

## Research Article

# A New Machine Learning Algorithm for Weather Visibility and Food Recognition

Young Im Cho\*, Akmaljon Palvanov

*Department of Computer Engineering, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, South Korea***ARTICLE INFO***Article History*

Received 31 October 2018

Accepted 15 December 2018

*Keywords*Atmospheric visibility  
convolutional neural networks  
CCTV  
graphic user interface  
recognition**ABSTRACT**

Due to the recent improvement in computer performance and computational tools, deep convolutional neural networks (CNNs) have been established as powerful class of models in various problems such as image classification, recognition, and object detection. In this study, we address two fundamentally dissimilar classification tasks: (i) visibility estimation and (ii) food recognition on a basis of CNNs. For each task, we propose two different data-driven approaches focusing on to reduce computation time and cost. Both models use camera imagery as inputs and works in real-time. The first proposed method is designed to estimate visibility using our new collected dataset, which consist of Closed-circuit Television (CCTV) camera images captured in various weather conditions, especially in dense fog and low-cloud. Unlikely, the second model designed to recognize dishes using artificially generated images. We collected a limited number of images from the web and artificially extended the dataset using data augmentation techniques for boosting the performance of the model. Both purposing models show high classification accuracy, requiring less computation power and time. This paper describes the complexity of both tasks and also other essential details.

© 2019 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).**1. INTRODUCTION**

Atmospheric visibility or meteorological optical range is a degree of ways nicely an observer can view a scene. This consists of how some distance you can see within the surroundings as well as the ability to see the textures and colors of the scene [1]. Observer can either be a human or a camera. Since fog and low-cloud is a collection of water-saturated fine particles or water droplets it acts to decrease the horizontal visibility of a scene significantly [2,3] and decreases the observer's ability to see through the air. Therefore, fog and low-cloud is the inverse of visibility that is the measure of the inability to view a scene [1].

Nobody would intentionally agree that poor air great is acceptable. The advantages of correct visibility are apparent. Visibility is a primary and enormously apparent indicator of preferred air excellent. Visibility is distinctly easy to recognize via public. Many humans regard suitable visibility as a figuring out thing for the satisfactory of out of doors existence. As an end result, the community can also decide the effectiveness of environmental control policies to improve air satisfactory by means of visibility. Experts use special equipment to provide measurement of visibility [1]. These tools themselves and their installation processes are ample expensive and measuring range is also limited. Hereby, we propose camera and Convolutional Neural Networks (CNNs) based approach to estimate visibility as the combination of these methods gives strong superiorities. Moreover, comparing to tools like visibility meters,

camera-based visibility measuring is more close to visual method. Also, the technique does not cost expensive.

On the other hand, extracting visibility features from an image, for representing visibility level, is too complex. Therefore, we have used state-of-the-art CNNs in the field of meteorological forecast. CNNs have the ability for solving hetero-associative problems and they can automatically extract effective features from the data. In addition, CNNs are good at classification and recognizing images as they can learn the relationships between inputs and outputs. The key factors for this ability is large labeled datasets, which support the learning process and novel network architectures with hundreds of millions of parameters. Although CNNs are good at understanding an image content from an image dataset, it is difficult to find suitable dataset always. Therefore, we have collected a large dataset of images captured by Closed-circuit Television (CCTV) camera as well as visibility values recorded by visibility sensors. The collected dataset was used for training our model to estimate visibility.

Unlike visibility estimation, dish recognition requires different approach as images of food might contain much information, many features than those of dense fog. Since input images different in terms of shape, texture, size and color as the Korean dishes lacks any kind of generalized layout. In contrast, to other types of food such as Indian or Italian, it is more complicated recognizing images of Korean dishes as the following reasons: images in the same category appear differently; it is complex for modeling dish images because of noise, likely, various backgrounds and textures; a number of high quality and proper images is limited. To evaluate effectiveness of our model on artificially extended dataset, we selected limited images of Korean dishes and

