<table>
<thead>
<tr>
<th>Title</th>
<th>Application of BW-ELM model on traffic sign recognition (Main Article)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Sun, Zhan-Li; Wang, Han; Lau, Wai-Shing; Steet, Gerald; Wang, Danwei</td>
</tr>
<tr>
<td>Date</td>
<td>2013</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/19911">http://hdl.handle.net/10220/19911</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2013 Elsevier B.V. This is the author created version of a work that has been peer reviewed and accepted for publication by Neurocomputing, Elsevier B.V. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<a href="http://dx.doi.org/10.1016/j.neucom.2012.11.057">http://dx.doi.org/10.1016/j.neucom.2012.11.057</a>].</td>
</tr>
</tbody>
</table>
Application of BW-ELM Model on Traffic Sign Recognition

Zhan-Li Sun\textsuperscript{a,c,*}, Han Wang\textsuperscript{b}, Wai-Shing Lau\textsuperscript{d}, Gerald Seet\textsuperscript{c} and Danwei Wang\textsuperscript{b}

\textsuperscript{a}School of Electrical Engineering and Automation, Anhui University, China
\textsuperscript{b}School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
\textsuperscript{c}School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore
\textsuperscript{d}School of Mechanical and Systems Engineering, Newcastle University, United Kingdom

Abstract

Traffic sign recognition is an important and active research topic of intelligent transport system. With a constant increasing of the training database size, not only the recognition accuracy, but also the computation complexity should be considered in designing a feasible recognition approach. In this paper, an effective and efficient algorithm based on a relatively new artificial neural network, extreme learning machine (ELM), is proposed for traffic sign recognition. In the proposed algorithm, the locally normalized histograms of the oriented gradient (HOG) descriptors, which are extracted from the traffic sign images, are used as the features and the inputs of the ELM classification model. Moreover, the ratio of feature’s between-category to within-category sums of squares (BW) is designed as a feature selection criterion to improve the recognition accuracy and to decrease the computation burden. Application on a well known database, German traffic sign recognition benchmark (GTSRB) dataset, demonstrates the feasibility and efficiency of the proposed BW-ELM model.

Key words: Traffic sign recognition, extreme learning machine, histograms of oriented gradient.

* Corresponding author, e-mail: zhlsun2006@126.com.
1 Introduction

With the development of intelligent techniques, driver assistance system (DAS) has been exploited as an important component of intelligent transportation system [1–3]. After a long travel or under a bad health condition, drivers may feel drowsy or lethargic, and can not pay attention to road conditions. Sometimes, a slight miss-concentration may cause deadly accidents. A timely warning message from the DAS will help the drives to be aware of the potential dangers and adopt necessary measures to avoid the possible accidents. Therefore, the DAS plays a very important role in preventing road accidents.

In the DAS, traffic sign detection and recognition is one major approach to acquire safety and precaution information. Under an ideal condition, the traffic signs are usually easily to be comprehensible because they are designed as simple pictograms and characters. In a real environment, it becomes a difficult task to recognize the traffic signs timely and accurately because the visibility of traffic signs may be decreased greatly by some unfavorable factors, such as terrible weather conditions (fog, rain, clouds, snow, etc.), variations of light conditions, motion blur, and so on.

So far, many algorithms have been proposed for traffic sign recognition [4–8]. In [4], a novel evolutionary version of Adaboost is proposed for sign detection, and a battery of classifiers are trained to split classes in an error-correcting output code framework. Constructed by SimBoost or a fuzzy regression tree framework, a robust sign similarity measure is proposed in [9] for road sign recognition. Support vector machines (SVM) [10] is a popular classifier and has been applied in many fields [11, 12]. An automatic road-sign detection and recognition system is presented in [13], in which a SVM is used for traffic sign detection while a Gaussian kernel SVM is adopted for traffic sign recognition. An eigen-based traffic sign recognition is proposed in [14] by using principal component analysis (PCA) algorithm [15–17] to choose the most effective components of traffic sign images to classify an unknown traffic sign. Boosted by the successful applications on handwritten digits recognition, convolutional neural network (CNN) has also been employed on traffic sign classification [5, 18]. In [5], instead of various features, a CNN is trained directly with the raw pixel values of traffic sign images. Moreover, a better result is obtained by integrating the results obtained by a CNN and a multilayer perceptrons (MLP).

