Journal of Theoretical and Applied Information Technology
15th July 2019. Vol.97. No 13

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING TRAFFIC SIGN DETECTION WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

HOANH NGUYEN

Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

E-mail: nguyenhoanh@iuh.edu.vn

ABSTRACT

Motivated by the observation that traffic signs are still not well-detected by deep convolution neural network-based methods because of the constraint of the size of feature maps, this paper is focused on improving the performance of traffic sign detection, especially for small-sized traffic signs. In this paper, a novel approach for traffic sign detection based on Faster R-CNN framework is proposed. First, Inception-ResNet model is used as the base network to provide a rich and discriminative hierarchy of feature representations. Next, a deconvolutional module is integrated into Faster R-CNN framework to bring additional context information which is helpful to improve the detection accuracy for small-sized traffic signs. Finally, atrous convolution is adopted in the region proposal network to enlarge the receptive field of the synthetic feature map. Experimental results on the German traffic sign detection benchmark show that the proposed approach obtained an accuracy comparable to other the state-of-the-art approaches in traffic sign detection.

Keywords: *Traffic Sign Detection, Convolutional Neural Network, Intelligent Transportation Systems, Object Detection, Deep Learning*

1. INTRODUCTION

Traffic sign detection plays an important role in intelligent transport systems, traffic management, and so on, and it has been an active area of research over the past decade. An automatic traffic sign detection system can assist the driver on the road such as reminding the driver of traffic constraints, stopping driver from performing inappropriate actions. Further, it could be integrated into an Automated Driving System (ADS) and Advanced Driver Assistance System (ADAS). Traditional approaches for traffic sign detection usually use color, texture, edge and other low-level features to detect the area or the edge of a traffic sign in an image such as adaboost [1], support vector machine (SVM) [2], Hough transform [3] and so on. These approaches based on low-level features do not detect or recognize traffic signs well because of variations in the traffic sign appearance due to different sign shapes or colors. Furthermore, objects such as trees and vehicles which may occlude the traffic signs make traffic sign detection become harder.

Recently, with fast development of deep learning [4, 5], a certain number of methods for traffic sign detection based on deep learning have been proposed. These methods show better performance than traditional methods. Deep CNN-based methods firstly create traffic sign candidates. Then, a deep CNN-based classifier is used to reject non-traffic sign candidates. Although these methods perform well in complex conditions, the small size of traffic signs makes them hard to detect, and useful context information is not exploited fully by CNN-based approaches. A popular solution to this problem in CNNs is to combine information from the background [6] or relationships among the objects [7], which combines finer details from multiple convolution layers with different local receptive fields. But it has been found that simply concatenating these feature maps does not significantly improve the accuracy due to over-fitting caused by curse of dimensionality.

A number of recent approaches have improved the feature extraction of small objects by using additional context information and increasing the spatial resolution of feature maps. DSSD [8] used deconvolution layers in combination with existing multiple layers to reflect the large-scale context. MS-CNN [9] applied deconvolution on shallow layers to increase the feature map resolution before

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

using the layers to extract region proposals and pool features. Recently, Long et al. [12] introduced the Fully Convolution Network (FCN), which demonstrated impressive performance in semantic segmentation [12, 13], and object detection [14]. In [14], the authors combined coarse high-layer information with fine low-layer information for semantic segmentation. Additionally, the atrous convolution represents a powerful and convenient tool to enlarge the field of view of filters and incorporate larger context without increasing the number of parameters or the amount of computations. However, these kinds of technology such as deconvolution and atrous convolution are less explored in traffic sign detection area.

Motivated by the above ideas, a novel effective traffic sign detection framework based on the Faster R-CNN [15] pipeline is introduced in this paper. This framework can achieve state-of-the-art performance on traffic sign detection, especially for small-sized traffic signs. The main contributions of this paper are summarized as follows:

1) A novel traffic sign detection framework is proposed by adding the deconvolutional module to the traditional Faster R-CNN network. The deconvolutional module can bring in more semantic context information to enhance the feature map, thereby improving the detection performance, especially for small-sized traffic signs.