\*Corresponding author. Email: [yicho@gachon.ac.kr](mailto:yicho@gachon.ac.kr)

artificially extended the dataset using data augmentation techniques. In this paper, we present an algorithm that generates multiple images from a single input image. The algorithm includes several data augmentation techniques to enlarge the dataset that consists of small number of images. Our model have achieved high performance on a large and diverse image datasets with non-uniform backgrounds. Our three major contributions in this study as following:

- We purpose two CNNs based models for different tasks: visibility estimation and recognizing Korean dishes from camera images in real-time.
- We collected a dataset of images each with corresponding visibility values. The dataset includes dense fog, low-cloud, and mist etc., captured images. Additionally, we have collected a new set of images that included pictures of well-known Korean daily life dishes.
- We purpose an algorithm that generates multiple image from single input efficiently using different data augmentation methods. The algorithm generates a maximum number of images with minimum similarity.

In the next section, a brief summary of related studies is reviewed. Then, Section 3 presents details of both models and experiments as well as achieved results is given in Section 4. Finally, Section 5 concludes the paper.

## 2. RELATED WORK

Shengyan et al. [4] proposed visibility estimation based on meteorological laws using digital images. They used deep learning model to artificially extract features from input images [1]. Another approaches for estimating visibility proposed by Sami et al. [5] essential properties of these studies are using single camera and evaluate visibility on a basis of daytime light from the input images [1]. In the field of meteorology, one of the main law is Koschmieder's law and many researchers used this function to estimate visibility in both cases, using digital images and using special equipment for visibility estimation, these works have been done in Du et al. [6]. Since all these methods use images, as an input quality of the images is vital to get better results, means that degradation of images strongly restricts accurately estimation [1]. This requires essential extra step prior estimate visibility the step is enhancement of the quality of input images. For this situation Dark Channel Prior (DCP) method is widely used [1]. For instance, in poor visibility conditions enhancement of color images proposed by Keong et al. [7], in this study, several visibility conditions tested using synthetic images taken via airborne camera. Another effective use of DCP can be seen in Kaiming et al. [8]. They used a haze-imaging model to estimate the thickness of the haze then recover quality of the image. CNN-based dehazing method Boyi et al. [9] relies on the re-formulated atmospheric scattering model. In next section, we will explain our approach and techniques in detailed.

Considering the food recognition problem, there has been widely investigated for many years and there have been wide variety of dish image recognition algorithms and datasets related to this problem. A system based on deep CNN namely DeepFood introduced by Chang et al. [10] analyzes the food image taken by mobile devices. The performance of the model outperformed of previous proposed dietary assessments. Team of Yoshiyuki [11] purposed a Japanese

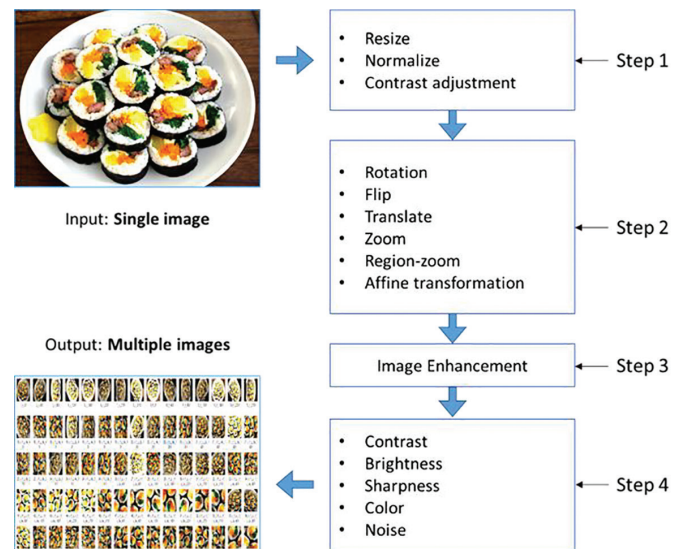
food recognition system by leveraging CNNs and compiled a dataset of 100 different foods in Japan. Their model obtain approximately 92% classification accuracy. Later on, the purposed model improved and have successfully applied other datasets such as Caltech-1014 and FOOD101/256. A model purposed by Yuzhen [12], obtained overall 90% accuracy using deep CNNs that consists of five convolutional layers in order to recognize food images. They used small dataset that has overall dish images 5822 with 10 classes. Another model FoodNet by Paritosh et al. [13] addressed food images recognition using multilayer CNN using the large dataset ETH Food-101. All models mentioned above take advantage of deep CNNs and require large sets of images. Therefore, we addressed to the problem of enlarging dataset from limited images.

