

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/327059637>

Transfer Learning Based Traffic Sign Recognition Using Inception-v3 Model

Article in *Periodica Polytechnica Transportation Engineering* · August 2018

DOI: 10.3311/PPtr.11480

CITATIONS

18

READS

918

5 authors, including:



Chunmian Lin

Beihang University (BUAA)

5 PUBLICATIONS 27 CITATIONS

[SEE PROFILE](#)



Lin li

Fujian Agriculture and Forestry University

17 PUBLICATIONS 129 CITATIONS

[SEE PROFILE](#)



Wenting Luo

Fujian Agriculture and Forestry University

17 PUBLICATIONS 128 CITATIONS

[SEE PROFILE](#)



Kelvin C. P. Wang

Oklahoma State University - Stillwater

167 PUBLICATIONS 2,441 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Multi-sensor Fused Perception for Autonomous Driving or Cooperative Vehicle Infrastructure System [View project](#)



ICC MDR408X [View project](#)

Transfer Learning based Traffic Sign Recognition Using Inception-v3 Model

Chunmian Lin¹, Lin Li^{1*}, Wenting Luo¹, Kelvin C.P. Wang², Jiangang Guo¹

Abstract

Traffic sign recognition is critical for advanced driver assistant system and road infrastructure survey. Traditional traffic sign recognition algorithms can't efficiently recognize traffic signs due to its limitation, yet deep learning-based technique requires huge amount of training data before its use, which is time consuming and labor intensive. In this study, transfer learning-based method is introduced for traffic sign recognition and classification, which significantly reduces the amount of training data and alleviates computation expense using Inception-v3 model. In our experiment, Belgium Traffic Sign Database is chosen and augmented by data pre-processing technique. Subsequently the layer-wise features extracted using different convolution and pooling operations are compared and analyzed. Finally transfer learning-based model is repetitively retrained several times with fine-tuning parameters at different learning rate, and excellent reliability and repeatability are observed based on statistical analysis. The results show that transfer learning model can achieve a high-level recognition performance in traffic sign recognition, which is up to 99.18% of recognition accuracy at 0.05 learning rate (average accuracy of 99.09%). This study would be beneficial in other traffic infrastructure recognition such as road lane marking and roadside protection facilities, and so on.

Keywords

Traffic Sign Recognition; Transfer Learning; Inception-v3 Model; Belgium Traffic Sign Database; traffic infrastructure maintenance

1 Introduction

Traffic sign recognition and classification are critical for advanced driver assistant system and road infrastructure survey (Li et al., 2016). Generally, traffic sign captured from visual cameras can be easily recognized and interpreted by human visual system. However, it is difficult to interpret for artificial machines due to some reasons such as occlusion, illumination, weather condition, exterior appearance of traffic signs, and so on. In order to eliminate influences of the existing issues on traffic recognition, substantial studies have been done to investigate automated traffic sign recognition methods, which can be roughly grouped into two categories: feature-based algorithm and deep learning-based technique.

Feature-based approach can be implemented by computer vision technique in two phases: (1) extracts useful features using the presented algorithms; (2) uses extracted features to classify traffic signs (Daugman, 1985; Lin et al., 2005; Dalal et al., 2005). Zhang et al. (2010) uses a binary tree of support vector machine (SVM) in local binary pattern (LBP) features for traffic sign recognition. Greenhalgh et al. (2012) extracts Histograms of Oriented Gradient (HOG) features from traffic sign images and utilizes a linear cascade of SVM for recognition and classification. Zaklouta et al. (2012) describes multi-scale HOG features and evaluates recognition performance using various classifiers, such as random forest classifier and SVM. Although feature-based approach can produce acceptable results on traffic sign recognition, it appears two limitations: (1) hand-engineering features need specialty-oriented knowledges and skills, which require both human expertise and labor; and (2) hand-engineering features are incapable to represent overall feature of traffic signs, resulting in unsatisfied recognition results.

With the development of deep learning technique, deep hierarchical neural network has drawn great attentions for traffic sign recognition. As German Traffic Sign Recognition Benchmark (GTSRB) is held by the International Joint Conference Neural Network (IJCNN) and IEEE Computational Intelligence Society (CIS), various deep learning-based models are designed and presented. (Tang et al., 2013; Tian et al., 2014; Mao et al., 2016). Notably, Ciresan et al. (2012) describes a multi-column architecture of deep neural

¹College of Transportation and Civil Engineering, Fujian Agriculture and Forestry University, Fuzhou, Fujian province, 350108, PR China

²School of Civil and Environmental Engineering, Oklahoma State University, Stillwater OK, USA74078

*(Corresponding author, e-mail: lilin531@gmail.com)

network, which yields the highest recognition accuracy of 99.46%. Sermanet et al. (2011) designs a multi-scale convolutional neural network, which reports a recognition accuracy of 99.17%. On contrast, the best recognition result from feature-based method is only 95.68% (Stallkamp et al., 2012). Although the result from deep convolutional neural network outperforms feature-based approach on traffic sign recognition, it still has two limitations: (1) deep learning model is generally designed by an iterative trial-and-error process, which requires a large amount of labeled data during training phase; and (2) a huge number of neuron connections would bring in heavy computation expense.