2) This paper proposes using a reduced network architecture, in which the prior layer is used as the initial feature map with a relatively large spatial resolution, instead of using the last layer as the output feature map. In addition, proper adjustment of the network has been made to avoid downsampling. The proposed architecture help retain more detailed information for traffic signs, especially with small-sized traffic signs.

3) Instead of using multi-layer feature map, in which low-level layers have less semantic information regarding small instances, a synthetic feature map that combines the initial feature map and the deconvolution layer with semantic information is proposed.

4) This paper proposes applying atrous convolution on the synthetic feature map to enlarge the receptive field and inject detailed context information. Larger receptive fields help the detection of large-sized instances and detailed context information helps the detection of small-sized instances of traffic signs. Therefore, application of atrous convolution can improve the detection accuracy of multi-scale object detection.

This paper is organized as follows: an overview of previous methods is presented in Section 2. Section 3 describes the detail of the proposed method. Section 4 demonstrates experimental results. Finally, the conclusion is made in Section 5.

2. RELATED WORK

In this section, this paper introduces previous work on license plate detection, including traditional methods and recently proposed methods based on deep CNN.

2.1 Traditional Approaches

Traditional approaches to traffic sign detection include a wide variety of algorithms and various representations [22]. Escalera et al. [16] took advantage of color and shape features to detect road traffic signs, while Shadeed et al. [17] used histogram equalization, light control and color segmentation to locate road signs. Later, Garcia-Garrido et al. [3] employed the Hough transform to get the information from the edges in the image, but the computational complexity was high so that it hindered the real-time application. To deal with the efficiency problem, Bahlmann et al. [1] detected traffic signs using a set of Haar wavelet features obtained from AdaBoost training [18]. To balance effectiveness and efficiency, Salti et al. [2] proposed an approach in which the regions of interest rather than the sliding window were extracted at first, and then a histogram of oriented gradients (HOG) in the regions of interest was extracted, to be the input feature of the SVM classifier [19]. Recently, Berkaya et al. [20] extended this approach by using an ensemble of features including HOG, local binary patterns (LBP) and Gabor features within an SVM classification framework. To improve results obtained by single view analysis, Timofte et al. [21] combined 2D and 3D techniques to generate and evaluate 3D proposals. Traditional methods can be processed in real-time on low-end systems, and these methods achieve good performance in limited conditions such as simple background and environment, fixed illumination and so on. However, the performance of these methods will significantly reduce in complex conditions.

2.2 Deep CNN-based Approaches

Recently, with fast development of deep learning, a certain number of methods for traffic sign detection based on deep CNN have been proposed. John et al. [23] used CNN to extract features and detect road traffic signs, making a saliency map containing the traffic light location. For network optimization, Jin et al. [24] suggested a hinge loss stochastic gradient descent (HLSGD) method to train a detection network. To perform fast and accurate traffic sign detection and recognition, Zhu et al. [25] employed a holistically nested edge detection network [26]. Deep CNN-based approaches are able to detect some traffic signs. However, constrained by feature map size, none show significant advantages.

Therefore, this paper proposes a new enhanced convolution neural network by using the Faster R-CNN framework, in which the deconvolutional module is added to bring in more semantic context information. A synthetic feature map is generated by combining the deconvolution layer and high-level feature map to enhance the feature representation. Furthermore, the receptive field of the proposed framework is extended by atrous convolution, which also helps to inject more context information. Compared with multi-layer feature extraction methods like MS-CNN, the proposed method reuses the higher-resolution maps in the feature hierarchy, instead of adding several feature maps for shallow layers. This approach is more effective for detecting small traffic signs.

3. APPROACH

Figure 1 shows the overall framework of the proposed approach. As shown in this Figure, the proposed method consists of the base network for generating feature map, the deconvolution module for generating synthetic feature map, and region proposal network generated with atrous convolution and classification. Each image is forwarded through the convolution layers to generate feature maps. Based on these feature maps, this paper applies deconvolution with the encoder-decoder structure, combining the deconvolution layer with feature maps to generate the synthetic feature map that collects additional context information. Finally, atrous convolution is applied to the synthetic feature map to generate region proposals. These proposals are then classified and adjusted with the detection module. Details of the proposed approach are explained in next sections.