A vast majority of approaches mentioned above classifies taking advantage of CNN and consumes significant amount of time and computer power. Unlike these methods we tried to implement a model using a few convolutional layers that can achieve a high performance with consuming less energy and time.

## 3. METHODS

### 3.1. Image Generator

To be able to learn different features and extract the link between input and output data, deep CNNs require a suitable set of examples and datasets with limited images arise a severe problem for deep CNNs. To tackle with this problem, we propose an algorithm that takes single image and applies several augmentation techniques sequentially in order to gain multiple images. The proposing algorithm can give desired number of unlike images with maximum efficiency. Figure 1 summarizes the purposing dataset enlarger algorithm. The algorithm has four steps. Initially, the system takes single or many images as input(s). Then, normalizes and adjusts the contrast and brightness of the input(s) in Step 1 (see Figure 1). Next, in Step 2 Figure 1, the normalized image(s) is/are applied data augmentation



**Figure 1** The proposing dataset enlarger algorithm. User enters a single image and the algorithm generates multiple images using data augmentation methods.

techniques (image generating functions can be chosen by users) for obtaining a new, unique appearance of the given input. Some generated inputs might lose their quality during the process in Step 2 [Figure 1](#), especially after zoomed/cropped/scaled or region-zoom applied. The region-zoom function crops different regions of the image then resizes. Therefore, images need to quality enhancement.

After gaining desired outputs, some images will be applied image enhancement methods in Step 3 [Figure 1](#), this stage is need only for some images. Pre-final, quality recovered images will be processed again in order to reduce the similarity of obtained images Step 4 [Figure 1](#). It is done with the contrast, sharpness, color and brightness changes as well as applying the random Gaussian noise.

### 3.2. Datasets

For the task of visibility estimation, we used a set of CCTV images taken by stationary surveillance cameras from four different points in South Korea, near the sea side [\[1\]](#). Atmospheric visibility measurements for those locations was also available. The visibility values are register in the range from 0 to 20,000 m. We assigned 21 classes for each 1 km of visibility (including 0) by dividing the visibility by 1000 and rounding the values. Because of the poor image quality during the night, we decided to use only images from daytime. Examples are shown in [Figure 2](#).

Since, there is no publicly available image datasets for Korean dish recognition, to train our complex food recognition model we have collected a new dataset. The resources for the dataset are taken from photos and web images. Our group in daily life collected some partial

of the images in this dataset. [Figure 3](#) shows example images from the dataset. We selected 50 images for each 23 classes. Images are artificially extended later using the proposed algorithm above, resulting in totally 92,000 images.

### 3.3. Networks

#### 3.3.1. Visibility estimation

Our first model estimates visibility from day time images. Especially, it can classify accurately if there is a foggy weather condition is illustrated. A logical structure of the model for visibility estimation can be seen in [Figure 4](#).



Figure 3 | Sample images from the food dataset.