To overcome the abovementioned limitations, transfer learning strategy is introduced for traffic sign recognition in this study. Transfer learning, namely transfer of learning, is primarily proposed to explore how individuals would transfer learning in one context to another similar context (Thorndike & Woodworth, 1901). Currently transfer of learning is usually described as: the process and the effective extent to which past experiences affect learning performances in a new situation (Ellis et al., 1965). That is, a pre-trained model can be transferred to implement a similar task by learning new data distribution and fine-tuning parameters across all layers of the model. Substantial studies have been done on transfer learning-based image recognition (Raina et al., 2007; Devikar et al., 2016; Esteva et al., 2017). In this paper, transfer learning-based traffic sign recognition is developed by using Inception-V3 model (Szegedy et al., 2016). Subsequently, Belgium Traffic Sign Database is chosen, and data augmentation method is employed to enrich the training data. Finally, the performance of transfer learning-based method is evaluated. The results indicate that the proposed approach is robust for traffic sign recognition and can effectively overcome limitations of existing methods.

The rest of the paper is organized in Fig.1. Firstly, the architecture of convolutional neural network is introduced, and then Inception-v3 model is presented in Section II. In the part of case study, data pre-processing technique is developed, and convolutional feature representations are investigated based on a visual analytics toolkit. In addition, the performance of transfer learning-based model is evaluated in Section III. Finally, conclusions and recommendations are made in Section IV, V, respectively

2 Model Architecture

In this section, convolutional neural network architecture and its key components are firstly presented. Subsequently, Inception-v3 model is introduced to explore the Inception architecture. Finally, transfer learning-based Inception-v3 model is used for traffic sign recognition.

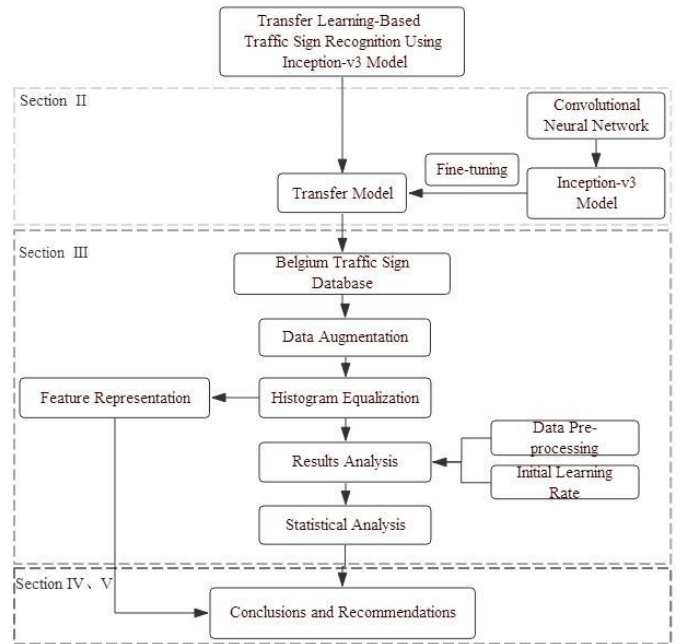


Fig.1 Paper organization architecture

2.1 Convolutional Neural Network

Convolutional neural network (CNN) is a multi-stage deep architecture that alternates convolutional layers with pooling or subsampling layers, followed by one or several fully connected layers. Its hierarchical architecture facilitates to learn invariant features and capture layer-wise feature representations from lower layers to higher layers. A standard convolutional architecture for digit recognition is presented in Fig.2 (Lecun et al., 1989). Herein, inputs feed forward through two-phase convolutional and subsampling operations to obtain feature representations, and then Gaussian classifier is used to produce probabilistic distribution. For convolutional neural network, it typically contains three key components, and is described below.

Convolutional operation produces the weighted sum of input pixel values by sliding a weighted window across the entire image, as presented in Fig.3(a). Subsequently, non-linear activation operation named activation function is applied to avoid learning trivial linear representations from the input. One of the most effective activation function is the Rectified Linear Unit (ReLU), which is a non-negative piecewise function that always obtains the maximum value between zero and the input, as mathematically described in Eq. (1):

$$f(x) = \max(x, 0) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (1)$$

Convolutional operation can be described as: the input (X) is convolved with a filter (W) of size $K_x \times K_y$. The resulting output (Y) is mathematically described in Eq. (2):

$$Y = f(\sum_{i=1}^n X_i * W + b_i) \quad (2)$$

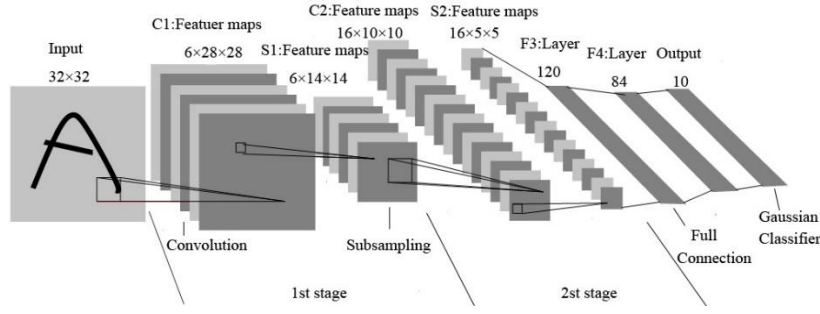


Fig. 2 Diagram of architecture of convolutional neural network (Lecun et al., 1989)

where n denotes the number of elements, $*$ is the sign of convolution operation, b_i is the bias of outputs, f is the activation function. In summary, 2D convolution operation is element-wise multiplication between input and weight matrix and calculate the weighted sum of input pixel values.

