JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Traffic Sign Detection Using Machine Learning

[1]K. Arun Babu, [2]A. Kumar Naidu, [3]K. Jahnavi ,[4]S. Siva Ram,[5]K. Sony Navya

[1]Asst. Professor, Department of Computer Science, Bapatla Engineering College,
Bapatla, India
[2]CSE, Department of Computer Science, Bapatla Engineering College,
Bapatla, India
[3]CSE, Department of Computer Science, Bapatla Engineering College,
Bapatla, India
[4]CSE, Department of Computer Science, Bapatla Engineering College,
Bapatla, India
[5]CSE, Department of Computer Science, Bapatla Engineering College,
Bapatla, India

Abstract : *Convolutional neural networks are currently the most popular deep learning methods for traffic signal classification[1], but because to the inherent limitations of the max pooling layer, they are unable to capture the position, view, and orientation of the images. The deep learning architecture known as capsule networks is used in this paper to provide a novel strategy for the detection of traffic signs that achieves exceptional performance on the German traffic sign dataset. A capsule network is made up of capsules, which are collections of neurons that use the dynamic routing and route by agreement algorithms to describe an object's position and orientation[2]. Our method reduces the manual labour and offers resistance to the spatial variations, in contrast to the earlier approaches of manual feature extraction and numerous deep neural networks with various parameters.*

***IndexTerms* - CNN, Capsule Neural Networks, Pose, Traffic sign, Dataset, GTSRB**

I. INTRODUCTION

Traffic sign detection is a real-life task that involves many obstacles and complexities. Even a small misalignment of road markings can cause disaster and even lead to loss of life. It is applied to advanced driver assistance systems and autonomous vehicles. A camera on the car's dashboard then captures a real-time video stream that is encoded with images and then transmitted to an embedded deep learning model. Car dashboard. depth is strong and reliable at all times. Cameras can capture road signs in different ways and situations, but the algorithm must be able to recognize the correct signs[4] and the capsule system is a good deep learning method that is perfect for solving this problem.

In general, convolutional neural networks are used for deep neural network algorithms [5] in many image-related tasks. Convolution extracts the spatial information of the image using a kernel function in the convolution layer. CNN has input, output and hidden levels. The hidden layers in the front contain the transformation, aggregation, full integration and configuration. CNNs work well for image-related operations, but they have limitations and drawbacks.
CNN fails[6] to capture space and orientation relationship.
CNN can be easily confused by the structure of the image or by the change in position.

Declarative information can be the orientation, thickness, inclination, exact position of the object. Asking the biggest question is a big drawback of CNN because it can not spread the spatial classification between simple and complex objects, which leads to distortion and prevents them from mapping the connections between the pixels of that thing. reduces the data and reduces the information about the space of the data passing to the next layer. To overcome this drawback, the capsnet architecture is developed, which performs the highest on the MNIST dataset [2] and performs better than CNN on the Multi MNIST dataset.

II. RELATED WORK

It is difficult to compare the research work done in the past in the field of traffic sign detection due to the large research efforts of researchers in this area and the use of different types of data. things to solve various problems including related research, classification. and tracking. Work.

A. Using computer vision feature extraction methods

It was one of the first of many algorithms and techniques proposed by computer scientists before the advent of machine learning. Colour gradients are calculated using different histograms created and measured. In this method, colour image gradients are calculated using different histograms created in the scale. Scale Invariant Feature Transformation (SIFT) [8] is used for classification and sliding window techniques to perform classification and detection tasks simultaneously.

B. Using machine learning

Many types of machine learning algorithms such as support vector machine, linear discriminant analysis[10], ensemble classifiers, random forest and kd-tree[11] have been used in the classification of road signs.

Linear Discriminant Analysis (LDA) [5] is based on the posterior estimation of a class. The class density is assumed to have a multivariate Gaussian matrix and a normal covariance matrix.

Random Forest is a classification system [1] that is based on the combination of unpruned decision trees. Each decision tree is constructed using randomly selected training data. based on majority voting considering decisions from all decision trees.

Support Vector Machines (SVM) is a classification algorithm that classifies data by dividing the data plane horizontally by surface for classification[9]. SVM can segment sparse data by transforming the segmentation plane into a higher dimension using a non-linear kernel function that uses a method called kernel manipulation for its implementation.

Machine learning techniques [12] have not been able to handle the different dimensions and image and feature measurements have to be manually processed which is time-consuming and an inconvenient process.