3.1 The Base Network

This paper uses Inception-ResNet as the base network to generate feature maps from input image. Inception-ResNet, proposed by Szegedy et al. [27], combines the optimization benefits conferred by residual connections with the computation efficiency of inception units. The inception module approximates a sparse CNN with a normal dense construction. It uses convolutions of different sizes (5x5, 3x3, 1x1) to capture details at various scales. The structure of Inception-ResNet module is shown in Figure 2. As shown, the Inception-ResNet replaces the filter concatenation stage of the Inception module with residual connections. Compared with VGG16, Inception-ResNet can be made more efficient by making the architecture deeper and wider. It also shows better performance than VGG- 16 in the ILSVRC 2014 classification and detection challenges.

Since this study uses only the convolution layers in Inception-Resnet architecture, the size of the input image does not have to be fixed. Followed the Inception-Resnet-v2 [27], a reduced Inception-ResNet is adopted in this paper. Figure 3 shows the architecture of the base network. In order to retain the high resolution of feature map, the output of the Inception ResNet block17 is taken as an initial feature map. Furthermore, this study sets the stride to 1 in the deep layers of the network Mixed_5b layer to the block17 layer, so downsampling can be avoid in order to retain as much information about the small objects as possible. The output shape of the initial feature map is [33 × 33 × 1088]. Finally, a deconvolution module is used in the base network to add the semantic information.

3.2 Deconvolutional Module

Pinheiro et al. [28] suggested that a factored version of the deconvolutional module for a refinement network has the same accuracy as a more complicated one and the network will be more efficient. Deconvolutional module is effective for small object detection [8]. Thus, to enhance the detection of traffic sign, adding extra deconvolution layers is proposed in this study. The deconvolutional module is built at the end of the base network. The structure of deconvolutional module is shown in Figure 4. As shown in this Figure, a 3 × 3 convolution layer and rectified linear activation are used. For the deconvolution branch, the encoder-decoder structure with 2×2 deconvolution is used followed by a 3×3 convolution. A batch normalization layer (BN) is added after each convolution layer. An intermediate feature map is extracted after Mixed_7a. Then, the deconvolution layer is added to enlarge the feature map size in order to match the size of the initial feature map. Finally, element-wise product is performed as a combination method, which is followed by rectified linear activation to generate the synthetic feature map.

3.3 Region Proposal Network (RPN)

First, the RPN takes the synthetic feature map and generates a set of anchor boxes. An anchor is centered at the sliding window and is associated with a scale and aspect ratio. Since the synthetic feature map do not have big enough receptive field. This paper designs the proposed network to apply atrous convolution onto the synthetic feature map to enlarge the receptive field and inject context information. Atrous convolution, which is a powerful tool in dense prediction tasks, allows us to effectively enlarge the field of view of filters to incorporate larger context without increasing the number of parameters or the amount of computation. Another advantage is that atrous convolution can be conveniently and seamlessly integrated to compute the responses of any layer. Figure 5 shows an example of feature extraction with atrous convolution. As shown, feature map a is produced from feature map b by an atrous convolution with rate $r = 2$. Feature map a corresponds to a receptive field of 9×9.

Atrous convolution with a rate r introduces $r - 1$ zeros between consecutive filter values, effectively enlarging the kernel size of a $k \times k$ filter to $k' \times k'$ without increasing the number of parameters or the amount of computation by using the following equation.

$$k' = k + (k - 1)(r - 1) \quad (1)$$

In this paper, the atrous convolution is used with a 3×3 kernel size and rate $r = 2$ leading to $k' = 5$. The output shape is [33 × 33 × 1088] after block17. The corresponding receptive field of each element is 47 × 47. After atrous convolution, the output shape is still [33 × 33 × 1088], but the receptive field of each element is 79 × 79. Thus, more context information can be obtained. To solve the multiple-scale detection problem this paper uses three scales and three aspect ratios for each anchor, different anchors are used with four scales [0.25, 0.5, 1.0, 2.0] and three aspect ratios [0.5, 1.0, 2.0].

4. RESULTS

In this section, this paper compares the performance of the proposed method with other state-of-the-art methods. The proposed method is implemented on a machine with Core I5 6400, 8GB of RAM, NVIDIA GTX 1050Ti GPU. This paper uses TensorFlow for implementing deep CNN frameworks.

