Simulated Traffic Sign Classification Using Cross-Connected Convolution Neural Networks Based on Compressive Sensing Domain

Jiping Xiong(✉), Lingfeng Ye, Fei Wang, and Tong Ye

College of Mathematics, Physics and Information Engineering,
Zhejiang Normal University, Jinhua, Zhejiang, People's Republic of China
xjping@zjnu.cn

Abstract. This paper proposes an algorithm of simulated traffic sign recognition based on compressive sensing domain and convolution neural networks for simulated traffic sign recognition. And the algorithm can extract the discriminative non-linear features directly from the compressive sensing domain. The image is transformed into compressive sensing domain by measurements matrix without reconstruction. This paper proposes a cross-connected convolution neural networks (CCNN) with an input layer, 6 six hidden layers (i.e., three convolution layers alternating with three pooling layers), a fully-connected layer and an output layer, where the second pooling layer is allowed to directly connect to the fully-connected layer across two layers. Experimental results show that the algorithm improves the accuracy of simulated traffic sign recognition. The recognition of the algorithm is possible even at low measurement rates.

Keywords: Compressive sensing domain · Convolution neural networks · CS measurements · Simulated traffic sign recognition

1 Introduction

Nowadays, Intelligent Simulation Transportation Systems attract more and more attention in research community and industry [1–4]. Traffic signs classification is one of the foremost important integral parts of simulated driving and advanced driver assistance systems (ADAS) [5–10]. Most of the time driver missed traffic signs due to different obstacles and lack of attentiveness. Automating the process of classification of the traffic signs would significantly help reducing accidents. Classification of traffic signs is not so simple task, images can be affected to adverse variation due to illumination, orientation, and the speed variation of vehicles etc. Normally wide angle camera is mounted on the top of a vehicle to capture traffic signs and other related visual features for ADAS. Such images are distorted due to several external factors including vehicles speed, sunlight, rain etc. Sample images from GTSRB dataset are shown in Fig. 1.

H. Song et al. (Eds.): SIMUtools 2019, LNICST 295, pp. 588–597, 2019.
https://doi.org/10.1007/978-3-030-32216-8_57

Fig. 1. Sample images of GTSRB dataset

Traditional computer vision and machine learning based methods were widely used for traffic signs classification [11, 12], but those methods were soon replaced by deep learning based classifiers. In deep learning, convolution neural networks (CNN) is developed in recent years and is a computer pattern recognition method which leads to an extensive attention [13, 14]. CNN is a multilayer preceptor special designed to recognize two-dimensional shapes. The network structure of CNN has highly invariance for translation, scaling, tilting or other forms of deformation. Because CNN does not require pre-processing images and can directly input the original image, it has been successfully applied to simulated traffic sign recognition [15, 16].

At the same time, the research shows that the comprehensive utilization of high level features and low level features are advantageous to improve the recognition performance of the visual system [17–19]. Therefore, based on the traditional convolution neural networks, by introducing the idea of cross-layer connections, the paper refers to a cross-connected convolution neural network model with a nine-layer structure, which is aimed at effectively combining low-level features and high-level features to build a better classifier.

In many computer visualization applications, however, the objective is not perfect recovery of the image, but to determine certain properties of the image. Because most of reconstruction algorithms of compressive sensing computationally expensive and the reconstruction results are poor at low measurement rates. In image recognition, such as simulated traffic sign recognition, we are interested in determining the category to which the object in the image belongs. Following some recent work in this emerging field of compressive sensing domain, we find further the possibility of performing effective high-level inference directly on compressive sensing measurements, without reconstruction. In [20], compressive sensing domain is applied to image recognition, without reconstruction. On the MNIST and ImageNet database, it got good recognition performance.

Therefore, the paper proposed a simulated traffic sign recognition approach using cross-connected convolution neural network based compressive sensing domain. The approach show that convolution neural network can be employed to extract discriminative non-linear features directly from compressive sensing domain. The image is compressed and is converted into the compressive sensing domain as the input of cross-connected convolution neural network, without reconstruction. Cross-connected convolution neural networks is a 9 layers framework with a input layer, 6 six hidden layers (i.e., three convolution layers alternating with three pooling layers), a fully-connected layer and an output layer, where the second pooling layer is allowed to directly connect to the fully-connected layer across two layers. Experimental results show that the algorithm improves the accuracy of simulated traffic sign recognition.

