ICIC Express Letters
Part B: Applications
Volume **15**, Number **4**, April **2024**

ICIC International ©2024 ISSN 2185-2766
pp. 381–388

DATA AUGMENTATION FOR OCCLUSION-ROBUST TRAFFIC SIGN RECOGNITION USING DEEP LEARNING

Andrew Dineley[1], Friska Natalia[2] and Sud Sudirman[1,*]

[1]School of Computer Science and Mathematics
Liverpool John Moores University
Liverpool, L3 3AF, United Kingdom
a.m.dineley@2020.ljmu.ac.uk; *Corresponding author: s.sudirman@ljmu.ac.uk

[2]Faculty of Engineering and Informatics
Universitas Multimedia Nusantara
Scientia Boulevard, Gading Serpong, Tangerang, Banten 15811, Indonesia
friska.natalia@umn.ac.id

Received July 2023; accepted October 2023

Abstract. *Traffic sign recognition is an essential feature for self-driving cars. It provides input to the decision-making process when maneuvering through traffic in real time. Correct identification and classification of traffic signs are a challenge because they may be occluded by natural entities, such as leaves and trees, or man-made such as graffiti. In this paper, we present the result of our study into achieving occlusion-robust traffic sign recognition by augmenting the data used to train deep learning models. The data augmentation is performed by applying random occlusion of varying coverage percentages to the traffic sign images. We investigated the performance of four different deep network architectures to recognize 11 German speed limit signs using transfer learning techniques on their respective pre-trained models (AlexNet, VGG19, ResNet50, and GoogLeNet). The results of our experiment show that our data augmentation technique improves the recognition accuracy at higher occlusion band (61%-70% occlusion) by 17% using GoogLeNet with a slight 2% hit in accuracy at lower occlusion band (1%-10% occlusion). Our study concludes that our data augmentation technique could significantly improve the recognition performance of all models when the traffic sign images are severely occluded.*

Keywords: Traffic sign recognition, Data augmentation, Deep learning, Occlusion, Computer vision

1. **Introduction.** The concept of cars that can drive autonomously is not new, but it is only recently that the technology that is required to make it happen is mature enough allowing us to see a small number of them deployed in real life. Self-driving cars use a combination of different technologies and hardware to operate. They sense the situation around and in front of them through different hardware sensors. The radar and LIDAR sensors detect nearby vehicles and humans. They use the Global Positioning System to locate where they are on the map and navigation software to decide which exit or turn to take. A computerized map helps the car navigate roads, traffic jams, and construction zones. By analyzing images taken by onboard cameras, the software also determines when it is safe to change lanes or turn corners.

Although still limited, the adoption of self-driving cars is growing in many developed countries as the technologies behind them are progressing rapidly. For example, the Waymo One Robotaxi service by Alphabet (Google's parent company) is slowly expanding across the US [1], whereas, in the UK, the automated lane-keeping systems [2] have been approved for hands-free use on 95% of Britain's major motorways. However, there are still many challenges that need to be addressed before self-driving cars become mainstream.

DOI: 10.24507/icicelb.15.04.381

One of these challenges is improving the performance of the software that is used to analyze the images taken by the onboard cameras to detect and recognize any traffic signs in all lighting and weather conditions.

In this paper, we present the result of our study into achieving occlusion-robust traffic sign recognition. Our methodology is based on our previous approach [3] in using data augmentation techniques to improve the performance of object detection and classification tasks. Our experimental results show that GoogLeNet is the best network to use based on absolute accuracy and performance difference compared to when the model is trained on the unmodified dataset. The major contribution of our paper is to show that although data augmentation has in general a positive impact on the performance of deep learning models, the extent of the improvement varies from model to model and our study finds the best model to use in the context of traffic sign recognition problem. The organization of the paper is as follows: in Section 2 we describe the research problem that we are tackling and provide a review of existing solutions, in Section 3 we describe the dataset and methodology that we use before presenting the experimental results in Section 4 and we conclude the paper and give future work in Section 5.