Gradient orientation is one kind of useful information in various object recognition, including the traffic signs. In [19], a novel local feature representation, the so-called histogram of oriented gradients (HOG), is initially proposed for pedestrian detection. Subsequently, HOG is adapted to traffic sign detection or traffic sign recognition in several works. Just as expected, the HOG de-
scriptors achieved relatively good performances in both traffic sign detection and traffic sign recognition, because of its fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in its overlapping descriptor blocks. MLP, a popular feedforward artificial neural network (ANN) [20], is adopted in [5] to classify the HOG features extracted from traffic sign images. In [21], the classification performance of K-d trees and random forests are evaluated for traffic signs with different sizes of HOG descriptors and distance transforms.

Recently, a relatively novel learning algorithm for single-hidden layer feedforward neural networks (SLFNs), called extreme learning machine (ELM), has been proposed [22,23] and widely applied in various fields [24,25]. In ELM, the input weights and hidden biases are randomly chosen, and the output weights are analytically determined by using Moore-Penrose generalized inverse. ELM not only learns much faster with a higher generalization performance than the traditional gradient-based learning algorithms but it also avoids many difficulties that are faced by gradient-based learning methods, such as stopping criteria, learning rate, learning epochs, local minima, and overtuning issues.

In this paper, a HOG-based ELM classification scheme is proposed for traffic sign recognition. In the proposed method, the HOG descriptors extracted from traffic sign images are used as the features. Moreover, the ratio of feature's between-category to within-category sums of squares (BW) is designed as a feature selection criterion to improve the recognition accuracy and to decrease the computation burden. Since ELM has a competitive classification performance to most popular classifiers, and the strategy of BW can effectively improve the recognition accuracy, the proposed BW-ELM model has a comparable performance to the state-of-the-art traffic sign recognition algorithms. Moreover, as ELM has a far less computation burden and only one parameter to be adjusted, the proposed BW-ELM model has a less computation complexity compared to the existing methods.

The remainder of the paper is organized as follows. The proposed method is presented in Section 2. Experimental results and related discussions are given in Section 3. Finally, conclusions are made in Section 4.

2 The BW-ELM Model

2.1 Feature Extraction of HOG

Fig. 1 shows the flowchart of the feature extraction of HOG. In this figure, the rectangle represents the input or output data while the ellipse represents the
operation. Before the feature extraction, all color images were scaled to a size of 40 \times 40 \text{ pixel} and converted to grayscale images. The image of a traffic sign is first divided into overlapping blocks. Then, each block is divided into non-overlapping cells. For each pixel of every cell, the gradients are computed by using Gaussian smoothing followed by a simple 1-D mask [-1, 0, 1]. A histogram of the gradient orientations of each cell is formed, and then weighted by the gradient magnitude. Finally, the histograms of the cells are concatenated to constitute a final descriptor of this block. The above parameter setting of HOG provided by [19] has been verified to be effective for traffic sign recognition [5].

2.2 Feature Selection with BW

The strategy of BW is to select the features with large between-category distances and small within-category distances. Assume that there are \( N \) samples \( \mathbf{x}_i, i = 1, \cdots, N \), where \( \mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{im})^T \in \mathbb{R}^n \). For the \( j \)th feature, the

![Figure 1. Flowchart of the feature extraction of HOG.](image-url)
mean of the samples can be computed as

\[ \mu_j = \frac{1}{N} \sum_{i=1}^{N} x_{ij}. \]  

(1)

The mean of the kth class samples can be obtained by

\[ \mu_{kj} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ij}, \quad x_i \in X_k, \]  

(2)

where \( X_k \) denotes a sample set of the kth class, \( n_k \) is the number of the samples in \( X_k \). With the obtained \( \mu_{kj} \), the sum of the distances within the kth class can be computed by

\[ s_{kj}^w = \sum_{i=1}^{n_k} (x_{ij} - \mu_{kj})^2, \quad x_i \in X_k. \]  