4.1 Dataset

The GTSDB database [10] is adopted in this paper to train and evaluate traffic sign detector. GTSDB is the most widely used dataset in traffic sign detection. This dataset contains 900 images and is divided into 600 training images and 300 testing images. Each image has the size of 1360×800 pixels. The traffic signs in the GTSDB database can be divided into four categories: 161 Prohibitory signs (usually of red color and circular shape), 49 Mandatory signs (usually of blue color and circular shape), 63 Danger signs (usually of red color and triangular shape), and other signs with different shapes and colors which cannot be classified into these three categories. Some examples of traffic sign in this dataset is shown in Figure 6. The sizes of traffic signs to detect in the GTSDB database vary from 16 × 16 to 128 × 128 and the size of traffic scenes is 800 × 1360.

4.2 Evaluation Criterion

In order to compare this work with other state-of-the-art methods on the same dataset, this paper uses the same widely used evaluation criteria in traffic sign detection, including precision and recall. A traffic sign is correctly detected only if the overlap between the detected bounding box and ground truth bounding box (Intersection over Union-IoU) is greater than 0.5. Precision is defined as ratio between the number of correctly detected traffic signs and the number of detected bounding box, while recall is defined as ratio between the number of correctly detected traffic signs and the number of ground truths. IoU is a threshold which measure the quality of detection.

4.3 Training

The model used in this paper is based on Inception-ResNet-v2 [27], which is pre-trained on the ILSVRC-CLS image classification dataset [10]. This study changes convolution stride of the Mixed_6a stage from 2 to 1 to increase the feature map resolution. However, the reduced stride also shrinks the receptive field. In order to offset, atrous convolution is used to enlarge the receptive field by increasing the dilation rate to 2. Then, the resulting model is fine-tuned using Stochastic Gradient Descent (SGD) with an initial learning rate of 0.0001, momentum of 0.9, and batch size of 1; the learning rate is reduced by a factor of 10 after 500,000 iterations and again after 700,000 iterations. Learning stops after 800,000 iterations.

With the fine-tuned network of region proposals, non-maximum suppression (NMS) is adopted to eliminate highly overlapped bounding boxes with lower scores. After using non-maximum suppression (NMS), a total of 100 proposals are generated for the second stage detection part.

4.4 Experimental Results

This paper compares the effectiveness of the proposed approach to other deep CNN-based approaches, including HOG+LDA+SVM [31], HOG+SVM [32], ROI+HOG+SVM [33], HOG+CNN [34], ROI+HOG+SVM [29] and ROI+Multi-task CNN [30]. Only three categories of GTSDB database (excluding the unique category which is illustrated in Figure 7) are evaluated in previous work. The proposed method can detect unique traffic signs, and the performance of detecting unique traffic signs has been evaluated as well. Figure 8 shows some examples of detection results on the GTSDB dataset. As shown in this figure, the proposed approach can locate exactly traffic signs in the wild, especially with small traffic sign. Table 1 shows a comparison to other state-of-the-art methods. As shown in this table, the proposed approach outperforms others on the GTSDB database, considering generality, reliability and run time.

Table 1 also shows a comparison report between different traffic sign detection methods in terms of their computation efficiency. Because the proposed method is implemented on a low-end machine, processing time of different methods cannot compare directly. Thus, Table 1 shows processing time of different methods based on their hardware configuration. While other methods were implemented on high-end GPU, the proposed method meets the requirement of real-time processing on low-end GPU. This shows that our method outperforms both in accuracy and in run time.

Figure 9 shows some unsuccessful detection results. As shown in this Figure, the proposed method cannot detect traffic sign that are blurred, tiny or backlit. The results could be improved by enlarging the amount of the training data and enriching the diversity of the samples (the GTSDB has only 600 training images). In the future, the proposed method will be extended to handle situations where the lighting is uneven, perspective irregular or the image is blurry.

