Road State Classification of Bangladesh with Convolutional Neural Network Approach

Conference Paper · October 2019

7 authors, including:

Taoseef Ishtiak
North South University

Sajid Ahmed
North South University

Hossain Mehreen
North South University

Rashed Shelim
North South University

Road State Classification of Bangladesh with Convolutional Neural Network Approach

Taoseef Ishtiak
Department of Electrical and Computer Engineering, North South University, Bangladesh
Dhaka, Bangladesh

Sajid Ahmed
Department of Electrical and Computer Engineering, North South University, Bangladesh
Dhaka, Bangladesh

Mehreen Hossain Anila
Department of Electrical and Computer Engineering, North South University, Bangladesh
Dhaka, Bangladesh

Saima Islam
Department of Electrical and Computer Engineering, North South University, Bangladesh
Dhaka, Bangladesh

Rashed Shelim
Department of Electrical and Computer Engineering, North South University, Bangladesh
Dhaka, Bangladesh

Tanjila Farah
Department of Electrical and Computer Engineering, North South University, Bangladesh
Dhaka, Bangladesh

ABSTRACT

Machine Learning, an important branch of Artificial Intelligence is now a widely used to detect features, predict future value, recognize patterns and detect anomalies from given data. Image classification, object detection is such massive fields where machine learning is being applied widely, especially the Convolutional Neural Network (CNN). Computer vision-based tasks have been in research interest for many years. But after Google introduced their own residual deep neural network architecture in Inception V-1, they have optimized the network further more through Inception V-4 to reduce computational cost. In this paper, we have classified roads of Bangladesh based on their surface condition from 2D images, inspired from residual network introduced in Inception V-4 architecture. Our research team modified the architecture to adjust for our own dataset of images and we have seen higher accuracy in result comparing to classical CNN based architecture with less computational cost and training time. The average condition of roads vary country wise. In developing and under developed countries; like Bangladesh, different roads are formed with different building materials. For the building material of roads, the surface condition and change in structural health of roads vary. CNN is very powerful for analyzing visual imagery. Based on our collected dataset of more than 5000 images, we applied CNN (Convolutional Neural Networks) based architecture and Inception-V4 model's derivation residual architecture to observe the differences in each performance to classify the roads based on their surface state condition. On the basis of the acquired dataset of the roads inside Bangladesh, our model has been trained to detect five classes of roads of Bangladesh based on the surface condition. The trained model has been validated and tested by roads of other countries and showed an average accuracy of 92% using our derivation of Inception V-4 model while classical CNN based architecture has given an average of 87% accuracy over all the five classes. The goal of the proposed model is to analyze the surface condition of streets and classify the image with residual deep convolutional networks that has been proposed in the Inception V-4 model [10].

Keywords*:* Machine learning, Convolutional Neural Network (CNN), Residual Deep Neural Network, Fully Connected Network (FCN), Inception V-1, V-2, V-3, V-4, Road Surface State Classification, Road Information System (RiS), Road Weather Information System (RWiS), Intelligent Transportation System (ITS).

