Autonomous Traffic Sign (ATSR) Detection and Recognition using Deep CNN

Article *in* Procedia Computer Science · January 2019

DOI: 10.1016/j.procs.2019.12.108

4 authors, including:

Ghazanfar Latif
University of Québec in Chicoutimi

Jaafar Alghazo
Virginia Military Institute

Loay Alzubaidi
Prince Mohammad University

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 163 (2019) 266–274

Procedia Computer Science

www.elsevier.com/locate/procedia

16th International Learning & Technology Conference 2019

Autonomous Traffic Sign (ATSR) Detection and Recognition using Deep CNN

Danyah A. Alghmgham[a], Ghazanfar Latif [a, b*], Jaafar Alghazo [a], Loay Alzubaidi [a]

[a]*College of Computer Engineering and Sciences, Prince Mohammad bin Fahd University, Khobar, Saudi Arabia.*

[b]*Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Malaysia*

Abstract

Automatic detection and recognition of traffic signs is very important and could potentially be used for driver assistance to reduce accidents and eventually in driverless automobiles. In this paper, Deep Convolutional Neural Network (CNN) is used to develop an Autonomous Traffic and Road Sign (ATRS) detection and recognition system. The proposed system works in real time detecting and recognizing traffic sign images. The contribution of this paper is also a newly developed database of 24 different traffic signs collected from random road sides in Saudi Arabia. The images were taken from different angles and including other parameters and conditions. A total of 2718 images were collected to form the database which we named Saudi Arabian Traffic and Road Signs (SA-TRS-2018). The CNN architecture was used with varying parameters in order to achieve the best recognition rates. Experimental results show that the proposed CNN architecture achieved an accuracy of 100%, thus higher than those achieved in similar previous studies.

Keywords: Arabic Traffic Signs; Convolutional Neural Networks (CNN); Traffic Sign Detection, Road Sign; Traffic Signs Recognition

1. Introduction

Traffic sign detection and recognition has gained importance with advances in image processing due to the benefits that such a system may provide. The recent developments and interest in self-driving cars has also increased the interest in this field. An automated traffic sign detection and recognition system will provide the ability for smart cars and smart driving. Even with a driver behind the wheel, the system may provide vital information to the driver reducing human errors that cause accidents. Certainly with such a system integrated into vehicles, it is expected that the number of car accidents will be reduced greatly saving human lives and the monetary value associated with car accidents. Automated systems will be able to control traffic on both open roads and intersections as well.

1877-0509
10.1016/j.procs.2019.12.108

The motivation behind developing such a system is clear due to the benefits of such a system in saving lives and saving cost. Therefore, the objective of this work is to develop an automatic Arabic traffic sign detection and recognition system based on deep learning algorithm. The proposed system has the ability to recognize the signs within images captured by cameras and processed by a Deep CNN network. Most car accidents are caused by human error either by drivers not noticing a certain sign or with drivers driving against the direction set by a certain traffic sign (i.e. traffic sign setting speed at 100 KM and driver driving at a greater speed).. For this reason, this paper carries the major objective of developing and improving the efficiency and robustness of the traffic sign detection system for Arabic Traffic Signs as well as handling associated issues, a recognition system should also classify traffic signs into different classes in real-time environment and avoid recognition errors.

Machine learning is divided into supervised learning, unsupervised learning, semi-supervised learning, and reinforced learning. In this paper, the choice of deep learning for an unsupervised learning approach is done by design because even though basic traffic signs are limited yet combined with road signs, street name signs, etc. the dataset becomes larger with endless possibilities. The ultimate goal is to have a system fitted into cars and that can detect and recognize any traffic sign to assist the driver or assist in the self-driving process. With deep learning algorithms, unlabeled data can be used and the system can extract features automatically without human intervention.

The rest of the paper is organized as follows: section 2 covers the related literature review. Section 3, details the proposed system. Section 4, explains the experimental data used in this paper. Section 5, shows the results along with discussion. Section 6, concludes the work and section 7 lists all the references used in this work.

