Abdesselam Bouzerdoum are with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, Australia (emails: {tivive; a.bouzerdoum} @uow.edu.au).

This work is supported in part by a grant from the Australian Research Council (ARC).

II. Previous Work on Pedestrian Detection

Most pedestrian detection systems have a recognition stage in which a classification algorithm is employed to discriminate a pedestrian from other objects, given a ROI. In the literature, various techniques have been reported for pedestrian classification; here, we only mention some of the popular algorithms that have been successfully applied in pedestrian detection systems (see [1], [2] for a comprehensive review on pedestrian detection). Papageorgiou and Poggio [3] developed a trainable pedestrian detection system, which uses an overcomplete dictionary of Haar wavelets for feature extraction and support vector machines (SVMs) for classification. Viola *et al.* [4], on the other hand, employed Harr-like features in conjunction with the adaptive boosting (AdaBoost) algorithm to develop a cascade of classifiers using motion and appearance information. Zhao and Thorpe [5] use stereo cameras and the disparities discontinuity algorithm to segment an image into sub-image object candidates. Then, a split-and-merge procedure is applied on these sub-images to form objects that satisfy pedestrian size and shape constrains. Finally, the gradient information of these objects is passed to a multilayer perceptrons for pedestrian classification. Another neural network-based algorithm for pedestrian detection was proposed by Wohler and Anlauf [6]. It is a time-delay neural network with spatio-temporal receptive fields to extract information directly from the temporal sequences of gray-scale image. Such an approach avoids hand-tuning the network structure parameters, as it is done during the training phase.

III. Convolutional Neural Networks

Convolutional neural networks are powerful hierarchical multilayered networks designed for visual pattern recognition problems, such as face detection [7], handwritten digit recognition [8], facial expression analysis [9], and video quality assessment [10], just to name a few. These neural networks are inspired by Hubel and Wiesel hierarchy model of the visual cortex [11], where the processing elements are structured in a hierarchical order to extract local features from the input image. The earliest model of CoNNs was proposed by Fukushima [12], called *Neocognitron*. Its network architecture consists of several stages, and each stage is made up of different layers, namely the S-layer

978-1-4244-1821-3/08/$25.00©2008 IEEE

(*simple* layer), the C-layer (*complex* layer) and the V-layer. In these layers, the neurons are arranged into planes to extract two-dimensional (2-D) information from the prior layer, and each layer has a specific task. The role of the S-layers is to extract simple features, whereas the C-layers have the responsibility of compressing the extracted features and blurring their locations. The V-layer provides a shunting mechanism, which enables the S-layer to extract not only learned features, but also distorted versions of these features. Across the Neocognitron network, the S- and C-layers are alternately arranged and the density of neurons in each layer decreases along the network. The Neocognitron, however, is a complex, massively connected network architecture that is often trained using unsupervised learning. LeCun and his colleagues [8], on the other hand, developed a convolutional neural network architecture, dubbed LeNet-5, which was trained with a supervised learning algorithm. Their network is built upon three architectural concepts: local receptive field, weight sharing, and spatial or temporal sub-sampling. Rather than having an S-layer followed by a C-layer, LeNet-5 consists of alternating convolutional and sub-sampling layers. Furthermore, the information processing units used in the convolutional layers are sigmoid-type neurons. Both LeNet-5 and Neocognitron share the following characteristics:

- Topographical mapping: the spatial topology of the input is well captured by the network structure, using the concept of local receptive fields.
- The feature extraction stage is integrated with the classification stage; the free parameters in both stages are generated by learning.
- The number of trainable parameters is reduced using the concept of weight sharing, leading to improved generalization.

Both the Neocognitron and LeNet-5 have complex network architectures with a large number of free parameters, despite the weight sharing mechanism. LeNet-5, which was initially developed for hand-written digit recognition, has over 300,000 connections and 60,000 free parameters. The Neocognitron, on the other hand, has a complex network architecture comprising several types of processing units. We have recently developed a CoNN architecture, called SICoNNet, which employs the concepts of weight sharing and receptive fields, but uses shunting inhibition for feature extraction [13]–[15]. In order to reduce the number of interconnections between layers, two partial connection schemes were developed: binary and toeplitz connections. SICoNNets have been applied successfully to a number of visual pattern recognition problems, namely face detection [13], gender recogntion [16], handwritten digit recognition [17], and texture segmentation [18]. In this paper, the SICoNNet architecture is modified so as to reduce further the number of free parameters, and the modified architecture is applied to pedestrian detection. The next section describes the modified SICoNNet architecture and its training algorithm.