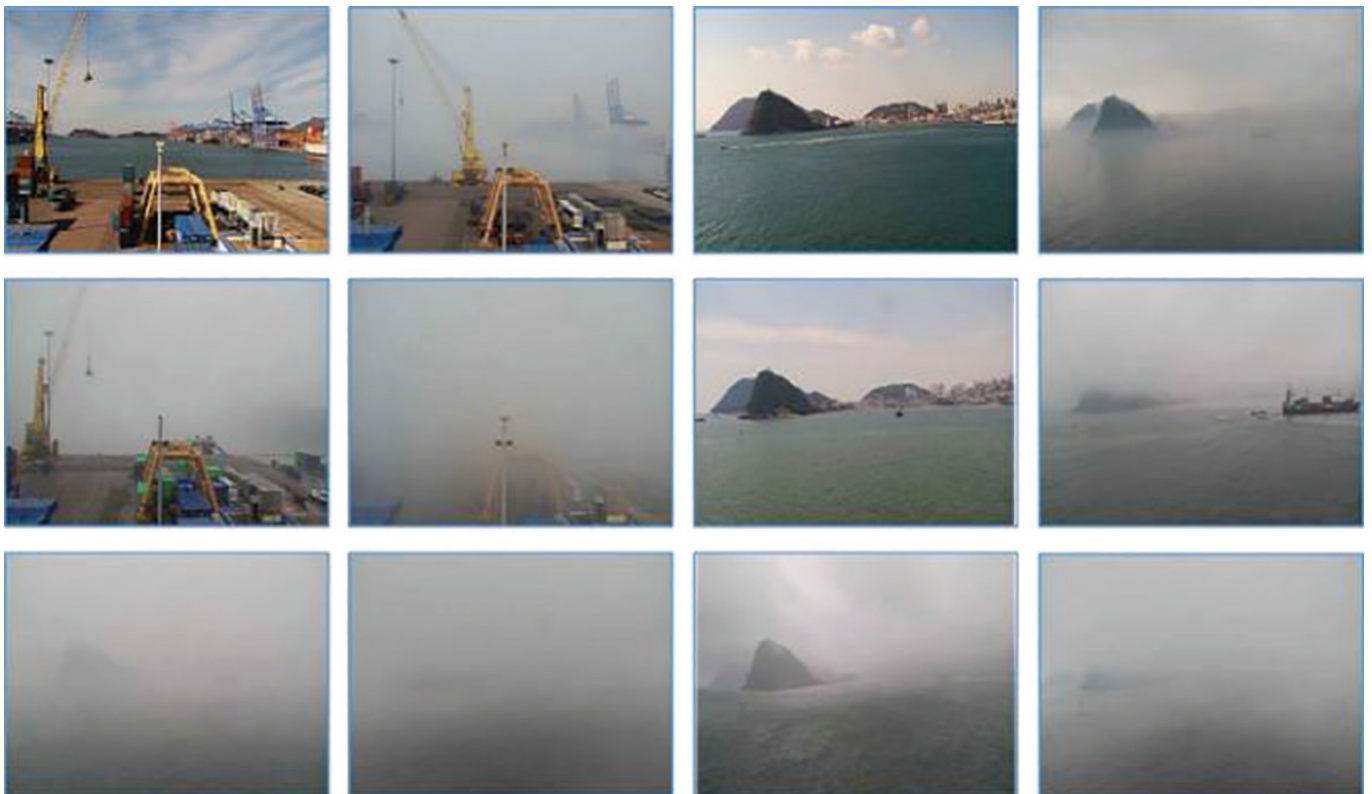


Figure 2 | Samples from the visibility dataset.



The first convolutional layer produces 32 feature maps using  $5 \times 5$  convolutional kernels followed by max-pooling operation. The second convolutional layer produces 64 feature maps using kernel size  $3 \times 3$ . After the last max-pooling operation, feature maps are sent to the dense layer, which includes 1024 neurons. Finally, the last layer is Softmax layer that produces the output. Visibility is complex phenomena as dense fog or low-visibility presented images have almost no features. Fog occupies entire image space as a repercussion everything is abstract, features of objects hidden. In this case, deep CNNs strongly overfits and produces wrong prediction. Therefore, we use only two convolution layers to avoid overfitting.

### 3.3.2. Food recognition

Schematic architecture of the model is given in Figure 5. Initially, the first layer receives  $150 \times 150$  image as an input and produces 32 feature maps by convolving with the kernel size  $9 \times 9$ . Next, max-pooling operation reduces the dimensionality.

The kernel size of  $7 \times 7$  is used to convolve an output of the previous layer, resulting 64 maps of features. Third and fourth convolution layers produce 128 and 256 feature maps respectively, each followed by max-pooling operation. After the last max-pooling layer, dense layers employed sequentially each with 512 neurons. Finally, there are 23 Softmax neurons in the output layer, which correspond to the classes of dishes.

