RES MILITARIS
REVUE EUROPEENNE D ETUDES EUROPEAN JOURNAL OF MILITARY STUDIES
Social Science Journal

TRANSFER LEARNİNG-BASED DEEP NEURAL NETWORK FOR TRAFFİC SİGN CLASSİFİCATİON

[1]Anusha Viswanadapalli, [2]Praveen Kumar Nelapati
[1]Department of Computer Science & Engineering, Vignan's Foundation for Science, Technology & Research, Vadlamudi, Guntur, Andhra Pradesh, India
[2]Research Scholar, Dept of SCOPE, VIT-AP, Amaravthi, Andhra Pradesh, India
Corresponding Author Mail: anusha6785@gmail.com

ABSTRACT

Traffic signs play a major role in human life while travelling ahead. Traffic signs are represented by specific symbols. These signs guide road conditions, provide instructions, give warnings to the drivers. As self-governing cars also came into form its mandatory for those cars to identify these signs. Now-a-days, the classification of traffic signs is a requisite for both the self-governing cars and driver assistance systems as the traffic signs vary from country to country. It is a field of computer vision. This system could guarantee the life of a human and enhance the road traffic safety situations. It assists the drivers or self-governing driving systems by detecting and classifying the traffic sign. It also alerts the driver or self-driving systems in a correct path by classifying the traffic sign boards. In this paper, Convolutional Neural Networks (CNN) is used for image classification. Here, German Traffic Sign Recognition Benchmark (GTSRB) Dataset is used which is open source. This dataset contains 50000 images as due to our system specifications, only 4300 images were taken and classified them into 43 classes where each class contains 100 images. This dataset is processed over two pre-trained models VGG-19 and VGG-16 along with CNN. The images which are categorized into 43 classes are passed as inputs and the output will show to which traffic sign the image belongs. Among these two pre-trained models VGG-16 gained 95.93% accuracy.

Keywords: *CNN, Deep learning, Traffic signs, VGG-16, VGG-19*

I. INTRODUCTİON

Recognition of context related objects from the images is the most crucial part included in computer vision applications. Traffic sign classification has become a vision problem due to different scenarios like climate, illumination changes, and speed limits, drivable lines etc. and so unable to identify the traffic sign. As the traffic signs vary from country to country it is difficult for a human being to memorize all the signs and its meanings. In Autonomous driving vehicles traffic sign classification is the crucial part to make a car drive through the correct path avoiding no. of accidents. So, advanced driver assistance systems have a large industrial demand.Traffic signs contribute a major part in Road Travelling to avoid accidents and also to ride safe. They provide very important information such as working condition of a road, speed limits, bumpy roads, school zones etc., which makes the driver alert and ride safer. Without all these useful traffic signs on the roads we may come across more accident's day by day as the driver is not aware of all the conditions of the way he is going through and also his speed limit. So, we came across the autonomous vehicles those can recognize and understand the traffic signs.Traffic Signs provide crucial visual content which controls the autonomous vehicles and also alerts the driver in the car. Every country doesn't follow a common pattern such as shapes, color, and sign of indication.

In this paper, we used the German Traffic Sign Recognition Benchmark (GTSRB) dataset for training and testing of data. GTSRB uses 2 different symbols which belong to mandatory category. This is an image classification dataset. This dataset is classified into 43 classes and has over 4300 images.The research in this field is increasing there are several methods developed and uses many features to extract the features automatically from an image. Among all the algorithms, Convolutional Neural Networks (CNN) gives the efficient results to recognize the traffic sign from the image given. CNN performs deep evaluation on the image and classifies the image. So, CNN gives more accuracy when compared to other algorithms in terms of image classification.

II. LİTERATURE SURVEY

In paper (Møgelmose et al., 2012), recognition of traffic sign for the visual system of driver is concerned with two stages:(1) Detection: The signs in the input images are located.(2) Classification: It includes shape, color, temporal of the image and determines the symbol which the system is detecting. Alternatively paying attention on detection and recognizing all the symbols belonging to some class. In self driving cars, the task is to find and highlight the signs that the driver was unable to see. The detection of the sign is done based on the color and shape. This needs to take into consideration of the drivers Centre of attention and interactivity business. The filter used in this paper is "Kalman" filter which will allow the structure to forecast a sign in the later frame and if it is away from the position then the candidate need to get rid of it.From paper (Houben et al., n.d.), three demonstrated recognition algorithms are provided: A Viola-Jones detector which is a linear classifier and model-based method. Hence, here the algorithm which contains real time ability is judged. On the opposition, the real-time and non-real-time algorithms develop the fast detectors. Here, the real-time approach is described by two properties: In first stage mostly sliding windows are evaluated in which there are only few classifiers. By these features while training simple Haar-like filters are evaluated by a pre-calculated image which contains gray values.In paper (Li et al., 2016), here this paper Centre and looks over the intensifying the performance of pre and post processing methods. The focal point is on the speed limit signs, the hardest super class in the U.S traffic sign net.