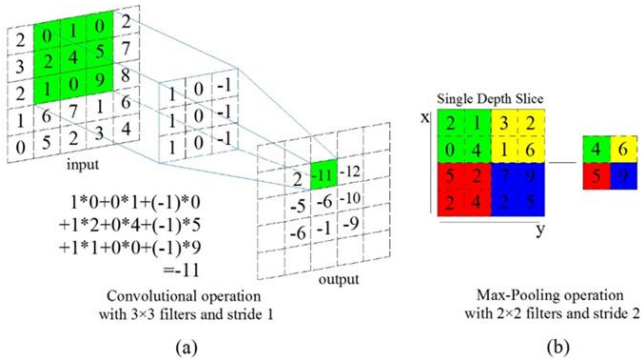


Fig. 3 Illustrations of convolution and max-pooling operation:

(a) convolutional operation; and (b) max-pooling operation.

Similar to convolution operation, pooling operation aggregates small pitches of pixel and subsamples features from the previous layer by sliding a weighted window across the pixel matrix. One of the most commonly used pooling operation is max-pooling, as illustrated in Fig.3(b). Max-pooling produces the maximum pixel value over the non-overlapping region of the weighted window. Pooling operation is widely applied for convolutional architecture to capture more significant feature representations, and to reduce computation expense.

Fully connected layers typically transform feature mappings into 1-D feature vectors by a series of affine transformations, and then a classifier is introduced to produce class-specific probabilistic distribution. For object recognition, SoftMax classifier is commonly utilized to normalize the label probability, as mathematically described in Eq. (3):

$$\text{softmax}(y_i) = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}} \quad (3)$$

Convolutional neural network has four unique keys, namely local connection, pooling operation, shared weight, and hierarchical architecture (Lecun et al., 2015). Previous studies

on visual cortex mechanism indicate that human cognition for real-world is from local to global. Hence, convolutional neural network is designed to mimic human visual mechanism: each neuron captures local features and integrates with local information to represent the whole image in higher layer. Through local connection, convolutional and max-pooling operations can obtain significant and unique feature representations of specific data. In addition, sliding weighted window in convolution and pooling operations are the use of sharing weights. The idea of sharing weight demonstrates that statistical characteristic for the whole image is spatial identity, that is, partial features can be shared. Therefore, the same weights can be used to extract features at each pixel location of the image. Moreover, to extract the layer-wise features, hierarchical architecture is introduced to explore correlations between neurons of adjacent layers. By contrast with fully connected network, distinctive architecture of CNN facilitates to extract effective features, and reduce parameter counts and neuron connections. In conclusion, convolutional neural network is the state-of-the-art technique on computer vision domain and is widely used for image recognition and classification purposes, especially for Inception-v3 model.

2.2 Inception-v3 Model

The annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an important competition on image recognition and classification, which contains 1.4 million images over 1000 object classes (Russklovsky et al., 2015). Krizhevsky et al. (2012) proposes the AlexNet model for object recognition and classification, and the remarkable progress is obtained. Subsequently several convolutional models are designed to reduce the Top-5 error rate of object recognition and classification. Table.1 shows detailed architectures of AlexNet, GoogleNet and ResNet are elaborated (Zeliet et al., 2014; Szegedy et al., 2015; He et al., 2016). Fig.4 shows Top-5 error rate of object recognition results based on ImageNet, and the outstanding recognition results can be observed for GoogleNet (Inception-v1). The findings indicate the deeper the model layer is, the better the recognition performance can be obtained.

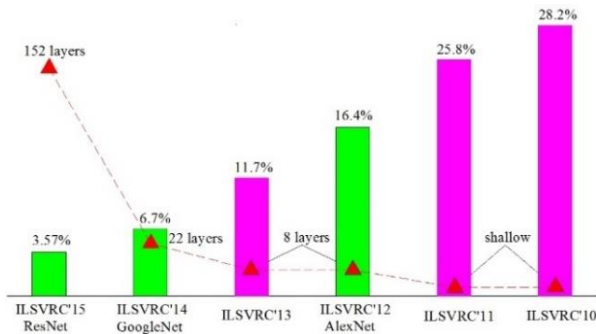
Table 1 Architectures of AlexNet, GoogleNet, ResNet

Model	Top-5 Rate	FCL & Size	Inception	DA	BN	LRN	DP
AlexNet	16.40%	3 & 4E+11	-	+	-	+	+
GoogleNet	6.70%	1 & 1000	+	+	-	+	+
ResNet	3.57%	1 & 1000	-	+	+	-	+

Comments: FCL & Size -- Full Connected Layer & Size; BN -- Batch Normalization

DA -- Data Augmentation; LRN -- Local Response Normalization;

DP-- Dropout;

**Fig. 4** Top-5 Error rate on ImageNet competition

Compared with GoogleNet (Inception-v1), Inception-v3 model has superior performance in object recognition. Specifically, Inception-v3 model includes three parts: the basic convolutional block, improved Inception module and the classifier. The basic convolutional block that alternates convolutional with max-pooling layers, is used for feature extraction. The improved Inception module is designed based on Network-In-Network (Lin et al., 2013), in which multi-scale convolutions are conducted in parallel and convolutional results of each branch are further concatenated. Due to the use of auxiliary classifiers, more stable training results and better gradient convergence are obtained, and simultaneously vanishing gradients and overfitting issues are alleviated as well.

