Regular paper

Journal of the KIECS. pp. 105-110, vol. 17, no. 1, Feb. 28. 2022, t. 111, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2022.17.1.105

교통 신호 인식을 위한 경량 잔류층 기반 컨볼루션 신경망

Kodirov Shokhrukh* · 류재홍**

Lightweight Residual Layer Based Convolutional Neural Networks for Traffic Sign Recognition

Kodirov Shokhrukh* · Jae Hung Yoo**

요 약

교통 표지 인식은 교통 관련 문제를 해결하는 데 중요한 역할을 한다. 교통 표지 인식 및 분류 시스템은 교통안전, 교통 모니터링, 자율주행 서비스 및 자율주행 차의 핵심 구성 요소이다. 휴대용 장치에 적용할 수 있는 경량 모델은 설계 의제의 필수 측면이다. 우리는 교통 표지 인식 시스템을 위한 잔여 블록이 있는 경량 합성곱 신경망 모델을 제안한다. 제안된 모델은 공개적으로 사용 가능한 벤치마크 데이터에서 매우 경쟁력 있는 결과를 보여준다.

ABSTRACT

Traffic sign recognition plays an important role in solving traffic-related problems. Traffic sign recognition and classification systems are key components for traffic safety, traffic monitoring, autonomous driving services, and autonomous vehicles. A lightweight model, applicable to portable devices, is an essential aspect of the design agenda. We suggest a lightweight convolutional neural network model with residual blocks for traffic sign recognition systems. The proposed model shows very competitive results on publicly available benchmark data.

키워드

Convolutional Neural Networks, Residual Blocks, Traffic Sign Detection
합성곱 신경망, 잔여 블록, 교통 표지 검출

Ⅰ. Introduction

Traffic sign recognition systems have been an interesting topic in the research community for the past 30 years. The main reason is the indispensability of protecting human life and saving more people on the road. From one study to another, scientists worked to improve the accuracy and recognition rates of these systems. The proposed models are mainly divided into two

* 전남대학교 컴퓨터공학과 석사과정 재학
(shoxrux.qodirov@gmail.com)
** 교신저자: 전남대학교 컴퓨터공학과
• 접 수 일 : 2022. 01. 03
• 수정완료일 : 2022. 01. 25
• 게재확정일 : 2022. 02. 17
• Received : Jan. 03, 2022, Revised : Jan. 25, 2022, Accepted : Feb. 17, 2022
• Corresponding Author : Jae Hung Yoo
Dept. of Computer Engineering, Chonnam Nat. Univ
Email : jhy@jnu.ac.kr

categories: non-automated feature extraction approaches and deep learning-based automated methods. Prior to deep learning, classical recognition models were applied along with manual labeling and feature extraction from specific color acceptance and machine learning-based models, dramatically reducing model efficiency in both accuracy and speed [1-2]. As is evident, manual labeling requires more work and the model does not guarantee the target performance.

In the post deep learning era, detection and classification methods have changed significantly. Neural network study has gradually become a popular field in the research community. Automation of models can eliminate the laborious manual annotation and automatically extract the features from the input data. In particular, Convolutional Neural Networks have shown improved accuracy in the fields of grid data classification and object detection [3].

The rest of the paper is arranged in that way: Section II review some related works on traffic recognition systems. Dataset is described in Section III. Section IV details proposed model architecture. In Section V, we show the results of conducted experiments. Finally, conclusion and future work are drawn in Section VI [4-23].

II. Related works

Methods for traffic sign recognition systems have been extended from color and shape based models to machine learning and finally deep learning based models. CNNs have attracted attention in the field of feature extraction and pattern recognition over the past decade and have been widely applied to tasks related to classification and detection [3].

One of the most common approaches is a color-based model. Different color spaces are used for segmentation of road images such as HSV, RGB, and HIS [5,8,10]. Others are shape-based models. The effectiveness of the circular traffic sign recognition system is studied by [1]. Symmetry information for circles, triangles, squares and octagons was adopted by [4]. The Hough feature extraction model was proposed by [7].