(3)

Given \( \mu_j \) and \( \mu_{kj} \), the mean distance between the kth class and all classes can be calculated as

\[ s_{kj}^b = (\mu_{kj} - \mu_j)^2. \]  

(4)

The ratio of feature’s between-category to within-category sums of squares can be given by

\[ \lambda_j = \frac{\sum_{k=1}^{m} s_{kj}^b}{\sum_{k=1}^{m} s_{kj}^w}, j = 1, \ldots, n. \]  

(5)

Finally, the features with large \( \lambda_j \) values are selected for classification.

2.3 Classification Using the ELM Model

Given \( N \) distinct samples \((x_i, t_i)\), where \((x_i = (x_{i1}, x_{i2}, \ldots, x_{in})^T \in R^n, t_i = (t_{i1}, t_{i2}, \ldots, t_{im})^T \in R^m)\), the standard SLFNs with \( \tilde{N} \) hidden nodes and activation function \( g(x) \) are mathematically modeled as

\[ \sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, j = 1, \ldots, N, \]  

(6)

where \( w_i = (w_{i1}, \ldots, w_{in})^T \) is the weight vector connecting the ith hidden node and the input nodes, \( \beta_i = (\beta_{i1}, \ldots, \beta_{im})^T \), is the weight vector connecting the ith hidden node and the output nodes, and \( b_i \) is the threshold of the ith hidden node. \( w_i \cdot x_j \) denotes the inner product of \( w_i \) and \( x_j \).

The SLFNs can approximate these \( N \) samples with zero error, i.e.,

\[ \|o_j - t_j\| = 0, \]  

(7)
The above $N$ equations can be written compactly as

$$H\beta = T,$$

where the hidden layer output matrix

$$H(w_1, \ldots, w_N, b_1, \ldots, b_N, x_1, \ldots, x_N)$$

$$= \begin{bmatrix}
g(w_1 \cdot x_1 + b_1) & \cdots & g(w_N \cdot x_1 + b_N) \\
\vdots & \ddots & \vdots \\
g(w_1 \cdot x_N + b_1) & \cdots & g(w_N \cdot x_N + b_N)
\end{bmatrix}_{N \times \tilde{N}}$$

is called the hidden layer output matrix of the neural network,

$$\beta = \begin{bmatrix}
\beta_1^T \\
\vdots \\
\beta_{\tilde{N}}^T
\end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad T = \begin{bmatrix}
t_1^T \\
\vdots \\
t_N^T
\end{bmatrix}_{N \times m}.$$
The activation function used in this paper is the sigmoidal function:

\[ g(x) = \frac{1}{1 + e^{-x}}. \]  

which has been demonstrated to be an efficient activation function of ELM. Finally, the smallest norm least squares solution of the linear system (9) can be given by

\[ \beta = H^\dagger T \]  

Where \( H^\dagger \) denotes the Moore-Penrose generalized inverse operation of \( H \). For the training data of the traffic sign dataset, one category label is provided for each sample. However, we can notice that the target in the ELM model is a target vector. Therefore, we should transform the category label into a target vector at first when we apply the ELM model on traffic sign classification. In the ELM model, the number of output neurons is set as the number of categories. Assume that the category number of the traffic sign dataset is \( m \), and the \( i \)th training sample \( x_i \) belongs to \( k \)th class, then the target vector \( t_i = (t_{i1}, t_{i2}, \ldots, t_{im})^T \) is a binary vector, in which the \( k \)th element is 1 while other elements are set as -1.

Another problem should be addressed is how to select the number of hidden neurons. Among various parameter selection methods, cross validation is one popular approach but suffers a heavy computation burden. For the traffic sign, the dataset size is usually very larger, and the problem of the computation burden becomes more serious. Therefore, cross validation is not a reasonable choice. In this paper, the number of hidden neurons is experimentally determined for the proposed BW-ELM model.