The rest of the paper is organized as follows. Section 2 details the proposed approach to simulated traffic sign recognition. Experimental results are illustrated in Sect. 3. Section 4 concludes the paper.

2 Simulated Traffic Sign Recognition Using Cross-Connected Convolution Neural Network Based on Compressive Sensing

This paper proposed a simulated traffic sign recognition approach using cross-connected convolution neural network based compressive sensing domain. First, the image is transformed into compressive sensing domain by measurements matrix without reconstruction. Then, the discriminative non-linear features can be directly extracted by cross-connected convolution neural network. Finally, traffic sign images are recognized by classifier.

2.1 Compressive Sensing Domain

Compressive Sensing

The compressive sensing (CS), $y \in R^m$, of an image $x \in R^n$ (ordered lexicographically), where $m < n$, are obtained using $y = \Phi x + e$, where $e \in R^m$ is the measurement noise. $\Phi \in R^{m \times n}$, called the measurement matrix with all entries and drawn from certain distributions such as Gaussian, Bernoulli etc. [21]. The problem of recovering x from y is generally ill-posed since the linear system is underdetermined. It has been proven by Candes et al. [22] and Donoho [23] that, by posing additional constraints that x is s-sparse in a basis Ψ, Φ being incoherent with Ψ and $m = O\left(s \log \frac{n}{s}\right)$, the solution to the linear system is unique and x can be recovered perfectly from y. This is typically achieved by solving an optimization problem of the form:

$$min_x \qquad \|\Psi x\|_1 \qquad \text{s.t} \qquad \|y - \Phi x\|_2 \le \epsilon \tag{1}$$

There exist variants of the optimization problem in (1) which are applicable to compressible signals (since images are not exactly sparse in the wavelet domain).

Compressive Sensing Domain

Over the past decade, a great number of algorithms have been designed to solve this problem such as Orthogonal Matching Pursuit [24] and Basis Pursuit [25]. However, these algorithms are computationally expensive in addition to belong ineffective at low measurement rates. Compressive sensing domain has gained momentum in recent years. Calderbank et al. [26] show that classifiers can be learned directly in the compressive sensing domain. Sankaranarayanan et al. [27] model videos as linear dynamical system (LDS) and use it to acquire and reconstruct CS videos. They also perform classification using the LDS parameters obtained directly from CS measurements. In [28], Matching Pursuit is modified and employed for reconstruction-free signal detection from CS measurements. In [29], a compressive sensing architecture is developed where, instead of perfect reconstruction of the CS images, only relevant parts of the scene i.e., the objects are reconstructed.

In simulated traffic sign recognition, we are interested in determining the category to which the object in the image belongs. Therefore, the paper applies compressive sensing domain to simulated traffic sign recognition. The traffic image is compressed by the measurement matrix. And the image is converted into the compressive sensing domain as the input of cross-connected convolution neural network, without reconstruction.

2.2 Cross-Connected Convolution Neural Network

Convolution Neural Network

The traditional convolution neural network is a special deep forward neural network model whose structure generally consists of input layer, multiple alternate convolution and pooling layer, full connection layer and output layer. The input layer is usually a matrix, such as an image. From the point of feed-forward network, convolution layer and pooling layer can be regarded as special hidden layers, in addition to output layer, and the full connection layer is the common hidden layer. In the convolution neural network, there are four basic operations which are defined as inner convolution, outer convolution, under-sampling and up-sampling.

Suppose A and B are matrixes, the size are $M \times N$ and $m \times n$, and $M \geq m$, $N \geq n$. Their inner convolution $C = A \check{*} B$'s elements are defined as below:

$$c_{ij} = \sum\nolimits_{s=1}^{m} \sum\nolimits_{t=1}^{n} a_{i+m-s,j+n-t} \cdot b_{st} \tag{2}$$

Where, $1 \leq i \leq M - m + 1$, $1 \leq j \leq N - n + 1$.

