SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

AIVP ADVANCES IN
IMAGE AND VIDEO PROCESSING

VOLUME 6, ISSUE 3
ISSN 2054-7412

Traffic Sign Detection and Recognition for Driving Assistance System

[1]Huei-Yung Lina, [2]Chin-Chen Chang, [3]Shu-Chun Huang
[1,3]Department of Electrical Engineering and Advanced Institute of Manufacturing with High-tech Innovations, National Chung Cheng University, Chiayi 621, Taiwan;
[2]Department of Computer Science and Information Engineering, National United University, Miaoli 360, Taiwan;
ccchang@nuu.edu.tw

ABSTRACT

In this paper, we present a traffic sign detection and recognition for a driving assistance system. The proposed approach consists of two subsystems for detection and recognition. First, the road sign detection subsystem adopts the color information to filter out most of irrelevant image regions. Image segmentation and hierarchical grouping are then used to select candidate regions of road signs. For the road sign recognition subsystem, a convolutional neural network (CNN) is adopted to classify traffic signs for candidate regions. Experimental results show that our approach can obtain the desired results effectively.

Keywords: Traffic Sign Detection; Traffic Sign Recognition; Color Feature; Neural Network.

1 Introduction

Because of different types of sensing and positioning technologies, driving assistance becomes a popular research topic. For unmanned vehicles and driving assistance systems, the safety problem is always the highest priority compared with the convenience or practicality for a project or system designer.

When driving a vehicle, a driver can get different messages according to local road signs. Traffic signs are often designed with eye-catching colors and easy-to-understand symbols. However, if a driver drives in a complex environment or a driver mental state is not well, this might cause the driver to overlook messages from traffic signs. If there is an automatic detection and recognition system for traffic signs [2, 7, 12, 13, 14, 19], it can report correct traffic signs quickly to the driver and also reduce the burden of the driver. When the driver ignores a traffic sign, the system can give a timely warning. If this system is used in an unmanned vehicle, it can help the automatic driving system to judge the road condition. Hence, the safety of the vehicle driving is greatly improved and the risk of accidents can be reduced.

In this paper, we focus on the detection and identification of traffic signs for driving assistance systems. In addition to testing the system performance, we also experiment with a combination of different system architectures. We first filter out most of the nonsign parts of an image using the color information. Then we extract regions where may have image blocks, and further extract candidate regions from the above image blocks. Finally, we use a convolutional neural network (CNN) to verify the candidate areas of non-road signs and identify the type of the traffic sign.

DOI: 10.14738/aivp.63.4603
Publication Date: 29th May, 2018
URL: http://dx.doi.org/10.14738/aivp.63.4603

This paper is structured as follows: Section 2 reviews related works. Section 3 describes our system structure. Section 4 describes the experimental results. Finally, Section 5 presents the conclusions.

2 Related Works

scalera et al. [7] proposed an approach for detecting traffic signs using a hue- saturation-intensity (HIS) color space. They first converted the color space of an image to an HSI color space, and then found the obvious red according to the range of hue and saturation. Similar approaches based on a hue-saturation-value (HSV) color space have also been applied [13, 14, 19]. Shaposhnikov et al. [21] and Vitabile et al. [22] used an HSI color space to define the red area outside the mark and made further retrieval. Miura et al. [18] first converted the color space of an image to a YUV color space, and to select the red and blue range. In terms of geometric and gradient characteristics, Broggi et al. [5] used pre-defined templates to align normalized color modules. If its similarity is greater than a threshold, the shape can be determined.

Escalera et al. [8] used a known angle mask and matched the pre-set search range to detect triangular traffic signs of the three vertices or circular traffic signs on its circumference after capturing a specific color. The slope with the absolute value 1 of the four specific circle points is used then determine the shape and range. Belaroussi and Tarel [3, 4] abandoned the use of color information and focused on the shape features for sign detection. Bahlmann et al. [2] used Haar features to deal with color information. They used AdaBoost and Bayesian classification training to achieve the purpose of road sign detection and identification. Kuo and Lin [16] adopted a Hough transform and corner detection and other methods to detect traffic sign locations, and then used an RBF neural network with a K-d tree to identify traffic signs.