After training, we can get the input weights \( w_i \), output weights \( \beta \), and the threshold \( b_i \). Given the test data, we can compute the corresponding hidden layer output matrix \( H \), and the target vector \( Y \):

\[ Y = H\beta. \]  

Further, the label of the test samples can be obtained by a maximum operation. Each column of \( Y \) corresponds to a target vector of one sample. The category of the sample is derived from the location of the maximum element. For example, if the \( k \)th element of the target vector is the maximum, the sample belongs to \( k \)th class.

The specific steps of the BW-ELM classification model can be summarized as follows:

**Step 1**: Extract the HOG feature of the training images and the test images;

**Step 2**: For the training data, compute the rates \( \lambda_j \) of all features; sort the features according to the rates in a decreasing order; select the features with a given percentage.
**Step 3**: According to the indices of the selected features, extract the new training data from the original training data, and the new test data from the original test data. After extraction, only the selected features are included in the new data.

**Step 4**: For the training data, initiate the input weights $w_i$ and the threshold $b_i$; compute the output matrix of the hidden layer $H$ according to Eq. (10), and then compute the output weights $\beta$ in terms of Eq. (13).

**Step 5**: Given $w_i$ and $b_i$, compute $H$ of the test data according to Eq. (10), and then compute $\beta$ in terms of Eq. (13).

**Step 6**: Compute $Y$ according to Eq. (14), and then determine the labels of the test samples via the maximum operation.

3 Experimental Results

3.1 Datasets

We evaluate the performance of the proposed BW-ELM model on the well known German traffic sign recognition benchmark (GTSRB) dataset [27,28]. The dataset contains more than 50,000 traffic sign images of 43 classes with unbalanced class frequencies. The sizes of these images vary between $15 \times 15$ and $250 \times 250$ pixels. For these images, each sample only contains one traffic sign. Different samples of one class reflect the strong variations in visual appearance of signs due to distance, illumination, weather conditions, partial occlusions, and rotations. The dataset is provided in the form of two sets: training set and test set. The training set consists of 39209 images, and the test set contains 12630 images. All simulations were conducted using MATLAB, running on an ordinary personal computer.

3.2 Experimental results

The HOG descriptors are first extracted from the traffic sign images of the GTSRB dataset in terms of the procedure described in Section 2.1, and used as the inputs of the BW-ELM classification model. For each traffic sign image, the length of the descriptor, i.e. the dimension of the feature, is 1568. Then, the features are sorted according to the BW criterion depicted in Section 2.2. After sorting, how to determine the number of features to be selected is still a difficult problem. In theory, the number of the selected features can vary from 1 to the number of all features, i.e., $n$. Since $n$ is relatively large, it is impossible to verify every possible value in the interval $[1, n]$. To simplify the

\[ \text{http://benchmark.ini.rub.de/} \]
computation, in experiments, we first set the percentages of features to be selected as 5%, 10%, 15%, ···, 100%. Then, the cross validations are applied on the training data for each percentage. After comparisons, we found that a good result can be achieved when the percentage of features is set as 15%.

![Figure 3](image1.png)

Figure 3. The classification rates of the GTSRB dataset for the BW-ELM model when different numbers of hidden neurons are set.

The number of hidden neurons ($\tilde{N}$) is another important parameter of the BW-ELM model. We can see from Eq. (9) that the SLFNs can approximate these $N$ samples with zero error when $\tilde{N}$ is equal to $N$. As the number of training samples ($N = 39209$) is very large, we should set the value of $\tilde{N}$ as large as possible to have a small estimation error. Based on the above consideration, $\tilde{N}$ is increased gradually. Unfortunately, the computation is out of the computer memory when $\tilde{N}$ is larger than 9000. Therefore, $\tilde{N}$ is set as 9000 in the experiments. Fig. 3 shows the classification rates of the GTSRB dataset for the BW-ELM model when different numbers of hidden neurons are set. We can see that the classification rates generally increase when $\tilde{N}$ increases constantly. The highest recognition rate is 97.19%, which is obtained when $\tilde{N}$ is set as 9000.