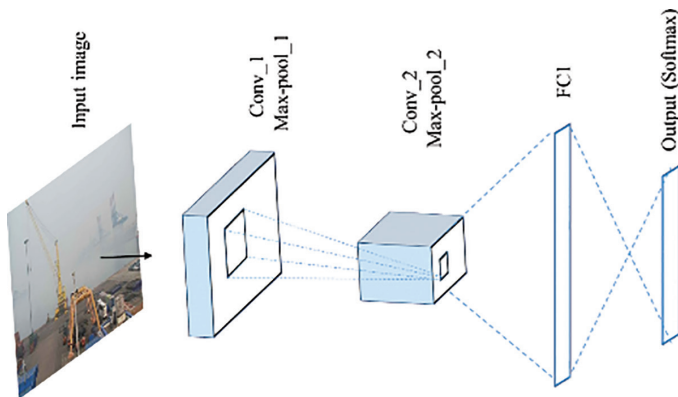


Figure 4 | The logical structure of the CNN model for visibility estimation.

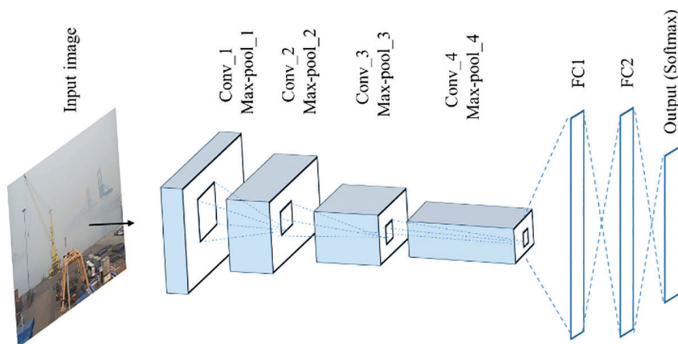


Figure 5 | The logical structure of the CNN model for food image recognition.

## 4. EXPERIMENTS AND RESULTS

Our experiment performed on a high-end server with 64 GB of Random-access Memory (RAM) equipped with two Nvidia GeForce GTX 1080 Ti GPUs. We conducted the experiment using Tensorflow machine learning framework. The training process is computationally costly and takes much time. It reads inputs from the memory then shuffles all images then starts training choosing random batches.

A training accuracy presents how perfectly fits every input image to the training data whereas a loss measures the inconsistency between predicted value and actual label. Thus, to obtain a higher performance in terms of classification, the accuracy function needs to be greater reaching up to approximately 100% whereas the loss function is required to reach the minimum value during the training and validation and testing. In Figures 6 and 7, the one can see that accuracy functions reached promising values and loss functions also minimized significantly. As we can see from Figure 6a, during training, the accuracy on random batches achieves 86% of successful visibility classification based in image similarity. Figure 6b shows the loss function, which decreased to 0.2. From Figure 7a, it is clear that accuracy function of the food recognition model reached over 87% and from Figure 7b, we can see the loss function reduced to approximately 0.11.

To perform the evaluation, we developed a Java application that selects images from the reference image database and performs an evaluation on each of them separately. Both proposing methods achieve about 81% classification accuracy on random sample