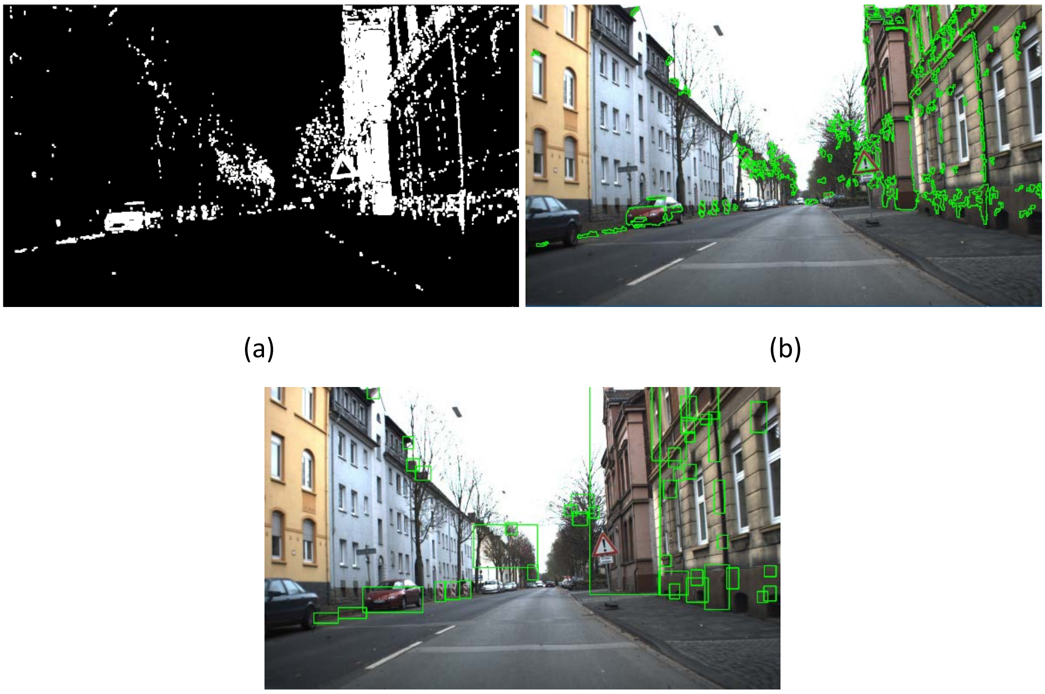
Greenhalgh and Mirmehdi [11] first converted the color space of an image to the HSI color space, and then used maximally stable extremal regions (MSER) to filter circle traffic sign locations. They used the histogram of oriented gradient (HOG) features and the SVM classifier to perform the traffic sign identification. Maldonado-Bascon et al. [17] first screened the images in an RGB color space, selected the red area with an aspect ratio to perform the restriction, and bound selected traffic signs of the candidate area. Then the distance to borders (DtB) features are extracted from the region, and the support vector machine (SVM) classifier was used to train and classify the traffic signs. Fang et al. [9] used a neural network as a basis and the image color and shape as features, and input into the two types of neural networks in order to achieve effect detection. The Kalman filter was then used to predict the possible position of the next frame under the traffic sign. There are several approaches using a deep learning to identify road signs. Some works [6, 20] used a CNN relying on the network to iterate the appropriate weight, and designed the traffic sign identification system.

3 The Proposed Approach

In our system, we first use a color filter and selective search as a detection subsystem. A large number of candidate areas will be detected and then input into the CNN for the final screening and identification.

3.1 Color Information

In our system, we choose HSI as our bases for color judgement. The reason is that this color space only uses one channel to represent the color interval. It has a range of values between 0 and 360 degrees. Because of the above properties, HSI is subject to minimal changes in light and shadow. The process is shown in Fig. 1.