Outer convolution is defined as:

$$A \hat{*} B = \widehat{A_B} \check{*} B \tag{3}$$

Where, $\widehat{A_B} = \widehat{a_{ij}}$ is a matrix which is got by using 0 for A. The size is $(M + 2m - 2) \times (N + 2n - 2)$, and:

$$\widehat{a_{i,j}} = \begin{cases} a_{i-m+1,j-n+1}, & m \leq i \leq M + m - 1, n \leq j \leq N + n - 1 \\ 0, & \text{else} \end{cases} \tag{4}$$

If matrix A is divided into blocks without overlapping. Suppose that the size of every block is $\lambda \times \tau$, and use $G^{A}_{\lambda,\tau}(i,j)$ to represent ijth block. The structure is as follows:

$$G^{A}_{\lambda,\tau}(i,j) = (a_{st})_{\lambda \times \tau} \tag{5}$$

Where, $(i-1) \times \lambda + 1 \leq s \leq i \times \lambda$, $(j-1) \times \tau + 1 \leq t \leq j \times \tau$. $G^{A}_{\lambda,\tau}(i,j)$' under-sampling is:

$$down\left(G^{A}_{\lambda,\tau}(i,j)\right) = \frac{1}{\lambda \times \tau} \sum\nolimits_{s=(i-1)\times\lambda+1}^{i\times\lambda} \sum\nolimits_{(j-1)\times\tau+1}^{j\times\tau} a_{st} \tag{6}$$

The under-sampling of the matrix A by the non-overlapping block with $\lambda \times \tau$ multiple is defined as:

$$down_{\lambda,\tau}(A) = down\left(G^{A}_{\lambda,\tau}(i,j)\right) \tag{7}$$

The up-sampling of the matrix A by the non-overlapping block with $\lambda \times \tau$ multiple is defined as:

$$up_{\lambda\times\tau}(A) = A \otimes 1_{\lambda\times\tau} \tag{8}$$

Where, $1_{\lambda\times\tau}$ is a matrix with all 1 elements. $\otimes$ represents Kronecker product.

Cross-Connected Convolution Neural Network

One drawback of the traditional convolution neural network is that it is difficult to efficiently integrate the lower level features with the higher features to construct a better classifier. Aimed at the problem. The paper introduces a cross-connected convolution neural network for the simulated traffic sign recognition. The network consists of a input layer x, three convolution layers (h_1, h_3, h_5), three pooling layers (h_2, h_4, h_6), a full connection layer (h_7) and a output layer o, as shown in Fig. 2. The network starts with an image as input. Then, using three interlaced convolution layer and pooling layer extracts the features of the image. Finally, the features extracted from two pooling layers (h_4 and h_6) are passed directly to the full connection layer for fusion processing and classification. In the network, the connection from pooling layer h_4 to pooling layer h_7 strides over two layers and is called cross-connection. At this point, the number of nodes in the full connection layer is the sum of the nodes of the pooling layer h_4 and h_6. The Table 1 presents the cross-connected convolution neural network, which consists of type, patch size, stride and output size.