IV. Proposed Pedestrian Detection System

A. Network Architecture

The proposed network model has some similarities with LeNet-5 and Necognitron in that it is also based on the three aforementioned architectural concepts of weight sharing, local receptive fields, and sub-sampling. The first layer is the input layer which is an array of input nodes of size $N \times M$. It is followed by three processing layers: two hidden layers for feature extraction and the output layer for classification. The second and third layers consist of planes of neurons (also called feature maps) whose role is to extract local features from the previous layer. The feature extraction mechanism can be described as follows. In each feature map, the neuron is connected to a local neighborhood of the previous input plane. This local neighborhood is the receptive field of the neuron and the set of weights linking the neuron to its receptive field is shared among all the neurons of the feature map, as shown in Fig. 1(a); in other words, the set of weights of the neuron acts as a filter kernel that is convolved with the previous input planes to extract local features. These features are then sent to the last layer for classification of the given input pattern. Instead of separating the feature extraction and sub-sampling operations into two successive layers as in LeNet-5, the receptive fields of adjacent neurons are shifted by two positions in both directions (vertically and horizontally). Therefore, the convolution and sub-sampling layers are merged into a single layer. To avoid a network with complex connections between layers, we let each feature map branches out onto two feature maps, similar to a binary tree. Figure 1(b) depicts the architecture of the proposed network. The only constraint is that the subsequent layer should have twice the number of feature maps as the preceding layer.

In LeNet-5 and its predecessors [7], [9], the processing elements employed in the network are sigmoid-type neurons, where the neuron performs a weighted sum of the input signals, which is added to a bias term before passing through a nonlinear activation function to generate a neural response. The Neocognitron, on the other hand, has three types of processing elements, including the V-cell that enables the network to be invariant to distortions and translations. In recent years, a new type of artificial neuron, based on the mechanism of *shunting inhibition*, has been introduced for pattern classification and regression [19], instead of the traditional sigmoid neuron. Shunting inhibitory neurons have the capability to produce more complex decision boundaries than the simple hyperplane; for example, a single shunting inhibitory neuron can solve linearly non-separable problems such as the 3-bit parity and XOR problems [20]. This type of neuron has been employed for feature extraction in the SICoNNet architecture. The response of a single static feed-forward shunting inhibitory neuron is given by

$$z = \frac{g\left(\vec{W} \cdot \vec{I} + b\right)}{a + f\left(\vec{C} \cdot \vec{I} + d\right)}, \tag{1}$$