2. Literature Review

Nowadays, recognition and classification of traffic signs are very important, especially for unmanned automatic driving. Extensive research has been done in the area of recognition and classification of traffic and road signs. In [1], the authors proposed a Convolutional Neural Network and Support Vector Machines (CNN-SVM) method for traffic signs recognition and classification. The coloring used in this method is YCbCr color space which is input to the convolutional neural network to divide the color channels and extracting some special characteristics. SVM is then used for classification. Their proposed method achieved a 98.6% accuracy for traffic signs recognition and classification. In [2], the authors proposed a color based segmentation method with Histogram Oriented Gradients (HOG) for feature extraction followed by SVM for classification. The model used CIECAM97 for color appearance, this model was applied to a segment to extract color information. Another model used for shape features is FOSTS [3] which achieved a 95% accuracy. In [4], the authors proposed feature extraction through HOG and local binary pattern (LBP) which are then input into an Extreme Learning Machine Network for classification and recognition. In [5], the authors propose a traffic sign recognition system based on extreme learning machine (ELM). Their method consists of feature extraction through extraction of histogram of the oriented gradient variant (HOGv) features followed by a single classifier trained by ELM.

In [6], the authors developed a new dataset consisting of 100,000 images and also proposed a traffic sign detection and classification method based on a robust end-to-end CNN. The method achieved 84% accuracy. In [7], the authors proposed a multi-scale deconvolution network (MDN) for localized traffic sign detection. The method achieved 99.1% accuracy. In [8], the authors presented a survey of available techniques for road traffic sign recognition systems using CNN. Their work presented the available proposed techniques in addition to the challenges faced by CNN methods in terms of time complexity and accuracy. They further proposed a method to overcome the challenges which utilized canny edge detection to highlight the edges of the traffic symbols which is then input into a CNN for classification. To enhance classification and recognition, the introduced fuzzy classification technique used in [9].

In [10], the authors proposed a traffic sign recognition and classification system based on scale-aware CNN. Their system consists of two CNN's; one for region proposals of traffic signs and the other for classification of each region. In addition, fully convolutional network (FCM) is utilized to achieve scale invariant detection. The system achieved 99.88% precision accuracy. In [11], the authors proposed a knowledge-based recurrent attentive neural network (KB-RANN) for small object detection. Their method achieved 81% accuracy. In [12], the authors proposed an efficient algorithm for traffic sign detection on low cost embedded systems. Their method consists of color thresholding, shape detection and sign validation. They utilized an efficient color thresholding technique based on the red-blue angle color transformation (RBAT) and the red color normalized. Ellipse fitting technique is also employed for detecting the circular signs. HOG is employed for validation. Their method achieved 97% precession accuracy. In [13], the authors analyzed the spatial transformers and stochastic optimization methods for deep neural network for traffic sign recognition. They finalized this with a proposed system that achieved 99.71% accuracy.

3. Proposed System

The contribution of this paper will be two folds; one is to develop a new database for Arabic Traffic and Road Signs and the other is develop and design a deep CNN architecture for Arabic Traffic sign recognition. Fig. 1 shows the high level view of the system. The collected data set is given as an input to the proposed CNN architecture for training, validation and testing. The detailed explanation of the CNN architecture is provided in the next section. Once the CNN is trained, it is ready to be used for classifying new images which were not part of the collected dataset.

The AATS system depends on a group of standard Arabic Traffic signs. In recent years, number of authors done work in this field but to the author's knowledge, this is the first time a complete database is developed for Arabic Traffic. A Deep CNN architecture is also proposed for Arabic traffic sign recognition Generally, CNNs consist of multiple hidden layers between the input and output layers [13]. The design of the proposed CNN is implemented using Python.