(a) (b)

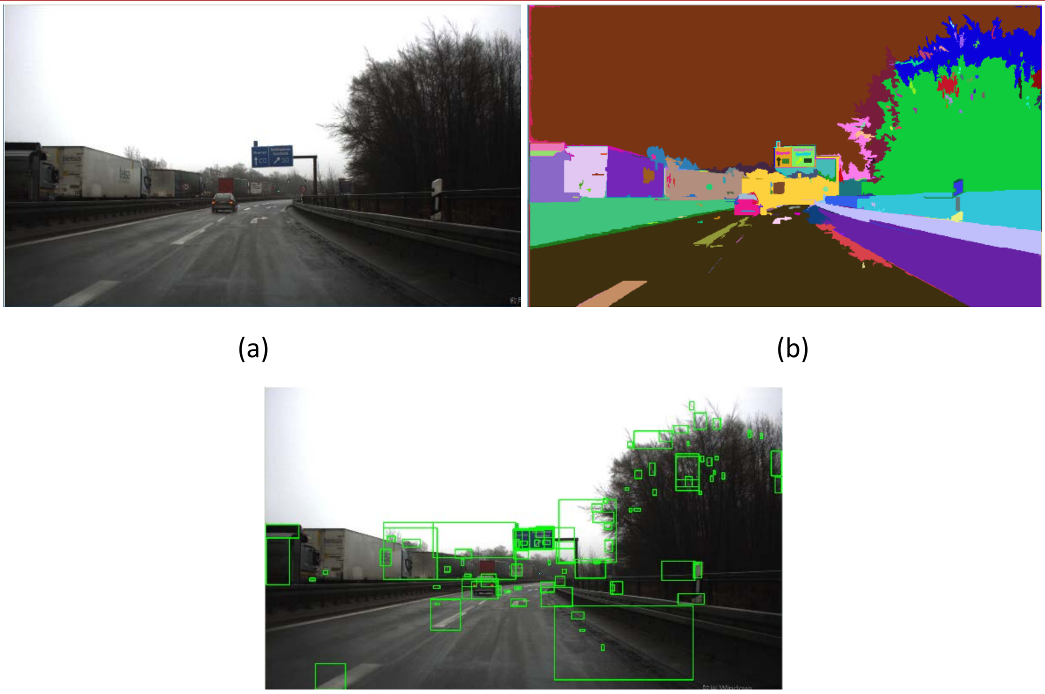
(c)

Figure. 1. Color selection process. (a) The binary image obtained according to the filter; (b) Select the object outline box based on the binary image; (c) Save a rectangular image according to the object contour

3.2 Selective Search

In general, a way to find objects from an image can be divided into two steps. First, select possible locations from an image. Second, extract features from the candidate area and recognize it. For the first step, we use selective search to identify possible regions. The calculation process of selective search is first to segment an image into a large number of super pixels as initial split areas. Hierarchical grouping is then used to combine the initial division areas. The merged large area is the candidate which will be used for identification in the later stage.

The existing algorithms usually perform the exhausted search, but it can only identify the subjects in the region within its kernel. To prevent the loss of any targets, the exhausted search uses a simple but violent way to solve this problem. It uses a mask to scan the entire area of the image. Because of the uncertainty of the target object size, the exhausted search needs to repeat search for all images with different kernel sizes. The computation complexity of the exhausted search is large. Therefore, its major drawback is the long processing time. The selective search has three advantages. First, the selective search can identify different sizes of objects with the strategies of image segmentation and hierarchical grouping. Second, the basis to distinguish includes color, texture, size and region similarity. Based on the parameters and weights, it can be implemented for most situations. Third, the processing speed is higher compared to the exhausted search. As a result, the selective search is able to produce a lot of candidate regions with high speed. The process is shown in Fig. 2.



(a) (b)

(c)

Figure. 2. The process diagram of the selective search. (a) The source image; (b) Image segmentation; (c) The results of the selective search

In addition, because the target objects have different characteristics, according to different situations, the selective search has some strategies to implement. Here, three different strategies are listed as follows: 1) Since the selective search algorithm requires hierarchical merging for the initial regions, it is important to use different initial algorithms according to the situations. 2) Use different color spaces to extract different color attributes. 3) In the case of region consolidation, we can change the estimation of region similarity with different situations.