1. INTRODUCTION

CNN based machine learning models are being very effectively used for pattern recognition. So far, the "AlexNet" architecture introduced by Krizhevsky et al [1] has been successfully applied in a number of computer vision tasks, for example in object detection [2], object tracking [3], segmentation, [4] video classification, [5] human pose estimation [6] and superresolution [7]. Classical deep Neural Networks like CNN and FCN extract image features finding key points and thus used for 2D image classification. But Google's Inception V-4 architecture has argued to reduce the computational cost using residual networks. Our motivation of this work is to apply such deep residual network for road image recognition and classification and prove its efficiency comparing to classical CNN based architecture. We have implemented the training methodology obtained by He at al. in [8] which is a residual connections introduction in computer vision field and the latest version of Google's Inception architecture [9]. The basic difference between Inception V-3 and Inception V-4 is in Inception V-4, they have designed a simpler architecture and more inception modules than their version 3 architecture [10]. The constraints contained in Inception V-3 had come from the need for partitioning model architecture with distributed training using DistBelief [11]. With migration to Tensorflow [12] those constraints have been lifted which allowed the reintroduced architecture to get simplified [10]. The goal of our research was to analyze the road surface condition of Bangladesh from 2D images and train a model to find the pattern to classify the roads based on surface. Based on our data preprocessing with canny edge detection, image thresholding, Fourier Transformation, we classified 5 classes of roads for our training namely: 1. Perfect Roads, 2. Good Roads, 3. Mild Bad Roads, 4. Severely Bad Roads 5. Water on Surface Roads. To achieve this, we have applied the part of convolutional neural network approach and a modified version of the architecture that has been proposed in Google's Inception V-4 model, to classify 2D images of roads. 2D image classification and image-processing based research such as: Food Classification [13], Gaussian noise detection [14] and in many other machine learning based projects as well CNN has been used successfully. We have taken motivation from such research project to apply CNN based architectures for a new classification dimension, which is the streets of a nation and make a comparison with deep residual network architecture. Our trained model analyzed the change of pixel intensity for the existence of cracks, garbage on road sides, pits, surface water, road-color change for hilly tracts and soil change. Our acquired dataset images were taken on almost keeping the same frame ration of images. This approach helped us let our machine learn about the width of the road as well and we labeled our dataset accordingly. Our project has been done based on a humanitarian ground where our trained model can be used for any government project where the road condition state can be automatically detected and the authority can be informed at once about the surface condition change of any locality. Smart RiS, RWiS, ITS systems are very crucially needed for countries like Bangladesh. We believe our model can be implemented for autonomous car technologies for countries like Bangladesh where, the vehicle can get idea to monitor its speed and direction analyzing the surface condition; and also, in advanced robotics industries. A portion of our dataset was provided by the Local Government Engineering Department of Government of People's Republic of Bangladesh and the rest were collected by our research team being physically present at the station over a time period of four months. The whole dataset thus contained 5,370 road images took on bright daylight of 53 stations of Bangladesh, including: local roads, highways, hilly tracts and of other criteria as well.

2. RELATED WORK

The classical CNN architecture cannot take inputs of arbitrary sizes where fully connected networks architecture like FCN, can take inputs of arbitrary sizes after the convolutional layer [15][16][17]. CNN based researches have been performed to detect road, semantic segmentation of road scenes [18][19][20][21], road-lane, road area extraction [22], rural roads detection [23], street signs detection and so on. The works are most commonly based on classification, object detection or semantic segmentation. For image recognition and object-detection based works some of the novel approaches have been introduced by Lin et al. [15], Simonyan et al. in VGGNet [16] and GoogleNet (Inception V-1) [17] by Szededy et al. But residual connections were introduced by He et al. in [1] where the authors emphasized for the practical advantages of using additive merging of signals in image processing and object detection. The use of residual connections boosts up the training speed. To build our own methodology, we have studied similar works on image processing, CNN based classification works and other object detection models. For computational power efficiency [23] is a very good approach for real-time joint denoising of images, vehicle detection and road segmentation. For surface health monitoring, in civil engineering projects, deep convolutional neural network approaches like; bounding box approaches is proposed in computer vision-based models. A crack detection model likewise; has been proposed in [24] based on semantic segmentation on concrete crack images which is also a fully connected network (FCN) approach. For detecting road surface condition [25][26][27] has used vibration and GPS sensor data [25] to achieve signals for road surface change and then trained the data with a simple artificial neural network-based model. For road surface condition change on environmental metrics like: sun heat, rainy days, snowfall, ice clogged condition [28] has proposed a combined approach with SVM (Support vector machine) algorithm connected with a grid search algorithm PSO (Particle Swarm Optimization) to optimize the kernel function factor and penalty factor of SVM. There GRU (Grated Recurrent Units) has been implemented with CNN for road segmentation [21]. Also, principal components analysis with PCA-net has been found very much equivalent to CNN based feature extraction and block matching and 3-D filtering method. Sparse denoising model was proposed in [23] meanwhile applied for adaptive learning for image denoising.

3. METHEDOLOGY

For data preprocessing, we reshaped our acquired images into 128 by 128 size and performed several image analysis methods like: canny edge implementation, black and white thresholding and Fourier Transformation of image signals. We only trained the images that were taken of similar frame size and during bright daylight. We fed our images into a 32 by 32 size 3 step CNN layers to train the model with our training set of 4.430 images. The corresponding CNN and residual model architecture is shown in Figure 2 and Figure 3. Our basic classification of training set was performed and the images were labeled based on the parameters mentioned in Table 1. We have labelled our training images to analyze the cracks and other anomalies present on road surface observing the RGB image and using openCV libraries. After image preprocessing, of the training set images, the images have been analyzed for variance over surface pixel values achieved through Fourier Transformation. Figure 2 and Figure 3 shows our feature learning and model training steps across the corresponding model. The imperfect roads that we considered from our dataset based on our image analysis mentioned above; are unfit for human transportation. Following the features of Table 1, we have analyzed the density of cracks, width of roads, surface smoothness, color of roads, environmental metric present on roads like: presence of water due to rainfall, sand or raw soil on road surface and trained our model on such criterion. For building the model, we have fed the images to a deep FCN as FCN performs faster than traditional CNN and can take any size of inputs. We went through trial and error approach with applying various range of epochs and lastly, we achieved our best result in context to our computational resource, for 100 epochs with an average accuracy of 87% for CNN model and 92% for Inception-V4 based residual model across all classes (detailed result mentioned in the Section 6).