where z is the activity of the neuron, $\vec{I}$ denotes the inputs

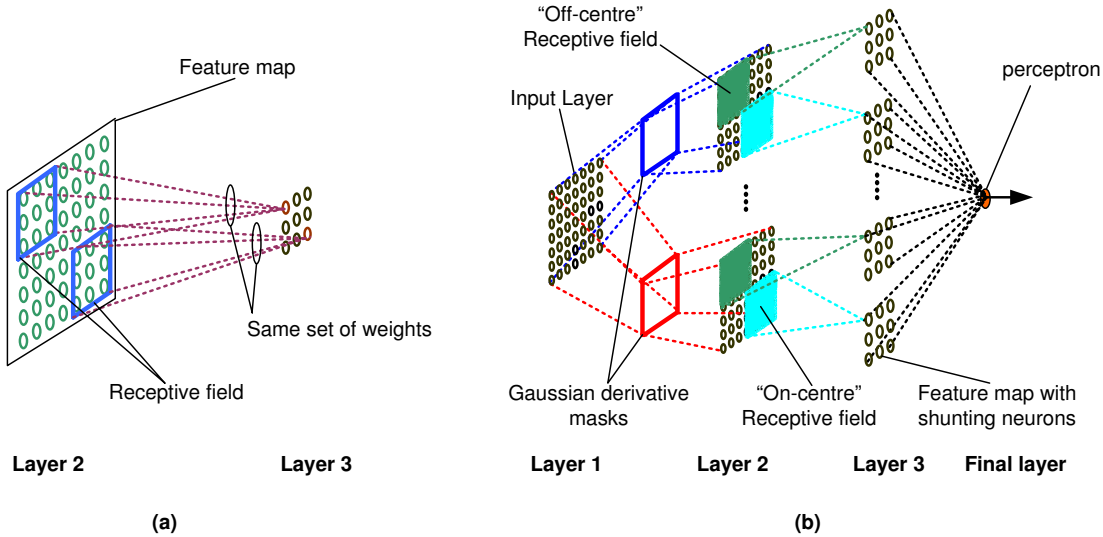


Fig. 1. Proposed network architecture: (a) the connection strategy between the neuron and its input plane, and (b) a schematic diagram of the network.

within the receptive field arranged in a vector form, a is the passive decay rate, $\vec{W}$ and $\vec{C}$ are the "excitatory" and "inhibitory" weights, respectively, b and d are constant biases, and f and g are the activation functions. One advantage the shunting neuron has over sigmoid and radial basis function neurons is that the input-output characteristic is adaptive; for example, using the logarithmic sigmoid for the activation functions f and g, and choosing randomly the weights and biases, we obtain the input-output transfer characteristics presented in Fig. 2. Clearly, this shows the input-output transfer characteristics of a shunting inhibitory neuron can be altered by varying the weights and biases only. This means that a single shunting neuron is able to generate complex decision boundaries if the right weights and biases can be found.

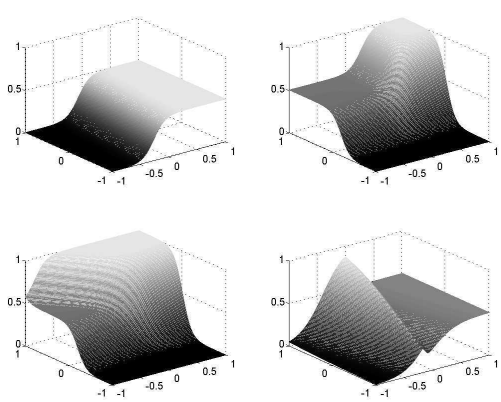


Fig. 2. The input-output transfer characteristics of a shunting inhibitory neuron.

For the final layer of the network, the processing elements can be simply linear or sigmoid neurons. More complex classifier can be employed, but the objective here is to investigate the efficiency of the network using shunting inhibitory neurons as feature detectors. The weights and biases in both hidden layers can be obtained with a training algorithm, but this may increase the complexity of the network and its training time. Therefore, we have used shunting inhibitory neurons with adjustable weights only in the second hidden-layer. In the first hidden-layer, shunting neurons are used as nonlinear bandpass filters, with predefined set of weights, to extract "coarse information" from the input layer. The information extracted by the first hidden-layer is refined by the second feature extraction layer. Bandpass filters were chosen because many approaches to pedestrian detection in the literature are based on edge information. In our approach, we use directional Gaussian derivative filters for the weight vector $\vec{W}$ in (1). These filters are obtained by convolving a Gaussian filter with 2-D derivative masks. These masks are the basic horizontal (M_H), vertical (M_V), and radial derivatives (M_R) shown below:

$$M_H = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}, \quad M_V = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$

$$M_R = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}.$$

Any directional derivative masks can be computed by combining the horizontal and vertical derivatives, i.e.,

$$M_\theta = \cos(\theta) \cdot M_H + \sin(\theta) \cdot M_V. \tag{2}$$

Overall, we have designed nine derivative masks: the radial derivative and eight directional derivative masks oriented

at 0^0, 22.5^0, 45^0, 67.5^0, 90^0, 112.5^0, 135^0, and 157.5^0 angles. While the weight vector $\vec{W}$ in (1) is replaced by the derivative of the Gaussian mask (G'), the weight $\vec{C}$ is simply a 7×7 Gaussian mask (G) with $\sigma = 1$. We should note that the same Gaussian mask is used to compute the directional derivative masks. The remaining parameters in (1), the bias and passive decay rate, are set to zero. Hence, the feature map response at the first hidden-layer is given by

