Traffic Sign Recognition & Detection using Transfer Learning.

Dinesh Kumar Padmanabhan, M.S.

Nima Zahadat, Ph.D.

The George Washington University

**ABSTRACT**

Traffic sign recognition feature is widely employed in industry today by researchers working in artificial intelligence and machine learning fields targeted to create an autonomous driving assistance system. This will help the driver’s human eye specifically the maneuvering capabilities which can’t be presumed to be consistent for high accuracy for all weather condition challenges. Road safety, in particular, hinges on the accurate human eye perception of road signs. This research paper illustrates a novel approach for traffic image classification using Deep Learning Model to distinguish the different classes of traffic signs using data augmentation offered by Keras Deep Learning Library for shifting traffic signs images. This will help us recreate a dataset similar to images captured during bad climatic, foggy or windy conditions or badly captured images due to fast moving cars with limitation in camera angles from roads. We will be introducing these pictures on a state-of-the-art architectures pre-trained transfer learning model such as VGG19, ALEXNET and RESNET50 as they provide better accuracy and operational speed to create a best fit model. The model generated a 98% accuracy in German sign recognition benchmark.

*Keywords:* Data Augmentation, Deep Learning, Keras, Transfer Learning

**INTRODUCTION**

As self-driving cars are being introduced in major cities, intelligent traffic signs recognition has become an essential part of any autonomous driverless vehicles. Transitioning from a vehicle with driver to a driverless vehicle should be gradual. The major issue encountered by drivers on the road is the failure to notice traffic signs immediately, owing to lapses in attentiveness caused by a myriad of factors. Automating the recognition of traffic signs through the installation of a safety harness could be beneficial, as it alerts the driver in due time . Addressing the lapses in driver attentiveness through modern safety gear can greatly reduce the number of casualties in major roads and highways. Orthodox computer vision techniques and machine learning-based architectures were popular for traffic sign classification but was yet to be successful in curbing road fatalities.

Recognition of traffic signs is a challenging real-world problem of high industrial relevance. Although commercial systems have reached the market and several studies on this topic have been published, systematic unbiased comparisons of approaches are missing, and comprehensive benchmark datasets are not freely available. Sign recognition is a multicategory classification problem with unbalanced class frequencies. Traffic signs show a wide range of variations between classes in terms of color, shape, and the presence of pictograms or text. However, there exist subsets of classes (e.g., speed limit signs) that are very similar to each other. The classifier has to cope with large variations in visual appearances due to illumination changes, partial occlusions, rotations, weather conditions, scaling, etc.

In this paper, we address the issue of learning and detecting a large number of traffic-sign categories for road-based traffic sign inventory management. As our main contribution, we propose a deep learning-based system for training a large number of traffic-sign categories using convolutional neural networks. We base our system on the state-of-the-art transfer learning models, which demonstrated great accuracy and speed in the field of detection.

The remainder of the paper is organized as follows. Section II provides the related work overview, Section III describes the data analysis and the research methodology in Section IV, Section V presents the key findings and discussion on recommendations for future research is provided in Section VI. The paper concludes with the discussion in Section VII.

**LITERATURE REVIEW**

Various methodical approaches were adopted for recognizing traffic signs in the field of computer vision. However, most of them were referenced to only one set of benchmark. In this work, we discuss such methodologies starting from classical brute force approaches to modern learning representations. Most of them tackle the fundamental classification, detection and localization challenges surrounding traffic sign recognition.

1. *Feature Extractor as Classification Technique*

Classical approaches include techniques such as Histogram Oriented Gradients (HOG), Scale Invariant Feature Transformation (Sift) and Sliding Window. HOG based techniques were used for visual salience and then for color exploitation for pedestrian detection. Moreover, gradients of red, green, and blue (RGB) images were computed along with different normalized, weighted histograms for finding the best detection algorithm to find pedestrians and signs. Furthermore, the Scale Invariant Feature Transform (SIFT) technique was used for classification of traffic signs, whereas the sliding window approach was used to find candidate Region of Interest (ROI) within a small-sized window, and then further verified within a large-sized window for higher accuracy in object detection.