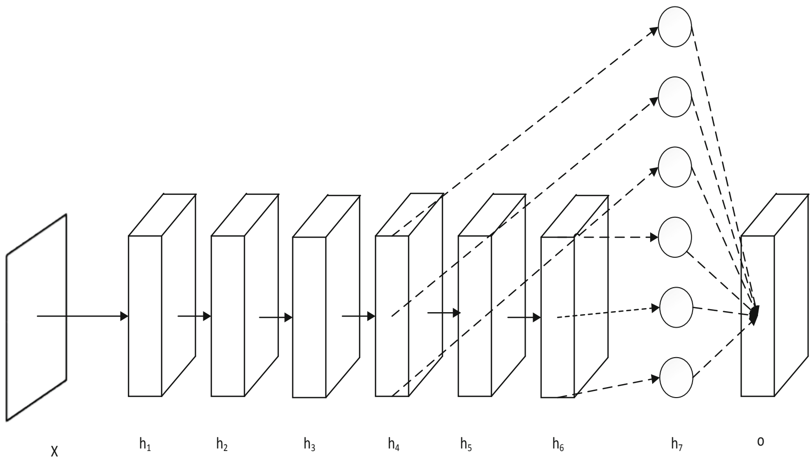

Fig. 2. The structure of the cross-connected convolution neural network

Table 1. The description of the cross-connected convolution neural network

Layer	Type	Patch size	Stride	Output size
x	Input			32×32
h_1	Convolution	5×5	1	$28 \times 28 \times 6$
h_2	Max pooling	2×2	2	$14 \times 14 \times 6$
h_3	Convolution	5×5	1	$10 \times 10 \times 12$
h_4	Max pooling	2×2	2	$5 \times 5 \times 12$
h_5	Convolution	2×2	1	$4 \times 4 \times 16$
h_6	Max pooling	2×2	2	$2 \times 2 \times 16$
h_7	Fully-connected			364
o	Output			2

3 Experimental Results

3.1 GTSRB Database

GTSRB (German Traffic Sign Recognition Benchmarks) database is one of the standard database of traffic sign recognition. GTSRB has 43 categories and 51839 images. Each category has 100–1000 images, including prohibitory signs, danger signs and mandatory signs. 39209 images are selected as the training data set, and the rest images is selected as the test data set. There is a distortion in the images, because perspective change, shade, color degradation, weather change and so on. In the database, the size of image is different. After the image converted into compressive sensing domain, the image size is adjusted to 32×32 as the input of network. Figure 3 is subsets of traffic signs in the GTSRB database.

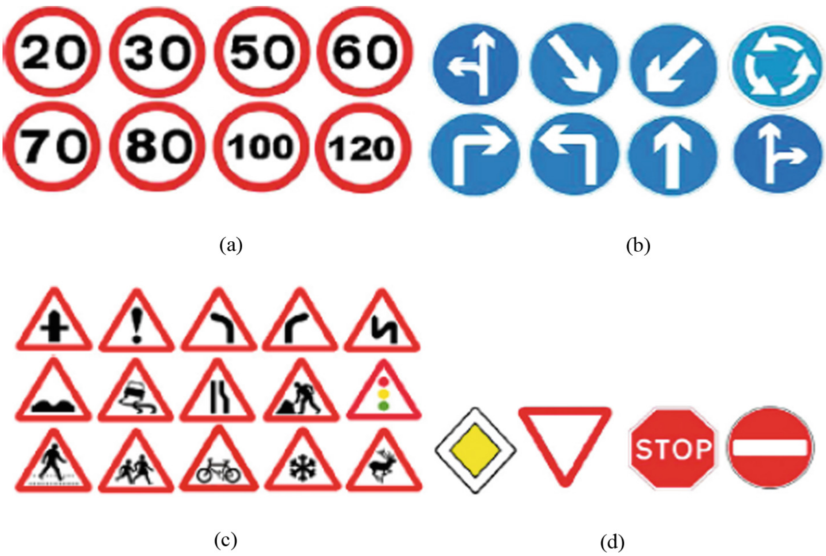

Fig. 3. Subset signs in the GTSRB data set. (a) Speed limit signs. (b) Derestriction signs. (c) Danger signs. (d) Unique signs.

3.2 Recognition Rate Comparing

In order to evaluate the effectiveness of the proposed method of simulated traffic sign recognition using cross-connected convolution neural network based on compressive sensing domain, the experiment is done on the GTSRB database. Cross-connected convolution neural network is trained and tested on caffe framework.

Table 2. Comparison of recognition of different method

Method	Speed limit	Prohibitory	Derestriction	Mandatory	Danger	Unique
[30]	99.47	99.93	99.72	99.89	99.07	99.22
[31]	99.86	**100.00**	**99.95**	99.89	99.89	99.87
[30]	98.61	99.87	94.44	97.18	98.30	98.63
[12]	98.82	98.27	97.93	96.86	96.95	**100.00**
[32]	95.95	99.13	87.50	99.27	92.08	98.73
Proposed method	**99.88**	**100.00**	99.93	**99.90**	**99.92**	99.89

From Table 2, compared with other method, the proposed method has some advantages. From the results of the six traffic signs, the proposed method has a higher recognition rate. In speed limit signs, prohibitory signs, mandatory signs and danger signs, the proposed method has obtained the highest recognition rate, and the recognition rate is 99.88%, 100.00%, 99.90% and 99.92%. In derestriction signs and unique signs, compared with the best method, the gap of the proposed method is not big.