![Figure 4](image2.png)

Figure 4. The computation times of the GTSRB dataset for the BW-ELM model when different numbers of hidden neurons are set.
Fig. 4 shows the computation times of the BW-ELM model when different numbers of hidden neurons are set. We can see that the computation times generally increase with the increasing of \( \tilde{N} \). Specially, the computation times increase sharply when \( \tilde{N} \) is larger than 6000.

### 3.3 Comparisons to Recently Reported Results

Table 1 shows the classification rates of the GTSRB dataset obtained by using the committee of CNNs [5], the human performance \(^3\), the multi-scale CNNs method [18], the random forests method [21], the LDA method based on three HOG features [27, 28], and the proposed BW-ELM model. We can see from Table 1 that the BW-ELM model has a comparable classification performance to the state-of-the-art methods.

As the implementations are not available, we can only present here some qualitative analyses and comparisons for the computation complexity of the various traffic sign recognition algorithms. We can learn from Section 1 that ANN is a popular classification tool in traffic sign recognition, such as MLP, CNN. Besides ANN, SVM is another widely used classifier in traffic sign recognition. In [22, 23], a series of experiments on artificial and real benchmark data indicated that, for the problems of regression and classification, the ELM algorithm has a far less computation time than ANN and SVM, because of its simple network structure and its simplified optimization scheme.

---

\(^2\) [http://benchmark.ini.rub.de/?section=gtsrb&subsection=results/](http://benchmark.ini.rub.de/?section=gtsrb&subsection=results/)

\(^3\) [http://benchmark.ini.rub.de/](http://benchmark.ini.rub.de/)
From Table 1, we can see that two kinds of CNN methods, the committee of CNNs and the multi-scale CNNs method, achieve a very good classification performance on traffic sign recognition. However, for the CNN, one problem is that there are so many parameters to be tuned, such as the numbers of the subsampling layers, the convolutional layers and the fully-connected layers, the number of the convolution kernels and their sizes, the subsampling rates, and so on. As we know, a heavy computation burden will be suffered if such a larger number of parameters are tuned during the training procedure. Moreover, it is difficult to find the optimal values if too many parameters are searched at the same time in the parameter selection methods, e.g., cross validation, evolutionary algorithm, etc.

In general, both the BW-ELM model and the recently reported methods have their own advantages and disadvantages. Considering the accuracy and the model complexity, the proposed BW-ELM model is competitive to the recently reported methods.

### 3.4 Discussions

![Figure 5](image.png)

Figure 5. Comparisons of the classification rates of the GTSRB dataset for the ELM method and the BW-ELM method when different numbers of hidden neurons are set.

In the proposed BW-ELM model, BW is designed as a feature selection criterion to improve the recognition accuracy and to decrease the computation burden. As a comparison, we also perform the traffic sign recognition only with the ELM algorithm. Fig. 5 shows the comparisons of the classification rates of the GTSRB dataset for the ELM method and the BW-ELM method when different numbers of hidden neurons are set. We can see that the classification rates of the BW-ELM method are obviously higher than those of the ELM method. Therefore, we can conclude that the feature selection strategy can significantly improve the recognition accuracy. On the other hand, Fig. 6 shows the comparisons of the computation times of the GTSRB dataset for
the ELM method and the BW-ELM method when different numbers of hidden neurons are set. We can see that the computation times of the BW-ELM model are less than those of the ELM algorithm. The results indicate that the computation times are reduced to some extent when the BW method is utilized in the classification model. As the computation burden is mainly caused by the dataset size, i.e., the number of the training samples, the reduction of the computation burden brought by the feature selection is not so obvious. In general, the strategy of BW not only can improve the recognition performance but also can reduce the computation burden.

Figure 7. Classification rates of ten trials when each fold of training set is used as the input of the BW-ELM model in turn.