In Inception-v3, 1×1 convolutional kernel is widely used to reduce the number of feature channels and accelerate training speed. In addition, the large convolution is decomposed into small convolutions, which reduces the number of parameters and computation expense. In summary, Inception-v3 has the state-of-the-art performance on object recognition, which

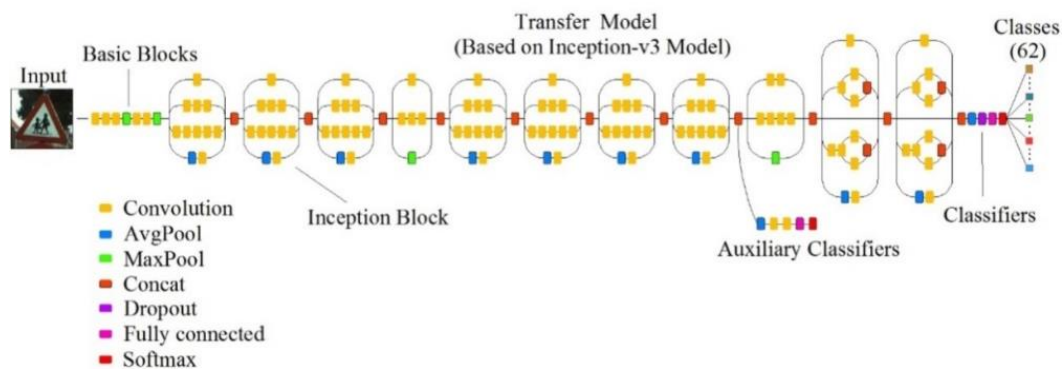
benefits from its unique Inception architecture. Therefore, this model is widely used for transfer learning.

2.3 Transfer Learning Model

Studies indicate that recognition and classification of a new image can be well implemented with Inception-v3 model by changing the architecture of fully-connected layers and reserving settings of all convolution layers (Raina et al., 2007).

Fig.5 shows the schematic illustration of the architecture of transfer learning-based model. The basic convolution block, the improved Inception modules, and the task-specific classifiers are sequentially concatenated based on Inception-v3 model. Specifically, low-level feature mappings are learned by basic convolutional operation with 1×1 and 3×3 kernels. In Inception module, multi-scale feature representations are concatenated to feed into auxiliary classifiers with diverse convolution kernels (i.e. 1×1 , 1×3 , 3×1 , 3×3 , 5×5 , 1×7 and 7×7 filters), which is used to produce better convergence performance. Followed by 11 Inception modules, fully connected layer is adopted to transform multi-scale feature vectors into one dimensional vector. Finally, Softmax classifier is used to produce one-hot vector, which is consistent to 62-class probability. The final classification result can be determined depending on the maximum value of 62-class probability.

Subsequently, case study is conducted on Belgium Traffic Sign Database to evaluate the performance of transfer learning-based model on traffic sign recognition.

**Fig.5** Illustration of transfer learning-based model

3 Case Study

First, Belgium Traffic Sign Database is augmented by data pre-processing technique. Subsequently, convolutional visualization results and relationships between adjacent convolutional layer are analyzed. Finally, extensive experiments are conducted to evaluate recognition performance of transfer learning-based model, and statistical analysis is developed to investigate the reliability and repeatability of this model.

3.1 Data Pre-processing

Belgium Traffic Sign Database is chosen for transfer learning-based model evaluation. It includes 4575 training images over 62 classes. Note that there are a number of bad samples due to various reasons such as illumination variations, distortion, motion-blur, obstacle blocks or color fading, as presented in Fig.6. The recognition of these poor data would be beneficial from transfer learning-based model since it would help the poor data to improve generalization and boost robustness.

Here, data augmentation method is employed to enlarge the database through randomly rotating, scaling and translating for original training data. After data augmentation, a new database containing 10775 training images and 4575 test images is obtained. In this study, the original Belgium Traffic Sign Database is denoted as NDA, and the new database after data augmentation is denoted as DA.



Fig.6 Bad or poor samples in database

Due to the limitation of mapping dimension, the image data are tailored into a constant resolution of 299×299 . Subsequently intensity contrast among samples are improved with the equalization method. In this study, gray intensity values are normalized into $[0,1]$ from $[0,255]$. For one gray image, the sum of pixels is denoted by N , and n_j denotes the number of gray intensity j . After normalization, pixel value s is mathematically described as Eq. (4)

$$s = \sum_{j=0}^k \frac{n_j}{N} \quad (4)$$

For RGB image, histogram equalization is conducted on three individual channels, and then the equalization result of

each channel is aggregated. Fig.7 (a) shows the frequency distribution after histogram equalization operation. It can be observed that the intensity contrast is smoother and more uniform than that in original figures. The similar findings are observed in Fig.7 (b).