The CNN architecture that has been implemented, consists of 32 by 32 convolutional layers containing 5 output layers. We used 32 size filter, 2D maxpooling with pool size of 2 by 2, 512 dense layers with activation function relu with softmax. The images' size was converted to 128 by 128 pixels. The configuration of our training resource is intel core i5 8th gen 6 core @4.00 GHz 32 GB DDR4 RAM PC NVidia GTX 1070 8 GB GDDR5 Memory: Dedicated 8GB + 8 GB Shared=16 GB GPU Memory in total. The training time for 100 epochs was 15.33 minutes for training 4,430 images.

In the proposed residual model architecture, the convolution block consists of total 13 layers and divided into 3 residual blocks. In the first block it has two 3 by 3 convolution layers with 32 filters and one 2 by 2 MaxPooling layer and a dropout of 20 percent. On the second block the model has one 3 by 3 Convolution layer with 64 filters and one 2 by 2 Maxpooling layer and a dropout of 20 percent. The third block has three fully connected layers with dense 512 neurons and a dropout of 45 percent. Finally, it has a classified layer to classify the images into 5 different classes. With "Softmax" activation function. After Each and every convolutional block we have used 'Relu' as the activation function. To get and Optimized output we have used 'Adamax' for better optimizer with 'categorical_crossentropy' loss function to calculate loss. As for the input the images' size was converted to 128 by 128 pixels. The configuration of our training resource for the residual architecture is same as mentioned above for CNN based FCN architecture.

To analyze the images perfectly and primary classification of training set, we applied black and white thresholding, canny edge detection, Fourier analysis and classified the images accordingly. This process has been observation based and images has been split into 5 classes by our research team. We considered the following parameters present in Table 1 to analyze our images.

The training set data has been fed into our Inception V-4 based deep neural network and the FCN network. The network learns how the surface key features from Table 1 are present on a particular image and learns the pattern of the roads.

Table 1. Parameters Considered to Detect the Road Condition

Sl No.	Parameters
1.	Cracks severity
2.	Road Width
3.	Surface Color
4.	Weather Effect on Road
5.	Ups and Downs Due to Cracks
6.	Drainage, Garbage and other anomalies presence

4. DATASET OVERVIEW

A) Brief Description:

Our dataset consists of 5,370 images. Among them 4, 430 images have been kept for training data split into 5 classes as mentioned in Table 2, 340 of the other remaining images have been kept for validation set and rest 600 images have been kept for test set. Among the whole dataset, 3,785 images were collected by our research team over a time period of 4 months, the remaining 1,285 images have been acquired from Local Government Engineering Department (LGED) of Government of People's Republic of Bangladesh [28] and 300 best quality images was collected from google images [29]. The collected dataset contains images of 53 different stations inside Bangladesh. For the classification of validation and testing set we considered all of the factors mentioned in Table 1. Figure 1, gives a brief idea about what types of images the dataset holds. The statistics of train, validation and test image set is compiled on Table 2.

B) Training Classes Description:

1) Perfect Roads: The best category concrete roads' pictures naming RCC and Rigid have been taken for training under this category. Roads that have a smooth surface condition and no cracks found have been taken for consideration. These roads are very wide and mostly of highway roads of the country.

Table 2. Dataset Split.

Classes	Training	Validation	Test
Perfect Road	865	68	150
Good Road	925	68	150
Bad Road	1,360	68	130
Severely Bad Road	700	68	100
Water on Surface	580	68	70

2) Good Roads: The local area roads' pictures which are good in condition but structurally not sustainable as RCC and Rigid roads are have been taken for training in this category. These roads are narrower than RCC and Rigid roads. Some roads of the hilly tracts of the country which are good for consideration have been taken in this class as well.