This paper adopts a graph-based image segmentation approach [10]. The algorithm uses image pixel as unit, according to the dissimilarity between the pixels, to determine whether two pixels are in the same region. The pixels are compared to the surrounding area in eight directions, and the similarity in eight directions is arranged from small to large, e_1, e_2, ..., e_N. The computation for its dissimilarity is as follows:

$$\sqrt{(r_1-r_2)^2+(g_1-g_2)^2+(b_1-b_2)^2}\ ,$$

where r_i, g_i, and b_i are the three channels of a color space. When the pixels make up a small area, repeat the calculation of dissimilarity so that we can spread to the entire image. In some different circumstances of the local area, the image-based segmentation does not use global thresholds, but adaptive thresholds. The adaptive thresholds are calculated by the interclass between different regions and intraclass in the same region.

About the regional consolidation, the region similarity can be calculated by the image features. We divide the feature similarity into color similarity, texture similarity, size similarity and filling similarity. The color similarity is calculated using three color channels such as hue, saturation and brightness. The texture similarity uses Gaussian function to calculate the differentiation in each direction of all channels. In the end, we can get a texture histogram. The size similarity is based on the number of pixels in two regions. This is to increase the combined rate of small areas. The texture similarity identifies the border around the region. If the border size of two regions is similar and the overlap area is large, then we merge these two regions. Using the four different similarities above, we can adjust individually according to the real situations. The selective search can produce a large number of candidate regions in a short time, but its disadvantage is also obvious. It has too many post-selected areas in the complex image, which decreases the speed of deep learning. In our system, we have already filtered out most regions with an HSI color space. Therefore, the influence of having too many candidate regions from the selective search can be reduced.

3.3 Convolutional Neural Network

We utilize two kinds of deep learning structures to train our system. One of the structures is Alexnet [15], in which the architecture mainly contains eight layers. The first five layers are convolutional layers and the latter three are all connected layers. In the last layer, we set a 43-way softmax layer to connect with the full connected layer. AlexNet can be considered as an example architecture of a deep convolutional network. There are many network architectures such as ZF-net, SPP-net, VGG and other networks, taking AlexNet as a prototype. AlexNet's success consists of several parts: Rectified Linear Unit (ReLU), pooling, local response normalization and dropout. It not only improves the training speed and accuracy, but also reduces the over-fitting problem.

The second deep learning architecture is GoogLeNet. In the case with a large amount of data, the easiest way to improve the performance of the network is to increase the depth and width of the network. However, it has two problems: 1) A large network generally needs more parameters, and using the fixed data, it is likely to cause the network over-fitting. 2) A large network needs more computing resources. For instance, increasing the number of convolution of the network will lead to increase the computing volume. In addition, if the increase of the network has not been effectively used such as the weight close to zero, it will cause a waste of computing resources. To solve these problems, Arora [1] proposed the construction of inception. The main idea of inception is to find and describe the dense components of the local sparse structure in the convolutional network. It is assumed that translation invariance is established by convolutional blocks and the repetition is spatially extended. We need to find the ideal local structure and expand the duplication in the space. GoogLeNet is also based on the above inception design concept.

4 Results

The experiments are divided into two parts, which are individually tested according to the detection and identification in the system. The datasets are used separately for detection and recognition. We use German Traffic Sign Detection Benchmark (GTSDB) as a dataset for detection and German Traffic Sign Recognition Benchmark (GTSRB) as a dataset for recognition.

For the detection part, we use 300 discrete images in GTSDB dataset for testing. In addition, because the system is designed for red traffic signs, we only record red signs as our ground truth. The correctness of the test is based on whether the candidate area is selected by the selective search and can be correctly extracted with a bounding box. There are two points to check with the benchmark. First, the target area has to be selected more than 80%. Second, the area of the bounding box containing the target should not exceed 1.5 times of the target area. If one of the conditions fails, then we consider the detection incorrect. According to the previous part of the selective search, the algorithm can use different color spaces as its strategy in different situations. Therefore, this experiment in addition to record the accuracy of the detection system, we also test the image segmentation for the selective search in different color spaces for the correction rate.

