

# Image based Road Surface Classification Method using CNN

Dong-Geol Choi

**Abstract:** Road detection and road surface classification in autonomous driving are the most basic and important issues. In this paper, we propose a data augmentation method for road surface classification using image information. We design an optimal network that can classify the type of road surface from the input image information and propose a data increase technique that can efficiently judge by using limited data to improve learning performance. To verify the proposed methods, many running images were used on the Internet. Experimental vehicle was developed and applied to verify the developed networks and it shows that they operate accurately in real time.

**Keywords :** road surface classification, CNN (Convolutional Neural Net), data augmentation, ResNet

## I. INTRODUCTION

Advances in deep learning [1], [2] have shown state-of-the-art performances in various recognition tasks [3-5]. Thanks to open-sourced deep learning frameworks, commercial applications [6], [7] based on deep learning are made possible. The deep learning algorithms are being applied to vehicles worldwide to enhance the safety, function, and quality of automobiles. Recently, cameras and techniques using them have been widely applied. The camera, which was mainly used in the black box mode in the aftermarket, is now being applied to commercial vehicles from vision-based LDWS (Lane Departure Warning System) to LKAS (Lane Keeping Assistant System) Research on the perception of the surroundings of vehicles and pedestrians is being actively carried out. Driving awareness and road surface judgment in autonomous driving are the most basic and important issues. Previous studies have used the sound, radar, and multispectral sensors to determine the road surface. However, the results of using the camera are limited to studies that recognize road surface damage, cracks, weather, and surrounding environment. If a camera image attached to a vehicle for various purposes is fused with a sensor installed in a vehicle and used for road surface judgment, a safety and a better driving experience can be provided to a user without increasing a cost. This study is a study to judge the road surface ahead of the vehicle using image information and machine learning. We designed an optimal network that can classify the type of road surface from the input image. The data to be used for network training were collected using real vehicles, and a running image that was published on the Internet was used to acquire a larger amount of data. In this

paper, we present a data augmentation method for road surface classification using CNN. We designed an optimal network that can classify the type of road surface from the input image. The data to be used for network training were collected using real vehicles, and a running image that was published on the Internet was used to acquire a larger amount of data. In the process of verifying the results on the actual vehicle, the road surface area was inferred using the vehicle speed and the steering angle obtained by the CAN bus. As a result, the types of road surface were classified in real time.

## II. METHODS

In this section, we introduce a research method to judge the road surface ahead of the vehicle using video information and machine learning. We design a network that can classify the type of road surface from video and apply it to the actual vehicle to confirm the performance.

For this purpose, the data to be used for network training were collected using actual vehicles, and the running images that were published on the internet were used to acquire a larger amount of data. In addition, we augmented the learning data for better learning performance. In order to verify the results on a real vehicle, a camera and a processing device were installed in the vehicle, and the kind of road surface was judged in real time and displayed on the screen.

### A. Decision of the types of road surface

The purpose of this study is to determine the type of the front road using machine learning based on the image information acquired by the vision sensor. For this purpose, the type of road to be classified must be selected. Recently, advanced SUVs optimized for off-road driving are equipped with devices that change vehicle behavior according to the type of road surface [8]. Table-1 shows a classification of road surface provided by three company representative SUVs. Based on pavement roads, slippery roads (snow, grass or pebbles), roads where the roads and car body are easy to sink down (mud, lots of snow), road surface is rough but the surface is rough, Easy way (rough rock road).

**Revised Manuscript Received on July 22, 2019.**

**Dong-Geol Choi**, Department of Information and Communication Engineering, Hanbat National University, Daejeon, 34158, South Korea. Email: dgchoi@hanbat.ac.kr

Table- 1: Initial Set of features used for the experimentation

Ford	Landrover	Jeep
Normal	Dynamic (On-road)	Auto
Snow	General Program	Sport
Sand	Grass/Gravel/Snow	Snow
Mud/Rut	Mud/Rut	Sand/Mud
	Sand	Rock
	Rock Crawl	

In this study, the road surface environment was classified into four kinds of pavement, sandy soil, mud and snow road considering the experimental conditions

**B. Data collection**

Training data for machine learning should be collected in sufficient quantities in various environments. To do this, we collect data from actual vehicles and collect running images from the Internet. The collected images were used for learning by increasing the proposed augmentation technique.