5. CONCLUSIONS

In this paper, a new method is proposed for traffic sign detection, especially for small-sized traffic signs. The proposed approach is based on the Faster R-CNN framework with a deconvolutional module and atrous convolution adopted to capture more context information. Furthermore, the state-of-the-art CNN model Inception-ResNet is used as a base network in the proposed approach to provide a rich and discriminative hierarchy of feature representations. Extensive experiments demonstrated that the proposed method is superior in detecting traffic signs, especially for small-sized traffic signs and achieves comparable or better performance relative to other state-of- the-art methods on GTSDB dataset. However, the proposed method is unsuccessful in cases of small size, blur of the image, and backlight. Thus, this paper will improve the detection performance on traffic signs that are blurred, tiny or backlit by enlarging the amount of the training data.

REFRENCES:

[1] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information", *Proc. IVS*, 2005, pp. 255–260.

[2] S. Salti, A. Petrelli, F. Tombari, N. Fioraio, and L. Di Stefano, "Traffic sign detection via interest region extraction", *Pattern Recognit.*, vol. 48, no. 4, 2015, pp. 1039–1049.

[3] M. A. Garcia-Garrido, M. A. Sotelo, and E. Martin-Gorostiza, "Fast traffic sign detection and recognition under changing lighting conditions", *Proc. ITSC*, 2006, pp. 811–816.

[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proc. IEEE*, vol. 86, no. 11, 1998, pp. 2278–2324.

[5] Hanafi, Nanna Suryana, Abd Samad Bin Hasan Basarideep, "Learning for recommender system based on application domain classification perspective: a review", *Journal of Theoretical and Applied Information Technology*, Vol. 96, No. 14, 2018, pp. 4513-4529.

[6] B. Yang, J. Yan, Z. Lei, and S. Z. Li, "CRAFT objects from images", *Proc. CVPR*, 2016, pp. 805–813.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] X. Chen and A. Gupta, "Spatial memory for context reasoning in object detection", *Proc. ICCV*, 2017, pp. 117–125.

[8] Fu C-Y, Liu W, Ranga A, Tyagi A, and Berg AC, "DSSD: deconvolutional single shot detector", arXiv:1701.06659 [cs.CV]. http://arxiv.org/abs/1701.06659. Accessed Jan 2017.

[9] Cai Z, Fan Q, Feris RS, and Vasconcelos N, "A unified multi-scale deep convolutional neural network for fast object detection", *Lecture Notes in Computer Science*, vol 9908, 2016.

[10] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German traffic sign detection benchmark", *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2013, pp. 1–8.

[11] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, and Bernstein M, "Imagenet large scale visual recognition challenge", *Int J Comput Vis*, 2015, pp. 211–252.

[12] Long J, Shelhamer E, and Darrell T, "Fully convolutional networks for semantic segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, pp. 640-651.

[13] Hariharan B, Arbeláez P, Girshick R, and Malik J, "Hypercolumns for object segmentation and fine-grained localization", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 447–456.

[14] Chen L-C, Papandreou G, Kokkinos I, Murphy K, and Yuille AL, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs", *IEEE Trans Pattern Anal Mach Intell*, 2017, pp. 834-848.

[15] Ren S, He K, Girshick R, and Sun J, "Faster R-CNN: towards real-time object detection with region proposal networks", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, pp. 1137-1149.

[16] A. de la Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, "Road traffic sign detection and classification", *IEEE Trans. Ind. Electron.*, vol. 44, no. 6, 1997, pp. 848–859.

[17] W. G. Shadeed, D. I. Abu-Al-Nadi, and M. J. Mismar, "Road traffic sign detection in color images", *Proc. ICECS*, vol. 2, 2003, pp. 890–893.

[18] G. Rätsch, T. Onoda, and K. Müller, "Soft margins for AdaBoost", *Mach. Learn.*, vol. 42, no. 3, 2001, pp. 287–320.

[19] C. Cortes and V. Vapnik, "Support vector machine", *Mach. Learn.*, vol. 20, no. 3, 1995, pp. 273–297.

[20] S. K. Berkaya, H. Gunduz, O. Ozsen, C. Akinlar, and S. Gunal, "On circular traffic sign detection and recognition", *Expert Syst. Appl.,* vol. 48, 2016, pp. 67–75.

[21] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation", *Mach. Vis. Appl.*, vol. 25, no. 3, 2014, pp. 633–647.