C. Using deep learning

To overcome the drawbacks of the traditional methods mentioned above, new implementations based on deep learning algorithms have replaced previous methods [13] in recent years with the increase in computing power and availability of structured data systems and access to big data. A convolutional neural network is a modern algorithm that achieves the highest accuracy rate. The LENET architecture [14] is the first CNN architecture for traffic signal processing.

A Convolutional neural network is a neural network architecture inspired by neural networks that learn invariant features. Each unit consists of a filter layer (convolution), a non-linear transformation layer, a spatial aggregation layer [15]. Spatial clusters transmit spatial information and act as complex cells in the visual cortex. An optimization based on gradient descent is used to train and improve each filter to reduce the loss function. The output of all the levels is transferred to the level to improve the planning process.

III. DATA SET

The German Traffic Sign Recognition Benchmark (GTSRB) is defined and presented visually. This is a publicly available dataset and was created from 10 hours of video of driving on various roads in Germany. The extraction of traffic signals is done using a software system based on the NISYS Advanced Development and Analysis Framework (ADAF) [16] module.

After cleaning and removing repeated frames, the dataset is reduced to 51,840 images from the 43 classes. All the images in the dataset are 32*32 in size, and the total dataset is divided into training data and testing data. A total of 39,209 frames as training data and 12,630 frames as testing data.

Fig. 1. 1 Sample images from GTSRB dataset

IV. CAPSULE NET ARCHITECTURE

Capsule networks are capsules instead of neurons. Capsule [17] is an artificial neural network that performs complex calculations on their inputs and stores the results in a small vector. Each capsule holds the relative position of the object and if the position of the object is changed, the direction of the output vector [18] is changed and therefore, makes them the same.

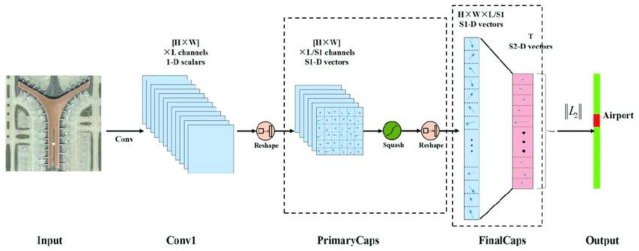

Fig. 2. 1 Capsule sign architecture for traffic sign detection

Caps Net has many layers and the first layer is called core where each cap receives a small part of the host field as input and tries to detect the position of a particular pattern. The capsule output is a vector and a dynamic conversion

method is used to ensure that the correct parents are sent to the layer which can be deduced from fig. 2.

A. Computation of capsule vector inputs and outputs

Capsule calculates the prediction vector[19] by multiplying the weight matrix (Wij) with its own output vector (ui). The correlation coefficient of the corresponding capsule
production increases the dot product and prediction [20] for the capsule production.

$$\hat{u}_{j|i} = W_{ij}u_i$$

where uj|i=prediction vector, Wij=weight matrix and ui=output vector.

B. Squash Function

In capsule networks, an invisible activation function called crush function is used[21]. This function converts the length of the output vector to the probability that the capsule connects to this object. It reduces the long output vectors to slightly less than unit length and the small output vector is almost zero.

$$v_j = \frac{||s_j||^2}{1+||s_j||^2}\frac{s_j}{||s_j||}$$

where s_j=Total Input, v_j=Vector Output of capsule j.

C. Routing Algorithm

CNN's scalar signal detector has been replaced with vector output capsules [18] and the maximum convergence and convergence method. The capsule segment increases with increasing levels due to the change from encoding (text in continuous space) to code number (text in) and the higher capsules represent the existing facilities. complex and have more degrees of freedom. The per-context method works better than the maximum package used in CNN.

$$S_j = \sum_i c_{ij}\hat{u}_{j|i}$$

where Sj=summation matrix,uj|i=prediction vector, and cij=coupling coefficients determined by iterative dynamic routing.

V. EXPERIMENTS

A. Data Preparation

The data is from the disk in pickled format.Pickling is the process of saving a file in a serialised format before writing it into the disk.Size of each image is 32*32 and total 34799 images are in the training dataset and 12630 in the testing dataset.

B. PREPROCESSING

The image brightness is enhanced with random uniform distribution of 0.6 to 1.5 and image contrast is also enhanced with random uniform distribution of 0.6 to 1.5.
The training dataset is augmented five-fold by replicating the available with data rotation ± 20, shear range of 0.2, width shift range of 0.2, horizontal flip which increased the training dataset size leading to better performance and regularizing thus avoiding the over fitting problem.
Augmenting the existing training dataset will increase the dataset size to 34799 X 5 = 173995 images.