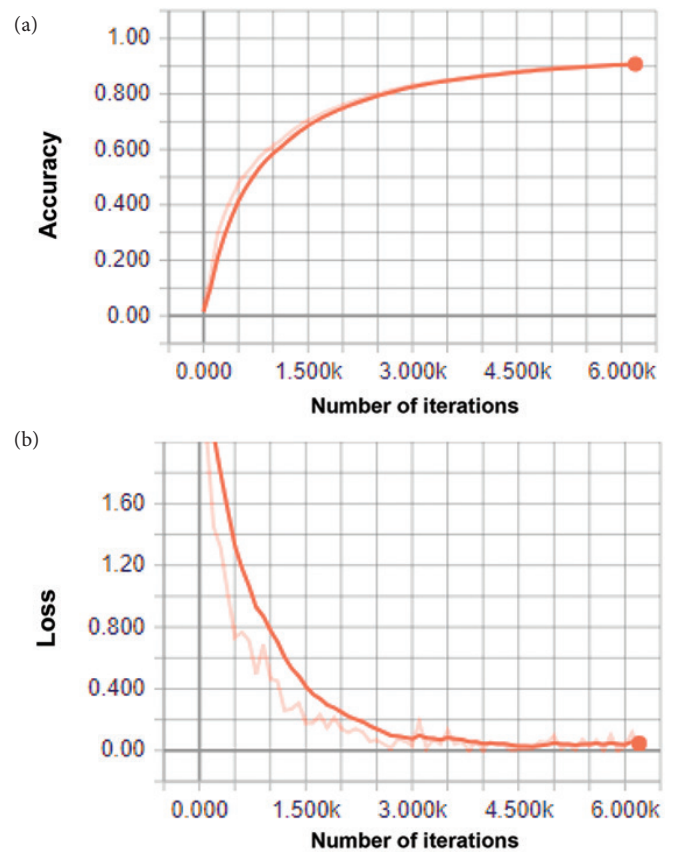
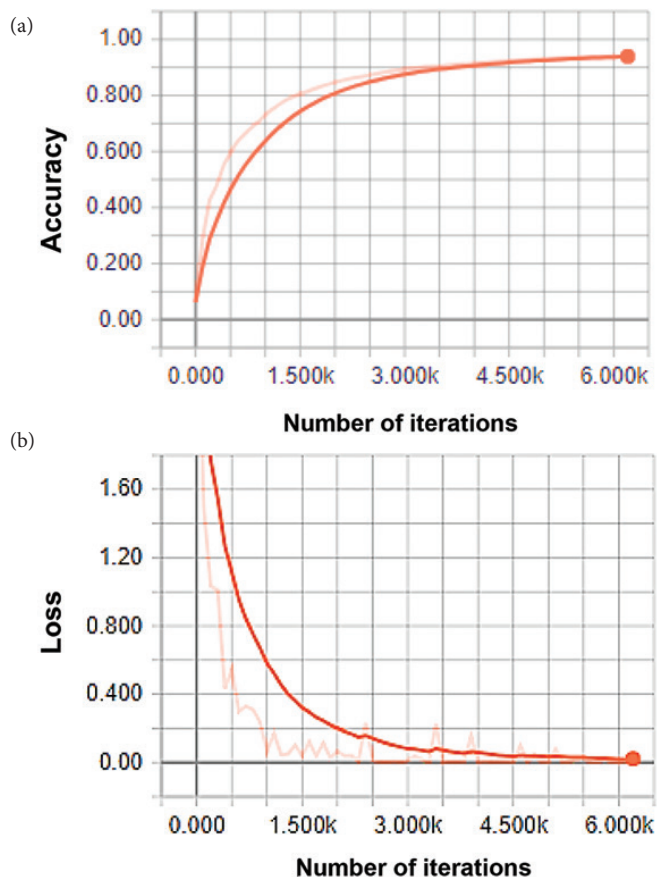


Figure 6 | The accuracy (a) and loss (b) functions of the visibility model. The horizontal axis represents number of iterations; the vertical axis indicates the accuracy (a) and the loss (b) functions.



**Figure 7** | The evolutions of accuracy (a) and loss (b) functions of food recognition model. The horizontal axis represents number of iterations; the vertical axis indicates the accuracy (a) and the loss (b) function.

images. After training, we exported models into a compressed file for deployment. We developed a Java-based offline graphic user interface as well as a Java EE-based web page for testing the model in real-time. A web-based application is designed to run on a remote server accessible over a TCP/IP network such as LAN or the Internet. The application takes a .jpeg or .png image as input, uploads it to the server, loads the trained neural network model and returns the inference result to the clients' web browser.