**B-1. Decision of the types of road surface**

In the afternoon on a clear or somewhat cloudy day, we actually traveled various roads and acquired images. We obtained driving images of roads such as asphalt, unpacked roads, mud troughs, mountain roads, snow roads, and gravel fields from pavement roads, roads, farm roads, construction sites, mountains, A vehicle which is easy to acquire data in various driving environments was selected and modified so as to be able to capture and store data. Additional batteries, DC-AC inverters, and PC shock absorbers were included for long-term testing as shown in Figure 1.

The camera to acquire the image is equipped with a 1280x960 resolution Ethernet camera on the top of the front glass inside the vehicle, and an LCD monitor, keyboard and mouse and a USB port for monitoring are installed in the passenger side of the vehicle.

The latest Linux long-term support version (16.04 LTS) and ROS (Robot OS) [9] have been installed in the SSD to facilitate data capture and system maintenance. Detailed system installation details are shown in Fig. In order to utilize the video camera in actual experiments, the CAN interface is installed so that the vehicle speed and yaw rate can be secured.

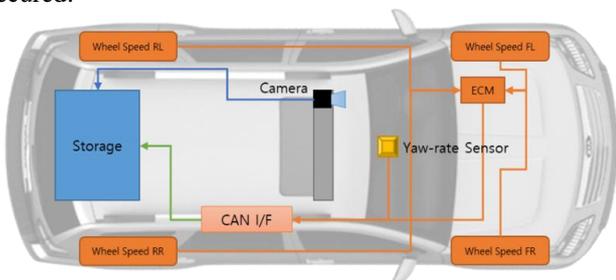


Fig. 1. Configuration of the test vehicle

**B-2. Image Data collection using internet**

The collection of data through driving is limited in terms of cost and time to obtain various data. In order to overcome this, various videos of the video of the In-dash camera which was released to YouTube were acquired. In particular, the

FIA's annual In-dash video of the World Rally Championship [10, 11] was conducted in various environments such as snowy roads and dirt roads, and it was worth the data. In addition, we used images of mountain bike rides and RC cars running in the grass. Figure 2 is part of the image collected on the Internet



Fig. 2. Examples of images collected by internet

**B-3. Configuration of the learning data**

Acquisition data in the real environment using the vehicle and traveling videos obtained from the internet were sampled at 1 Hz and used as learning data

Table- 2: Data distribution by class

Class	Number of images
Off-Road (Dirt road)	6794
Asphalt	16216
Snow	12911
Mud	718

Data imbalance in which data is collected in a specific class when data is acquired Table 2 is an unavoidable phenomenon in large-scale data collection.



This imbalance has an adverse effect on learning when learning a learning-based classifier. In order to mitigate this, we set the sampling rate to be two times for unpacked road and 15 times for mud road in learning process, so that the data sampling probability for each class is uniform.

#### B-4. A data augmentation method

In general, a classification network receives images of fixed input size. However, the acquired images and the images acquired online have the same aspect ratio and resolution, respectively, depending on the shooting environment. The most common way to match this is to extract the center of the image to fit the input size, or adjust the entire image to fit the input size. However, this method may not be learned robustly for images with some difficulty levels

In the general situation, roads are seen in the middle, but roads are not visible in the uphill or high-speed section, and only the background is caught in the image. If the network is learned by central extraction or size adjustment, the network may be learned by deflecting on the data of the road, and it may be difficult to cope with the situation shown in Figure 3. We want to introduce data enhancement techniques that can cope with two things:

1. A situation where various image sizes come in as inputs
2. Learn how to network together with information about the background as well as the road