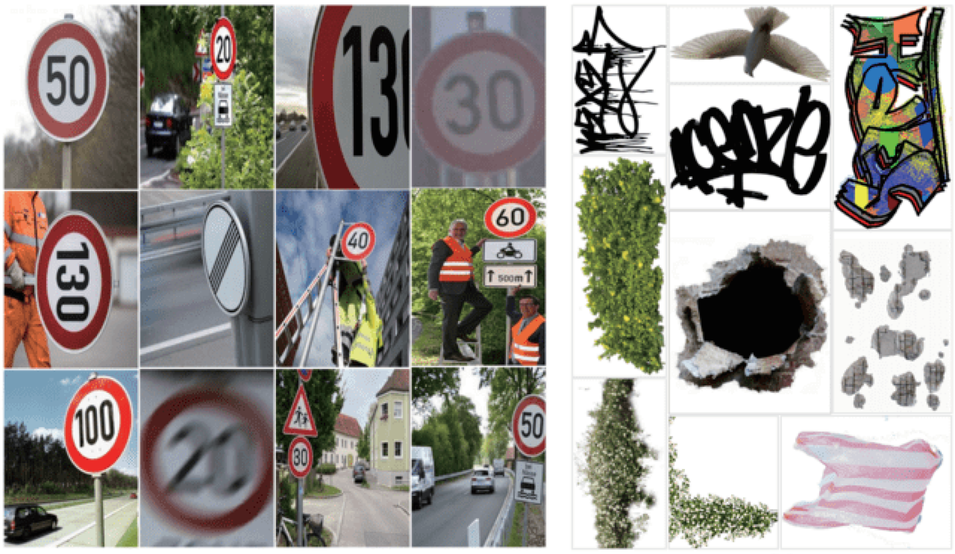
2. **Problem Statement and Literature Review.** The development of systems that are focused on imitating a standard human driver via vision systems is the primary area of research with self-driving vehicles. The advancement in artificial intelligence technology, and more specifically in deep learning, allows many objects in video frames captured using a dashboard-mounted camera, to be detected and recognized in real time. Deep learning has been proposed in the literature to solve a wide range of problems from human action recognition [4] to sign language recognition [5] and to pest monitoring in agriculture [6]. In [3] for example, the authors proposed using generative adversarial networks for data augmentation to reduce data imbalance with applications in car damage detection. Deep learning has also been used for semantic segmentation and depth estimation of urban road scenes in the context of self-driving vehicle technology [7]. Deep learning, however, is very computationally expensive and requires a high-specification computer system to run. As highlighted in [8], from the hardware side, this is made possible with the advancement in CPU and GPU technologies that resulted in faster and more powerful processing capabilities in computer devices, not to mention the introduction of AI-specific hardware, such as the Google's Tensor Processing Unit, that is designed to produce much higher throughput than off-the-shelf hardware.

From the software side, traffic sign recognition remains one of the most difficult tasks to be reliably carried out in practice because of the many external factors that can affect the quality of the image captured by the onboard camera. Some of the challenges include deteriorating visibility due to lighting inconsistency and weather conditions, deteriorating physical condition of the signs which could fade or rust over time, and random external factors such as strong wind and storms that make the signs tilt, rotate, flip, or just position wrongly, as well as occlusion. Occlusion is a common problem that can greatly impact the accuracy of traffic sign recognition systems. Occlusion can occur due to natural entities, such as leaves and trees, or man-made such as graffiti, and the resulting incomplete information makes sign recognition difficult, and finding a robust solution to this problem remains an open research question [9]. A field study on highway sign damage [10] found around 2.3% of all traffic signs surveyed were "damaged to the point of needing replacement", with even more signs having poor retro-reflectivity which decreases the visibility of the signs. Damage such as this to signs makes it challenging for computer vision models to accurately identify them. To address this issue, it is essential to train models with these non-ideal signs in mind. By doing so, the effectiveness of such models can be improved. Therefore, studying the effects of training computer vision models on non-ideal signs is crucial to enable the correct identification of signage in less-than-ideal conditions, which could have significant practical implications in the real world.