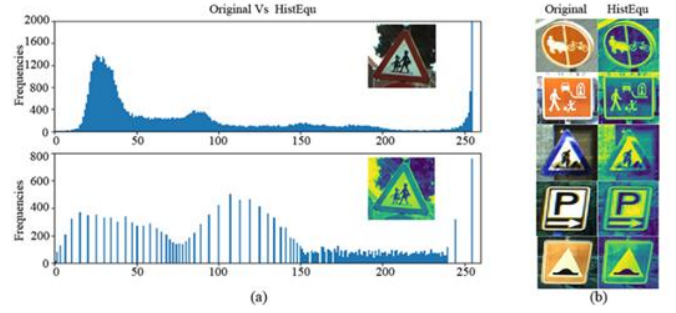


Fig.7 Frequency distribution of pixel intensity

Findings indicate data augmentation is a practical processing technique to increase the number of labeled training samples by geometrical transformation, and histogram equalization is powerful method to eliminate effects of poor samples on image recognition and classification.

3.2 Feature Representation

To better understand convolutional mechanism, a visual analytics toolkit is used to examine feature representation for each convolutional operation, as shown in Fig.8. Firstly, RGB image with a resolution of 299×299 , is used as the inputs of transfer learning-based model. The three feature mappings represent three types of color channels, namely red, green, and blue channels.

As abovementioned, transfer learning-based model includes three components: the basic convolutional block, Inception modules and Softmax classifier. The basic convolutional block contains 5 convolution layers, from conv2d_1 to conv2d_5. In conv2d_1 layer, 32 filters with a size of 3×3 are used to extract low-level features, and thus 32-channel feature maps with a size of 149×149 are produced in conv2d_1 layer. Similarly, 32, 64, 80 and 192-channel feature representations are obtained in the subsequent convolution layers, with a size of 147×147 , 147×147 , 73×73 and 71×71 , respectively, as illustrated in Fig 8.

Inception module is the core block of transfer learning-based model. 256-channel feature representations with a resolution of 35×35 are presented in the first Inception module 'mixed0'. Subsequently, the layer-wise convolutional operations are further developed, and multi-scale feature representations are obtained in the following Inception modules. In general, more abstract feature representations are observed with the increment of convolutional layer, and 2048-channel feature mappings with a size of 8×8 are presented in the last Inception module 'mixed9', as illustrated in Fig 8.