1. *Orthodox Machine Learning Techniques*

Machine learning methods like Support Vector Machines (SVM), Linear Discriminant Analysis (LDA) , KD-trees , Maximally Stable Extremal Regions and Random Forest minimized the brute force approaches in traffic sign recognition. Concurrently, LDA is based on maximum likelihood estimate or maximum posteriori estimation between classes and class densities are represented by multivariate Gaussian and common covariance matrix. However, discriminant function analysis is very similar to logistic regression, and can both be used to answer the same research questions. Logistic regression does not have as many presumptions and restrictions as LDA. Nevertheless, when discriminant analysis suppositions are met, it is more superior and stronger than logistic regression. In Random Forest, a set of non-pruned random decision trees are used to make an ensemble architecture through which the best classification scores are achieved. The decision trees are made with features selected randomly from the training set. For traffic sign recognition, the test data is validated by all the decision trees and the categorical output and probability scores are based on majority voting. Support Vector Machines (SVM) is used for classification problems and it classifies the data by dividing the n-dimensional data plane with a hyper plane. SVM can transform the classification plane to higher dimensions using kernel trick. The method separates the non-linearly scattered data using a non-linear kernel function. The major problem for the above-mentioned techniques is that the features need to be hand-engineered and machine learning is heavily dependent on human-intervention. These approaches cannot handle variable length images, and neither can it converge better with data augmentation and low pre-processing.

1. *Data augmentation*

An important factor to consider when learning deep models is the size of the training set. Due to millions of learnable parameters, the system becomes undetermined without a sufficient number of training samples. We partially address this issue with a pre-trained model, one learned on 1:2 million images of ImageNet, but with additional data augmentation. The nature of the traffic-sign domain allows us to construct a large number of new samples using artificial distortions of existing traffic-sign instances.

1. *Deep Learning Approaches*

To eliminate the drawback of the above-mentioned techniques, new architectures based on deep learning algorithms emerged, due to the increase in the amount of available computing resources and access to huge, annotated data. Currently, almost all the state-of-art architectures for traffic signs are convolutional neural networks. The first of its kind was the introduction of LeNet Architecture. The tower-like structure associated with convolutional neural networks makes it more accurate and easily implementable, which enables it to process information and learn features in-depth. The variations in each block and layers includes: (1) the convolution layer which is the main feature extractor that uses filters with small receptive field to process input; (2) the pooling layer which is used to reduce spatial dimension; and the(3) dense or fully-connected layer which takes input from all the neurons of the previous layers and shares the information to connected layers. A loss function is defined that is subsequently reduced by Back-propagation. Moreover, a new type of convolution called dilated convolution has replaced vanilla convolution in latest architectures.. Dilated convolution involves stacking the layers vertically which increases the receptive field exponentially, whereas the stacking involved in the vanilla convolutional layer increases the receptive linearly.

**DATA**

This paper uses the German Traffic Sign Recognition Benchmark (GTSRB), which was presented at the 2011 International Joint Conference on Neural Networks (IJCNN). The internal traffic signs are collected from the real road traffic environment in Germany, and it has become a common traffic sign dataset used by experts and scholars in computer vision, self-driving and other fields. The GTSRB comprises 51,839 images which are divided into training and testing sets. A total of 39,209 and 12,630 images are provided in the training and testing sets, accounting for approximately 75% and 25%, respectively. Each image contains only one traffic sign, which is not necessarily located in the center of the image. The image size is unequal; the maximum and smallest images are 250X250 and 15X15 pixels, respectively.

The traffic sign images in GTSRB are taken from the video captured by the vehicle-mounted camera. As shown in Figure 7, GTSRB includes 43 classes of traffic signs, and the number of different types of traffic signs varies. Each type of traffic sign corresponds to a catalogue, which contains a CSV file annotated with a class label and a single image of multiple tracks. The traffic signs include different resolutions, illumination conditions, weather conditions, occlusion degree, tilt levels and other images, making the dataset more in line with the actual road scenes.

**DATA ANALYSIS**

GTSRB contains 43 classes/types of different traffic signs, with some class containing more than 2010 images while some other class contains 210 images, and such a dataset is called class imbalance or “The Class Skew”. The unequal distribution of samples across classes in such dataset will result in overfitting which in turn will affect the generalization of the entire network model.