3. **Material and Method.** In this study, we use a subset of the Road Signs Dataset (version 5) [11] that contains primarily speed limit road signs. This dataset consists of images of German speed limit road signs which are made up of stock images as well as still images taken from a dashboard camera recording of a drive from one German city to another. The dataset has already been split into training and testing sets, which we used in this study instead of splitting the dataset randomly ourselves. Each image in the dataset is annotated with a label from 17 classes, each corresponding to the speed limit (e.g., 20, 130, or unlimited). However, due to significant population underrepresentation in some classes, only 11 are used in this study. These classes and the number of images in each class are shown in the two leftmost columns of Table 1. To provide some ideas of the type of images in the dataset, several of them are shown in Figure 1(a).

Table 1. Number of images per class in the training and testing sets

Classes	Training set (Unmodified)	Training set (Augmented)	Testing set (per occlusion band)	Testing set (Total)
20	23	560	70	490
30	64	326	140	980
40	15	910	90	630
50	69	359	170	1190
60	206	1085	590	4130
70	60	311	150	1050
80	186	991	520	3640
100	48	255	140	980
120	111	561	210	1470
130	30	100	100	700
No limit	33	328	80	560
Total	845	5786	2260	15820



(a)

(b)

Figure 1. A collage of (a) images in the Road Signs Dataset and (b) images used to occlude the traffic signs

Our method is based on the premise of using data augmentation to improve the accuracy of image classification in machine learning and deep learning that has been reported in [3,12]. The data augmentation process we propose is through artificially occluding the images in the dataset with natural images (shown in Figure 1(b)). We experimented using seven bands of occlusion levels ranging from minimal occlusion (1%-10%) to severe occlusion (61%-70%). The range of occlusion levels for each band is summarized in Table 2.

Table 2. The minimum and maximum occlusion in each occlusion band

Occlusion band	Min occlusion %	Max occlusion %
Band 1	1	10
Band 2	11	20
Band 3	21	30
Band 4	31	40
Band 5	41	50
Band 6	51	60
Band 7	61	70

The data augmentation process is performed for each occlusion band. It starts by retrieving the first image in the dataset and the metadata containing the location and size of a rectangle bounding the traffic sign object in the image. After the area of the bounding box is calculated, it then randomly selects an obstruction image. The width and height of the obstruction image are then scaled by a factor S, which is calculated as $S = \sqrt{R}$, where R is the occlusion level, defined as the ratio between the area of the obstruction image, A_o, and the area of the traffic sign bounding rectangle, A_S, i.e., $R = A_o/A_S$. In practice, the value of R is selected randomly between the minimum and maximum occlusion level of the occlusion band. After the scaling is applied, the obstruction image is then randomly rotated and placed within the traffic sign bounding box and the modified image is saved. The process repeats for every image in the dataset as depicted in Figure 2. The data augmentation process is performed on both the training and testing sets. A histogram is shown in Figure 3 to provide a picture of the population distribution of the classes in the training and testing sets, including that in the original, i.e., unmodified, training set. The augmented dataset is then used to train four deep-learning models of different architectures. They are AlexNet [13], VGG19 [14], ResNet50 [15] and GoogLeNet [16].