[22] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey", *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, 2012, pp. 1484–1497.

[23] V. John, K. Yoneda, B. Qi, Z. Liu, and S. Mita, "Traffic light recognition in varying illumination using deep learning and saliency map", *Proc. ICITS*, 2014, pp. 2286–2291.

[24] J. Jin, K. Fu, and C. Zhang, "Traffic sign recognition with hinge loss trained convolutional neural networks", *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, 2014, pp. 1991–2000.

[25] Y. Zhu, C. Zhang, D. Zhou, X. Wang, X. Bai, and W. Liu, "Traffic sign detection and recognition using fully convolutional network guided proposals", *Neurocomputing*, vol. 214, 2016, pp. 758–766.

[26] S. Xie and Z. Tu, "Holistically-nested edge detection", *Proc. ICCV*, 2015, pp. 1395–1403.

[27] Szegedy C, Ioffe S, Vanhoucke V, and Alemi AA, "Inception-v4, Inception-ResNet and the impact of residual connections on learning", *Proceedings of the Thirty-First Conference on Artificial Intelligence*, 2017, pp. 4278–4284.

[28] Pinheiro PO, Lin TY, Collobert R, and Dollár P, "Learning to refine object segments", *ECCV*, 2016.

[29] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards real-time traffic sign detection and classification", *IEEE Trans. Intell.*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Transp. Syst., vol. 17, no. 7, 2016, pp. 2022–2031.

[30] H. Luo, Y. Yang, B. Tong, F. Wu, and B. Fan, “Traffic sign recognition using a multi-task convolutional neural network”, *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, 2018, pp. 1100–1111.

[31] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, “A robust, coarse to fine traffic sign detection method”, *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2013, pp. 1–5.

[32] M. Liang, M. Yuan, X. Hu, J. Li, and H. Liu, “Traffic sign detection by roi extraction and histogram features-based recognition”, *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2013, pp. 1–8.

[33] S. Salti, A. Petrelli, F. Tombari, N. Fioraio, and L. Di Stefano, “A traffic sign detection pipeline based on interest region extraction”, *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2013, pp. 1–7.

[34] Y. Wu, Y. Liu, J. Li, H. Liu, and X. Hu, “Traffic sign detection based on convolutional neural networks”, Proc. IEEE Int. Joint Conf. Neural Netw., 2013, pp. 1–7.