Table 1 shows the evaluation results. The detection system only focuses on the correct capture of the traffic sign location. The selected region of a road sign is integrated into the identification system. Thus, only the values of true positive, false negative and accuracy are reported in the table. According to the experimental results in Table 1, although the HSI color space is used to reduce the impact of light and shadow changes in the image, it still cannot fully reduce its influence. The reflective light and back-light will cause the detection failure. In different color spaces, although the CIE XYZ gives the highest correction rate, it still cannot distinguish the background from the traffic sign due to their similarity.

Table 1 Selective search in the color space with the accuracy and recall rates, and the red filter screening results

	True positive	False negative	Precision
RGB	238	19	92.6%
HSI	215	42	83.6%
YCrCb	230	27	89.4%
CIE XYZ	241	16	93.7%
CIE Lab	230	27	89.4%
CIE Luv	224	33	87.1%
Red filter	246	11	95.7%

For the recognition part, the part of deep learning is based on Caffe. We have 39209 training images from GTSRB as our training data, which can be divided into 43 road sign categories. The test data are randomly selected 1000 images from GTSRB. The results of training and testing via GoogLeNet are shown in Fig. 3. We use a gradient descent method for the training. The parameters include the basic learning rate set as 0.01, the number of iterations as 100000 times, and the weight attenuation as 0.0002. Fig. 4 shows the results of using AlexNet. The base learning rate is set as 0.01, the number of iterations is 600, and the weight attenuation is 0.0005. With the above parameters, when the iteration is more than 600 times, the model will incur the over-fitting issue.

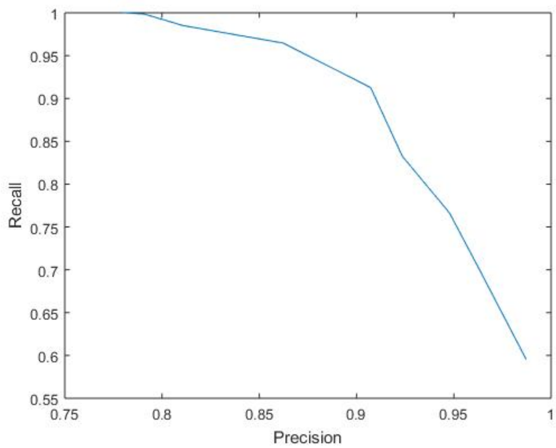


Figure. 3. Test result of GoogLeNet

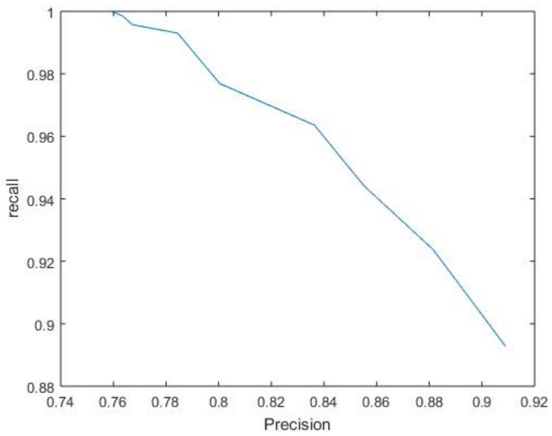


Figure. 4. Test result of AlexNet

5 Conclusion

We have presented traffic sign techniques for driving assistance systems. The system can be divided into road sign detection and identification subsystems. The road sign detection system uses the color, the image segmentation and the hierarchical grouping methods to select candidate areas of the road sign. The candidate area from detection is used for identification. We then use a CNN for the road sign recognition system. Results show that the proposed algorithm can obtain the desired results effectively.

ACKNOWLEDGEMENTS

The support of this work in part by the Ministry of Science and Technology of Taiwan under Grant MOST 104-2221-E-194-058-MY2 is gratefully acknowledged.