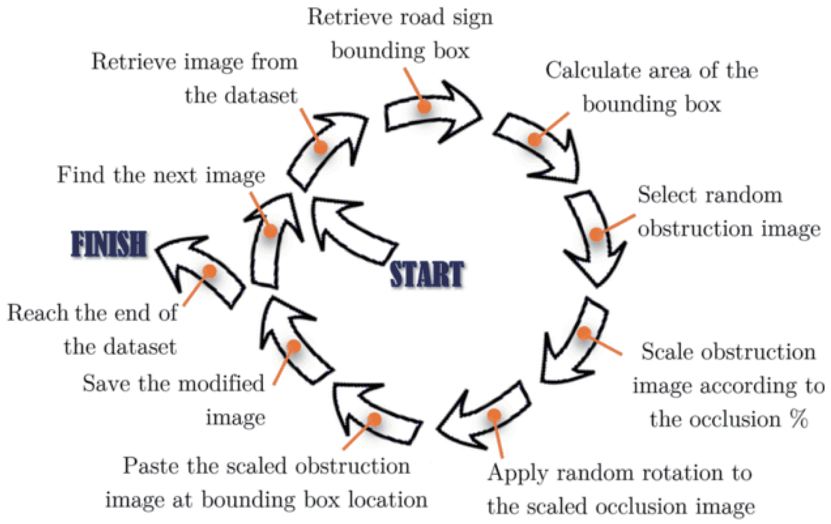

Retrieve road sign bounding box
Retrieve image from the dataset
Calculate area of the bounding box
Find the next image
Select random obstruction image
FINISH
START
Reach the end of the dataset
Scale obstruction image according to the occlusion %
Save the modified image
Paste the scaled obstruction image at bounding box location
Apply random rotation to the scaled occlusion image

