Negative-Supervised Cascaded Deep Learning for Traffic Sign Classification

Kaixuan Xie[1,2], Shiming Ge[2(✉)], Rui Yang[1,2], Xiang Lu[2], and Limin Sun[2]

[1] University of Chinese Academy of Sciences, Beijing, China
[2] Beijing Key Laboratory of IOT Information Security,
Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{xiekaixuan,geshiming,yangrui,luxiang,sunlimin}@iie.ac.cn

Abstract. In this paper, we propose a novel deep learning framework for object classification called negative-supervised cascaded deep learning. There are two hierarchies in our cascaded method: the first one is a convolutional neural network trained on positive-only samples, which is used to select supervisory samples from a negative library. The second one is inherited from the trained first CNN. It is trained on positive and negative samples, which are selected from domain related database by utilizing negative-supervised mechanism. Experiments are applied this idea to traffic sign classification using two classic convolutional neural networks, LeNet-5 and AlexNet as baselines. Classification rates improved by 0.7% and 1.1% with LeNet-5 and AlexNet respectively, which demonstrates the efficiency and superiority of our proposed framework.

Keywords: Convolutional neural network · Deep learning · Negative-supervised · Object classification · Traffic sign classification

1 Introduction

Traffic signs play an important role in our daily life. They define a visual language providing useful information, which makes the driving safe and convenient. In intelligent transportation system (ITS), traffic sign recognition is a critical step for advance driver assistance system (ADAS) and autonomous intelligent vehicles[1]. Traffic sign recognition has two tasks: finding the locations and sizes of traffic signs in natural scene images (traffic sign detection) and classifying the detected traffic signs to their specific sub-classes (traffic sign classification)[2]. In this paper, we focus on traffic sign classification.

Due to the complex outdoor environment such as viewpoint variations, bad lighting conditions, motion-blur, occlusions, sun glare, physical damage, colors fading, clustered backgrounds, low resolution and so on, traffic signs are rotated, blurred, damaged and degenerated, which is shown in Fig. 4. Traffic sign recognition is more challenging in comparison with indoor object classification tasks, such as character and face recognition. The uncertainty and ambiguity of traffic signs do not stop the pace of research. To address traffic signs classification, many methods have been proposed according to existing survey literatures[3][4].

H. Zha et al. (Eds.): CCCV 2015, Part I, CCIS 546, pp. 249–257, 2015.
DOI: 10.1007/978-3-662-48558-3_25

Traditional traffic sign classification has been approached with a number of popular machine learning methods, such as support vector machines[5], linear discriminant analysis[6], etc. These methods need to extract of features, such as Histogram of Gradients (HoG)[7], Local Binary Pattern (LBP)[8], Integral Channel Features[9] first. Recently, deep learning such as Convolutional Neural Networks (CNNs), was proposed for handling various classification tasks and has given state-of-the-art results in many vision tasks[10][11][12].

The existing deep learning based traffic sign classification methods[10][11][12] mainly learn a model to classify the detected sign to one specific class of traffic signs, and do not consider auxiliary information or extra knowledge such as background and attributes. It is argued that auxiliary information or extra knowledge could be used to augment model learning and improve classification performance. In [13], Kumar et al. presented two high-level visual features, attribute and smile, for face verification. In [14], in order to learn more discriminative representation for face classification task, Sun et. al. introduced verification signal to supervise deep learning face identification.

Motivated by [13][14], we propose a novel negative-supervised cascaded deep learning framework constructed by two hierarchies. The first deep learning module is trained with the positive samples (proposals belong to one class of objects) and then is used to select negative samples (proposals not belong to any class of objects) from domain related image database. The selected negative samples are combined with positive samples and then are trained with the second deep learning module to form final result. The major advantage of our method over those standard approaches is that we transfer knowledge from domain related data. We introduce the negative-supervised mechanism for traffic sign classification to verify the efficiency. Experimental results show that the proposed learning framework could improve classification performance without sacrificing computation performance in test stage.

The remainder of the paper is organized as follows. In Section 2, we review some closely related work and some relative knowledge. We propose negative-supervised cascaded deep convolution neural network framework in Section 3. Experiments are detailed in Section 4 and results are illustrated and analyzed in Section 5. At last we give conclusions in Section 6.