Data augmentation techniques are often used to regularize models which work with images in neural networks and other learning algorithms. With the labelled original training dataset, synthetic images can be created by various transformations to the original images. Keras Image Data Generators is the tool used for generating more training data from the original data to avoid model overfitting. It is conducted online by looping over in small batches during each optimizer iteration. There are some graphic parameters (e.g. rotation, shift, flip, add Gaussian noises) to help generate artificial images.

Generating an artificial dataset aims to construct a new artificial sample by randomly sampling from the value space of each attribute feature of the same sample type. Therefore, the number of different kinds of traffic signs is as equal as possible to solve the problem of sample data imbalance.

**RESEARCH METHODOLOGY**

1. *Model Implementation Environment*

To execute this project, Windows 10 64-bit operating system was used for creating scripts and AWS for hosting the model. JetBrains PyCharm 2019.3.5, Anaconda libraries, TensorFlow 2.4.0, Python 3.7.4 64-bit were used as software environments. Amazon AWS: NVIDIA Corporation GM204GL [Tesla M60] (rev a1) was used as computation environment.

1. *Methodology*

Traffic sign classification and recognition experiment can be divided into two stages, namely, the network training and testing stages. In the network training stage, the training set samples of GTSRB are taken as input. By performing thousands of network iterations, parameters, such as network weights and offsets, are continuously updated on the basis of forward learning and back propagation mechanisms until the loss function is reduced to the minimum, thereby classifying and predicting traffic signs. In the network testing stage, the testing set samples of GTSRB are inputted into the trained network model to test the accurate recognition rate of the training network.

In the Network Training Stage, the training set samples are preprocessed; the artificial dataset is generated, and the dataset order is disrupted. The Gabor kernel is used as the initial convolutional kernel, and the convolutional kernel size is 5 x5, as activated by the ReLU function. The training set samples are forwardly propagated in the network model, and a series of parameters are set. The BN is used for data normalization after each pooling layer, and the Adam method is used as the optimizer algorithm. The dropout parameter is set to 0.5 in the fully-connected layers, and the Softmax function is outputted as a classifier. The gradient of loss function is calculated, and the parameters, such as network weights and offsets, are updated on the basis of the back-propagation mechanism. The error between the real and the predicted value of the sample is calculated, when the obtained error is lower than the set error or reaches the maximum number of training, training is stopped and final the classification test is conducted in the network model. The subordinate categories of traffic signs in the GTSRB are predicted and compared with the real categories. The classification prediction results of traffic signs are counted, and the correct prediction rate is calculated.

In the network testing stage, several images are randomly selected from the testing set samples, and the images are inputted into the trained network model after preprocessing. The recognition results are outputted through the network model, thereby showing the meaning of traffic signs with the highest probability and the output results are compared with the actual reference meanings, and the statistical recognition results are obtained.

**KEY FINDINGS**

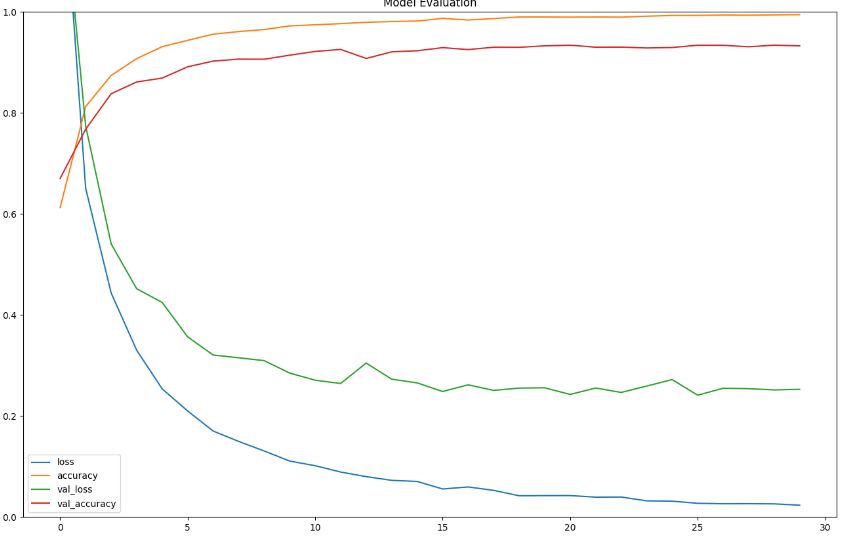
The model performance was evaluated with accuracy and loss function for the training, validation and test datasets in the absence or presence of applying multiple data augmentation techniques. The characteristics derived from confusion matrices are used to compare individual class predictions.