Finally, multi-scale feature representations are flattened to

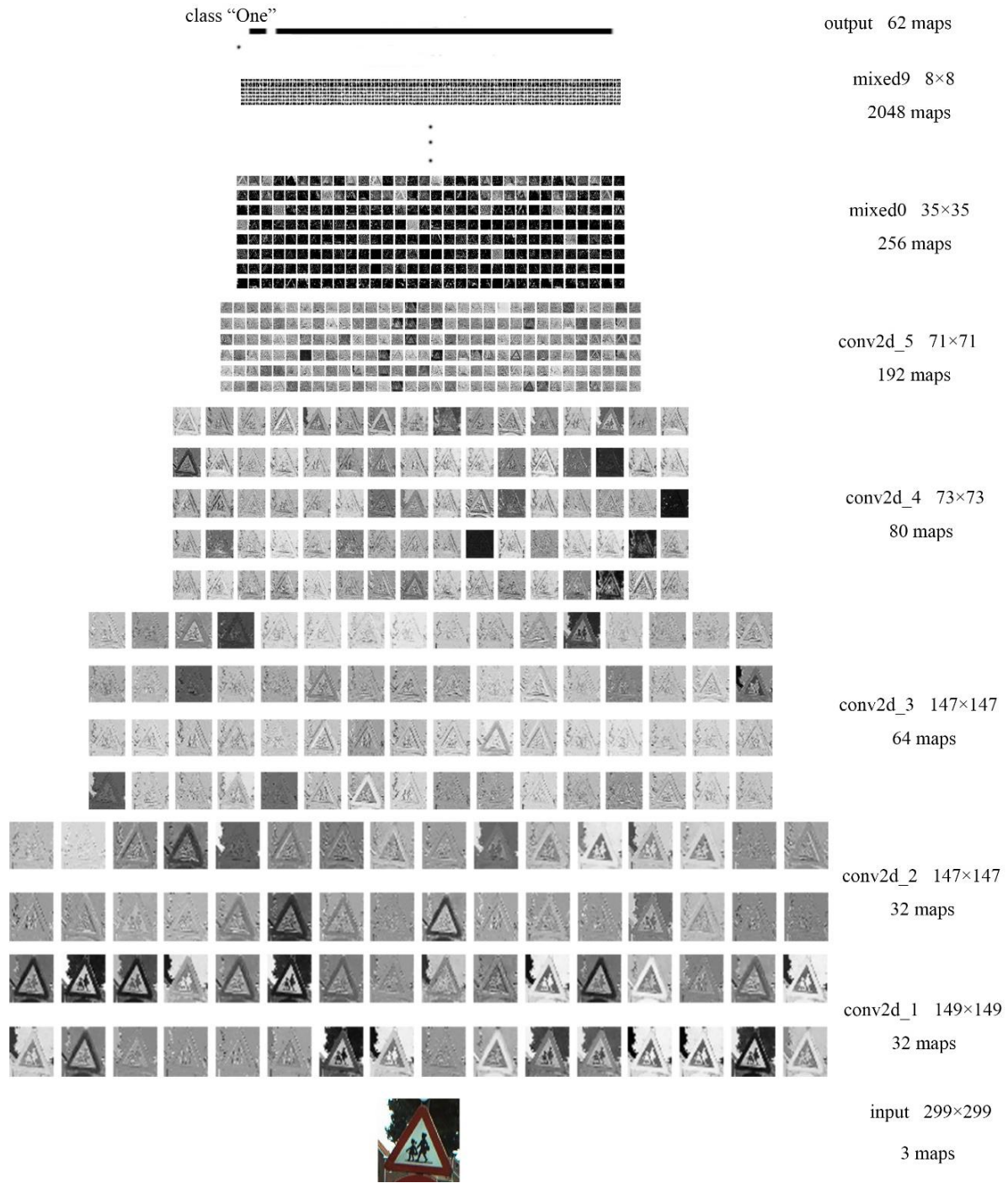


Fig.8 Visualization result of convolutional operation and feature mapping

1-D vector, and 62-neuron output corresponds to 62-class probability. The neuron with maximum probability would be the class label that the tested traffic sign belongs to, as shown

in Fig.8. In addition, parameter information in convolutional visualization result is consistent to architecture configuration, as shown in Table.2.

Table 2 The architecture configuration of transfer learning-based model

Layers	Maps & Size	Filter Size	Layers	Maps & Size	Filter Size
input	3&299×299	-	mixed3	768&17×17	1×1,3×3
conv2d_1	32&149×149	3×3	mixed4	768&17×17	1×1,3×3,1×7,7×1
conv2d_2	32&147×147	3×3	mixed5	768&17×17	1×1,3×3,1×7,7×1
conv2d_3	64&147×147	3×3	mixed6	768&17×17	1×1,3×3,1×7,7×1
conv2d_4	80&73×73	1×1	mixed7	768&17×17	1×1,3×3,1×7,7×1
conv2d_5	192&71×71	3×3	mixed8	1280&8×8	1×1,3×3,1×7,7×1

mixed0	256&35×35	1×1,3×3	mixed9	2048&8×8	1×1,3×3,1×3,3×1
mixed1	288&35×35	1×1,3×3,5×5	mixed1	2048&8×8	1×1,3×3,1×3,3×1
mixed2	288&35×35	1×1,3×3	output	62&62×1	-

Comments: The prefix conv2d_5: convolutional layers in basic convolutional block.

The prefix mixed: 11 Inception modules.

3.3 Training Details

Two databases (NDA and DA) are used to re-train transfer learning model based on Tensorflow machine learning framework (Abadi et al., 2015). The model is repetitively trained 5000 epochs. Each porch is one round of forward and backward propagation iteration. To evaluate the model performance, different initial learning rates are given, and the exponential decayed rate is 0.94. In this study, 10 percentage of training samples are randomly chosen for validation samples, and the model are validated at each 100 epochs. Through 5000 training epochs, the performance of model can be evaluated.

3.4 Result Analysis

Fig.9 shows the validation and test accuracy of the transfer learning-based model. Note that the better validation and test

accuracy is obtained when the training is conducted on the new database (DA) at the same initial learning rate. The gray areas in Fig10 represent the difference of validation accuracy trained based on two databases (NDA and DA). Besides, findings indicate different initial learning rates result in different recognition results. The recognition accuracy increases with the increase of the initial learning rate when the learning rate is less than 0.05. In this study, the best test accuracy of 99.18% is observed at the initial learning rate of 0.05.

Moreover, other approaches or algorithms are also used for traffic sign recognition based on the same database, and the best recognition accuracy is 80.00%, which is 18% less than the transfer learning-based method. It can be concluded that transfer learning-based method is powerful for traffic sign recognition.