2 Related Work

Traffic sign classification is one of representatives of object classification and its purpose is to classify the specific sign to the corresponding class. Many machine learning methods are introduced to traffic sign classification, like k-d trees and random forests[15], SVM[5][16][17] and LDA[6]. These methods usually share the common process pipeline that consists of image pre-processing, feature extraction like HoG[7] or LBP[8]. And the processing time and classification accuracy of these methods are very dependent on the specifically designed features. By contrast, neural networks need less pre-processing and no independent feature extraction. In [10], Ciresan et. al. proposed a committee of CNNs and a

Multi-layer Perceptron (MLP) to perform traffic sign classification. In [11], they improved the method with multi-column deep neural network. Jin et al. [12] described an ensemble of CNNs and proposed a Hinge Loss Stochastic Gradient Descent (HLSGD) method to train CNNs. Most of traffic sign classification approaches modify their network architecture to improve their performance but they use only positive samples and ignore other information.

Traffic sign classification belongs to artificial intelligence. It can be regarded as a problem of knowledge mining, analysis and understanding, where transfer learning can truly be beneficial[18]. We are motivated by the transfer learning techniques[18][19]. Some knowledge is specific for individual domains or tasks and some knowledge may be common between different domains such that they may help improve performance for the target domain or task. We realize that various patches of scene images can be considered as common knowledge. We make them one specific negative class after scientific selection thus forming negative-supervised learning.

Taking all above into consideration, we proposed a negative-supervised cascaded deep learning framework. It consists of two cascaded deep convolutional neural networks (CNNs). We train the first CNN positive samples, and use it to select negative images from random patches as one component of training set in the second neural network. Experiments prove that negative-supervised mechanism improves the separability and recognition of the classifier.

3 Our Method

This section introduces our negative-supervised cascaded framework for classification via deep CNNs. Overall architecture of pipeline is illustrated in Fig. 1.

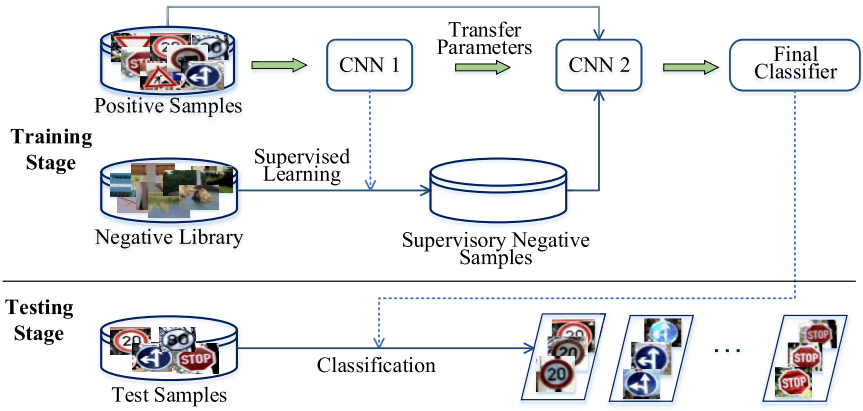


Fig. 1. Pipeline of Negative-supervised Cascaded Deep Learning

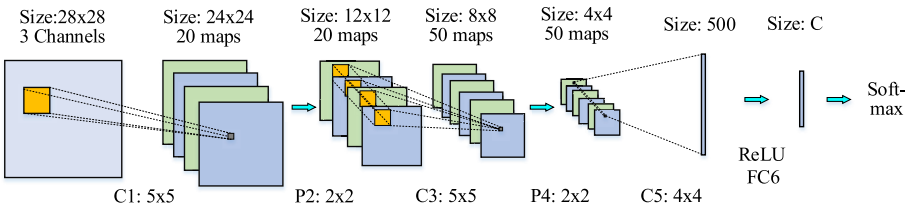
The first CNN is trained on positive samples. We use it to select discriminative negative samples from domain related negative sample library and train the second convolutional neural network on both positive and selected negative samples. The learned parameters of the first CNN are transferred to the second CNN for training.

In the following subsections, we first introduce two baseline CNNs [20][21], which are capable of achieving record-breaking results on a highly challenging data set using purely-positive learning. Then we detail our negative-supervised deep learning mechanism. Finally we construct a cascaded CNN to perform the task of traffic sign classification.