Fine grained classification applying different methods through a pipeline of three stages: feature extraction, dimensionality reduction and classification was proposed by [13]. They merged grayscale values of traffic sign images and Histogram of Oriented Gradients (HOG) based features, reducing the dimensionality through Iterative Nearest Neighbors-based Linear Projections (INNLP) and classifying with Iterative Nearest Neighbors. Shaped based detection algorithms to recognize traffic signs was proposed by [6]. The authors chose convolutional neural network for the purpose of classification. After simulation they obtained higher accuracy result of the benchmark.

III. Dataset

We applied German Traffic Sign Recognition Benchmark (GTSRB) [9] for training and testing the proposed model. Input images were randomly obtained from the camera in a real-time scene and was first provided by the International Joint Conference on Neural Networks (IJCNN) in 2011. The dataset includes 51838 images in 43 categories of each image resolution is dynamically changed from 15x15 to 250x250 pixels. The images with constantly changing scales increase the richness of the data and can improve the fitting effect of the network. Table 1 shows the distribution of GTSRB dataset for training, validation, and testing. The validation data was chosen as 15% of the given training data set.

Table 1. Cardinality of GTSRB dataset

Training	Validation	Testing	Total
33326	5882	12630	51838

Figure 1 shows some examples of the dataset. As it is clear from the examples that dataset has different outside variations.

Fig. 1 Example images from GTSRB dataset.

Ⅳ. Proposed Architecture

We introduce a traffic sign recognition system that carries out fine-grained classification to traffic sign images through a lightweight CNN whose main blocks are convolutional and residual block modules. In order to find an accurate and efficient model for our purpose, we first study and discuss the two steps: the effect of processing the input data prior to model computation and the structure of the proposed model architecture. Finally, we discuss weight initialization for the model. Figure 2 shows our overall architecture.

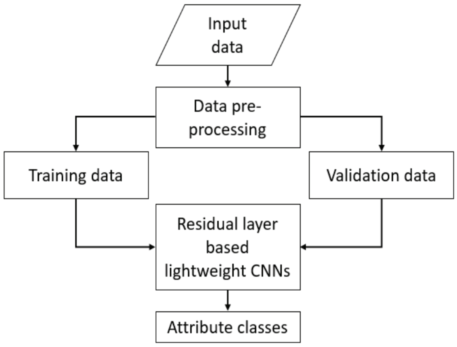

Fig. 2 Flowchart of proposed model architecture.

4.1 Data Pre-processing

Traffic sign samples from GTSRB are raw RGB and sizes vary from 15x15 to 250x250 pixels. To make compatible input data to our model process, all input data are down-sampled or up-sampled to 32x32 pixel size. Furthermore, we also use data normalization method, a specific way to normalize the image pixel value from (0, 255) to (0, 1), that has a similar data distribution for the entire dataset before feeding to the CNNs part. Image normalization process helps the network training to converges faster and better, accelerating the model process speed. The process is run by subtracting the dataset mean from the input and divide it by the standard deviation of that feature as well for each RGB channels:

$$z = (x - mean)/Stdev + 0.5\ . \tag{1}$$

4.2 CNN with Residual Blocks

Our CNN architecture is consisted of overall six feature extraction layers, of which two of them are residual blocks and final layer for classification. Full view of the CNN model is given in Figure 3.

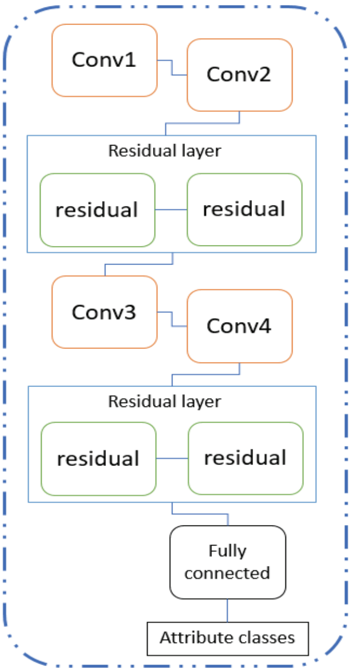

Fig. 3 CNN part overview.

The first convolution layer is a 3x3 kernel filter applied. Max pooling is not applied to preserve initial features. In the second layer, the kernel filter is equivalent to an additional 2x2 max-pooling layer to reduce the dimensions of the previous layer. A two-part residual layer is then applied to the skip connection, especially to avoid performance degradation. After that, two convolutional layers with max-pooling and one residual block are connected to the model. Finally, we perform a dimensional transformation in a fully connected layer on the extracted output features and apply a softmax activation function to classify one of 43 traffic signs.