Table 3. Comparison of the average recognition rate of different method

Method	Average	Method	Average
[30]	99.55	[12]	98.14
[31]	99.88	[32]	95.44
[30]	97.84	Proposed method	**99.92**

From Table 3, in the results of the six traffic signs, the proposed method gets highest average recognition rate 99.92%. Compared with the best method, the average recognition rate increased by 0.04%. It shows that the proposed method improves the robustness of the simulated traffic sign recognition.

3.3 Comparison of Different Measurement Rates

We train different networks, with the same architecture, at five different measurement rates of MR = 1, 0.25, 0.1, 0.05. If the number of pixels in original traffic sign images is n = 784, these MRs correspond to m = 784, 196, 78, 39 and 8 CS measurement respectively. For the five cross-connected convolution neural networks, the parameters such as learning rate, momentum and weight decay are fixed in order to be compared. The recognition rate on the test set at different measurement rates are shown in Table 4. It shows that the transformation of image to the compressive sensing domain can improve the performance of traffic sign recognition.

Table 4. The recognition rate on the test set at different measurement rates

Measurement rate (MR)	1	0.25	0.10	0.05	0.01
Recognition rate	99.92	98.43	97.35	94.77	60.20

4 Conclusion

Based on the development of the compressive sensing domain theory and the simulated traffic sign recognition of convolution neural network, the paper proposed a simulated traffic sign recognition using cross-connected convolution neural network based on compressive sensing domain. Using cross-connected convolution neural network directly extracts features from the compressive sensing domain of the image for simulated traffic sign recognition. The image is transformed into compressive sensing main by measurements matrix without reconstruction as the input of cross-connected convolution neural network. The proposed method can extract discriminative non-linear features directly from compressive sensing domain for simulated traffic sign recognition. Experiment on standard traffic sign database GTSRB, it is found that the method improves the recognition accuracy and enhances the robustness of simulated traffic sign recognition.

References

1. Chhabra, R., Verma, S., Rama Krishna, C.: A survey on driver behavior detection techniques for intelligent transportation systems. In: 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence. IEEE (2017)
2. Jiang, D., Huo, L., Lv, Z., Song, H., Qin, W.: A joint multi-criteria utility-based network selection approach for vehicle-to-infrastructure networking. IEEE Trans. Intell. Transp. Syst. **PP**(99), 1–15 (2018)
3. Jiang, D., Xu, Z., Wang, W., Wang, Y., Han, Y.: A collaborative multi-hop routing algorithm for maximum achievable rate. J. Netw. Comput. Appl. **57**(2015), 182–191 (2015)
4. Jiang, D., Wang, Y., Yao, C., Han, Y.: An effective dynamic spectrum access algorithm for multi-hop cognitive wireless networks. Comput. Netw. **84**(19), 1–16 (2015)
5. Haloi, M., Jayagopi, D.B.: A robust lane detection and departure warning system. In: IEEE Intelligent Vehicles Symposium (IV), pp. 126–131 (2015)
6. Jiang, D., Xu, Z., Li, W., Chen, Z.: Network coding-based energy-efficient multicast routing algorithm for multi-hop wireless networks. J. Syst. Softw. **104**(2015), 152–165 (2015)
7. Jiang, D., Li, W., Lv, H.: An energy-efficient cooperative multicast routing in multi-hop wireless networks for smart medical applications. Neurocomputing **220**(2017), 160–169 (2017)
8. Jiang, D., Wang, Y., Han, Y., et al.: Maximum connectivity-based channel allocation algorithm in cognitive wireless networks for medical applications. Neurocomputing **220** (2017), 41–51 (2017)
9. Jiang, D., Xu, Z., Li, W., et al.: An energy-efficient multicast algorithm with maximum network throughput in multi-hop wireless networks. J. Commun. Netw. **18**(5), 713–724 (2016)
10. Jiang, D., Zhang, P., Lv, Z., et al.: Energy-efficient multi-constraint routing algorithm with load balancing for smart city applications. IEEE Internet of Things J. **3**(6), 1437–1447 (2016)
11. Le, T.T., Tran, S.T., Mita, S., Nguyen, T.D.: Real time traffic sign detection using color and shape-based features. In: Intelligent Information and Database Systems, pp. 268-278. IEEE (2010)
12. Haloi, M.: A novel pLSA based Traffic Signs Classification System. https://arxiv.org/. Accessed 2015
13. Zhao, Z.H., Yang, S.P., Ma, Z.Q.: The study of license character recognition based on the convolution neural network LeNet-5. J. Syst. Simul. **22**(3), 638–641 (2010)
14. Xu, S.S., Liu, Y.A., Xu, S.: Wood defect recognition based on the convolution neural network. J. Shandong Univ.: Eng. Sci. **43**(2), 23–28 (2013)
15. Mrinal, H.: Traffic Sign Classification Using Deep Inception Based Convolutional Networks. https://arxiv.org/. Accessed 2016
16. Yang, Y., Luo, H., Xu, H., Wu, F.: Towards real-time traffic sign detection and classification. IEEE Trans. Intell. Transp. Syst. **17**, 2022–2031 (2016)
17. Zhong, S.H., Liu, Y., Ren, F.F., Zhang, J.H., Ren, T.W.: Video saliency detection via dynamic consistent spatio-temporal attention modelling. In: Proceedings of the 2013 AAAI Conference on Artificial Intelligence, pp. 1063–1069. AAAI, Bellevue (2013)