## 5. CONCLUSION

To sum up, in this study we show two of the possibilities of neural network models based on CNNs. In addition, we purposed a new, efficient algorithm for enlarging dataset with limited number of images to train complex CNN-based models. Our algorithm receives single image and generates multiple images using data augmentation methods. Moreover, we have compiled two datasets for providing experiments on two different tasks. For each dataset we have implemented two models on a basis of deep CNNs and evaluated their performance in terms of an accuracy and loss functions. Finally, our models have obtained very promising results.

## CONFLICTS OF INTEREST

There is no conflicts of interest.

## ACKNOWLEDGMENTS

This paper is supported in part by NRF (no: 2018R1D1A1A09084151) and in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2017-0-01630) supervised by the IITP (Institute for Information & communications Technology Promotion).

## REFERENCES

- [1] A. Palvanov, A. Giyenko, Y.I. Cho, Development of visibility expectation system based on machine learning, in: K. Saeed, W. Homenda, (Eds.), *Computer Information Systems and Industrial Management, CISIM 2018, Lecture Notes in Computer Science*, Vol. 11127, Springer, Cham, 2018, pp. 140–153.
- [2] S.G. Narasimhan, S.K. Nayar, Contrast restoration of weather degraded images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (2003), 713–724.
- [3] R.T. Tan, Visibility in bad weather from a single image, 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Anchorage, AK, USA, 2008, pp. 1–8.
- [4] S. Li, H. Fu, W.-L. Lo, Meteorological visibility evaluation on webcam weather image using deep learning features, *Int. J. Comput. Theor. Eng.* 9 (2017), 455–461.
- [5] S. Varjo, J. Hannuksela, Image based visibility estimation during day and night, *Computer Vision – ACCV 2014 Workshops, Lecture Notes in Computer Science, Revised Selected Papers Part III*, Springer International Publishing, Singapore, 2015, pp. 277–289.
- [6] K. Du, K. Wang, P. Shi, Y. Wang, Quantification of atmospheric visibility with dual digital cameras during daytime and nighttime, *Atmos. Meas. Tech.* 6 (2013), 2121–2130.
- [7] K. Tan, J.P. Oakley, Enhancement of color images in poor visibility conditions, *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, IEEE, Vancouver, BC, Canada, 2002, pp. 788–791.
- [8] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (2011), 2341–2353.
- [9] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, AOD-Net: all-in-one dehazing network, 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, Italy, 2017, pp. 4780–4788.
- [10] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, Y. Ma, DeepFood: deep learning-based food image recognition for computer-aided dietary assessment, *International Conference on Smart Homes and Health Telematics*, Springer, Cham, 2016, pp. 37–48.
- [11] B.T. Nguyen, D.-T. Dang-Nguyen, T.X. Dang, T. Phat, C. Gurrin, A deep learning based food recognition system for lifelog images, 7th International Conference on Pattern Recognition Applications and Methods, Scitepress – Science and Technology Publications, Funchal, Madeira, Portugal, 2018, pp. 657–664.
- [12] Y. Lu, Food image recognition by using convolutional neural networks (CNNs), arXiv, CSE803 Final Project, 2016, arXiv:1612.00983.
- [13] P. Pandey, A. Deepthi, B. Mandal, N.B. Puhon, FoodNet: recognizing foods using ensemble of deep networks, *IEEE Signal Process. Lett.* 24 (2017), 1758–1762.

## Authors Introduction

### Young Im Cho



She received the B.S. and M.S. and PhD degrees in computer science from Korea University, in 1988, 1990 and 1994, respectively. She received the postdoctoral degree in computer science from University of Massachusetts, in 2000. She is currently the chairman of Korea Intelligent System Society, and the chief of AI and Smart City Laboratory at Gachon University, and Korea-

Kazakhstan ICT Cooperation Center and Intelligent Service Robot Automation System Society, and Smart City Forum in Korea.

### Akmaljon Palvanov



He received the B.S. degree in telecommunication technologies from Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Tashkent, Uzbekistan, in 2017. He is currently pursuing the M.S. degree in computer engineering at Gachon University, South Korea. His research interest includes the development of data processing

and analysis techniques, deep neural networks and machine learning algorithms.