REFERENCES

[1] S. Arora, A. Bhaskara, R. Ge, and T. Ma, "Provable bounds for learning some deep representations," in Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32, ser. ICML'14. JMLR.org, 2014, pp. I–584–I–592.

[2] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in IEEE Proceedings. Intelligent Vehicles Symposium, 2005., June 2005, pp. 255–260.

[3] R. Belaroussi and J.P. Tarel, "A real-time road sign detection using bilateral chinese transform," in Proceedings of the 5th International Symposium on Advances in Visual Computing: Part II, ser. ISVC '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 1161–1170

[4] R. Belaroussi and J.P. Tarel, "Angle vertex and bisector geometric model for triangular road sign detection," in 2009 Workshop on Applications of Computer Vision (WACV), Dec 2009, pp. 1–7.

[5] Broggi, P. Cerri, P. Medici, P.P. Porta, and G. Ghisio, "Real time road signs recognition," in 2007 IEEE Intelligent Vehicles Symposium, June 2007, pp. 981–986.

[6] D. Cirean, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in The 2011 International Joint Conference on Neural Networks, July 2011, pp. 1918–1921.

[7] De La Escalera, J.M. Armingol, J.M. Pastor, and F.J. Rodriguez, "Visual sign information extraction and identification by deformable models for intelligent vehicles," Trans. Intell. Transport. Sys., vol. 5, no. 2, pp. 57–68, Jun. 2004.

[8] de la Escalera, L.E. Moreno, M.A. Salichs, and J.M. Armingol, "Road traffic sign detection and classification," IEEE Transactions on Industrial Electronics, vol. 44, no. 6, pp. 848–859, Dec 1997.

[9] C.Y. Fang, S.W. Chen, and C.S. Fuh, "Road-sign detection and tracking," IEEE Transactions on Vehicular Technology, vol. 52, no. 5, pp. 1329–1341, Sept 2003.

[10] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient graph-based image segmentation," Int. J. Comput. Vision, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[11] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs," IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 4, pp. 1498–1506, Dec 2012.

[12] S.C. Huang, H.Y. Lin, C.C. Chang. An in-car camera system for traffic sign. Proceedings of the Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems, Detection and Recognition, Otsu, Japan, June 2017.

[13] W.C. Huang and C.H. Wu, "Adaptive color image processing and recognition for varying backgrounds and illumination conditions," IEEE Transactions on Industrial Electronics, vol. 45, no. 2, pp. 351–357, Apr 1998.

[14] D.L. Kellmeyer and H.T. Zwahlen, "Detection of highway warning signs in natural video images using color image processing and neural networks," in Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on, vol. 7, Jun 1994, pp. 4226–4231.

[15] Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 1097–1105.

[16] W.J. Kuo and C.C. Lin, "Two-stage road sign detection and recognition," in 2007 IEEE International Conference on Multimedia and Expo, July 2007, pp. 1427–1430.

[17] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," IEEE Transactions on Intelligent Transportation Systems, vol. 8, no. 2, pp. 264–278, June 2007.

[18] J. Miura, T. Kanda, and Y. Shirai, "An active vision system for realtime traffic sign recognition," in ITSC2000. 2000 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.00TH8493), 2000, pp. 52–57.

[19] L. Pacheco, J. Batlle, and X. Cufi, "A new approach to real time traffic sign recognition based on colour information," in Intelligent Vehicles'94 Symposium, Proceedings of the, Oct 1994, pp. 339–344.

[20] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in The 2011 International Joint Conference on Neural Networks, July 2011, pp. 2809–2813.

[21] W.D. Shaposhnikov, D.G. Shaposhnikov, L. N, E.V. Golovan, and A. Shevtsova, “Road sign recognition by single positioning of space-variant sensor,” in Proc. 15th International Conference on Vision Interface, 2002, p. 2002.

[22] S. Vitabile, A. Gentile, and F. Sorbello, “A neural network based automatic road signs recognizer,” in Neural Networks, 2002. IJCNN ’02. Proceedings of the 2002 International Joint Conference on, vol. 3, 2002, pp. 2315–2320.

.