Figure 2. A flowchart depicting the proposed data augmentation process

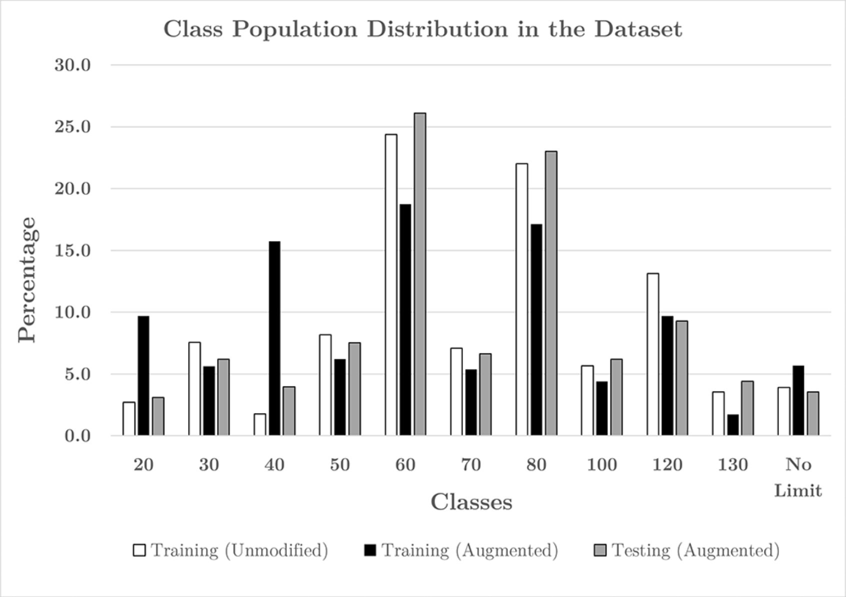


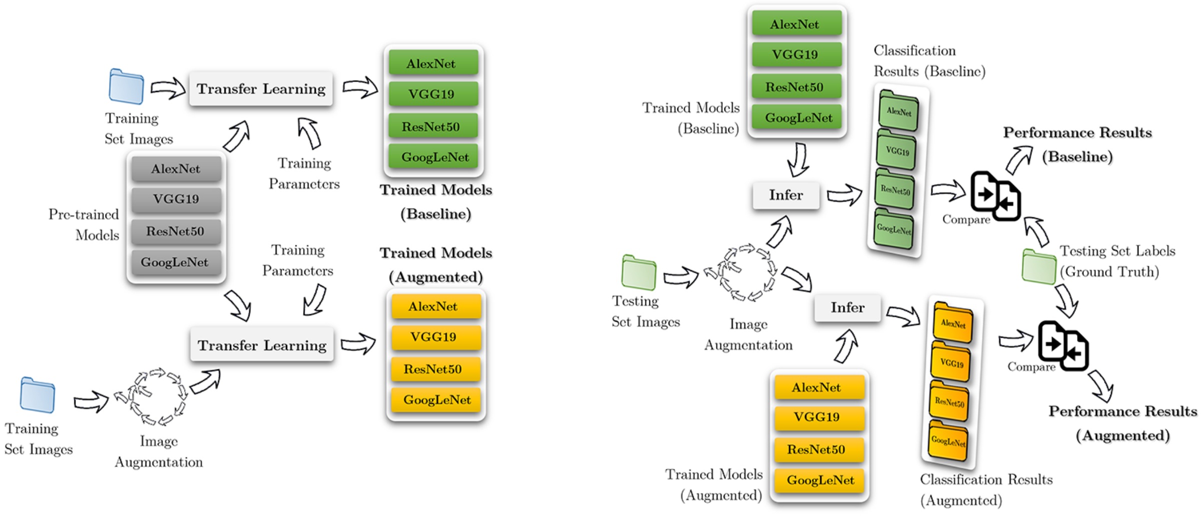
Figure 3. A histogram showing the class population distribution of the dataset

Each of these models has its unique strengths and weaknesses that make them suitable for inclusion in this study.

We employ the widely accepted best method to use deep learning models which is through transfer learning of pre-trained models [17]. The transfer learning method leverages the knowledge gained in a deep learning model after it has been pre-trained on a large image dataset. Training deep learning from scratch requires very powerful hardware and a very large dataset and can take a long time. By using the weights in a pre-trained model as initial weights and replacing just the classification layer before retraining it using our dataset, we can tune the model to learn features from the new dataset. To allow us to measure the relative performance of the approach, we also carry out transfer learning using images from the unmodified training dataset to produce their respective baseline models. A flowchart depicting the development of the deep learning models using both augmented and unmodified training sets is shown in Figure 4(a). The models are then used in the inferring process to classify images in the augmented testing set as illustrated in Figure 4(b).

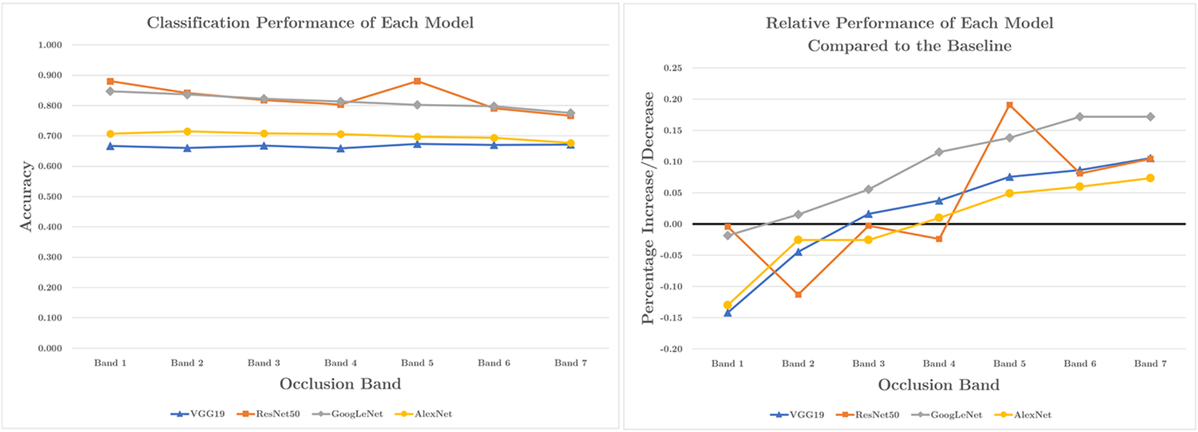
4. **Experimental Results, Analysis, and Discussion.** We use the accuracy metric [18] to measure the classification performance of the models. It is calculated as the ratio between the number of correctly classified images and the size of the testing set for all classes. The summary of the models' performance is shown in Figure 5. Figure 5(a) shows the absolute accuracy of the models using the proposed method whereas Figure 5(b) shows their performance relative to their baselines (black line). In that figure, any points below this line signify a decrease in performance and vice versa.