3) Mild Bad Roads: Roads' images which surface carpeting have started decaying but not that much bad in condition have been taken for training under this class. The crack parts are visible in the images and which have been confirmed with black and white thresholding as shown in Figure 1. using OpenCV library.

4) Severely Bad Road: These roads surface condition has been considered unfit for transportation. The concrete carpet has been totally removed by environmental and heavy transportation reasons. There are big cracks, pits muddy condition visible in the images for such condition.

5) Water on Surface Road: Roads that have standstill water on surface have been considered for training. The images taken for this training class have visual water on the whole road surface, but not on the pits only. Similar work has been done in [30][31] but for our proposed model roads that have whole road area drowned beneath water surface so that the usual road surface color does not get matched with the surface water color as shown in Figure 1 and result is shown in Figure 5. in section 7.

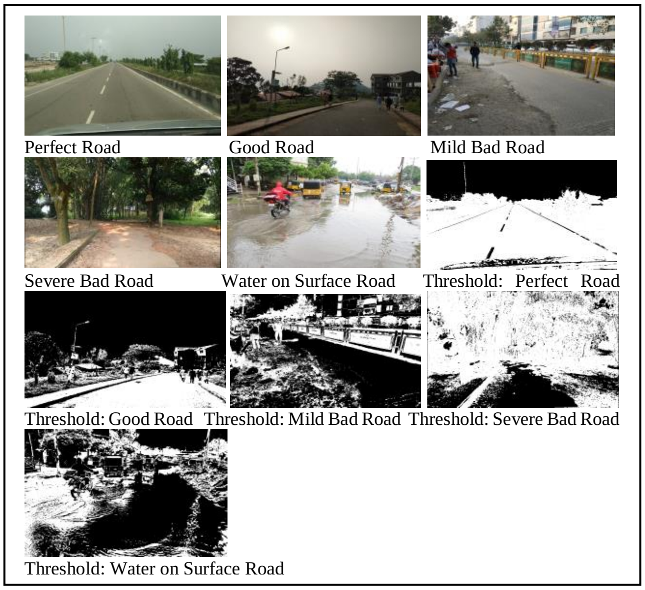

Figure 1. Dataset abstract view: Sample 5 classes of Roads (during preprocessing and labeling steps).

5. MODEL ARCITECTURE

Classical CNN gets trained with image feature learning. With steps by steps of convolution layer and with proper activation function the network extracts the key features from image pixel boxes and starts to recognize the pattern.

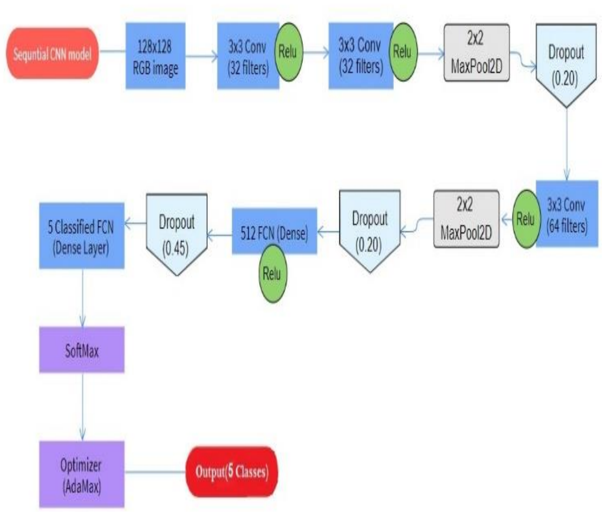

Figure 2. Modified Resnet Model Architecture.

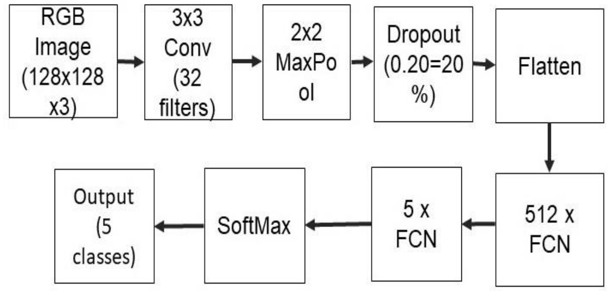

Figure 3. Modified CNN Model Architecture.