$$Z_{L2} = \frac{G' * I}{G * I}. \tag{3}$$

Each of these feature maps is then connected to two feature maps in the subsequent layer, without sub-sampling. Furthermore, the "on" and "off" responses are segregated into separate channels; i.e., in the second hidden-layer half of the feature maps process positive responses, whereas the other half process the negative responses:

$$z^{+}_{L3} = \frac{g\Big(\vec{W} \cdot max(\vec{Z}_{L2}, 0) + b\Big)}{a + f\Big(\vec{C} \cdot max(\vec{Z}_{L2}, 0) + d\Big)}, \tag{4}$$

$$z^{-}_{L3} = \frac{g\Big(\vec{W} \cdot min(\vec{Z}_{L2}, 0) + b\Big)}{a + f\Big(\vec{C} \cdot min(\vec{Z}_{L2}, 0) + d\Big)}, \tag{5}$$

where the receptive field $[W]$ and $[C]$ have size of 5×5. Before passing the feature to the classification stage, the feature maps undergo a sub-sampling operation by shifting the receptive fields of adjacent neurons by two positions, horizontally and vertically. Here only one output neuron is required, and we simply use a linear neuron to perform classification. More complex classifiers could be used at the expense of simplicity and with increased computational cost.

B. Training Method

Most CoNNs, such as LeNet-5, are trained with an online algorithm, i.e., the parameters of the network are adjusted after every input pattern. This is mainly due to the complexity of the network architecture and the large number of free parameters. Although online algorithms can process a large number of input patterns relatively quickly, they are not always able to optimize the cost function. Since the proposed network has a relatively simple architecture—the only free parameters are the weights and biases of the second hidden-layer and the output neuron—the Levenberg-Marquardt algorithm [21] has been used to train the new network architecture. During the initialization phase, the network is initialized with random values taken from a uniform distribution in the range $[-1, 1]$. The passive decay rate constant, however, is constrained so as to avoid dividing by zero in (1), i.e.,

$$a \geq \varepsilon - \inf(f), \tag{6}$$

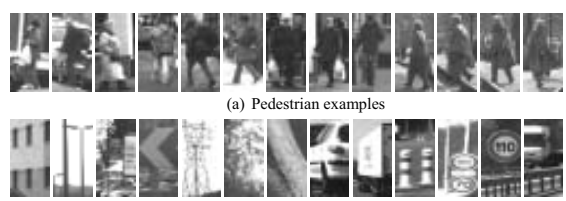
where ε is a small positive constant and $\inf(f)$ denotes the *infimum* of the activation function f—we should note that f must be bounded from below. Therefore, the hyperbolic tangent function is chosen as the activation function f, whereas the exponential function is used for g. The above constraint in (6) is enforced during the weight initialization and training phases. The target outputs for pedestrian and non-pedestrian patterns are set to 1 and -1, respectively.

V. Experimental Procedure

In this section we introduce the experimental procedure and present the experimental results. First the database employed for training and testing the pedestrian detection system is described in the next subsection, followed by the experimental results and a discussion.

A. Pedestrian Detection Database

Many researchers have used their own databases for training and testing their pedestrian detection systems; these databases often consist of only few hundreds of pedestrian images and a set of scenery images that do not contain people standing in an up-right position, e.g., the MIT pedestrian database [3] and the INRIA database [22]. Recently, the DaimlerChrysler research center has established a benchmark database for further research in pedestrian classification—the database is publicly available on the Web[1]. It contains three training sets and two test sets. Each set contains 4800 segmented pedestrian images of size 18×36, i.e., the height and width are 36 and 18 pixels, respectively. Instead of randomly collecting windows from the scenery images which can be "easy" non-pedestrian patterns when the scenery image has uniform background like the sky or pavement, the researchers from the DaimlerChrysler research center employed a bootstrapping strategy with a pedestrian detector to select negative samples for the training and test sets [23]. Some examples of the pedestrian and non-pedestrian images from DaimlerChrysler database are shown in Fig. 3.