Figure 5(a) shows that when the proposed method is used, the ResNet50 and GoogLeNet models outperform AlexNet and VGG19 models by around 10%. The figure also shows that the performance of the ResNet50 and GoogLeNet models is relatively similar except for the abrupt jump in performance in Band 5 for the ResNet50 model. Currently, we cannot ascertain the underlying reason behind this but we speculate that it is due to the random element in the augmentation process. Figure 5(b) shows that all models perform worse than their baselines at lower occlusion level bands but progressively outperform the baseline at high occlusion level bands. This means that the proposed method results in a drop in accuracy when no occlusions are present in the dataset. The amount of drop is the worst for AlexNet and VGG19 (around 15% drop) but much better for ResNet50 and



(a) (b)

Figure 4. Flowcharts depicting (a) the development of the deep learning models and (b) the inferring process to classify the images in the augmented testing set



(a) (b)

Figure 5. Graphs showing the performance of each model when the images are occluded with varying levels of occlusion, (a) shows the absolute accuracy of the models, and (b) shows the accuracy compared to their baseline (black line). Points below this line signify a decrease in performance and vice versa.

GoogLeNet (less than 2%). However, as more and more severe occlusions are introduced to the dataset all models outperform their baselines. That means using the proposed method does improve the performance regardless of what model is used. The amount of improvement and the speed of improvement vary from model to model. Our experiment shows that the GoogLeNet model is the quickest in outperforming its baseline (already outperforms its baseline from Band 2), whereas the ResNet50 model is the slowest (only starts outperforming its baseline from Band 5). More importantly, our experimental results show that GoogLeNet has the best improvement among all four models considered in this study.

The experimental results show that, generally, training the models using augmented training data can produce better classification performance than without, at a high occlusion level. This is in line with the consensus of the research community as reported in

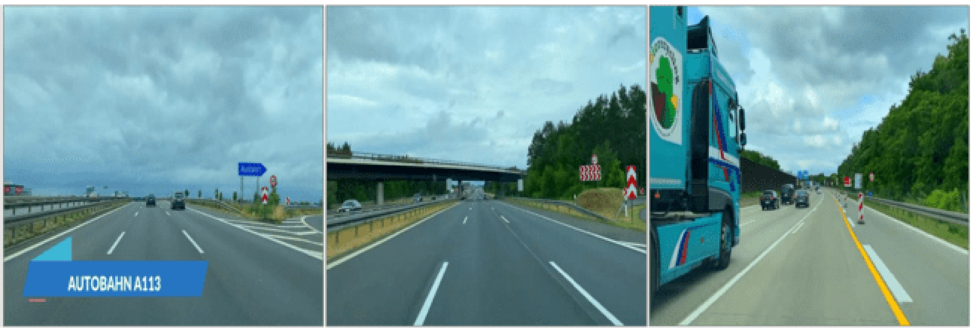


Figure 6. Examples in which traffic signs only occupy a tiny fraction of the entire image

the literature. The hit in performance at the lower occlusion level band is expected as the models are no longer tuned to only perfectly visible signs. However, it appears that some models (e.g., ResNet50) suffer from this more than others (e.g., GoogLeNet). In addition, we would also like to address the fact that none of the models produces accuracy above 90%. While we cannot for certain find the reason for this, our understanding is that the dataset contains images, which were mainly captured from dashboard cameras, where the traffic signs are exceedingly small and occupy only a few pixels, as shown in Figure 6.

We also would like to note a potential problem with our approach regarding the possibility of the models to start recognizing the type of obstruction images during training and inferring the input images using the type of obstruction image used instead of the traffic sign itself. In this study, we use ten different obstruction images and there are eleven different traffic signs to classify. If we were to use this approach to recognize significantly more traffic signs, it is therefore important to diversify and increase the number of obstruction images.