C.Network Architectuure

The structure used to detect road signs with an input layer and the first convolutional layers are part of the main capsules and send the vector of the original capsule to the road sign capsules.

1) Input Layer: The input layer consists of input training images and the dimension is equal to the total training images.

2) Primary Capsule Layer: The first layer that follows the input layer is the primary capsule layer and for calculating the output first two convolution layers were used. The first convolution layer consists of kernel size 9 and 256 filters and padding was not used.Rectified Linear unit (ReLU) was used as the non linear activation function and a drop out of 0.7 which is fixed to be optimal after testing with different values.

Output is repeated to obtain the output vectors of the first capsules. Since the first capsule layer is connected to the road layer, the output vectors will be rewritten using the rewrite function. A small epsilon value is added to the rewrite function to avoid smoothing problems during training. Now, the production of the squash work goes through the capsule layer of the logo.

3) Traffic Sign Capsule Layer: To calculate the output of road sign capsules, calculate the predicted output vectors for each road sign capsules and implement the path-by-negotiation algorithm.

The road sign capsule contains 43 capsules each representing one class of the German road sign dataset with a size of 32 capsules each. For each capsule in the first layer, predict the equation and output vector of each capsule j in the second layer.

D. Reconstruction

A decoder network is added to the road sign capsule network which constitutes a fully connected network layer that helps in reconstructing the input image by optimizing the output of the road sign capsule network. This feedback mechanism will allow the network to store the information needed to reconstruct traffic signals throughout the network. This is done as a regularization and it avoids the over-fitting of the data and helps with the good generalization of traffic signs.

1) Mask: For the reconstruction of the input traffic sign only that particular output vector corresponding to predicted traffic sign is sent and all remaining outputs should be masked.Masking function is used to avoid all the other output vectors during training phase. The reconstruction mask is realized using the one-hot function.For the target class its value will be one and for all the other classes its value will be zero.

2) Decoder: The Decoder consists of a non linear activation layer of ReLU followed by a sigmoid activation layer.

E. Losses

1) Margin Loss: The length of the instantiation output vector represent the probability of the respective capsule's entity exists or not.The digit class k has the longest vector output only if that traffic sign is present in the input image.

For every traffic sign capsule k the margin loss is separate and it is given as k

$$L_k = T_k \max(0, m^+ - ||v_k||)^2 + \lambda(1 - T_k)\max(0, ||v_k|| - m^-)^2$$

the value of T_k is 1 if a traffic sign of class k is present and here m+ = 0.9 and m- = 0.1.λ is a regularization parameter which stops the learning from shrinking the activity vector of all traffic sign capsules.

2) Reconstruction Loss: It is the difference between the squares of the input image and reconstructed image

$$R = (\text{Input image})^2 - (\text{Reconstructed image})^2$$

where R = Reconstruction Loss

3) Final Loss: The Final Loss is the sum of Margin loss and Reconstruction Loss scaled to a factor λ which acts as a scaling factor and it should be very much less than one.

$$F = (\text{Margin Loss}) - \Lambda(\text{Reconstruction Loss})$$

where F= Final Loss, λ= 0.0005

Margin loss should always dominate the Reconstruction loss in comparison.If reconstruction loss is more in the final loss then the model tries to exactly match output image with the input image of training dataset which lead to overfitting of the model to the training data.

F. Results

The model is evaluated using the testing data set of 12,630 testing images.Accuracy is computed as the ratio of the correctly identified traffic signs by the total number of traffic signs.[22]

$$Accuracy = \frac{\sum correctly\ identified\ traffic\ signs}{Total\ number\ of\ traffic\ signs}$$

with a batch size of 50 obtained an accuracy of 97.6 percent and a final loss of 0.0311028 evaluated on the testing dataset.

The performance evaluation is based on the correct classification rate(CCR) and binary loss (0 or 1) which means by counting the number of misclassification's.The

test set is

TABLE I

COMPARISON OF CORRECT CLASSIFICATION RATE FOR DIFFERENT METHODS

CCR (%)	Method
97.62	using Capsule networks
96.14	Random Forests [1]

95.68	LDA(HOG 2) [23]
93.18	LDA(HOG 1) [8]
92.34	LDA(HOG 3) [23]

unbalanced in terms of the number of samples for a particular class but assume that all the classes are equally important and carry equal weightage.

Keras and tensor flow deep learning libraries with CUDA and CUDNN libraries for GPU accelerated training to implement this capsule network model. The model is trained using the system with Intel i7 7500U Processor 2.7G,3M Processor speed 8GB Memory RAM, 1 TB Hard Disk, NV GT 940 MX 2G DDR3 Nvidia GeForce GPU. It took 10 hours for training the model and can be decreased by using a high performance GPU like Nvidia K80.