Over past few decades, as sensor technology has developed cars are provided with Advanced Driver Assistance System [ADAS] in different forms are common in today's cars. Recognition of Traffic sign consists two layers. They are1. Detection layer, 2. Classification layer. As U.S traffic signs has been detected by a current study and has accepted new benchmark based on ACF [AGGREGETED CHANNEL FEATURES] object detector .The results of VIVA challenge in 2015 and 2016 are shown in the Table 1:

Table 1. ***Results of VIVA challenge***

Method used	No Turn	Limit of Speed	Cease	Deterrent	Avg. AUC
ACF	96.17%	84.31%	96.11%	98.98%	93.89%

The paper (Yang et al., 2016) holds text-based traffic symbols and the symbol, these are divided into seven categories, which are compared to the before used symbol-based signs. Instead of stable images, every sample in the dataset holds a video of 5~20 low standard frames which are seized from a camera inside the vehicle. In Japan in 1984 (Habibi Aghdam & Jahani Heravi, 2017), the initial step on recognition of traffic sign was carried out and to resolve the problem many suggestions were given through distinct techniques (Aghdam et al., 2017). The majority one is Support Vector Machines (SVM) (Yang et al., 2016) and (Maldonado Bascón et al., 2010), template matching (Piccioli et al., 1996) and (Gao et al., 2006) and recently CNNs. On traffic sign classification CNNs have better performance than human performance (Stallkamp et al., 2012). However, there are different architectures in CNN. The training of last layers of pre-trained models discussed in (Salluri et al., 2020).

Model-based methods (Barnes et al., 2005) and (Belaroussi & Tarel, 2009) rely on robust edge detection and linking regular polygons to circles, generally via a Hough-like voting mechanism or template matching.The authors discussed about the Object detection using convolutional neural neworks (Salluri et al., 2020).The images in the dataset were transformed to RGB colour and saved in PPM file format using the constant-hue Demosaicking technique (Wang et al., 2005) and (Ramanath et al., 2002).On the LISA-TS dataset, the world's largest public traffic sign dataset for the United States, (Mogelmose et al., 2014) implemented ICF (Integral Channel Features) and ACF (Aggregated Channel Features) (Appel et al., 2014). (Mohan et al., 2015).Advanced Driver Assistance Systems (ADAS) are widely utilised in today's automobiles, and research in autonomous cars are soaring (Maurer et al., 2016) and (Galceran et al., 2015).

III. METHOD

As Traffic signs are very useful while travelling ahead, in order to recognize the sign board which is placed in side of the road, we had presented an automatic traffic sign classification using deep neural networks which will guarantee the life of a human and enhance the road traffic safety situations. Deep learning is a subset of machine learning that is based on Artificial Neural Networks. We discussed artificial neural networks briefly in the first section. Later in second section, we had discussed about the pre-processing and in third section, discussed about the pretrained models of VVG-19 and VGG-16.

ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks (ANN) function similar to a human brain where they receive, process, and signal a neuron linked to it. In neural network, when an image is send as an input it will process the image and then gives an output. Starting with input layer followed by the two hidden layers and then generates the output layer as shown in Figure 1.

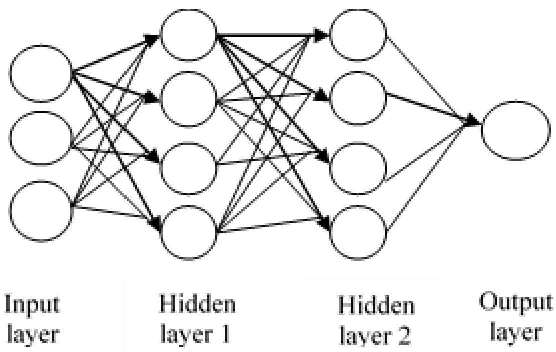

Figure 1. Sample model of Deep Neural Network

PRE-PROCESSING

The dataset includes 50,000 images of traffic signs.These images are divided into 43 classes, with 100 images in each class. The dataset contains different sizes of images, so make it uniform first all the images are converted into equal size. In pre-processing the image is resized to 224×224. After pre-processing rotation and mirroring of images is done. The GTSRB dataset contains 50,000 images among which 4300 images were taken for processing. 3440 images were used for training and 860 images were used for testing out of the 4300 total images.

PRE-TRAINED MODELS

The Pre-trained models work similar to CNNs. When we use image data to train a CNN, general feature extraction is done and further extraction of specific features is done. For pre-trained models we do not to train them from the beginning because they were already aware of feature extraction which is referred as Transfer learning.Transfer learning is used on all layers of the proposed pre-trained models, except for the last layer.Only the last layer, namely the Dense layer, drop out, global average pooling, and applied softmax layer, is replaced by our own predicting layers, while the remaining layers are frozen.Now, the pre-trained models have been trained on the predicting layers, which are combined to produce the output.This method will consume less time and produce more accurate results.