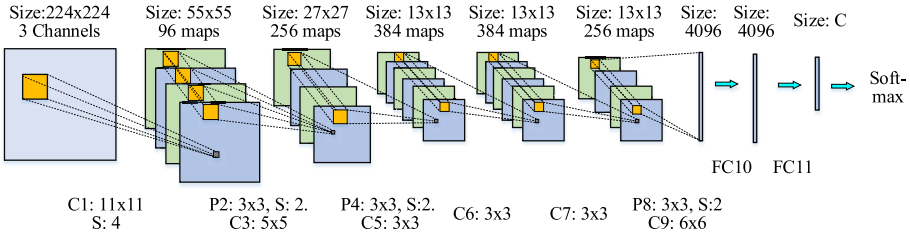
3.1 Deep Convolutional Neural Networks

CNNs are designed to recognize visual patterns directly from raw image pixels with minimal preprocessing. They can recognize patterns with variability and robustness. A CNN usually contains convolutional layers, pooling layers, fully-connected layers and finally a soft-max layer. We use the LeNet-5[20] and AlexNet[21] for our two baseline networks, which is shown in Fig.2.

LeNet-5 is originally designed for handwritten and machine-printed character recognition. It contains three convolutional layers and one fully-connected layer and ultimately a soft-max layer. Rectified Linear Units (ReLUs) $f(x) = max(0, x)$ is follow the convolutional layers and used to promote nonlinearity. Max-pooling layer is applied over neighboring neurons and used to lower in



(a) Architecture of LeNet-5[20]



(b) Architecture of AlexNet[21]. Each convolutional layer follows a ReLU layer.

Fig. 2. Architecture of Deep CNN. C, P, FC are short for convolutional layer, max-pooling layer, fully-connected layer, respectively. S is short for stride and omitted if stride is 1.

dimension thus decreasing over-fitting. The fully-connected layers have 500 neurons each and followed by the soft-max layers and input images need to be scaled to 28×28 pixels.

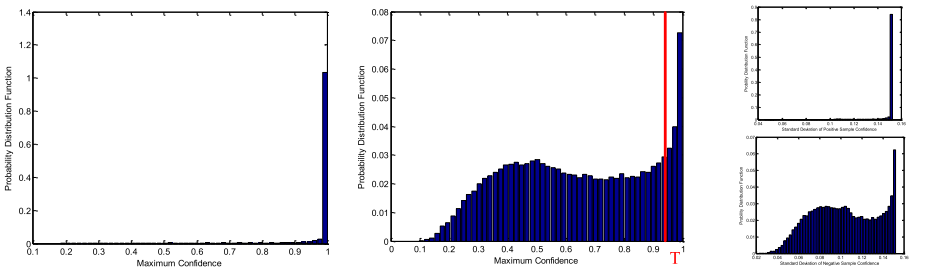
The architecture of AlexNet is similar with LeNet-5 but much deeper. It contains five convolutional layers, two fully-connected layers and ultimately a soft-max layer. Additional response-normalization layers are applied across feature channels. The fully-connected layers have 4096 neurons each and followed by the soft-max layer. Similarly, the input images need to be scaled to 224×224 pixels to adapt to the neural network.

3.2 Negative-Supervised Learning

The goal of negative-supervised learning is to boost the discriminative ability of CNN by retraining another CNN with the supervisory negative samples. The supervisory negative samples refer to the samples that may have discriminative information in negative library. These supervisory negative samples provide additional information related to positive samples, thus making classification task more accurate.

We select them according to the confidence generated by the first CNN and extract the maximum confidence MC of each positive sample after the soft-max layer. Knowing that the maximum confidence corresponds to the specific assigned category so we obtain the frequency of them. Then we put the same order of magnitude negative samples to the network and do the same procedure.

The probability density are illustrated as Fig.3. (a)(b) are the distributions of maximum confidence of positive and negative samples respectively. (c) is the standard deviation confidence of samples and it is shown that they have the same distribution. So we choose the supervisory samples based on maximum confidence. Maximum confidences are normalized between 0 and 1, most positive samples are convergent to a peak while negative samples are distributed scattered. It is obvious that the larger the maximum confidence is and the larger the standard deviation is, the more supervisory information the negative samples have. We select the supervisory negative samples that have $MC > T$, in which



(a) Maximum Confidence of Positive Samples

(b) Maximum Confidence of Negative Samples

(c) Standard Deviation of Confidence

Fig. 3. Statistics of Standard Deviation of Confidence

T is the threshold deduced from the confidence of validation set generated by the first CNN.

3.3 Traffic Sign Classification

In this subsection we describe the application of negative-supervised learning and realize it on traffic sign classification.