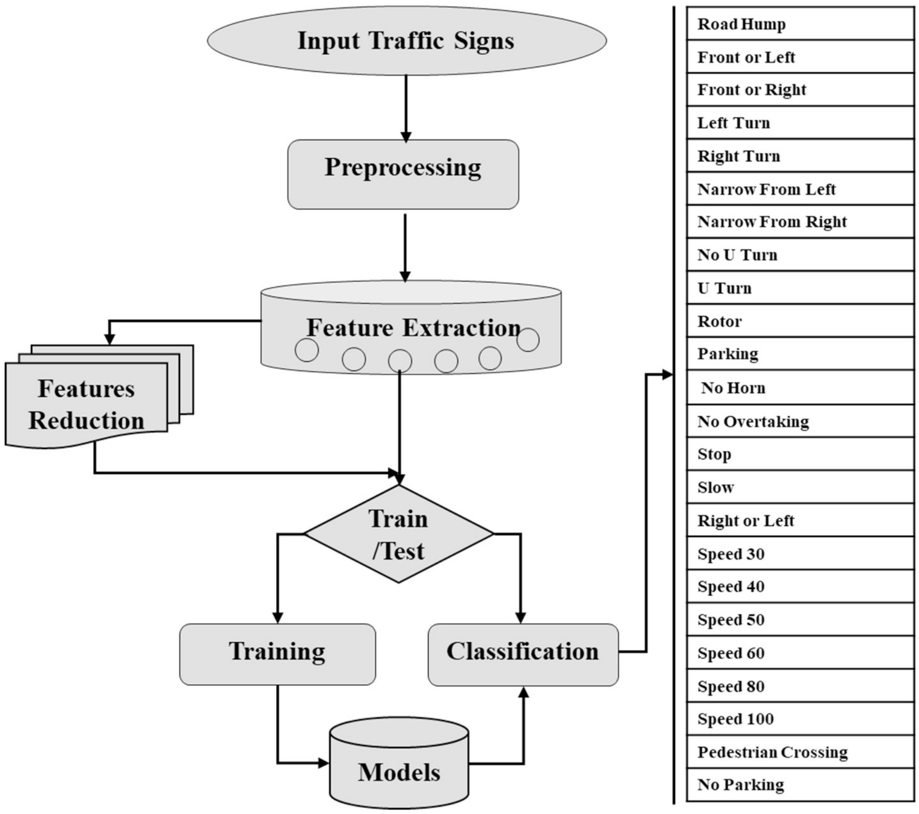

Fig. 1 . Block diagram of supervised learning

3.1. Experimental Data

In this research, a new dataset for Arabic Traffic Signs is developed. As part of this research, 24 most common Arabic traffic signs are selected. The dataset consists of 2,728 images captured for 24 traffic signs. The images are captured from three connected cities (Khobar, Dammam and Dhahran) in the Eastern Province of Saudi Arabia. For each traffic sign, sign boards ranging 20 to 30 location were found and for each location, 5 photos were captured with different angles and distances. Table 1 describes the list all the selected traffic signs visual shape, their names in English and Arabic and the total number of images captured for each sign as part of the new database.

Table 1. Description of 24 Traffic Signs with total 2728 images

#	Sign Name in English	Sign Name in Arabic	Sign Image	No. of Images	#	Sign Name in English	Sign Name in Arabic	Sign Image	No. of Images
1	Road Hump	مطب صناعي		124	13	No Overtaking	التجاوز محظور		112
2	Front or Left	الإتجاه الى الأمام او اليسار		144	14	Stop	قف		117
3	Front or Right	الإتجاه الى الأمام او اليمين		144	15	Slow	تمهل		107
4	Left Turn	منحنى الى اليسار		118	16	Right or Left	المرور على احد جانبي الطريق		104
5	Right Turn	منحنى الى اليمين		117	17	Speed 30	ممنوع تجاوز السرعة 30 في الساعة		132
6	Narrow From Left	الطريق يضيق من اليسار		105	18	Speed 40	ممنوع تجاوز السرعة 40 في الساعة		114
7	Narrow From Right	الطريق يضيق من اليمين		103	19	Speed 50	ممنوع تجاوز السرعة 50 في الساعة		110
8	No U Turn	ممنوع الدوران للخلف		104	20	Speed 60	ممنوع تجاوز السرعة 60 في الساعة		109
9	U Turn	الإتجاه الى الأمام او الخلف		102	21	Speed 80	ممنوع تجاوز السرعة 80 في الساعة		101
10	Rotor	الإتجاه مستدير		102	22	Speed 100	ممنوع تجاوز السرعة 100 في الساعة		80
11	Parking	مواقف		126	23	Pedestrian crossing	عبور المشاة		111
12	No Horn	ممنوع التزمير		132	24	No Parking	ممنوع الوقوف		110

3.2. Preprocessing

The images were RGB images with different dimensions which led to the need of pre-processing the images before inputting them to the CNN network. The images are transformed into Greyscale images and the dimension is also reduced to 30x30 pixels. Moreover, the total number of output classes is 24 classes each represents an Arabic traffic sign so all the images are carefully labeled and placed in their corresponding folders. The number of images per class differs from one class to another as shown in Table 1. Fig. 2 shows a sample of each Traffic Sign after the pre-processing.

The newly developed dataset consisting of 2728 images was randomly partitioned into 80% percent training set (2183 images) and 20% percent testing set (545) images. The training set was further partitioned; so 20% percent of the images (436 images) were used for validation.