1. *Basic CNN Model for Image Classification*

In a class skew dataset, it is always a good practice to understand where the model stands without doing any preprocessing (augmentation) as that would help you establish a score for the model, which you could improve upon each iteration.

Model performance is evaluated with train and test accuracy and loss curves. These curves are plotted together for better comparison. Accuracy is defined as the ratio of correctly predicted observation to the total observations, and loss is the categorical cross-entropy to predict class probabilities (i.e. the difference between the predicted value and the true value). The basic CNN model is highly overfitted even after several epochs of model training. The two loss curves are also widely separated (Figure 1). This suggests that more layers and neurons should be added into the CNN architecture for better model performance. Some regularizations should also be employed, such as dropout layers, kernel regularizes, and batch normalization layers. Image misclassification happens when an image is incorrectly predicated with higher probability of other classes. For example, the “turn right ahead” in the first image first row column 4 is misclassified as “ahead only” (Figure 2).

*Figure 1: Accuracy and loss curves*

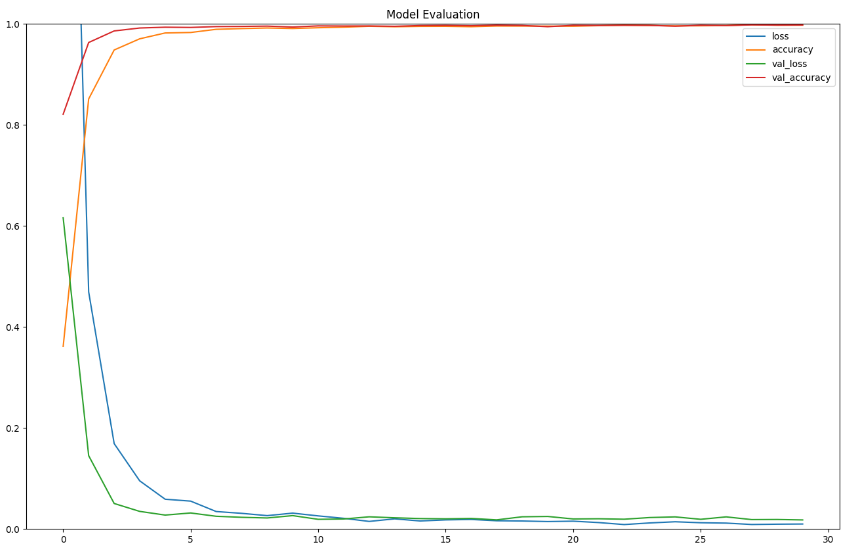


*Figure 2: Misclassified image samples by the basic CNN model*

1. *Data Augmentation for Image Classification*

Compared with the above model, training with the augmented data in the model requires the training data to pass through a Keras ImageDataGenerator. These data are looped over in mini-batches indefinitely. Different image data augmentation techniques and parameters are applied. General techniques like rotation, horizontal flip and translation are used in all other two models.

Training with augmentation allows the model to reach a better accuracy as there is no difference between the training and test accuracy(Figure 3), and thus the Image classification happens where there is no image that is predicted incorrectly (Figure 4).