(a) Pedestrian examples

(b) Non-pedestrian examples

Fig. 3. Examples of pedestrian and non-pedestrian images contain in the training and test sets.

B. Experimental Results and Discussion

In [24], Munder and Gavrila employed the Daimler-Chrysler pedestrian database to study several classification methods for pedestrian detection. These classifiers are the convolutional neural network developed by Wohler and Anlauf [6], SVMs, and the Adaboost classifier. Moreover, they

[1]The DaimlerChrysler pedestrian classification benchmark dataset is available at: http://www.gavrila.net

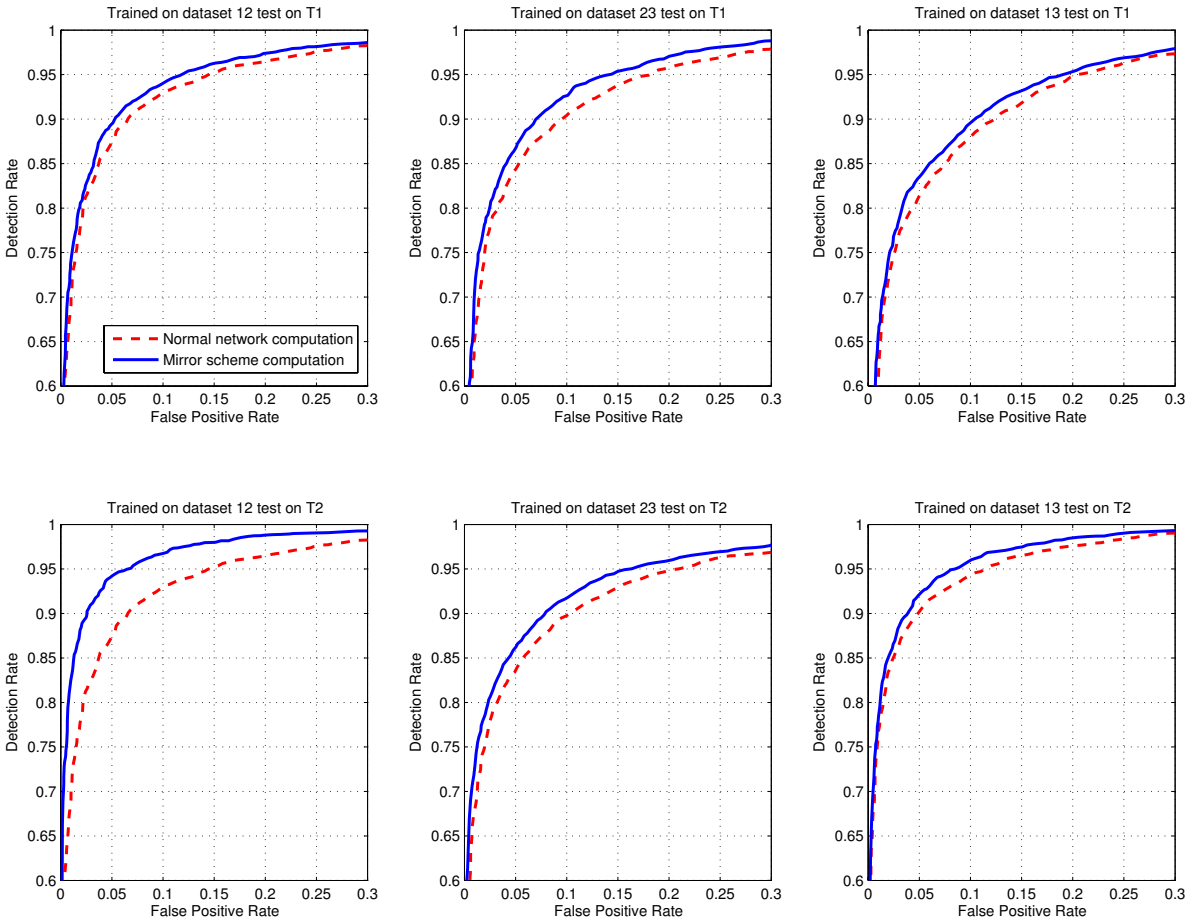


Fig. 4. Detection vs false positive rate of six classifiers trained on different combination of training sets 1, 2, and 3 and and evaluated on test sets T1 and T2. The solid ROC curves illustrate the performance of the system with the average response to the input pattern and its mirror-image, whereas the dotted curves are obtained by using only the response to the input pattern.