Fig. 2. Sample of the Traffic Signs after the pre-processing

As mentioned above, the data set was partitioned into training, validation and testing sets. This partitioning was done for cross validation which is a statistical approach to evaluate the performance of Machine Learning (ML) algorithms. We used the Hold-out cross-validation technique which basically splits the data set into three mutually independent sets. This technique is commonly used due to its efficiency and ease of implementation [14]. Since there is no fixed rule on the percentage in which the data set is partitioned by, we used one of the commonly used partitioning percentages.

3.3. Deep CNN Architecture

The proper choice of different design hyper-parameters, such as non-linearity and pooling variants, directly affects the performance of the network. Since there is no clear guidance on how to choose the CNN hyper-parameters, many researchers tend to use trial and error experimentation in order to discover good settings [15]. We analyzed previous research studies that have been carried in the area of deep learning generally in order to choose the hyper-parameters setting that is more likely suitable for this research. So by analyzing the work done in [16-18], the hyper-parameters setting in this work is as follows.

Pooling Layer: It helps in reducing the amount of computations needed in the training process by reducing the dimension of the image, and therefore the overfitting problem is reduced. The max-pooling technique was used since the convergence rate is faster as compared to other subsampling techniques, and thus has a better generalization performance. The maximum pixel value in a non-overlapping region, equals to the window of the pooling, is the output of this layer; this is beneficial in creating position invariance.

Non-Linearity: Since the relation between the images and their classes is not linear, introducing non-linearity in the CNN is needed. This is achieved by using non-linear activation so that the construction of the non-linear relation between the images and their classes is possible by the CNN. The Rectified Linear Unit (ReLU) is a widely used activation function in CNN. ReLU has the advantage that the CNN trains faster as compared to other functions. One of the ReLU variants was used which is the leaky ReLU since it overcomes the problem of dead neurons that is faced if the original ReLU is used. Basically, leaky ReLU does not output zero when the input values are less than zero, instead it outputs negative value. So after each convolution layer, and before the pooling layer, we added a leaky ReLU activation function.

Dense Layer Activation Function and Loss Function: As mentioned above, we used the leaky ReLU activation function multiple times in the proposed network. However, the output of the fully connected layer, which is one dimensional vector, is passed to a softmax activation function to predict the label of that particular input. The input vector is transformed to a vector of the same size but the values range from zero to one only and the summation is always equals to 1. Softmax function outputs a probability distribution that is then converted to one-hot encoding vector. So the element that has the largest portion of probability distribution will have the value one in the one-hot vector and all other elements will have the value zero. To estimate the loss of softmax, the categorical cross-entropy function is used. Basically, categorical cross-entropy is used to measure the difference between the output of the softmax function and the one-hot encoding of the actual class. It is used as a part of gradient descent to evaluate CNN performance as an error measure between the true distribution and the predicted one. Softmax function and categorical cross-entropy are widely used in computer vision tasks when multiple classes are involved.

After setting the needed hyper-parameters, the network was compiled using the Adam optimizer. The need of using optimization algorithms is the nature that CNN weights and biases are set automatically and they are of great importance in the field of machine learning. They work on optimizing a given function; whether maximizing or minimizing it with respect to its parameters. Since the loss function in CNN is differentiable with respect to its parameters, gradient descent based algorithms are often used; first order partial derivatives are also fast to compute. Adam is an optimization algorithm that is used to update network weights. It combines the advantages of other classical stochastic gradient descent which are Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). It is computationally efficient, requires less memory and has straightforward implementations. Given different parameters, this algorithm computes each parameter's adaptive learning rate by estimating the gradient's first and second moments [19]. Table 1 shows the summary of the parameters in each layer as well as the total parameters in the proposed network. This CNN network will be referred to as Model 1 in the results section.

Table 2. Summary of the Initial Proposed CNN Network (Model 1)

Layer (Type)	Output Shape
conv2d_1 (Conv2D) input (30x30 image)	(None, 28, 28, 32)
max_pooling2d_1 (MaxPooling2)	(None, 14, 14, 32)
conv2d_2 (Conv2D)	(None, 28, 28, 32)
dropout_1 (Dropout)	20%
conv2d_3 (Conv2D)	(None, 28, 28, 32)
max_pooling2d_2 (MaxPooling2)	(None, 7, 7, 32)
flatten_1 (Flatten)	(None, 1568)
dense_1 (Dense)	(None, 128)
dropout_2 (Dropout)	(None, 128)
dense_2 (Dense)	(None, 64)
dense_3 (Dense)	(None, 32)