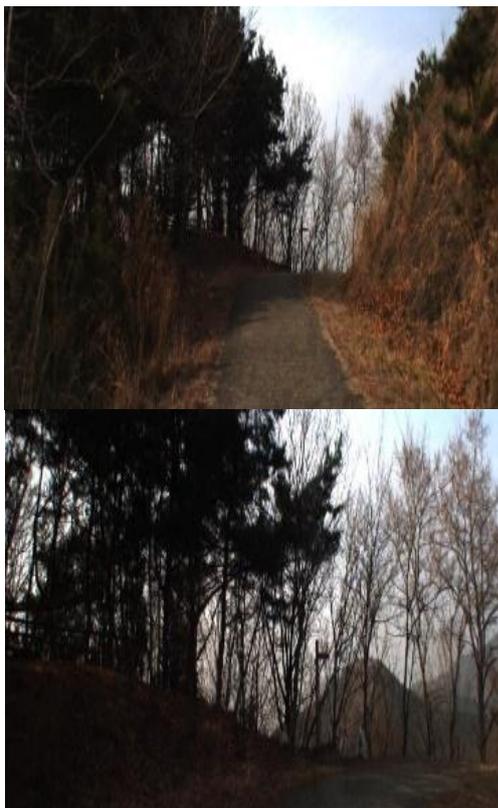


Fig. 3. Normal condition (Top) and special condition (Down)

If the input size of the network is  $n$  by  $n$ , adjust the size by  $1.2n$  times based on the height of the input image. After that, the  $n$  by  $n$  image is extracted using the sampling probability distribution based on the image center and applied to the network learning

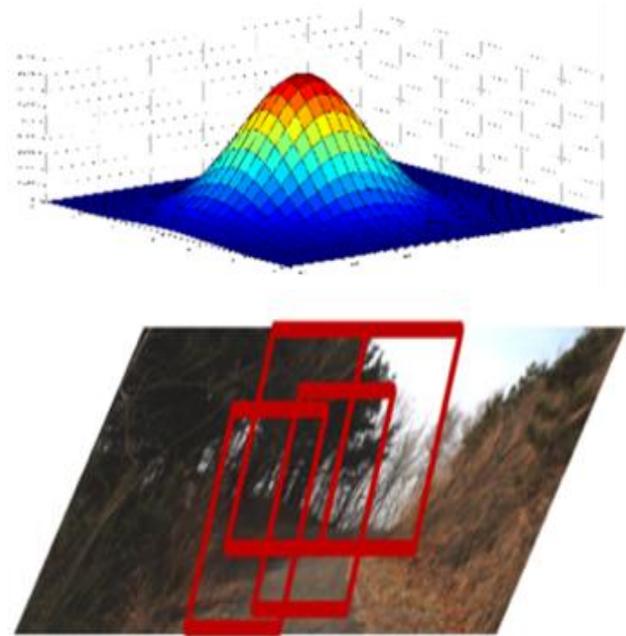


Fig. 4. Data augmentation method. (Top image) Region extraction probabilistic distribution, (Down image) original image

The sampling probability distribution can use a normal distribution or a uniform distribution depending on the situation. In this paper, uniform distribution is used for learning, as shown in Figure 4.

### III. RESULTS AND DISCUSSION

In order to divide the data collected in this paper into learning sets and evaluation sets, the first half of the collected video was used for learning and the other half was used for evaluation. Each input image is edited according to the method described in Section 2.2.3 and input to the network.

Table- 3: Processing time per each Network : Total time for editing and recognition after loading image file

Network	Off-Road	On-Road	Snow	Mud
SqueezeNet v1	97.92	98.77	98.65	97.47
SqueezeNet v2	98.25	99.10	98.57	97.76
AlexNet	97.89	98.47	98.88	96.36
VGG16	99.08	99.21	99.08	97.20
ResNet18	<b>99.78</b>	99.25	<b>99.74</b>	98.32
ResNet34	<b>99.78</b>	<b>99.56</b>	99.66	<b>99.44</b>

In order to measure the speed of the road surface recognition system proposed in this paper, 20006 test images were set to batch size 1 and the average time required was calculated (see Table 3). Computers used for speed measurement are equipped with Inter Core i-7-6800k (CPU 3.40GHz) and one Titan-X (pascal). The average recognition speed was about 200 fps at 5 ~ 6ms. It is about 6 times faster than general real-time algorithm and can operate well without delay even in high-speed camera environment.