18. Jiang, D., Nie, L., Lv, Z., et al.: Spatio-temporal Kronecker compressive sensing for traffic matrix recovery. IEEE Access **4**, 3046–3053 (2016)
19. Qi, M.-B., Tan, S.-S., Wang, Y.-X., Liu, H., Jiang, J.-G.: Multi-feature subspace and kernel learning for person reidentication. Acta Automatica Sinica **42**(2), 299–308 (2016)
20. Suhas, L., Kuldeep, K., Pavan, T.: Direct inference on compressive measurements using convolutional neural networks. In: Image Processing (ICIP), pp. 1913–1917 (2016)
21. Sun, Z.J., Xue, L., Xu, Y.M.: Review of deep learning research. Comput. Appl. Res. **29**(8), 2807–2810 (2012)
22. Candès, E.J., Wakin, M.B.: An introduction to compressive sampling. Signal Process. Mag. **25**(2), 21–30 (2008)
23. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. **59**(8), 1207–1223 (2006)
24. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
25. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inf. Theory **53**(12), 4655–4666 (2007)
26. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. **20**(1), 33–61 (1998)
27. Sankaranarayanan, A.C., Turaga, P.K., Baraniuk, R.G., Chellappa, R.: Compressive acquisition of dynamic scenes. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6311, pp. 129–142. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15549-9_10
28. Duarte, M.F., Davenport, M.A., Wakin, M.B., Baraniuk, R.G.: Sparse signal detection from incoherent projections. In: Acoustics, Speech and Signal Processing, vol. 3, pp. 305–308 (2006)
29. Mahalanobis, A., Muise, R.: Object specific image reconstruction using a compressive sensing architecture for application in surveillance systems. IEEE Trans. Aerosp. Electron. Syst. **45**(3), 1167–1180 (2009)
30. Sermanet, P., LeCun, Y.: Traffic sign recognition with multi-scale convolutional networks. In: Neural Networks (IJCNN), pp. 2809–2813. IEEE (2011)
31. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: The German traffic sign recognition benchmark: a multi-class classification competition. In: Neural Networks (IJCNN), pp. 1453–1458. IEEE (2011)
32. Zaklouta, F., Stanciulescu, B., Hamdoun, O.: Traffic sign classification using KD trees and random forests. In: Neural Networks (IJCNN), pp. 2151–2155. IEEE (2011)