4. Results and Discussions

The improved optimized CNN network layers are shown in Table 3. As observed in the optimized CNN architecture, the number of layers have been reduced without compromising the accuracy of the network. This CNN will be referred to as Model 2. Table 3 shows the experimental results performed on the proposed CNN architecture. The experiments were performed on epochs size of 10, 50, 100, and 150. For each epoch, the batch sizes of 50,100, 200 and 400 are used. The table for the experiment shows four columns with Validation Accuracy (Val. Acc), Validation Loss (Val. Loss), Test Accuracy (Test Acc) and finally Test Loss (Test Loss). Different design parameters were changed and the effect of the changes was tested by evaluating accuracy. A total of 16 different experiments were performed on the CNN architecture in order to identify an optimized design. Note that all the preprocessing steps remain the same as described above. The proposed CNN models are shown in the Table 1 and Table 3.

Experimental results show that the number of epochs and number of batch size directly affects the test accuracies achieved for the models. The number of epochs is directly proportional with test accuracy; increasing epoch size increases the test accuracy. While the opposite is true for batch size as it is observed that increasing batch size decreases test accuracy. Overall, the test accuracy obtained for the improved CNN network-model 2 is better than those achieved for the initial CNN-model 1. In fact, model 2 results for epoch 150 achieved 100% accuracy for all batch sizes. The result is better than those reported in recent literature, however, it is pointed out here that the CNN is applied for Arabic Traffic Sign dataset and no recent literature was found targeting Arabic Traffic sign detection and recognition.

No relationship was found for validation accuracy for both models in relation to epoch size and batch size.

Table 3. Summary of the Improved CNN Network with less number of layers (Model 2)

Layer (type)	Output Shape
conv2d_1 (Conv2D) input (30x30 image)	(None, 28, 28, 64)
max_pooling2d_1 (MaxPooling2)	(None, 14, 14, 64)
dropout_1 (Dropout)	20%
max_pooling2d_1 (MaxPooling2)	(None, 12, 12, 32)
max_pooling2d_2 (MaxPooling2)	(None, 6, 6, 32)
flatten_1 (Flatten)	(None, 1152)
dense_1 (Dense)	(None, 128)
dropout_2 (Dropout)	(None, 128)
dense_2 (Dense)	(None, 32)

Table 4. Comparison of Accuracies for different CNN Parameters

CNN Model	Epochs	Batch Size	Val. Acc.	Val. Loss	Test Acc.	Test Loss
1	10	50	100.00	13.50	91.61	36.11
		100	100.00	12.82	77.53	24.17
		200	100.00	11.89	72.07	30.54
		400	100.00	10.80	64.02	41.84
	50	50	99.98	14.76	99.59	0.62
		100	99.98	14.64	99.56	0.69
		200	99.99	13.83	98.56	8.27
		400	100.00	13.61	98.44	9.31
	100	50	99.97	14.88	99.62	0.04
		100	99.98	14.62	99.59	0.24
		200	99.99	14.96	99.55	0.53
		400	99.98	14.60	99.4	1.63
	150	50	99.97	14.93	99.7	0.01
		100	99.97	15.01	99.65	0.05
		200	99.98	14.97	99.54	0.17
		400	99.99	14.85	99.49	1.28
2	10	50	100.00	14.09	93.26	11.98
		100	100.00	13.61	81.23	22.53
		200	100.00	12.41	78.23	29.38
		400	100.00	11.99	77.03	39.68
	50	50	99.98	15.14	99.8	0.04
		100	99.97	15.17	99.79	0.16
		200	99.99	15.07	99.71	1.29
		400	99.98	14.88	99.69	9.57
	100	50	99.98	15.28	99.9	1.22
		100	99.98	15.32	99.89	0.02
		200	99.98	15.36	99.87	0.08
		400	99.99	15.29	99.84	0.66
	150	50	99.97	15.20	100.00	0.00
		100	99.98	14.72	100.00	0.70
		200	99.98	15.34	100.00	0.03
		400	99.98	15.22	100.00	0.15

5. Conclusion

Automatic Arabic Traffic Sign (AATS) recognition system was designed using Convolutional Neural Networks (CNN). The training and testing dataset consists of more than 2,728 sign samples collected for 24 different traffic signs. The dataset went through a preprocessing stage before inputting it to the network. It got partitioned into training, testing and validating datasets. The initial design was based on similar previous work which was used as a base for the subsequent improved design. The final Deep CNN architecture proposed in this work consists of two convolutional layers, two maxpooling layers one dropout layer and 3 dense layers. 100% accuracy was obtained for epoch 150 for all batch sizes. We demonstrated the usefulness of the designed CNN architecture by implementing a practical system that can take real time Traffic Sign via camera attached to vehicle, classifies them to the corresponding sign, and then give voice notification to driver or take automatic decision for autonomous cars. Future work will include increasing the size of the dataset and publishing it so that it can be used by other researchers for benchmarking purposed. Work will also continue on developing more robust and computationally low cost recognition systems.