From the learning results, as shown in Figure 5, we could confirm the high recognition rate of 98% or more on all networks used in experiment.

There are two reasons for the high recognition rate. First, the data distribution of the learning set and the evaluation set is very similar, and the data increment method presented in this paper effectively learned both road and background information.



Fig. 5. Example results

In terms of network performance, AlexNet and SqueezeNet showed poor performance, while ResNet showed the highest performance. In the case of AlexNet, SqueezeNet is similar to AlexNet, but its performance is similar to that of AlexNet, though it has the advantage of being lightweight as a model that dramatically reduces the number of parameters to optimize for a mobile environment. The VGG family of networks is a model presented after AlexNet, which increases performance by significantly increasing network size. ResNet uses the residual learning module to show the best performance of ImageNet. In this experiment, we can confirm that the performance difference of each network is similar to that of ImageNet performance Table 4.

Table- 4: Classification performance per each Network

Network	Processing time (ms)	Speed (fps)
SqueezeNet v1	5.136	194.7
SqueezeNet v2	5.123	195.2
AlexNet	4.675	213.9
VGG16	5.316	188.1
ResNet18	5.141	194.5
ResNet34	6.088	164.3

One thing to note here is the performance difference between ResNet18 and ResNet34. It is generally known that stacking multiple layers of a network increases performance. However, in this experiment, we confirmed that ResNet34 is deeper than ResNet18, but there is no significant difference in performance. This is due to the fact that there are not many classes to classify in the road surface classification (1000 classes in ImageNet), and the variety of road surface data is not large. Based on this experiment, it can be concluded that ResNet networks (ResNet 34, ResNet50, ResNet101, ResNet151), which are deeper than ResNet18, increase the

computation volume but do not have a significant improvement in performance.

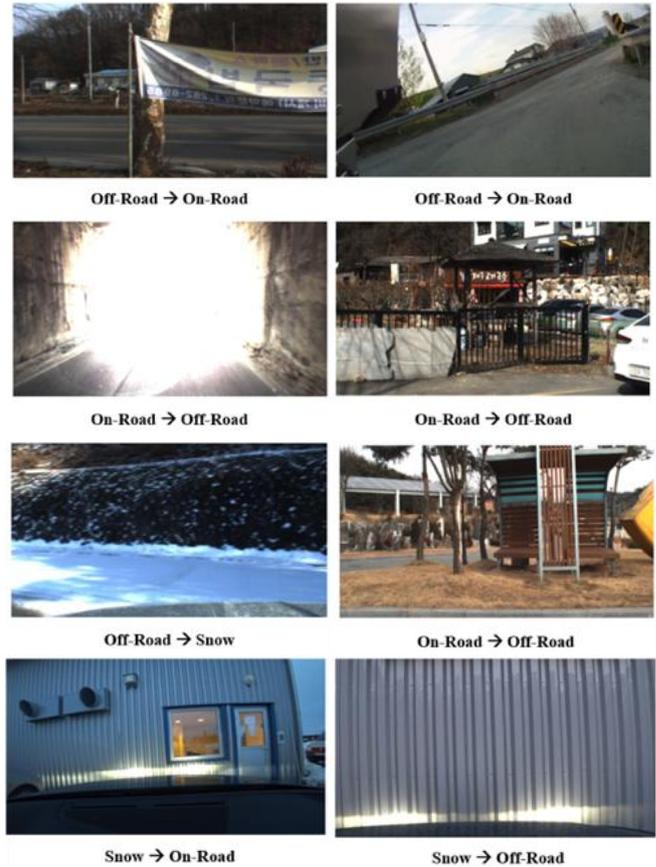


Fig. 6. Failure cases (Ground Truth Label → Expected result)

Through failure cases (see Figure 6), some limitations of the image-based road surface judgment algorithm can be grasped. Even if the road surface on which the vehicle is treading is packed or unpacked, the image is taken ahead of the bottom of the vehicle. As you can see from the failure case, if you see a dirt road or a mountain in front of you on a pavement road, on the contrary, if you see a pavement or a building in front of the dirt road, Another failure factor is that the classification of the road surface present in the dataset is not clearly visible. There are images that are classified into several snowy roads, which are difficult to distinguish from pavement due to melting snow and difficult to distinguish from general unpaved roads due to low moisture among the roads classified by clay. We could identify cases in which the network failed for images whose road surface classification was unclear.