5. **Conclusion.** We have presented the result of our study into achieving an occlusion-robust traffic sign recognition by augmenting the data used to train deep learning models with images that have been artificially occluded using AlexNet, VGG19, ResNet50, and GoogLeNet models. The results of our experiment show that our data augmentation technique improves the recognition accuracy at the higher occlusion band with a slight hit in accuracy at the lower occlusion band. This shows that our method could significantly improve the recognition performance when the traffic sign images are severely occluded. We plan in the future, to expand the work to include more road signs and using videos to measure the methodology's suitability to be used in real time.

REFERENCES

[1] D. Lim and H. Hwangbo, UX design for holistic user journey of future robotaxi, in *Advances in Usability, User Experience, Wearable and Assistive Technology. AHFE 2021. Lecture Notes in Networks and Systems*, T. Z. Ahram and C. S. Falcão (eds.), Cham, Springer, 2021.

[2] S. Wei, P. E. Pfeffer and J. Edelmann, State of the art: Ongoing research in assessment methods for lane keeping assistance systems, *IEEE Transactions on Intelligent Vehicles*, DOI: 10.1109/TIV.2023.3269156, 2023.

[3] M. Mahyoub, F. Natalia, S. Sudirman, P. Liatsis and A. H. J. Al-Jumaily, Data augmentation using generative adversarial networks to reduce data imbalance with application in car damage detection, *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, pp.480-485, 2023.

[4] U. A. Usmani, J. Watada, J. Jaafar, I. A. Aziz and A. Roy, Particle swarm optimization with deep learning for human action recognition, *International Journal of Innovative Computing, Information and Control*, vol.17, no.6, pp.1843-1870, 2021.

[5] M. Mahyoub, F. Natalia, S. Sudirman and J. Mustafina, Sign language recognition using deep learning, *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, pp.184-189, 2023.

[6] L. Liu, R. Wang, C. Xie, P. Yang, S. Sudirman, F. Wang and R. Li, Deep learning based automatic approach using hybrid global and local activated features towards large-scale multi-class pest monitoring, *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol.1, pp.1507-1510, 2019.

[7] M. Mahyoub, F. Natalia, S. Sudirman, A. H. J. Al-Jumaily and P. Liatsis, Semantic segmentation and depth estimation of urban road scene images using multi-task networks, *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, pp.469-474, 2023.

[8] X. Feng, Y. Jiang, X. Yang, M. Du and X. Li, Computer vision algorithms and hardware implementations: A survey, *Integration*, vol.69, pp.309-320, 2019.

[9] S.-H. Yen, C.-Y. Shu and H.-H. Hsu, Occluded traffic signs recognition, *Advances in Information and Communication*, pp.794-804, 2020.

[10] V. P. K. Immaneni, W. J. Rasdorf, J. E. Hummer and C. Yeom, Field investigation of highway sign damage rates and inspector accuracy, *Public Works Management & Policy*, vol.11, no.4, pp.266-278, 2007.

[11] T. Sommer, Road Signs Dataset v5, *Roboflow Universe*, Roboflow, 2022.

[12] D. Kim, J. Joo and S. C. Kim, Fake data generation for medical image augmentation using GANs, *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp.197-199, 2022.

[13] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, pp.1097-1105, 2012.

[14] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *International Conference on Learning Representations*, 2015.

[15] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, 2016.

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich et al., Going deeper with convolutions, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[17] F. Natalia, J. C. Young, N. Afriliana, H. Meidia, R. E. Yunus and S. Sudirman, Automated selection of mid-height intervertebral disc slice in traverse lumbar spine MRI using a combination of deep learning feature and machine learning classifier, *PLoS One*, vol.17, no.1, e0261659, 2022.

[18] B. J. Erickson and F. Kitamura, Magician's Corner: 9. Performance metrics for machine learning models, *Radiology: Artificial Intelligence*, vol.3, no.3, e200126, 2021.