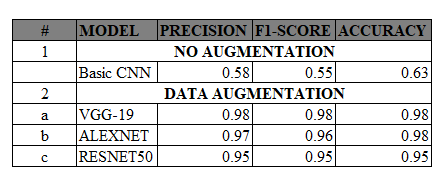
*Figure 3: Accuracy and Loss Curves*



*Figure 4: Classified Images*

1. *Accuracy Reports for Transfer Learning*

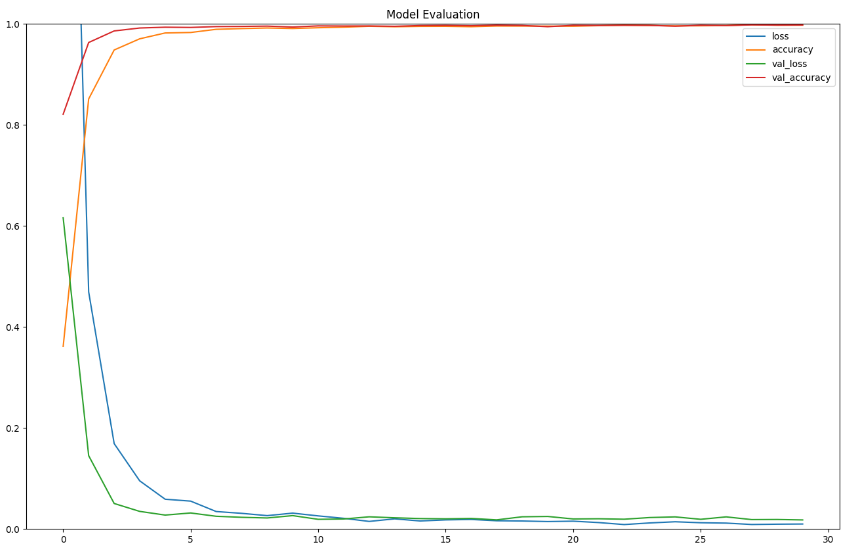
From the accuracy and loss curves below, the overfitting is easy to spot. As you cansee from the accuracy curve, when training with augmentation, the accuracy onthe test set are at levels while the accuracy on the training set keepsdecreasing. There is a small gap between those two curves, which clearly shows thatstill there is a small degree of overfitting (see Figures 5 to 10).

The proposed algorithm is compared with other algorithms adopted in various literature to verify the performance of traffic sign recognition algorithms. Table 1 lists the comparison of statistics in algorithm performance based on the GTSRB dataset.