But the old inception model was trained in a partitioning manner. Each replica of that network was partitioned into sub-networks for the memory to fit the whole network. With the introduction of Tensorflow [12], the most recent Inception models can be trained without the replica partitioning. Recent optimizations of memory with backpropagation enabled in part and it achieved considering what tensors are needed for gradient computation and trying to reduce such tensors by structuring the computation. As our research team was experimenting on how a residual deep network works comparing to another classical CNN architecture, with the better results achieved by residual network, described in section 6, the Figure 2. demonstrates our modified resnet architecture and Figure 3. demonstrates the CNN architecture that we ended up considering based on our achieved best results.

6. RESULT ANALYSIS

The model was experimented for 0 to 100 epochs and the accuracy sharply raised from 5th epoch and flattened till 100 epochs. The cost or loss function gradually decreased in loss from 5th epoch. Figure 4. shows the performance of epoch and loss function of our residual network model. Figure 5. demonstrates our model’s result view.

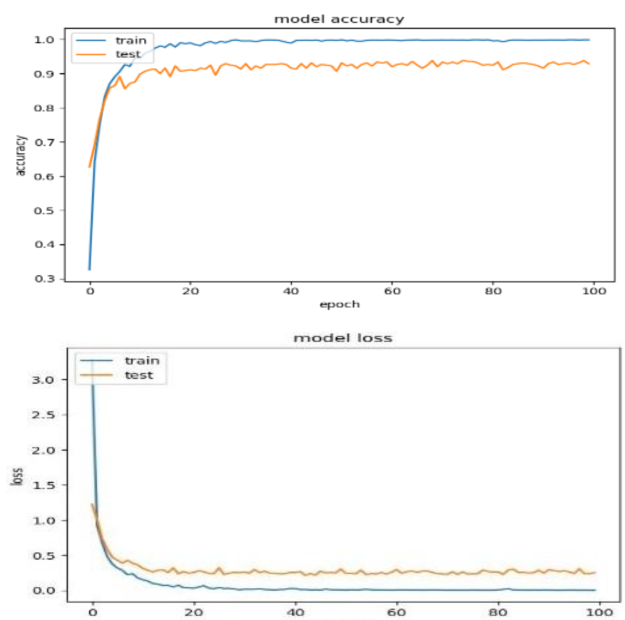

Figure 4. Resnet model accuracy performance with respect to epoch and loss function.

7. CONCLUSION AND DISCUSSION

From Table 3, we observe the test accuracy of our modified Residual model staying around 87-96%, whereas for CNN architecture it is 82-92% for all the classes which we believe a realistic satisfactory result for such approach. We are developing a web interface to open the model for public use, so that more images data can be achieved by our research team to check the model performance across verities of image framing and daylight-based conditions.

The motivation of this research work is to build an efficient machine learning image recognition model that have a decent accuracy to classify the roads in Bangladesh and more or same roads of other developing and developed countries. Our dataset is believed to hold ideal good concrete Rigid and RCC roads and also rural roads of Bangladesh. The situation of cracks and surface color is more or less same in other countries also. We tested our model by few other countries' road images taken from social media and google images. To our satisfaction we have seen the both tested models give good test results for those images. But the resolution of the image set is a concern. But inception blocks-based model gives better results than classical CNN based model as proven above. The model is not being able to take bad quality pictures which is understandable because for those images the dimensionality highly reduce and it creates problem for the model to perform its calculations.

Table 3. Model Architecture-wise Result Summary for Training and Testing Across All Classes.

	CNN Architecture		Residual Architecture	
Classes	**Train Accuracy**	**Test Accuracy**	**Train Accuracy**	**Test Accuracy**
Perfect Road	94%	85%	97%	96%
Good Road	92%	92%	92%	92%
Mild Bad Road	92%	85%	91%	89%
Severely Bad Road	90%	82%	95%	87%
Water on Surface	95%	89%	91%	95%

Our future goal contains to add more classes based on the context of Bangladesh and some other countries as well; like: we have plan to train our model for rural roads made of just soils, roads made with bricks and roads covered with grass. We also have motivation to develop the model to detect the lane and edge detection of such roads.

Figure 5. Test Set results of classified roads.

8. REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", **In Advances in neural information processing systems, pages** 1097–1105, 2012

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),** 2014.

[3] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking", In ***Advances in Neural Information Processing Systems,*** pages 809–817, 2013.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,** pages 3431–3440, 2015.

[5] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks", In **Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1725–1732. IEEE,** 2014.

[6] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks", In **Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference** on, pages 1653–1660. IEEE, 2014.