VGG-19

VGG-19 is a model made up of 19 layers.Convolutional layers make up 16 of the 19 layers and input image size is 224×224×3. After sending the images as input,its features are extracted in convolution layer and then followed by max pooling layer where the image size is reduced. The same process is repeated until the image size is reduced to 14×14×512, at which point it is forwarded to Fully connected layers for image classification.Except the last layer all the layers that are present in pre-trained models are fed to transfer learning by frozing all the layers.The final layer will be removed and replaced with a Dense layer, drop out, global average pooling, and an applied softmax layer of our own predicting layers, and the output will be produced by categorising the images, as shown in Figure 2.

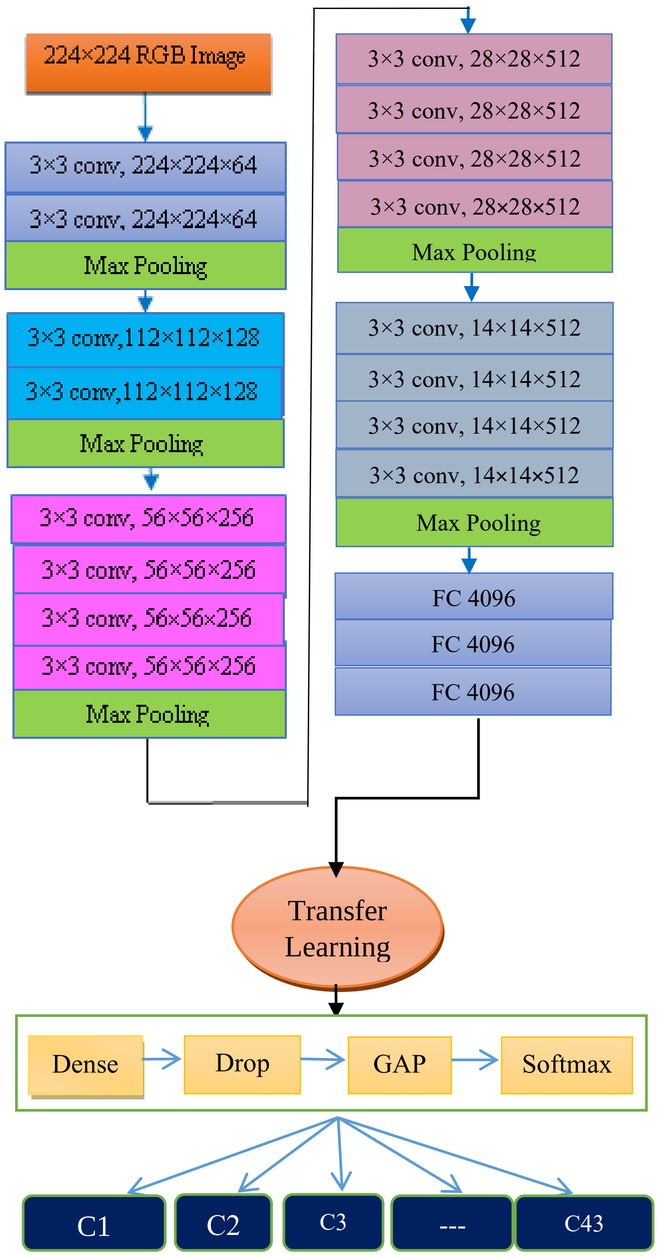

Figure 2. Proposed VGG-19 model

It is comprised of convolutional layers, sub inspecting layers and completely associated layers. The dataset on which we are pre-training these models contains 43 classes as shown in the Figure 2 and the categories of the traffic signs are shown in Figure 3.

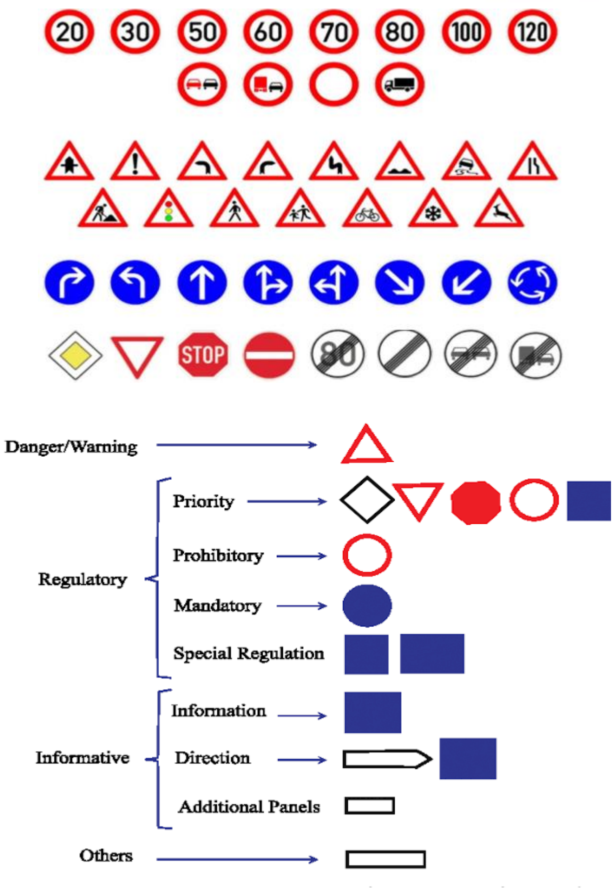

Figure 3. Definition of European traffic sign classifications.The primary category, subcategories, and most common shapes are listed from left to right.