We use the traditional strategy to train the first CNN. we directly apply five convolutional layers and following two fully-connected layers and then fed to a soft-max layer corresponding to the number of positive classes defined as C. After the soft-max layer, we get C values meaning the distinguish confidence. Then we apply the negative-supervised principle presented in Section 3.2 to select negative samples. Next we use both C positive classes and additional $C+1$ class to train the following cascaded CNN.

4 Experiments

In this section, we introduce the positive training database and negative library. describe two baseline experimental results in comparison with our proposed method. The first experiment is trained on only positive samples. The second is trained on positive samples and randomly selected negative samples.

4.1 Database and Negative Library

The database, German Traffic Sign Recognition Benchmark(GTSRB)[22] shown in Fig.4 was created from approximately 10 hours of video that was recorded while driving on different road types in Germany during daytime. It is a large lifelike data set of more than 50,000 traffic sign images in 43 classes and contains images of more than 1700 traffic sign instances. The size of the traffic signs varies between 15×15 and 222×193 pixels [22].

Fig. 4. Partial Traffic Signs of GTSRB

The negative database is randomly selected from the SUN database[23], which is a comprehensive collection of annotated images covering a large variety of environmental scenes, places and the objects within. It has various outdoor scenes and using it as negative library can provide domain information for classification. We randomly cut patches from the SUN database and selected 100,000 patches that varies between 15×15 and 222×193 pixels, thus keeping correspondence with the positive database.

4.2 Training Convolutional Neural Networks

We apply the two baseline CNNs AlexNet and LeNet-5 for traffic sign classification on the benchmark GTSRB. The GTSRB data are split into two subsets. Set I is training data set, which contains 39209 training images in 43 classes. Set II is testing data set, which contains 12630 test images. Firstly, we use 43 classes positive samples in training set, directly apply five convolutional layers and following two fully-connected layers and then fed to a 43-way soft-max layer. Then we put nearly 300,000 negative samples to the network and get the confidence after soft-max layer. Next we make the threshold $T = 0.98$ and select negative samples whose maximum confidence more than T, thus we have approximately 40,000 positive samples and 100,000 negative samples plunged into the second CNN.

4.3 Experimental Results

This subsection give the classification rate of our proposed approach presented in Section 3 compared with positive-only samples training and additional randomly selected negative samples training. The only purpose of our experiments is to demonstrate that adding negative samples is useful to classification and if selecting negative samples scientifically can be more efficient. Our framework is compared with the basic CNNs and does not modify their architecture. For each basic CNN, we perform three schemes: positive samples only (Pos-only), positive and randomly selected negative samples (Neg-randomly), positive and negative-supervised negative samples (Neg-supervised).

Table 1. Classification Rate

Training Network	Pos-only	Neg-randomly	Neg-supervised
LeNet-5 [20]	87.2%	87.5%	87.9%
AlexNet [21]	91.7%	92.4%	92.8%

The classification rates are shown in Table.1. Firstly we simply use the two baselines LeNet-5 and AlexNet on only positive samples, the results are 87.2% and 91.7% respectively. Then we add negative samples generated randomly and the classification rate improve 0.3% and 0.7% each. Finally, we applied our negative-supervised mechanism to select negative samples and the rate reach

87.9% using LeNet-5 and 92.8% using AlexNet. These experiments show that useful information from domain related negative library can really help obtain related information and improve classification rate.

5 Conclusion and Discussion

This paper presented a negative-supervised cascaded deep convolutional neural network architecture. We proposed a negative sample selecting method to supervise CNN training. Experiments showed that only positive training set were not sufficient and can be improved by adding negative samples, especially specific selected negative samples. Negative-supervised mechanism can help utilize data more sufficient. And our framework can be utilized by all deep learning strategy and extended to other applications.

Acknowledgments. This work is supported in part by the National Natural Science Foundation of China (No.61402463), the Excellent Young Scientist Foundation of Institute of Information Engineering Chinese Academy of Sciences (No.1102008202), and the "Strategic Priority Research Program" of the Chinese Academy of Sciences (No.XDA06040101).