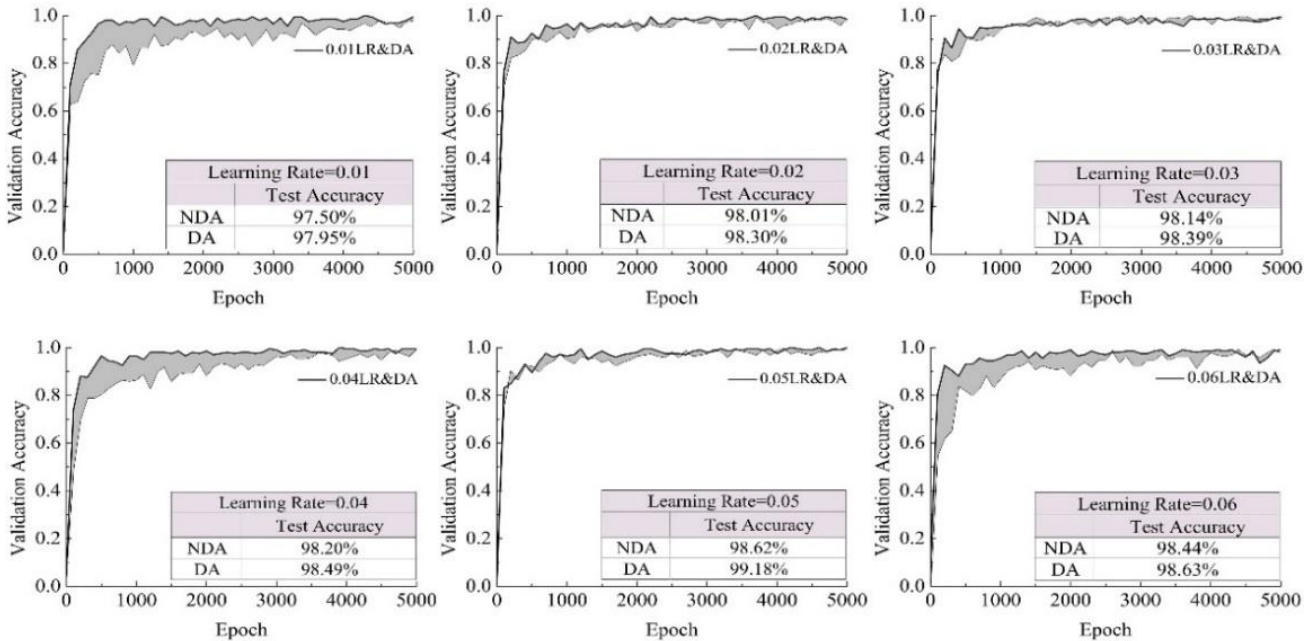


Fig. 9 Recognition performance at different learning rate on different databases

To further investigate the reliability and repeatability of transfer learning-based method, ten repetitive experiments are conducted at eight different learning rates, respectively. Table.3 shows statistical results of recognition accuracy for the ten repetitive experiments. Note that the average test accuracy of 99.06% is obtained when the training is conducted at the learning rate of 0.05. In addition, other statistical indices (e.g. variance and CoV) indicate the transfer learning-based method can produce the reliable and repetitive recognition result.

4 Conclusions

In this paper, transfer learning-based method is used for traffic sign recognition. First, data pre-processing technique is used to enhance Belgium Traffic Sign Database, including data augmentation and histogram equalization. Subsequently, data feature representation is demonstrated by a visual analytics toolkit, in which layer-wise convolutional feature representation is analyzed. Based on Belgium Traffic Sign

Table. 3 Statistical analysis results of recognition accuracy at different learning rates

LR	0.01	0.02	0.03	0.04	0.045	0.05	0.055	0.06
Samples	10	10	10	10	10	10	10	10
Max	0.9807	0.9830	0.9839	0.9849	0.9890	0.9918	0.9900	0.9879
Min	0.9788	0.9818	0.9829	0.9840	0.9868	0.9895	0.9878	0.9846
Mean	0.9795	0.9824	0.9835	0.9845	0.9877	0.9906	0.9889	0.9860
Variance	3.46E-07	1.39E-07	9.56E-08	6.89E-08	5.99E-07	4.88E-07	4.90E-07	6.98E-07
Qd	0.0009	0.0005	0.0005	0.0004	0.0013	0.0001	0.0014	0.00095
CoV	6.01E-04	3.79E-04	3.14E-04	2.67E-04	7.84E-04	7.05E-04	7.08E-04	8.47E-04
Kurtosis	2.4409	2.3799	2.1531	2.5520	1.9146	2.2670	2.1158	4.0345
Skewness	0.6374	0.4365	-0.3491	-0.3019	0.4567	-0.0794	-0.3933	0.5916

Comments: Qd -- Quartile Deviation; CoV -- Coefficient of Variation

Database, transfer learning-based model is retrained for 5000 epochs at different learning rates. The accuracy test results indicate that the transfer learning-based method are powerful for traffic sign recognition, with the best recognition accuracy of 99.18% at the learning rate of 0.05. Moreover, the repetitive experiments are conducted at different initial learning rates, and findings indicate the reliable and repetitive recognition results can be obtained using transfer learning-based method. Therefore, the transfer learning-based method is robust and powerful in traffic sign recognition, which would be beneficial in other traffic infrastructure maintenance such as lane marking and roadside protection facilities.

5 Recommendations

Hyperparameter optimization would be explored to improve model efficiency in future study. Besides, different traffic sign databases would be employed to further validate the accuracy and reliability performance of transfer learning-based method.

Acknowledgements

This work was supported by ‘Digital Fujian’ Key Laboratory of Internet Things for Intelligent Transportation Technology, and funded by Chinese National Natural Fund for Young Scholars under grant No: 51608123, Fujian Natural Science Funds under grant No: 2017J01475 and 2017J01682.