Convolution Layer

This layer is the first and basic step in CNN. We extract certain characteristics from the input image in this layer.Filters are the terms used to describe the extracted features.The filter's aim is to move all over the image and execute a dot product, after which the features are saved. Not only one filter is chosen similarly other filters are also taken from the input image and the same process is done to see the difference and the result.

ReLu Layer

This layer acts as an activation function. In this function the values have less than 0 i.e the negative value is replaced by 0 whereas the positive value is kept as same.

Pooling Layer

It is also called as sub sampling layer. Pooling is of three types: Max. Pooling, Min. Pooling and Avg. Pooling. This layer reduces the size of the image to perform Max. Pooling on the image firstly select one stride. Next scroll the stride over the filtered image and from that select one Max. Value and store it in the feature map. By performing this operation, the dimensionality of the image reduces. Similarly, to perform Min. or Avg. Pooling the same process is done.

Fully Connected Layer

This is the last layer in CNN. In this layer all the neurons have connected to the activations present in the previous layer. The neurons present in this layer capture elements of the input image and then predict it.

VGG-16

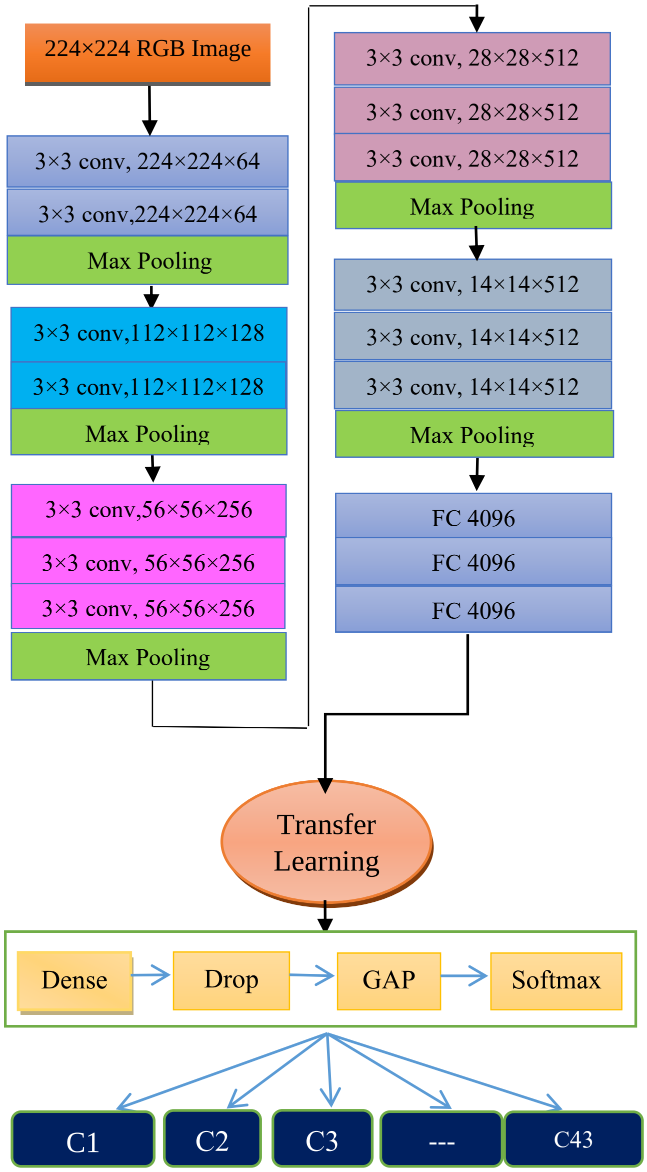

Figure 4. Proposed VGG-16 model

The VGG-16 model is a 16-layer pre-trained model.13 of these 16 layers are convolutional layers.The input image size is 224×224×3. After sending the images as input,its features are extracted in convolution layer and then followed by max pooling layer where the image size is reduced. The same process is repeated until the image size is reduced to 14×14×512 which then forwarded to Fully connected layers where classification of image is done.Except the last layer all the layers that are present in pre-trained models are fed to transfer learning by frozing all the layers.The final layer will be replaced by a Dense layer, drop out, global average pooling, and an applied softmax layer of our own predicting layers, and the output will be produced by categorising the images, as shown in Figure 4.

IV. RESULTS, DİSCUSSİON AND RECOMMENDATİONS

VGG-19 architecture is one of the deep neural networks with input size 224×224 was used to build the classification network. The network was pre-trained on the GTSRB dataset with 4300 images for classification which are categorised into 43 classes. It is a single image, multi-class classification problem. The network is implemented on Anaconda framework. The images that are present in GTSRB dataset are used as input where they were processed and the image category is send as output. Table 2 shows accuracies of two pre-trained models. The same process applied to the VGG-16 model also and comparing the accuracy between them.