References

[1] Lai, Y., Wang, N., Yang, Y., & Lin, L. (2018). Traffic signs recognition and classification based on deep feature learning. In *7th International Conference on Pattern Recognition Applications and Methods (ICPRAM), Madeira, Portugal* (pp. 622-629).

[2] M. Teague, "Image analysis via the general theory of moments," *J. Opt Soc. Am.*, vol. 70 (8), pp. 920–930, 1980.

[3] Lee, A. (2015). Comparing Deep Neural Networks and Traditional Vision Algorithms in Mobile Robotics. Swarthmore College.

[4] Srinivas, S., Sarvadevabhatla, R. K., Mopuri, K. R., Prabhu, N., Kruthiventi, S. S., & Babu, R. V. (2016). A taxonomy of deep convolutional neural nets for computer vision. Frontiers in Robotics and AI, 2, 36.

[5] Huang, Z., Yu, Y., Gu, J., & Liu, H. (2017). An efficient method for traffic sign recognition based on extreme learning machine. *IEEE transactions on cybernetics*, *47*(4), 920-933.

[6] Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S. (2016). Traffic-sign detection and classification in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2110-2118).

[7] Pei, S., Tang, F., Ji, Y., Fan, J., & Ning, Z. (2018). Localized Traffic Sign Detection with Multi-scale Deconvolution Networks. *arXiv preprint arXiv:1804.10428*.

[8] Hatolkar, Y., Agarwal, P., & Patil, S. (2018). A Survey on Road Traffic Sign Recognition System using Convolution Neural Network.

[9] Pei, S., Tang, F., Ji, Y., Fan, J., & Ning, Z. (2018). Localized Traffic Sign Detection with Multi-scale Deconvolution Networks. *arXiv preprint arXiv:1804.10428*.

[10] Yang, Y., Liu, S., Ma, W., Wang, Q., & Liu, Z. (2018). Efficient Traffic-Sign Recognition with Scale-aware CNN. *arXiv preprint arXiv:1805.12289*.

[11] Yi, K., Jian, Z., Chen, S., Chen, Y., & Zheng, N. (2018). Knowledge-based Recurrent Attentive Neural Network for Traffic Sign Detection. *arXiv preprint arXiv:1803.05263*.

[12] Soetedjo, A., & Somawirata, I. K. (2018). An Efficient Algorithm for Implementing Traffic Sign Detection on Low Cost Embedded System. *International Journal of Innovative Computing Information and Control*, *14*(1), 1-14..

[13] Hoo-Chang, S., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, *35*(5), 1285.

[14] Kim, B., Park, S., Kim, K., Lim, J., & Nahm, K. (2018). Systematic process to determine DNBR limit of CHF correlation with repetitive cross-validation technique. *Journal of Nuclear Science and Technology*, 1-9.

[15] Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*.

[16] Latif, G., Alghazo, J., Alzubaidi, L., Naseer, M. M., & Alghazo, Y. (2018, March). Deep Convolutional Neural Network for Recognition of Unified Multi-Language Handwritten Numerals. In *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)* (pp. 90-95). IEEE.

[17] Latif, G., Iskandar, D. A., Alghazo, J., Butt, M., & Khan, A. H. (2018). Deep CNN based MR Image Denoising for Tumor Segmentation using Watershed Transform. *International Journal of Engineering & Technology*, *7*(2.3), 37-42.

[18] Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., ... & Hodjat, B. (2019). Evolving deep neural networks. In *Artificial Intelligence in the Age of Neural Networks and Brain Computing* (pp. 293-312). Academic Press.

[19] Xiong, F., Xiao, Y., Cao, Z., Gong, K., Fang, Z., & Zhou, J. T. (2018). Towards good practices on building effective cnn baseline model for person re-identification. *arXiv preprint arXiv:1807.11042*.