References

- Abadi, M., Agarwal, A., Barham, P., et al. (2016) Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems. *arXiv preprint arXiv:1603.04467*.
- Ciresan, D., MERIER, U., MASCI, J. (2012) Multi-Column Deep Neural Network for Traffic Sign Classification. *IEEE Conference on Computer Vision and Pattern Recognition*, Rhode Island, USA. DOI: [10.1109/CVPR.2012.6248110](https://doi.org/10.1109/CVPR.2012.6248110)
- Daugman, J. G. (1985) Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-dimensional Visual Cortical Filters. *Journal of the Optical Society of America*, 2(7), pp. 1160-1169. DOI: [10.1364/JOSAA.2.001160](https://doi.org/10.1364/JOSAA.2.001160)
- Dalal, N., Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA. DOI: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177)
- Devikar, P. (2016) Transfer Learning for Image Classification Various Fog Breed. *International Journal of Advanced Research in Computer Engineering & Technology*. 5(12), pp. 2278-2323.
- Esteva, A., Kuprel, B., Novoa, R. (2017) Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature*. 542, pp. 115-118. DOI: [10.1038/nature21056](https://doi.org/10.1038/nature21056)
- Greenhalgh, J., Mirmehdi, M. (2012) Real-time Detection and Recognition of Road Traffic Signs. *IEEE Intelligent Transportation Systems*, 13(4), pp. 1498-1506. DOI: [10.1109/TITS.2012.2208909](https://doi.org/10.1109/TITS.2012.2208909)
- He, K.M., Zhang, X. Y., Ren, S. Q., et al. (2016) Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA. DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)
- Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*. Tahoe, USA.
- Lecun, Y., Boser B., Denker J. S., et al. (1989) Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*. 1(4), pp. 541-551. DOI: [10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541)
- Lecun, Y., Bengio, Y., Hinton, G. E. (2015) Deep Learning. *Nature*. 521, pp. 436-444. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539)
- Li, L., Wang, K. C. P. (2016). Bounding box-based technique for pavement crack classification and measurement using 1 mm 3d laser data. *Journal of Computing in Civil Engineering*, 30(5), 04016011. DOI: [10.1061/\(ASCE\)CP.1943-5487.000056](https://doi.org/10.1061/(ASCE)CP.1943-5487.000056)
- Lin, M., Chen, Q., Yan, S. (2013) Network in Network. *arXiv preprint arXiv:1312.4400*.
- Liu, C. L., Koga, M., Fujisawa, H. (2005) Gabor Feature Extraction for Character Recognition: Comparison with Gradient Feature. *International Conference on Document Analysis and Recognition*. Seoul, Korea. DOI: [10.1109/ICDAR.2005.119](https://doi.org/10.1109/ICDAR.2005.119)
- Mao, X. H., Hijazi, S., Kaul, P., et al. (2016) Hierarchical CNN

- for Traffic Sign Recognition. *IEEE Intelligent Vehicles Symposium*. pp. 130-135. DOI: [10.1109/IVS.2016.7535376](https://doi.org/10.1109/IVS.2016.7535376)
- Raina, R., Battle, A., Lee, H., et al. (2007) Self-Taught Learning: Transfer Learning from Unlabeled Data. *Proceedings of International Conference on Machine Learning*. Corvallis, OR, USA. DOI: [10.1145/1273496.1273592](https://doi.org/10.1145/1273496.1273592)
- Russakovsky, O., Deng, J., Su, H. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*. 115(3), pp. 211-252.
- Sermanent, P., Lecun, Y. (2013) Traffic Sign Recognition with Multi-Scale Convolutional Network. *Proceedings of International Joint Conference on Neural Networks*. San Jose, CA, USA. DOI: [10.1109/IJCNN.2011.6033589](https://doi.org/10.1109/IJCNN.2011.6033589)
- Stallkamp, J., Schlipsing, M., Salmen, J., et al. (2012) Man vs Computer: Benchmarking Machine Learning Algorithms for Traffic Sign Recognition. *Neural network*, 32: pp. 323-332. DOI: [10.1016/j.neunet.2012.02.016](https://doi.org/10.1016/j.neunet.2012.02.016)
- Szegedy, C., Liu, W., Jia, Y. Q. (2015) Going Deeper with Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA. DOI: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594)
- Szegedy, C., Vanhoucke, V., Ioffe, S. (2016) Rethinking the Inception Architecture for Computer Vision. *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA. DOI: [10.1109/CVPR.2016.308](https://doi.org/10.1109/CVPR.2016.308)
- Tang, S. S., Huang, L. L. (2013) Traffic Sign Recognition Using Complementary Features. *Asian Conference on Pattern Recognition*. Okinawa, Japan. DOI: [10.1109/ACPR.2013.63](https://doi.org/10.1109/ACPR.2013.63)
- Thorndike, E. L., Woodworth, R. S. (1970) The Influence of Improvement in One Mental Function Upon the Efficiency of Other Functions. *Psychological Review*. 8, pp. 384-395.
- Tian, T., Sethi, I., Patel, N. (2014) Traffic Sign Recognition Using a Novel Permutation-Based Local Image Feature. *Proceedings of International Joint Conference on Neural Networks*. Beijing, China. DOI: [10.1109/IJCNN.2014.6889629](https://doi.org/10.1109/IJCNN.2014.6889629)
- Zaklouta, F., Stanculescu, B. (2012) Real-Time Traffic-Sign Recognition Using Tree Classifiers. *IEEE Transaction on Intelligent Transportation Systems*. 13(4), pp. 1507-1514. DOI: [10.1109/TITS.2012.2225618](https://doi.org/10.1109/TITS.2012.2225618)
- Zelner, M., Fergus, R. (2014) Visualizing and Understanding Convolutional Networks. *The 13th European Conference on Computer Vision*. 8689, pp. 818-833.
- Zhang, Y., Hong, C. (2010) An Efficient Real Time Rectangle Speed Limit Sign Recognition. *IEEE Intelligent Vehicles Symposium*, California, USA. DOI: [10.1109/IVS.2010.5548140](https://doi.org/10.1109/IVS.2010.5548140)