Table 2. ***Accuracy of VGG19 & VGG16 models***

Model	Accuracy	Precision	Recall	F1-score
VGG19	95.23	96.00	96.00	96.00
VGG16	95.93	96.00	97.00	96.00

Performance Evaluation

Based on the accuracy test the images which are affected and the images which are not affected were determined.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

TP - True Positive, FN - False Negative, TN - True Negative, and FP - False Positive.

The accuracies that obtained from VGG-16 and VGG-19 ares shown below.

VGG-19

It consists of 19 layers. 224×224 is the image input size of this model. Here, GTSRB dataset is used which contains 50,000 images with 43 different classes. Among these 50,000 images only 4300 images were taken for processing and 100 images were present in each class. Among these 4300 images, 3440 images were used for training and 860 images for testing. Our proposed model consists of 43 classes. Softmax layer is used to modify the pre-trained model by our own predicting layers. For classifying the images 5 epochs were used and single CPU system is used. The accuracy that is obtained from this model is 95.23%. Training loss and testing loss in figure 5, training and testing accuracy in figure 6. The accuracy level of VGG-19in table 3 and confusion matrix in figure 7.

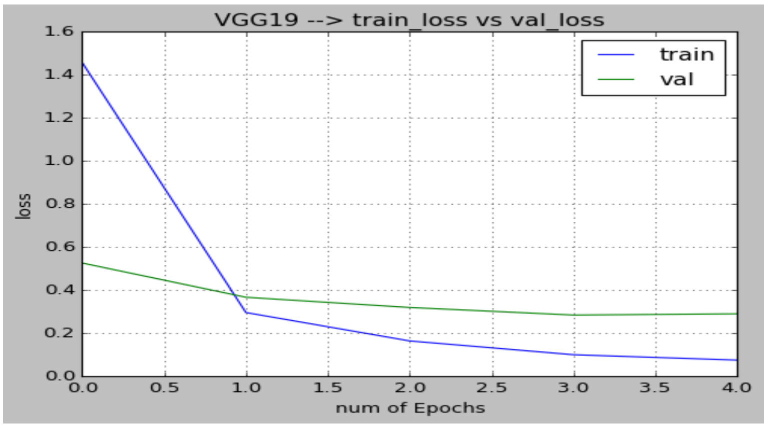

Figure 5.train_loss Vs val_loss

Training loss: The error that occurred during data training.

Validation loss: When data is processed by a neural network, an error occurs.

Dropout: It is used to solve network's overfitting issue. 0.3% dropout is used.

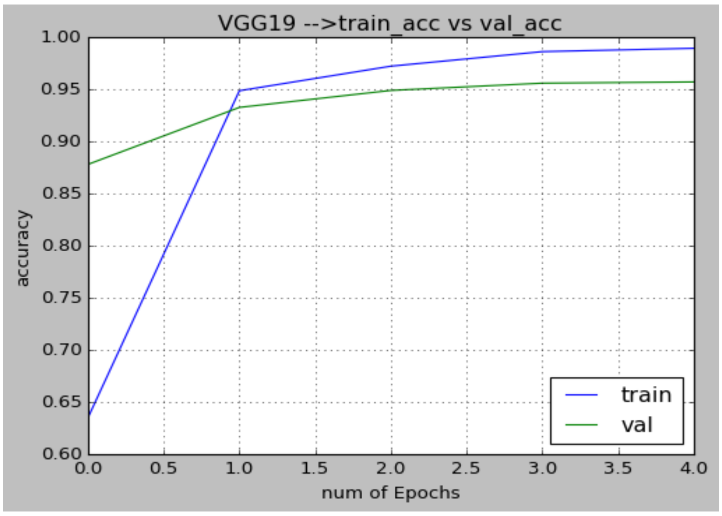

Figure 6.train_acc Vs val_acc

Training accuracy: The accuracy which is obtained from the trained data.

Validation accuracy: The accuracy which is obtained after neural networks processing.

Table 3. ***Accuracy level of VGG-19***

Class	Precision	Recall	F1-score	Support
0	0.95	1.00	0.98	21
1	0.74	0.53	0.62	32
2	0.93	0.93	0.93	15
3	0.91	1.00	0.95	20
4	1.00	0.88	0.94	25
5	0.96	0.92	0.94	24
6	1.00	1.00	1.00	20
7	0.89	1.00	0.94	17
8	0.95	1.00	0.98	21
9	0.95	0.95	0.95	21
10	1.00	1.00	1.00	22
11	1.00	1.00	1.00	21
12	0.94	1.00	0.97	16
13	0.94	1.00	0.97	15
14	1.00	1.00	1.00	16
15	0.88	1.00	0.94	15
16	0.96	1.00	0.98	24
17	1.00	1.00	1.00	22