From Table 1 we can conclude that capsule network is superior and performing better than other methods mentioned.

G. Conclusion

Traffic sign detection is a difficult task and capsule networks use their natural ability to detect pose and space transitions more efficiently compared to CNN and capsule networks increase reliability and accuracy by performing the task of processing and identifying images correctly, even if they are distorted, distorted and distorted. image. photo.

H. Acknowledgement

Thanks to Dr. K.Arun Babu for his constant motivation and for cultivating curiosity in machine learning.

REFERENCES

[1] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. The german traffic sign recognition benchmark: a multi-class classification competition.
[2] Geoffrey Hinton, Nicholas Frosst, and Sara Sabour. Matrix capsules with em routing. 2018.
[3] Jiawei Su, Danilo Vasconcellos Vargas, and Kouichi Sakurai. One pixel attack for fooling deep neural networks. CoRR, abs/1710.08864, 2017.
[4] N. Deepika and V. V. Sajith Variyar. Obstacle classification and detection for vision based navigation for autonomous driving. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pages 2092–2097, Sept 2017.
[5] Yihui Wu, Yulong Liu, Jianmin Li, Huaping Liu, and Xiaolin Hu. Traffic sign detection based on convolutional neural networks. In Neural Networks (IJCNN), The 2013 International Joint Conference on, pages 1–7. IEEE, 2013.
[6] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial Transformer Networks. ArXiv e-prints, June 2015.
[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 1, pages 886–893 vol. 1, June 2005.
[8] David G. Lowe. Object recognition from local scale-invariant features. In Proceedings of the International Conference on Computer VisionVolume 2 - Volume 2, ICCV '99, pages 1150–, Washington, DC, USA, 1999. IEEE Computer Society.
[9] Jung-Guk Park and Kyung-Joong Kim. Design of a visual perception model with edge-adaptive gabor filter and support vector machine for traffic sign detection. Expert Systems with Applications, 40(9):3679 – 3687, 2013.
[10] Rabia Malik, Javaid Khurshid, and Sana Nazir Ahmad. Road sign detection and recognition using colour segmentation, shape analysis and template matching. In Machine Learning and Cybernetics, 2007 International Conference on, volume 6, pages 3556–3560. IEEE, 2007.
[11] Fatin Zaklouta and Bogdan Stanciulescu. Real-time traffic sign recognition in three stages. Robotics and autonomous systems, 62(1):16–24 , 2014.
[12] Andreas Mogelmose, Mohan Manubhai Trivedi, and Thomas B Moeslund. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. IEEE Transactions on Intelligent Transportation Systems, 13(4):1484–1497, 2012.
[13] C. K. Chandni, V. V. S. Variyar, and K. Guruvayurappan. Vision based closed loop pid controller design and implementation for autonomous car. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pages 1928–1933, Sept 2017.
[14] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural networks, 32:323–332, 2012.
[15] Pierre Sermanet and Yann LeCun. Traffic sign recognition with multiscale convolutional networks. In IJCNN, pages 2809–2813. IEEE, 2011.
[16] Sebastian Houben, Johannes Stallkamp, Jan Salmen, Marc Schlipsing, and Christian Igel. Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. In International Joint Conference on Neural Networks, number 1288, 2013.
[17] Edgar Xi, Selina Bing, and Yang Jin. Capsule network performance on complex data. arXiv preprint arXiv:1712.03480, 2017.
[18] Geoffrey E Hinton, Alex Krizhevsky, and Sida D Wang. Transforming auto-encoders. In International Conference on Artificial Neural Networks, pages 44–51. Springer, 2011.
[19] Dilin Wang and Qiang Liu. An optimization view on dynamic routing between capsules. 2018.
[20] Geoffrey F Hinton. A parallel computation that assigns canonical object-based frames of reference. In Proceedings of the 7th international joint conference on Artificial intelligence-Volume 2, pages 683–685. Morgan Kaufmann Publishers Inc., 1981.
[21] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In Advances in Neural Information Processing Systems, pages 3859–3869, 2017.

[22] Jung-Guk Park and Kyung-Joong Kim. Design of a visual perception model with edge-adaptive gabor filter and support vector machine for traffic sign detection. Expert Syst. Appl., 40(9):3679–3687, July 2013.
[23] David G. Lowe. Distinctive image features from scale-invariant
keypoints. Int. J. Comput. Vision, 60(2):91–110, November 2004.