used principal component analysis and wavelet transform for feature extraction. To compare with these approaches, we employ the same training and testing procedure; that is, we generate three classifiers, and each classifier is trained with two out of three training sets and the remaining set is used for validation. All three classifiers are then tested on the two test sets to produce six different receiver operating characteristic (ROC) curves. Each curve represents the detection rate of the classifier versus false positive rate on a given test set. In other words, at a certain acceptable false positive rate, the detection rate of the classifier can be determined from the ROC curve. Finally, the six ROC curves are averaged to produce the final curve that is used to compare the performance of the proposed method with the results obtained in [24].

From our previous studies on human gender recognition [16], [25], we found that by taking the average response to the input pattern and its mirror-image, a slight improvement in the classification accuracy is obtained; in other words, the input pattern and its folded version along the Y-axis are sent to the network and both responses are averaged to produce the final score of the input pattern. Here, we also investigate the performance of the pedestrian detection system with and without the network response. In the DaimlerChrysler pedestrian database, the training sets are labelled as 1, 2, 3 and the test sets as T1 and T2. Therefore, each network is trained on a combination of two training sets, i.e., 12, 13 and 23, and evaluated on T1 and T2. The experimental results presented here are obtained from a single trial, i.e., each network is trained once only. The ROC curves for each network are shown in Fig. 4, where each figure has two ROC curves: the dotted curve is generated directly from the network response to the input pattern, whereas the solid ROC curve is computed from the average response to the input pattern and its mirror-image. The ROC curves in the first row of Fig. 4 are for networks evaluated on test set T1, and those in the second row are obtained from T2. The ROC curves show that the data set T1 is slightly more difficult to classify than T2, and the patterns in training set 12 are

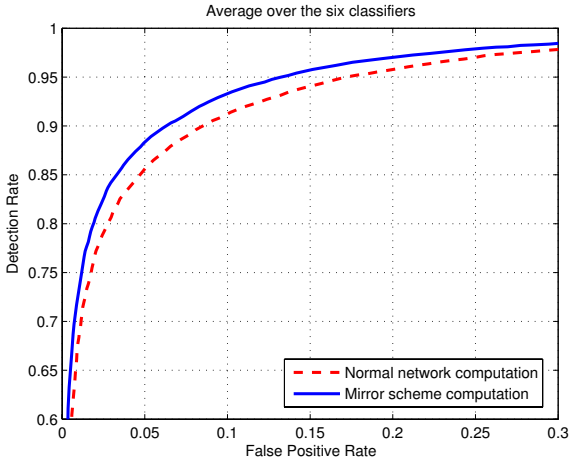


Fig. 5. ROC curves averaged over the six classifiers.

more representative than those in the training sets 23 and 13. Across all six figures, the area under the dotted ROC curve is smaller than the area under the solid ROC curve; this shows that using the input pattern and its mirror-image improves the detection performance slightly.

To benchmark the proposed method with other existing approaches, we compare the detection rate of our method with those of SVMs, CoNN and Adaboost classifier, given in [24]. In Fig. 8 of [24], the detection rate of the eight-stage cascade Adaboost classifier is around 85% at 10% false positive rate. At similar false positive rate, the SVM method achieves detection rates in the range of 75% to 90% (see Fig. 5(c) and (d) of [24]), and the detection performance of the CoNN developed by [6] is around 70%. Compared to these methods, the proposed system achieves a higher classification accuracy, with a detection rate exceeding 90% at the same false positive rate as shown in Fig. 5. These results are obtained using a single hyperplane at the classification stage, which means that the proposed system achieves efficient feature extraction using the shunting inhibitory neuron as feature detector.