18	0.94	1.00	0.97	15
19	1.00	1.00	1.00	11
20	1.00	1.00	1.00	23
21	0.89	1.00	0.94	24
22	1.00	0.95	0.97	19
23	0.95	0.95	0.95	20
24	1.00	0.94	0.97	16
25	0.96	1.00	0.98	23
26	1.00	1.00	1.00	24
27	1.00	1.00	1.00	21
28	1.00	1.00	1.00	16
29	1.00	0.92	0.96	13
30	1.00	1.00	1.00	18
31	1.00	0.95	0.97	19
32	0.95	0.95	0.95	22
33	0.95	0.95	0.95	21
34	0.93	0.81	0.87	16
35	0.86	1.00	0.92	18
36	1.00	1.00	1.00	19
37	1.00	0.86	0.92	28
38	1.00	1.00	1.00	24
39	0.96	1.00	0.98	26
40	0.86	1.00	0.92	12
41	0.92	0.96	0.94	25
42	1.00	1.00	1.00	20
accuracy			0.96	860
macro avg	0.96	0.96	0.96	860
weighted avg	0.96	0.96	0.96	860

Precision: The positive values are divided into two groups: correctly classified positive values and incorrectly classified positive values.

Recall: A distinction is made between positive values and relevant positive samples.

F1-score: The accuracy that is gained from precision and recall test.

Support: The positive samples present in the class.

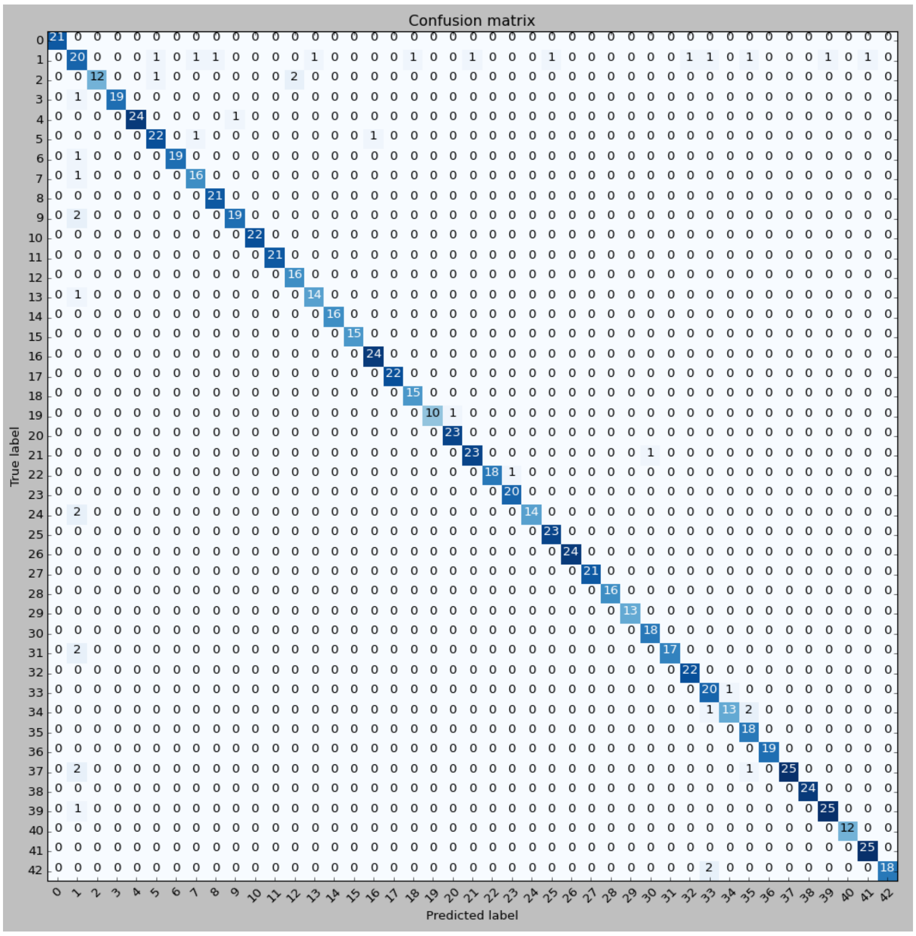

Figure 7. Confusion matrix

VGG-16

It consists of 16 layers.The image input size of this model is 224×224. Here, GTSRB dataset is used which contains 50,000 images with 43 different classes. Among these 50,000 images only 4300 images were taken for processing as 100 images were presented in each class. Among these 4300 images, 3440 images were used for training and 860 images for testing. Our proposed model consists of 43 classes. Softmax layer is used to modify the pre-trained model by our own predicting layers. For classifying the images 5 epochs were used and single CPU system is used. The accuracy that is obtained from this model is 95.93%. Training loss and testing loss in figure 8, training and testing accuracy in figure 9. The accuracy level of VGG-16 in table 4 and confusion matrix in figure 10.