The first layer has 64 filters, the second layer has 192 filters, the initial residual block has 192 filters, and each subsequent layer has the same 256 layers for the next residual block. The model uses only a fraction of the memory in terms of the total trainable weights.

Rectified Linear Units (ReLU) were applied non-linearly through each layer and a dropout of 20% was applied as a normalization step.

4.3 Weight Initialization

We train the proposed model from scratch without additional data or pre-trained weights. A random value with a mean Gaussian distribution and zero standard deviation initializes the weights for both convolutional and fully connected layers.

Ⅴ. Experimental Results

The performance of the model is analyzed by conducting experiments on the GTSRB dataset. PyTorch open source deep learning framework is adopted for implementation. For the model cost function, multi-class Cross-entropy loss is applied.

$$Loss = \sum_{i=1}^{n}[y_i \log a_i + (1-y_i)\log(1-a_i)] \quad . \qquad (2)$$

Here y_i is target a_i is activation of the logistic function.

All layers are adjusted using Adam (Adaptive moment estimation) optimization technique [11] with 256 batch size. Learning rate is 1e-4, respectively.

Plot of training and validation accuracy is given in Figure 4. There is no extra fluctuation of validation accuracy even if the validation data is randomly selected without stratification. This can be explained by the robustness of Adam gradient descent [11]. We compares the accuracy on the GTSRB dataset in the Table 2.

Accuracy may be not the best metric to measure the model efficiency. Thus we evaluated our model furthermore using the classwise weighted average of precision, recall and F1 score in the Table 3.

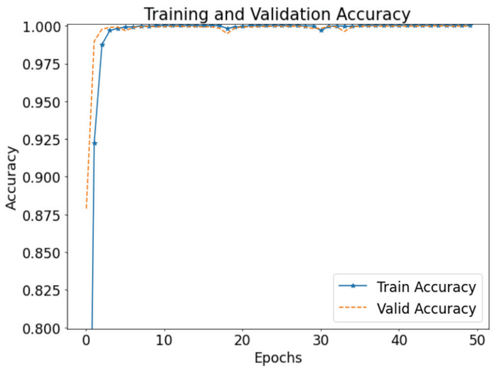

Fig. 4 Train and valid accuracy vs epochs.

Table 2. Accuracy comparison on GTSRB

Models	Accuracy(%)
Modified LeNet-5 [14]	95.20
LeNet architecture [13]	96.23
Small CNN [16]	97.40
Improved LeNet-5 [15]	98.12
Deep CNN [18]	98.96
Efficient CNN [17]	99.66
Proposed model	99.09

Table 3. Classwise Weighted Average Metrics

Precision	Recall	F1 score
0.991	0.991	0.991

Ⅵ. Conclusions

We propose a lightweight residual layer-based convolutional neural network for traffic sign recognition problems. The proposed model is very fast compared to other deep CNN models and uses only a fraction of the memory. Our future work will be modifying and testing the proposed model on other small sized classification problems.

References

[1] S. K. Berkaya, H. Gunduz, O. Ozsen, C. Akinlar, and S. Gunal, "On circular traffic sign detection and recognition," *Expert Systems with Applications*, vol. 48, 2016, pp. 67-75.

[2] Y. Yang, H. Luo, H. Xu, and F. Wu, "Toward real-time traffic sign detection and classification," *IEEE Trans. on Intelligent transportation systems*, vol. 17 no. 7, 2015, pp. 2022-2031.

[3] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech and time series," In M. A. Arbib, ed., *The handbook of brain theory and neural networks*, Cambridge, MA, USA, MIT Press, 1995, pp. 276-279.

[4] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," *2004 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 1, 2004, pp. 70-75.

[5] W. G. Shadeed, D. I. A, A. Nadi, and M. J. Mismar, "Road traffic sign detection in color images," *IEEE conf.: electronics circuits and systems*, vol. 2, 2003, pp. 14-17.

[6] W. J. Yang, C. C. Luo, P. C. Chung, and J. F. Yang, "Simplified Neural Networks with Smart Detection for Road Traffic Sign Recognition," *Lecture Notes in Networks and Systems*, vol 69, 2020, pp. 237-249.