VI. Conclusion

In this article, we have proposed a biologically inspired pedestrian detection system. It comprises two feature extraction layers, followed by a single unit as classifier. The feature extraction layers contain feature detectors that are based on the biophysical mechanism of shunting inhibition. In contrast to existing systems, such as CoNNs, our system has a much simpler architecture with fewer free parameters, as its first processing layer contains pre-defined nonlinear derivative filters. The experimental results demonstrate that the proposed approach has the capability to extract useful local features directly from the input image, features that can be classified with a single hyperplane. The performance of the proposed system compares favorably with systems that use wavelets for feature detection and SVMs and Adaboost for classification.

References

[1] D. M. Gavrila, "Sensor-based pedestrian protection," in *IEEE Intelligent Systems*, vol. 6, 2001, pp. 77–81.

[2] M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, and M. Porta, "Artificial vision in road vehicles," *Proc. of the IEEE*, vol. 90, no. 7, pp. 1258–1271, 2002.

[3] C. Papageorgiou and T. Poggio, "Trainable pedestrian detection," in *Proc. of the International Conference on Image Processing*, vol. 4, 1999, pp. 35–39.

[4] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Ninth IEEE International Conference on Computer Vision*, vol. 2, 2003, pp. 734–741.

[5] L. Zhao and C. Thorpe, "Stereo- and neural network-based pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 148–154, 2000.

[6] C. Wohler and J. K. Anlauf, "An adaptable time-delay neural-network algorithm for image sequence analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 6, pp. 1531–1536, 1999.

[7] C. Garcia and M. Delakis, "Convolutional face finder: a neural architecture for fast and robust face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1408–1423, 2004.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[9] B. Fasel, "Robust face analysis using convolutional neural networks," in *Proc. of the Sixteenth International Conference on Pattern Recognition*, vol. 2, Quebec, Canada, 2002, pp. 40–43.

[10] P. L. Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1316–1327, 2006.

[11] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat," *Journal of Neurophysiology*, vol. 28, pp. 229–289, 1965.

[12] K. Fukushima, "Interpolating vectors for robust pattern recognition," *Neural Networks*, vol. 20, no. 8, pp. 906–916, 2007.

[13] F. H. C. Tivive and A. Bouzerdoum, "A face detection system using shunting inhibitory convolutional neural networks," *Proc. of the International Joint Conference on Neural Networks*, vol. 4, pp. 2571 – 2575, 2004.

[14] ——, "Efficient training algorithms for a class of shunting inhibitory convolutional neural networks," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 541– 556, 2005.

[15] ——, "A brain-inspired visual pattern recognition architecture and its applications," in *Pattern Recognition Technologies and Applications: Recent Advances*, B. Verma and M. Blumenstein, Eds. IGI Global Press.

[16] ——, "A shunting inhibitory convolutional neural network for gender classification," in *Proc. of the 18th International Conference on Pattern Recognition (ICPR'06)*, 2006, pp. 421–424.

[17] ——, "Application of siconnets to handwritten digit recognition," *International Journal of Computational Intelligence and Applications*, vol. 6, no. 1, pp. 45–59, 2006.

[18] ——, "A nonlinear feature extractor for texture segmentation," in *Proc. of the 2007 IEEE International Conference on Image Processing (ICIP 2007), Vol. 2*, 2007, pp. II–37–II–40.

[19] A. Bouzerdoum, "Classification and function approximation using feed-forward shunting inhibitory artificial neural networks," in *Proc. of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, 2000, pp. 613–618.

[20] G. Arulampalam and A. Bouzerdoum, "A generalized feedforward neural network architecture for classification and regression," *Neural Networks*, vol. 16, no. 5-6, pp. 561–568, 2003.

[21] M. T. Hagan and M. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, pp. 989–993, 1994.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.

[23] D. M. Gavrila and V. Philomin, "Real-time object detection for "smart vehicles," in *Proc. of the Seventh IEEE International Conference on Computer Vision*, vol. 1, 1999, pp. 87–93.

[24] S. Munder and D. M. Gavrila, "An experimental study on pedestrian classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1863–1868, 2006.

[25] F. H. C. Tivive and A. Bouzerdoum, "A gender recognition system using shunting inhibitory convolutional neural networks," in *Proc. International Joint Conference on Neural Networks*, 2006.