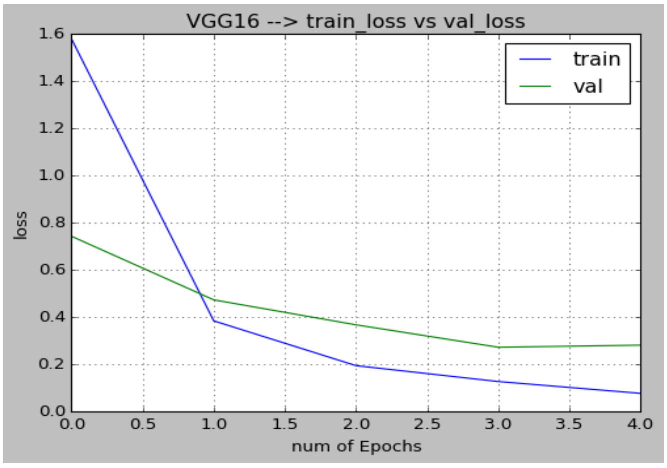

Figure 8.train_loss Vs val_loss

Training loss: The error that occurred during data training.

Validation loss: The error that occured in neural network when the data is processed.

Dropout: It is used to solve network's overfitting issue. 0.3% dropout is used.

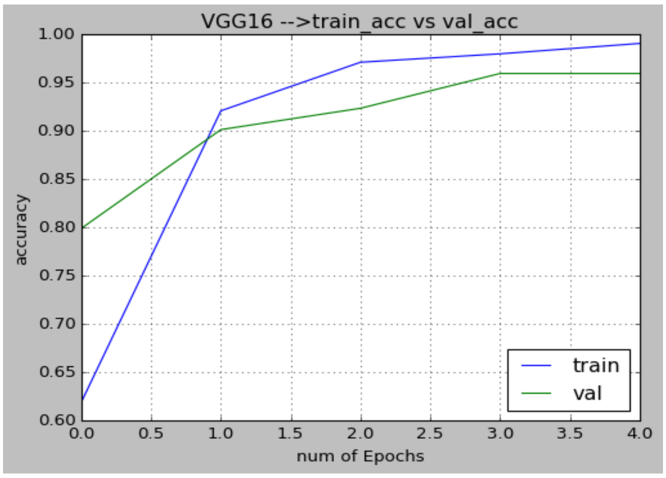

Figure 9.train_acc Vs val_acc

Training accuracy: The accuracy which is obtained from the trained data.

Validation accuracy: The accuracy which is obtained after neural networks processing.

Table 4. ***Accuracy level of VGG-16***

Class	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	21
1	0.77	0.62	0.69	32
2	0.94	1.00	0.97	15
3	1.00	1.00	1.00	20
4	1.00	0.96	0.98	25
5	0.96	1.00	0.98	24
6	1.00	1.00	1.00	20
7	0.94	1.00	0.97	17
8	0.91	1.00	0.95	21
9	1.00	0.86	0.92	21
10	0.96	1.00	0.98	22
11	1.00	1.00	1.00	21
12	1.00	1.00	1.00	16
13	0.94	1.00	0.97	15
14	1.00	0.88	0.93	16
15	1.00	0.87	0.93	15
16	1.00	1.00	1.00	24
17	0.96	1.00	0.98	22
18	0.94	1.00	0.97	15
19	1.00	1.00	1.00	11
20	1.00	1.00	1.00	23
21	0.96	0.92	0.94	24
22	0.95	1.00	0.97	19
23	0.95	1.00	0.98	20
24	1.00	1.00	1.00	16
25	0.96	1.00	0.98	23
26	1.00	0.96	0.98	24
27	1.00	1.00	1.00	21
28	0.94	1.00	0.97	16
29	1.00	1.00	1.00	13
30	1.00	1.00	1.00	18
31	0.94	0.89	0.92	19
32	0.95	0.95	0.95	22
33	0.88	1.00	0.93	21
34	1.00	0.94	0.97	16

35	0.90	1.00	0.95	18
36	1.00	0.95	0.97	19
37	1.00	0.93	0.96	28
38	0.96	0.96	0.96	24
39	0.93	0.96	0.94	26
40	0.86	1.00	0.92	12
41	0.88	0.92	0.90	25
42	1.00	0.95	0.97	20
accuracy			0.96	860
macro avg	0.96	0.97	0.96	860
weighted avg	0.96	0.96	0.96	860

Precision: The positive values are divided into two groups: correctly classified positive values and incorrectly classified positive values.

Recall: A distinction is made between positive values and relevant positive samples.

F1-score: The accuracy that is gained from precision and recall test.

Support: The positive samples present in the class.