We can learn from Section 3.2 that, a limitation for the BW-ELM model is, the computation is out of the computer memory when the number of hidden neurons is too large. For the intelligent computation techniques, such as CNN, SVM, ELM, etc., the training data set should contain enough samples to achieve a good performance. However, a high memory size is generally required for these techniques when the training data size is large. When the
computation is out of the computer memory, a feasible approach is to use the technique of classifier committee learning (CCL). In order to investigate the classification performance of the CCL, the training set of the GTSRB dataset is divided into 10 folds. Each fold of training set is then used as the input of the BW-ELM model in turn. Ten sets of predicted labels of the test samples are obtained after ten trials. Finally, the labels of the test samples are determined by the majority-vote rule. Fig. 7 shows the classification rates of ten trials when each fold of training set is used as the input of the BW-ELM model in turn. The classification rate of the CCL method is 0.9633. Therefore, the classification accuracy of the CCL method is not as high as a direct application of the BW-ELM model. We think that there may be two reasons for this. One the one hand, the classification rates of ten weak classifiers are not so high, as shown in Fig. 7. On the other hand, after checking, we found that the predicted labels of ten weak classifiers are not diverse enough. As we know, the diversity of the weak classifiers’ outputs is a necessary condition to achieve a good classification performance for the CCL method. Similar results are also observed when the training data set are divided into more or less folds.

4 Conclusions

In this paper, an effective and efficient algorithm based on BW and ELM is proposed for traffic sign recognition. The strategy of BW is verified that it not only can improve the recognition performance but also can reduce the computation burden to some extent. Application on the GTSRB dataset demonstrated that the proposed BW-ELM model has a comparable classification performance to the state-of-the-art methods. Further, some qualitative analysis and comparisons indicated that the BW-ELM model has a less computation complexity than the popular traffic sign recognition methods.

References


**Zhan-Li Sun** received the Ph.D. degree from the University of Science and Technology of China, in 2005. Since 2006, he has worked with The Hong Kong Polytechnic University, Nanyang Technological University, and National University of Singapore. He is currently a Professor with School of Electrical Engineering and Automation, Anhui University, China. His research interests include machine learning, and image and signal processing.

**Han Wang** was graduated from Northeast Heavy Machinery Institute (Yanshan University) in 1982, and obtained his PhD from the Leeds University in 1989. He has been working as a teacher in Shanghai, research associate in Oxford, lecturer, senior lecturer and associate professor in Nanyang Technological University since 1992. His research interests include 3D computer vision, recognition and robot navigation. He has published over 200 papers in international conferences and journals. He is a senior member of IEEE, and currently serves as the chairman for the IEEE, RAS, Singapore chapter.

**Lau-Wai Shing** is now a Senior Lecturer in the School of Mechanical and Systems Engineering, Newcastle University, United Kingdom. Before joining University of Newcastle (UK), He was with the School of Mechanical and Aerospace Engineering, Nanyang Technology University, Singapore. His research interests include mobile robotics, underwater robotics vehicles, and mechatronics systems interfacing.

**Gerald Seet** received his Ph.D. degree from University of Aston in 1985. He is currently an Associate Professor with the School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore. He holds a concurrent appointment as Director of the Robotics Research Center. His research interests include mechatronics and field robotics, underwater mobile robotics, fluid power systems, Unmanned Aerial Vehicles (UAV) and collaborative robotic systems.

**Danwei Wang** received his Ph.D and MSE degrees from the University of Michigan, Ann Arbor in 1989 and 1984, respectively. He received his B.E degree from the South China University of Technology, China in 1982. Since 1989, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Currently, he is a professor in the Division of Control and Instrumentation. He has served as general chairman, technical chairman and various positions in international conferences, such as International Conference on Control, Automa-
tion, Robotics and Vision (ICARCVs) and IROS conferences. He is an associate editor for the International Journal of Humanoid Robotics and served as an associate editor of Conference Editorial Board, IEEE Control Systems Society from 1998 to 2005. He was a recipient of Alexander von Humboldt fellowship, Germany. His research interests include robotics, control theory and applications. He has published widely in technical areas of iterative learning control, repetitive control, robust control and adaptive control systems, manipulator/mobile robot dynamics, path planning, and control, as well as model-based fault diagnosis and satellite formation flying. (Personal home page: http://www.ntu.edu.sg/home/edwwang)