[7] N. Barnes, G. Loy, and D. Shaw, "The regular polygon detector," *Pattern Recognition*, vol. 43, 2010, pp. 592-602.

[8] D. L. Escalera, L. Moreno, M. Salichs, and J. Armiongol, "Road traffic sign detection and classification," *IEEE Trans. on Industrial Electronics*, vol. 44, 1997, pp. 848-859.

[9] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: a multi-class classification competition," In *IEEE. Int. joint conf. on neural networks*, San Jose, CA, USA, 2011, pp. 1453-1460.

[10] 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 Trans. on Intelligent Transportation Systems*, vol. 8, 2007, pp. 264-278.

[11] D. P. Kingma and L. J. Ba, "Adam, A method for stochastic optimization," In *Proc. of the Int. Conf. for Learning Representations*,

San Diego, CA, USA, May 2015, pp. 1-15.
[12] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, "Traffic sign recognition – How far are we from the solution?," *The 2013 Int. Joint Conf. on Neural Networks,* Dallas, TX, USA, 2013, pp. 1-8.
[13] A. Bouti, M. A. Mahraz, J. Riffi, and H. Tairi, "A robust system for road sign detection and classification using LeNet architecture based on convolutional neural network," *Soft Computing*, vol. 24, no. 9, 2020, pp. 6721-6733.
[14] D. Yasmina, R. Karima, and A. Ouahiba, "Traffic signs recognition with deep learning," in *2018 Int. Conf. on Applied Smart Systems,* Medea, Algeria, Nov. 2018, pp. 1-5.
[15] W. Li, X. Li, Y. Qin, W. Song, and W. Cui, "Application of improved LeNet-5 network in traffic sign recognition," in *Proc. of the 3rd Int. Conf. on Video and Image Processing,* Shanghai China, Dec. 2019, pp. 13-18.
[16] W. Li, D. Li, and S. Zeng, "Traffic sign recognition with a small convolutional neural network," *IOP Conf. Series: Materials Science and Engineering,* vol. 688, article 044034, 2019.
[17] J. Li and Z. Wang, "Real-time traffic sign recognition based on efficient CNNs in the wild," *IEEE Trans. on Intelligent Transportation Systems*, vol. 20, no. 3, 2019. pp. 975-984.
[18] S. Yin, J. Deng, D. Zhang, and J. Du, "Traffic sign recognition based on deep convolutional neural network," in *Computer Vision: Second CCF Chinese Conf., CCCV 2017,* Tianjin, China, 2017, pp. 685-695.
[19] S. Salimov and J. Yoo, "A Design of Small Scale Deep CNN Model for Facial Expression Recognition using the Low Resolution Image Datasets," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 16 no. 1, 2021, pp. 75-80.
[20] J. Yoo, "An Extension of Unified Bayesian Tikhonov Regularization Method and Application to Image Restoration," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 15, no. 01, 2020, pp. 161-166.
[21] J. Yoo, "A Unified Bayesian Tikhonov Regularization Method for Image Restoration," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 11, no. 11, 2016, pp. 1129-1134.
[22] J. Yoo, "Self-Regularization Method for Image Restoration," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 11, no. 1, 2016, pp. 45-52.
[23] S. Park, "Optimal QP Determination Method for Adaptive Intra Frame Encoding," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 10, no. 9, 2015, pp. 1009-1018.

저자 소개(Authors)

2018년 Tashkent University of Information, Software Engineering 공학과 졸업(공학사)
2021년 현재 전남대학교 대학원 컴퓨터공학과 재학

※관심분야: 심층신경망, 영상처리, 컴퓨터비전

류재홍(Jae-Hung Yoo)

1981년 한양대학교 기계공학과 졸업(공학사) (BE in Mechanical Engineering from Hanyang Univ. in 1981)
MA in Computer Science from Univ. of Detroit in 1986
PhD in Computer Science from Wayne State Univ. in 1993
Joined as a faculty member in the Dept. of Computer Engineering, Yosu Nat. Univ. in 1994
Became a faculty member in the Dept. of Computer Engineering, Chonnam Nat. Univ. in 2006
※(Main research areas: Artificial Neural Networks, Pattern Recognition, Machine Learning, Image Processing and Computer Vision)