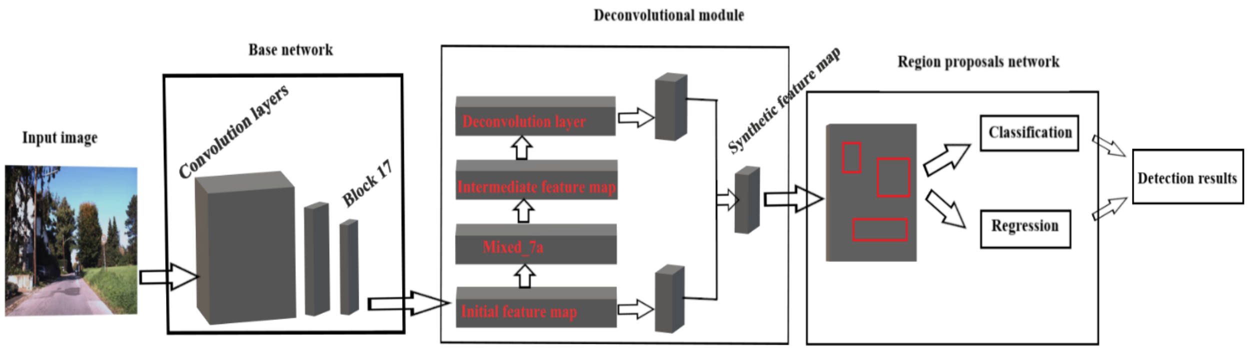

Figure 1: The Overall Framework of The Proposed Method

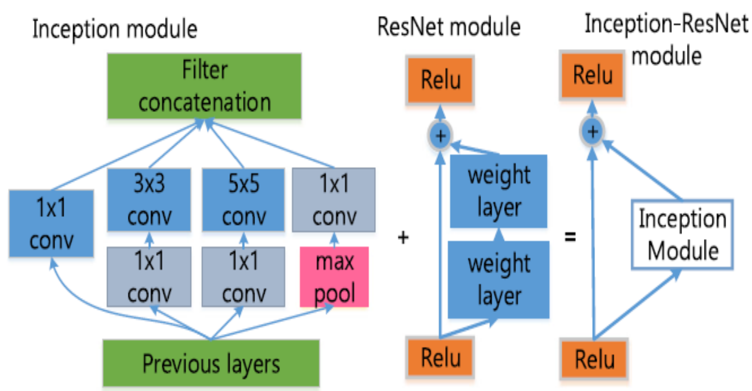

Figure 2: The Structure of Inception-ResNet Module

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

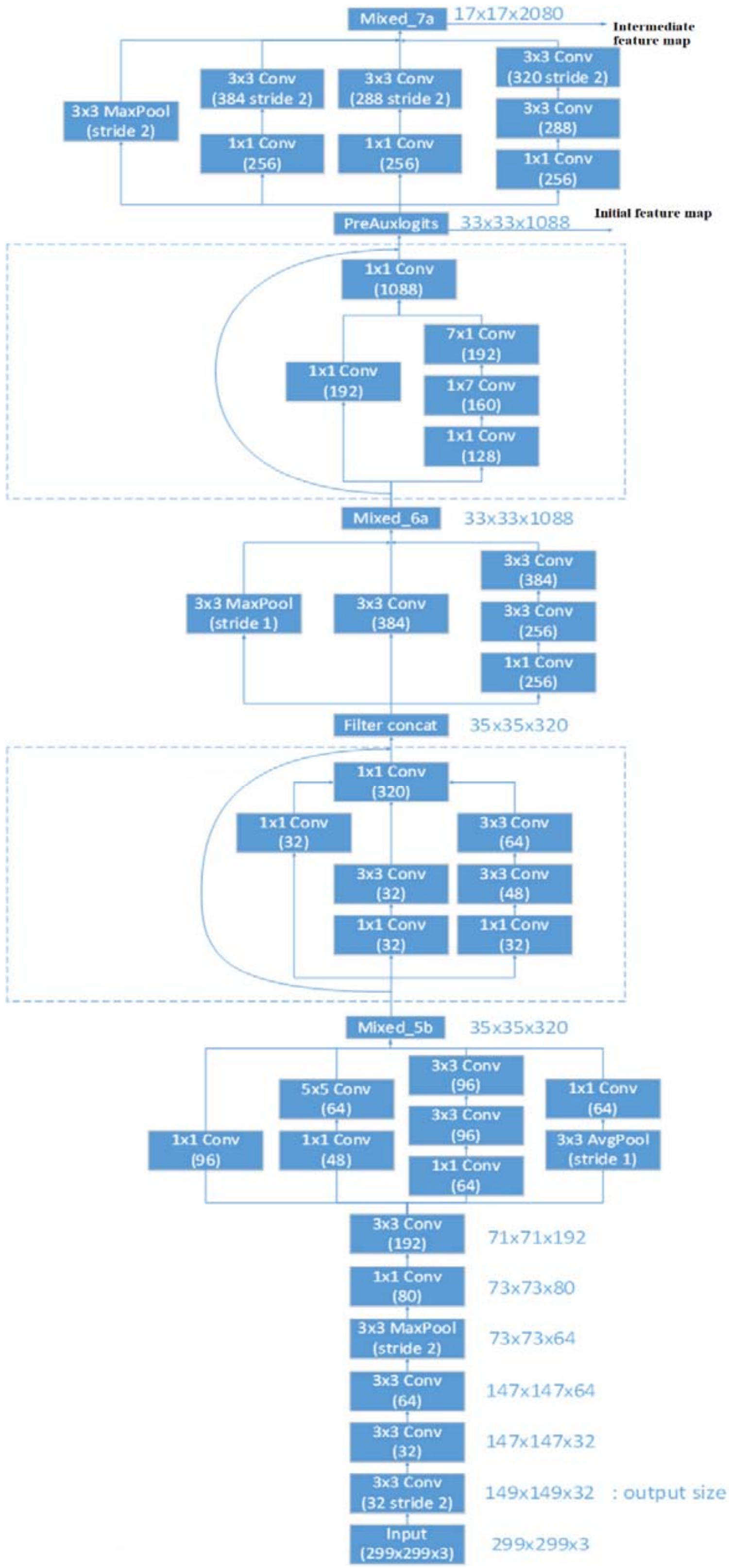

Figure 3: The Architecture of The Base Network

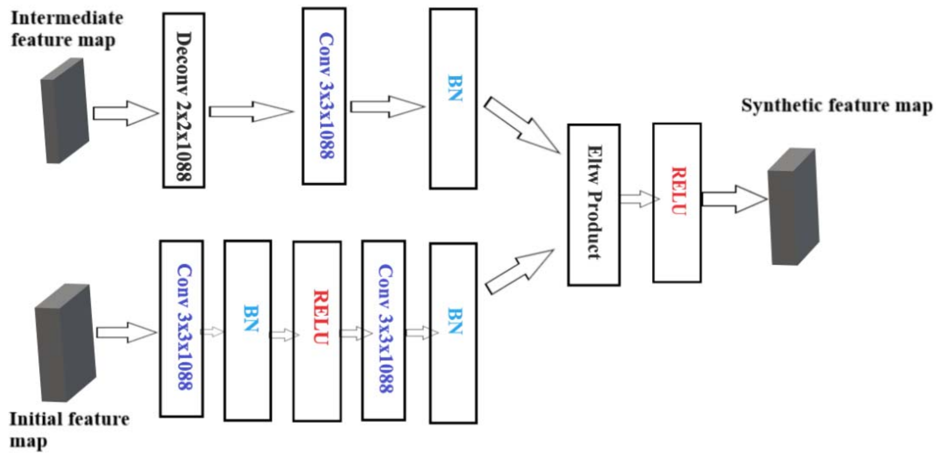

Figure 4: Deconvolutional Module

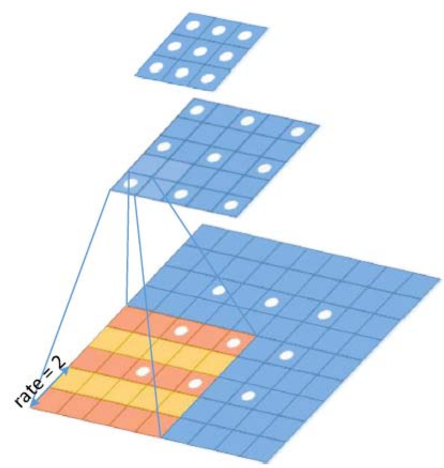

Figure 5: Example of Feature Extraction with Atrous Convolution

Figure 6: Examples Traffic Sign Images in The GTSDB Dataset

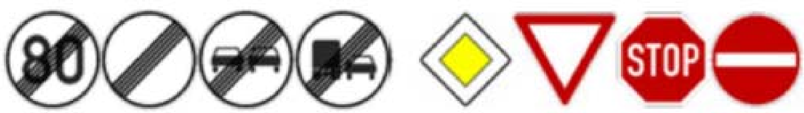

Figure 7: Examples of Unique Traffic Signs

Figure 8: Examples of Detection Results on The GTSDB Dataset

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 9: Examples of Unsuccessful Detection

Table 1: Experimental Results on GTSDB Dataset

Method	Prohibitory signs (161)	Mandatory signs (49)	Danger signs (63)	Unique signs (92)	Processing Time (s)
HOG+LDA+SVM [31]	100%	100%	99.91%	-	3.533
HOG+SVM [32]	100%	96.98%	100%	-	0.4 ~ 1
ROI+HOG+SVM [33]	99.98%	95.76%	98.72%	-	3.032
HOG+CNN [34]	-	97.62%	99.73%	-	12 ~ 32
ROI+HOG+SVM [29]	99.29%	96.74%	97.13%	-	0.162
ROI+Multi-task CNN [30]	99.99%	98.72%	98.34%	-	0.366 ~ 0.450
Proposed method	98.76%	100%	100%	99%	0.830 (low-end machine)