[7] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution", In **Computer Vision–ECCV 2014, pages 184–199. Springer,** 2014.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", **arXiv preprint arXiv:1512.03385,** 2015.
[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", **arXiv preprint arXiv:1512.00567,** 2015.
[10] C. Szegedy., S. Ioffe, Vincent V., A. Vanhoucke. "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", **arXiv preprint arXiv:1602.07261v2 [cs.CV] 23 Aug 2016**
[11] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q. V. Le, et al. "Large scale distributed deep networks", In **Advances in Neural Information Processing Systems,** pages 1223–1231, 2012.
[12] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, ´ R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. War- ´den, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. **TensorFlow: Large-scale machine learning on heterogeneous systems, 2015.** Software available from tensorflow.org.
[13] A. Goswami and H. Liu, "Deep Dish: Deep Learning for Classifying Food Dishes,"(2017)[Online],Available: **https://www.semanticscholar.org/paper/Deep-Dish-%3A-DeepLearning-for-Classifying-Food-Goswami Microsoft/2e596abd9d4616f75c9da45d88787cb2fd5bee81.** [Accessed April 29, 2019].
[14] J. H. Chuah, H. Y. Khaw, F. C. Soon and C. Chow, "Detection of Gaussian Noise and Its Level using Deep Convolutional Neural Network", **Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia,** November 5-8, 2017.
[15] M. Lin, Q. Chen, and S. Yan, "Network in network", **arXiv preprint arXiv:1312.4400,** 2013.
[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", **arXiv preprint arXiv:1409.1556, 2014**.
[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,** pages 1–9, 2015.
[18] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, "Pyramid Scene Parsing Network", **arXiv preprint arXiv:1612.01105v2 [cs.CV] 27 Apr 2017.**
[19] T. Pohlen, A. Hermans, M. Mathias, B. Leibe, "Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes", **arXiv:1611.08323v2 [cs.CV] 6 Dec 2016.**
[20] M. Teichmann, M. Weber, M. Zöllner, R. Cipolla and R. Urtasun, "MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving", **IEEE Intelligent Vehicles Symposium, IV 2018, Changshu, China,** June 26-30, 2018, pp. 1013-1020.
[21] Y. Lyu, and X. Huang, "Road Segmentation Using CNN with GRU", **22nd International Conference on Information Visualization,** 2018.
[22] Z. Zhang, Q. Liu, Y. Wang. "Road extraction by deep residual U-net", **IEEE Geosci. Remote Sens. Lett., vol. 15, no. 5, pp. 749-753,** May 2018.
[23] J. G. Haran, J. Dillenburg, and P. Nelson, "Realtime Image Processing Algorithms for the Detection of Road and Environmental Conditions", **Ninth International Conference on Applications of Advanced Technology in Transportation, AATT 2006, Chicago, Illinois,** August 13-15, 2006, pp. 55-60.
[24] C. V. Dung and L. D. Anh, "Autonomous concrete crack detection using deep fully convolutional neural network", **Automation in Construction,** Volume 99, Pages 52-58, March 2019.
[25] J. Zhao, H. Wu, and L. Chen, "Road Surface State Recognition Based on SVM Optimization and Image Segmentation Processing," **Journal of Advanced Transportation,** Vol 2017, pp. 1-21. 2017.
[26] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring," In **Proceedings of the 6th international conference on Mobile systems, applications, and services (MobiSys '08), New York, USA,** June 17 - 20, pp. 29-39.
[27] S. Yoda, H. Okabe, J. Takagi, and T. Yamashita, "Road surface recognition sensor using an optical spatial filter", in **Proceedings of the IEEE Conference on Intelligent Transportation System,** pp. 253–257, 1997.
[28] Local Government Engineering Department (LGED), **http://www.lged.gov.bd/** [Accessed November 11, 2018].
[29] Google Images, **http://www.images.google.com** [Accessed February 25, 2019].
[30] K. Ueda, I. Horiba, and K. ledaga, "A detecting method of wet condition on road using image processing", **Information Processing Society of Japan,** vol. 35, no. 6, pp. 1072–1080, 1994.
[31] M. Yamada, K. Ueda, I. Horiba, S. Yamamoto and S. Tsugawa, "Detection of Wet-Road Conditions from Images Captured by a Vehicle-Mounted Camera", **Journal of Robotics and Mechatronics,** Vol.17 No.3 pp. 269-276 doi: 10.20965/jrm. 2005.p0269.