*Table 1: Performance comparison*

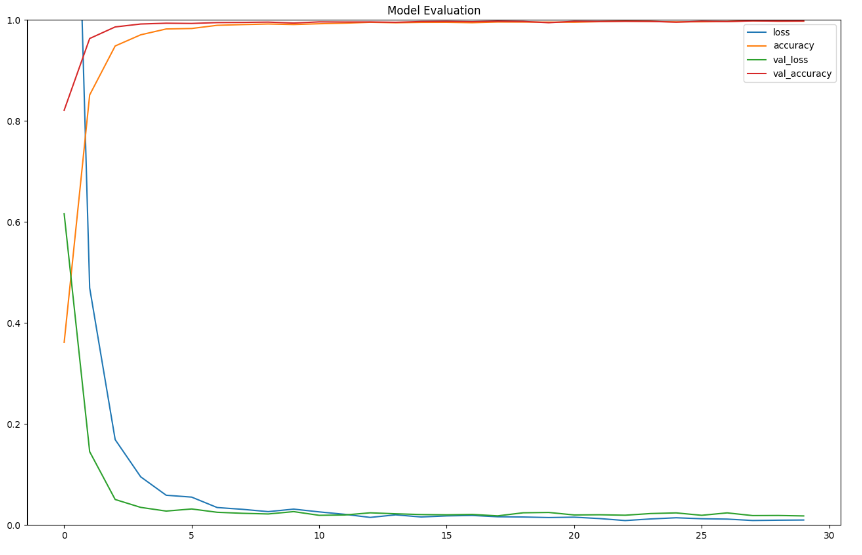


*Figure 8: Vgg-19 Image Misclassification*

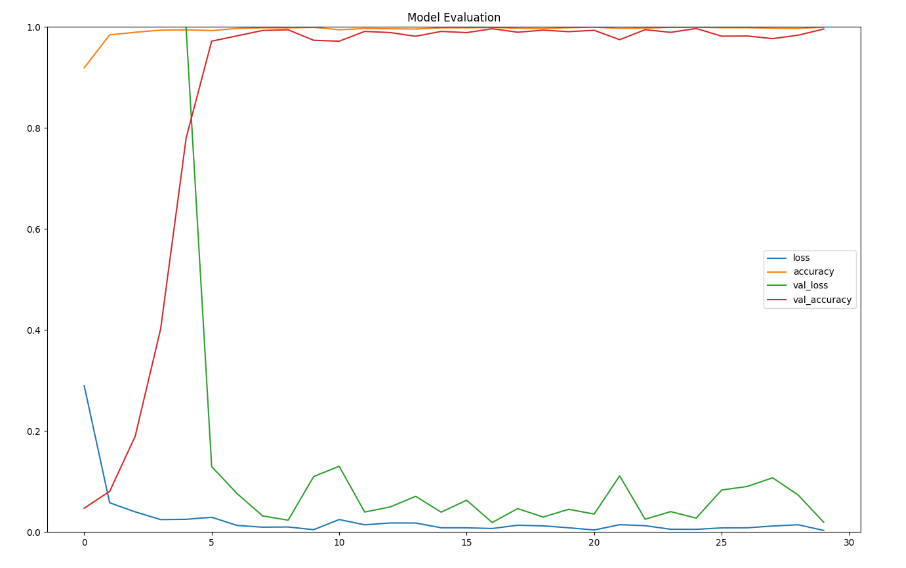


*Figure 5: VGG-19 Accuracy and loss curves*

*Figure 9: AlexNet Image Misclassification*



*Figure 6: AlexNet Accuracy and loss curves*



*Figure 7: Resnet50 Accuracy and loss curves*



*Figure 10: ResNet50 Image Misclassification*

**RECOMMENDATIONS FOR FUTURE RESEARCH**

Model patching enables automating the process of model maintenance and improvement when a deployed model exhibits flaws. This methodology is becoming a ground breaking area that would alleviate the major problem in safety-critical systems, including healthcare (e.g. [improving models to produce MRI scans free of artifact](https://ai.facebook.com/blog/fastmri-leverages-adversarial-learning-to-remove-image-artifacts/)) and autonomous driving (e.g. improving perception models that may have poor performance on irregular objects or road conditions).

Also YOLO for object detection with bad and challenging video to get real-time prediction on video captured in extremely bad weather conditions and suggestions for safety on roads will be the next challenge to work on.

**CONCLUSION**

In this study, an improved traffic sign detection and recognition algorithm is proposed forintelligent vehicles. Firstly, the basic model with no augmentation and traffic signs were not effectively detected due to overfitting. Secondly, this model is considerably improved on the basis of data augmentation by selecting Adam method as the optimizer algorithm. Finally, the traffic sign classification and recognition experiments were run on state-of-the-art transfer learning models that generated a favorable prediction and accurate recognition of traffic signs and achieved through the continuous training and testing of the network model. The experimental results show that the accurate recognition rate of traffic signs reaches 98%. The proposed algorithm has more admirable accuracy, better real-time performance, stronger generalization ability and higher training efficiency than other algorithms.

From the viewpoint of traffic sign recognition accuracy and algorithm time-consuming, the proposed traffic sign detection and recognition algorithm has remarkable advantages. Considerably enhancing the driving safety of intelligent vehicles in the actual driving environments and effectively meeting the real-time target requirements of smart cars are conducive. Furthermore, a strong technical guarantee is provided for the steady development of intelligent vehicle driving assistance. In the future, the inclusiveness and anti-error recognition of the traffic sign recognition algorithm can be further optimized and improved to exploit the overall performance of the algorithm.

Despite excellent performance of the proposed approach there is still room for improvement. Our analysis revealed that the ideal performance is still not achieved, mostly due to several missed detections that were lost by the classification network. Future improvements should focus on improving this part of the system.

**BIOGRAPHY**

**Dinesh Kumar Padmanabhan** is a graduate student in the Data Science Program at The George Washington University. His interests include business analytics and data science consulting. He has worked in software industry in various analytical positions for eight years.

Dinesh also enjoys biking, cooking, and photography.

**Dr. Nima Zahadat** is a Professor of Information Systems and Computer Science. His research interests include data mining, information visualization, digital forensics, mobile security, risk management and information security. Dr. Zahadat has worked extensively in both the public and private sectors and has held positions such as Chief Security Officer, Chief Information Officer, Director of Security, Director of Training Solutions, Dean of Computer Science, Program Chair of Information Systems, and Director of Operations.