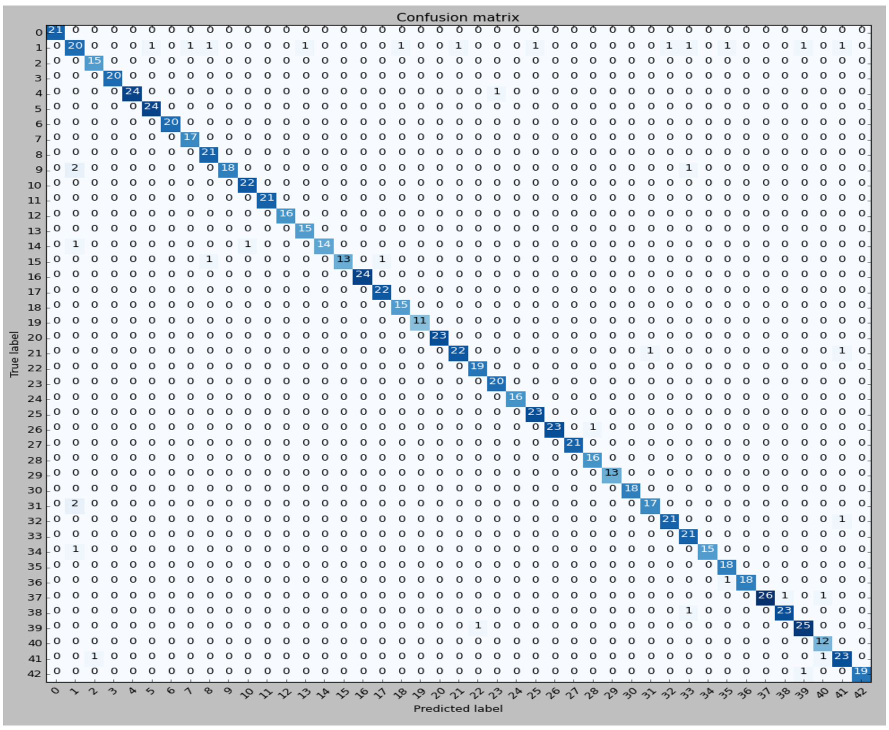

Figure 10. Confusion matrix

V. CONCLUSION

Image classification of Traffic sign images which contains 43 classes are chosen for testing and validation using deep learning. For classification the convolutional neural networks are used. From the experiment it is observed that the images are classified correctly, and deep learning algorithm effectiveness is shown. For training and testing purposes these datasets used CNN. VGG-16 has gained highest accuracy of 95.93%. As, here we had taken only 4300 images in future, we will increase the dataset and apply the same model to increase the accuracy.

VI. REFERENCES

Aghdam, H. H., Heravi, E. J., & Puig, D. (2017). A Practical and Highly Optimized Convolutional Neural Network for Classifying Traffic Signs in Real-Time. *International Journal of Computer Vision*, *122*(2), 246–269. https://doi.org/10.1007/s11263-016-0955-9

Galceran, E., Olson, E., & Eustice, R. M. (2015). Augmented vehicle tracking under occlusions for decision-making in autonomous driving. *IEEE International Conference on Intelligent Robots and Systems*, *2015-Decem*, 3559–3565. https://doi.org/10.1109/IROS.2015.7353874

Gao, X. W., Podladchikova, L., Shaposhnikov, D., Hong, K., & Shevtsova, N. (2006). Recognition of traffic signs based on their colour and shape features extracted using human vision models. *Journal of Visual Communication and Image Representation*, *17*(4), 675–685. https://doi.org/10.1016/j.jvcir.2005.10.003

Habibi Aghdam, H., & Jahani Heravi, E. (2017). Traffic Sign Detection and Recognition. *Guide to Convolutional Neural Networks*, 1–14. https://doi.org/10.1007/978-3-319-57550-6_1

Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., & Igel, C. (n.d.). *Houbenetal_Gtsdb_Ijcnn2013.Pdf*.

Li, Y., Møgelmose, A., & Trivedi, M. M. (2016). Pushing the “speed limit”: High-accuracy US traffic sign recognition with convolutional neural networks. *IEEE Transactions on Intelligent Vehicles*, *1*(2), 167–176. https://doi.org/10.1109/TIV.2016.2615523

Maldonado Bascón, S., Acevedo Rodríguez, J., Lafuente Arroyo, S., Fernndez Caballero, A., & López-Ferreras, F. (2010). An optimization on pictogram identification for the road-sign recognition task using SVMs. *Computer Vision and Image Understanding*, *114*(3), 373–383. https://doi.org/10.1016/j.cviu.2009.12.002

Maurer, M., Gerdes, J. C., Lenz, B., & Winner, H. (2016). Autonomous driving: Technical, legal and social aspects. In *Autonomous Driving: Technical, Legal and Social Aspects*. https://doi.org/10.1007/978-3-662-48847-8

Møgelmose, A., Trivedi, M. M., & Moeslund, T. B. (2012). 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. https://doi.org/10.1109/TITS.2012.2209421

Piccioli, G., De Micheli, E., Parodi, P., & Campani, M. (1996). Robust method for road sign detection and recognition. *Image and Vision Computing*, *14*(3), 209–223. https://doi.org/10.1016/0262-8856(95)01057-2

Salluri, D. K., Bade, K., & Madala, G. (2020). Object detection using convolutional neural networks for natural disaster recovery. *International Journal of Safety and Security Engineering*, *10*(2), 285–291. https://doi.org/10.18280/ijsse.100217

Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C. (2012). *Benchmarking Machine Learning Algorithms for Traffic Sign Recognition.pdf*.

Yang, Y., Luo, H., Xu, H., & Wu, F. (2016). Towards Real-Time Traffic Sign Detection and Classification. *IEEE Transactions on Intelligent Transportation Systems*, *17*(7), 2022–2031. https://doi.org/10.1109/TITS.2015.2482461