References

1. Fleyeh, H., Dougherty, M.: Road and traffic sign detection and recognition. In: Proceedings of the 16th Mini-EURO Conference and 10th Meeting of EWGT, pp. 644–653 (2005)
2. Yang, Y., Luo, H., Xu, H., et al.: Towards real-time traffic sign detection and classification. In: 2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), pp. 87–92. IEEE, Qingdao (2014)
3. Fu, M.Y., Huang. Y.S.: A survey of traffic sign recognition. In: Proceedings of the 2010 International Conference on Wavelet Analysis and Pattern Recognition(ICWAPR), pp. 119–124. IEEE, Qingdao (2010)
4. Mogelmose, A., Trivedi, M.M., Moeslund, T.B.: Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey. J IEEE Transactions on Intelligent Transportation Systems **13**, 1484–1497 (2012)
5. Maldonado-Bascn, S., Lafuente-Arroyo, S., Gil-Jimenez, P., et al.: Road-sign detection and recognition based on support vector machines. J. IEEE Transactions on Intelligent Transportation Systems **8**, 264–278 (2007)
6. Stallkamp, J., et al.: Man vs. computer: benchmarking machine learning algorithms for traffic sign recognition. Neural Networks **32**, 323–332 (2012)
7. Hasan, F., Janina, R.: Benchmark evaluation of HOG descriptors as features for classification of traffic signs. International Journal for Traffic and Transport Engineering **3**(4), 448–464 (2013)
8. Liu, C., Chang, F., Chen, Z.: Rapid multiclass traffic sign detection in high-resolution image. IEEE Transactions on Intelligent Transportation System **PP**(99), 1C–10 (2014)

9. Yang, Y., Luo, H., Xu, H., et al.: Towards real-time traffic sign detection and classification. In: IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), pp. 87C–92 (2014)
10. Ciresan, D., Meier, U., Masci, J., et al.: A committee of neural networks for traffic sign classification. In: The 2011 International Joint Conference on Neural Networks (IJCNN), pp. 1918–1921. IEEE, California (2011)
11. Ciresan, D., Meier, U., Masci, J., et al.: Multi-column deep neural network for traffic sign classification. J. Neural Networks **32**, 333–338 (2012)
12. Jin, J., Fu, K., Zhang, C.: Traffic sign recognition with hinge loss trained convolutional neural networks. J. IEEE Intelligent Transportation Systems Society **15**(5), 1991–2000 (2014)
13. Kumar, N., Berg, A.C., Belhumeur, P.N., et al.: Describable visual attributes for face verification and image search. In: 2013 IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 33(10), pp. 1962–1977. IEEE (2011)
14. Sun, Y., Chen, Y., Wang, X., et al.: Deep learning face representation by joint identification-verification. In: Advances in Neural Information Processing Systems (NIPS) (2014)
15. Zaklouta, F., Stanciulescu, B., Hamdoun, O.: Traffic sign classification using kd trees and random forests. In: The 2011 International Joint Conference on Neural Networks (IJCNN), pp. 2151–2155. IEEE, California (2011)
16. Greenhalgh, J., Mirmehdi, M.: Real-time detection and recognition of road traffic signs. J IEEE Transactions on Intelligent Transportation Systems **13**, 1498–1506 (2012)
17. Wang, G., Ren, G., Wu, Z., et al.: A hierarchical method for traffic sign classification with support vector machines. In: The 2013 International Joint Conference on Neural Networks (IJCNN), pp. 1–6. IEEE, Texas (2013)
18. Pan, S.J., Yang, Q.: A survey on transfer learning. J IEEE Transactions on Knowledge and Data Engineering **22**, 1345–1359 (2010)
19. Xu, C., Cetintas, S., Lee, K.C., et al.: Visual Sentiment Prediction with Deep Convolutional Neural Networks. J. arXiv preprint arXiv:1411.5731 (2014)
20. LeCun, Y., Bottou, L., Bengio, Y., et al.: Gradient-based learning applied to document recognition. J. Proceedings of the IEEE **86**, 2278–2324 (1998)
21. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
22. Stallkamp, J., Schlipsing, M., Salmen, J., et al.: The German traffic sign recognition benchmark: a multi-class classification competition. In: 2011 IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1453–1460. IEEE, California (2011)
23. Xiao, J., Hays, J., Ehinger, K.A., et al.: Sun database: large-scale scene recognition from abbey to zoo. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3485–3492. IEEE, San Francisco (2010)
24. Sermanet, P., LeCun, Y.: Traffic sign recognition with multi-scale convolutional networks. In: 2011 IEEE International Joint Conference on Neural Networks (IJCNN), pp. 2809–2813. IEEE, California (2011)