#### IV. CONCLUSION

In this paper, we propose a series of design process from data set acquisition to deep network learning to make a deep learning based road surface judgment system. In order to obtain road surface data in various environments, the vehicle was modified accordingly and data was collected online to supplement the missing data.



Data enhancement techniques have been proposed for effective learning and six networks with excellent performance in ILSVRC have been verified for performance verification. It is experimentally confirmed that ResNet-based network, which shows excellent performance of 95% or more on all networks and shows the best classification performance on ImageNet, also shows good performance on the road surface. Using the system presented in this paper, the recognition performance is close to 99%, and if a large variety of large-scale data is collected, a very robust road surface judgment system will be possible.

autonomous robotics, and artificial intelligence issues. He received a fellowship award from Qualcomm Korea R&D Center in 2013. He was a member of Team KAIST, which won First Place at the DARPA Robotics Challenge Finals in 2015. He is a member of the IEEE

## ACKNOWLEDGMENT

This research was supported by the research fund of Hanbat National University in 2018.

## REFERENCES

1. K. Alex, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012. Jan: 1097-1105.
2. H. Kaiming, Z. Xiangyu, R. Shaoqing, S. Jian, "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. Jun: 770-778.
3. D. Jia, D. Wei, S. Richard, L. Li-Jia, L. Kai, F. Li., "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009. June: 248-255
4. L. Tsung-Yi, M. Michael, B. Serge H. James, P. Pietro, R. Deva, D. Piotr, Z. C Lawrence, "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014. Sep: 740-755.
5. A. Stanislaw, A. Aishwarya, L. Jiasen, M. Margaret, B. Dhruv, L. Zitnick, C., P. Devi, "Vqa: Visual question answering." Proceedings of the IEEE international conference on computer vision. 2015. May: 2425-2433
6. Google Cloud Vision [Internet]. [place unknown]; [updated 2019 Feb 14; cited 2019 March 15]. Available from: <https://cloud.google.com/vision/>
7. papago [Internet]. [place unknown]; [updated 2019 Jan 01; cited 2019 March 15]. Available from: <https://papago.naver.com/>
8. Roverguide [Internet]. [place unknown]; [updated 2011 Dec 18; cited 2019 March 15]. Available from: <http://www.roverguide.com/12999/compare-three-selectable-terrain-m-management-systems/>
9. ROS.org [Internet]. [place unknown]; [updated 2017 July 25; cited 2019 March 15]. Available from: <http://wiki.ros.org/kinetic/Installation/Ubuntu>
10. Official Live Stream & Video Channel of the WRC [Internet]. [place unknown]; [updated 2019 March 10; cited 2019 March 15]. Available from: <http://www.wrcplus.com>
11. FIA World Rally Championship [Internet]. [place unknown]; [updated 2019 March 2; cited 2019 March 15]. Available from: <https://www.youtube.com/user/wrc/videos?app=desktop>
12. I. Forrest N., H. Song, M. Matthew W., A. Khalid, D. William J., K. Kurt, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size." arXiv preprint arXiv:1602.07360. 2016.
13. S. Karen, A. Zisserman, "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556. 2014.

## AUTHORS PROFILE



**Dong-Geol Choi** (M'16) received the B.S. and M.S. degrees in electric engineering and computer science from Hanyang University, in 2005 and 2007, respectively, and the Ph.D. degree in the Robotics Program from Korea Advanced Institute of Science and Technology (KAIST) in 2016.

After working at KAIST as a Postdoctoral Researcher and Research Assistant Professor until August 2018, he is currently an Assistant Professor with the Department of Information and Communication Engineering, Hanbat National University. His research interests include sensor fusion,