Dr. Zahadat also enjoys biking, photography, travel, skiing, and writing.

**REFERENCES**

Stanford Artificial Intelligence Laboratory. [online] Available at: http://ai.stanford.edu/ [Accessed 1 Aug. 2019].

Ramler, R. and Ziebermayr, T. (2017). What You See Is What You Test - Augmenting Software Testing with Computer Vision. 2017 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW).

Nguyen, B. T., Ryong, S. J., & Kyu, K. J. (2014). Fast traffic sign detection under challenging conditions. 2014 International Conference on Audio, Language and Image Processing.doi:10.1109/icalip.2014.7009895

Changzhen, X., Cong, W., Weixin, M. and Yanmei, S. (2016). A traffic sign detection algorithm based on deep convolutional neural network. *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*.

Mahatme, M. and Kuwelkar, M. (2017). Detection and recognition of traffic signs based on RGB to red conversion. *2017 International Conference on Computing Methodologies and Communication (ICCMC)*.

Tastimur, C., Karakose, M., Celik, Y. and Akin, E. (2016). Image processing-based traffic sign detection and recognition with fuzzy integral. *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*.

Yanjun Fan and Weigong Zhang (2015). Traffic sign detection and classification for Advanced Driver Assistant Systems. *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*.

Liu, C., Chang, F., Chen, Z. and Liu, D. (2016). Fast Traffic Sign Recognition via High-Contrast Region Extraction and Extended Sparse Representation. *IEEE Transactions on Intelligent Transportation Systems*, 17(1), pp.79-92.

van Opbroek, A., Ikram, M., Vernooij, M. and de Bruijne, M. (2015). Transfer Learning Improves Supervised Image Segmentation Across Imaging Protocols. *IEEE Transactions on Medical Imaging*, 34(5), pp.1018-1030.

CENGIL, E. and CINAR, A. (2018). A Deep Learning Based Approach to Lung Cancer Identification. *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*.

Chou, L., Tseng, C., Lai, M., Chen, W., Chen, K., Yen, C., Ou, T., Tsai, W. and Chiu, Y. (2018). Classification of Malicious Traffic Using TensorFlow Machine Learning. *2018 International Conference on Information and Communication Technology Convergence (ICTC)*.

Choi, S., Jeong, H., Park, K. and Ha, Y. (2019). Efficient Driving Scene Image Creation Using Deep Neural Network. *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*.

Ling Shao, Fan Zhu and Xuelong Li (2015). Transfer Learning for Visual Categorization: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, 26(5), pp.1019-1034. 24

Bruha, I. (2000). From machine learning to knowledge discovery: Survey of preprocessing and postprocessing. *Intelligent Data Analysis*, 4(3-4), pp.363-374

Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. (2019). In: *KDD '19 Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: ACM New York, NY, USA ©2019, p.3253.

Gertosio, C. and Dussauchoy, A. (2004). Knowledge discovery from industrial databases. *Journal of Intelligent Manufacturing*, 15(1), pp.29-37.

Mair, B., Wilson, D. and Reti, Z. (1996). Deblurring the discrete Gaussian blur. *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*.

En.wikipedia.org. (2019). *Gaussian blur*. [online] Available at: https://en.wikipedia.org/wiki/Gaussian\_blur [Accessed 9 Aug. 2019].

Seetala, K., Birdsong, W. and Reddy, Y. (2019). Image Classification Using TensorFlow. *16th International Conference on Information Technology-New Generations (ITNG 2019)*, pp.485-488.

*Deep Learning*. [online] Available at: http://deeplearning.net/ [Accessed 9 Aug. 2019].

Medium. (2019). *Custom Image Augmentation*. [online] Available at: https://towardsdatascience.com/image-augmentation-14a0aafd0498 [Accessed 9 Aug. 2019].

*Very Deep Convolutional Networks for Large-Scale Image Recognition*. [online] arXiv.org. Available at: https://arxiv.org/abs/1409.1556 [Accessed 9 Aug. 2019]

TensorFlow. (2019). *TensorFlow Core | TensorFlow*. [online] Available at: https://www.tensorflow.org/tutorials/ [Accessed 2 Aug. 2019].