J Inf Process Syst, Vol.16, No.4, pp.809~819, August 2020
https://doi.org/10.3745/JIPS.02.0137

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Voting and Ensemble Schemes Based on CNN Models for Photo-Based Gender Prediction

Kyoungson Jhang*

Abstract

Gender prediction accuracy increases as convolutional neural network (CNN) architecture evolves. This paper compares voting and ensemble schemes to utilize the already trained five CNN models to further improve gender prediction accuracy. The majority voting usually requires odd-numbered models while the proposed softmax-based voting can utilize any number of models to improve accuracy. The ensemble of CNN models combined with one more fully-connected layer requires further tuning or training of the models combined. With experiments, it is observed that the voting or ensemble of CNN models leads to further improvement of gender prediction accuracy and that especially softmax-based voters always show better gender prediction accuracy than majority voters. Also, compared with softmax-based voters, ensemble models show a slightly better or similar accuracy with added training of the combined CNN models. Softmax-based voting can be a fast and efficient way to get better accuracy without further training since the selection of the top accuracy models among available CNN pre-trained models usually leads to similar accuracy to that of the corresponding ensemble models.

Keywords

Majority Voting, Softmax-based Voting, Ensemble Scheme, Gender Prediction, CNN models

1. Introduction

Photo-based age/gender prediction systems are used in commercial terminals and kiosks to provide age/gender-appropriate ads. As diverse and efficient convolutional neural network (CNN) models have been developed, gender prediction performance also improves significantly. For example, Adience dataset [1], which is the most challenging dataset in recent years, has a prediction performance of around 90% [2-5]. Commercial sighthound APIs [3] and Microsoft APIs also show slightly higher than 90% gender prediction accuracy [3]. In terms of gender prediction accuracy, AlexNet [6], Caffe reference model [7], GoogLeNet [8], and VGG-16 [9] are compared with the several options such as weight initializations with ImageNet data [10] or IMDB-WIKI [11] data, face alignment methods, etc. [4]. VGG-16 initialized with ImageNet data showed the best accuracy, i.e., more than 92% gender prediction accuracy [4].

Voting is one of the classifier ensemble methods and is also commonly used in improving hardware reliability such as in triple module redundancy (TMR). In this paper, majority voting is employed as a

Manuscript received August 27, 2019; first revision April 9, 2020; accepted April 14, 2020.

Corresponding Author: Kyoungson Jhang (sun@cnu.ac.kr)
* Dept. of Computer Science and Engineering, Chungnam National University, Daejeon, Korea (sun@cnu.ac.kr)

method to improve gender prediction performance by using well developed CNN models. The majority voting generally takes an odd number of models as input and outputs the voting result of the models. In the majority voting, the required number of models should be odd. The classifier output of CNN models is not suitable as inputs for voters since the output is not normalized. The proposed softmax-based voter utilizes the outputs of CNN models converted by softmax function [12]. The softmax-based voting eliminates the need to make the required number of models odd, because it is based on the addition of the softmax outputs of CNN models. Also, its prediction accuracy can be gradually improved by adding as many similar accuracy CNN models as possible. Voting does not require further training while the ensemble method necessitates further training, since the ensemble model usually adds the final fully-connected layers whose weights have not yet been determined. With experiments, it is observed that the softmax-based voting always shows better accuracy than that of the corresponding majority voting. Besides, the softmax-based voting shows similar accuracy to that of the corresponding ensemble model which requires further training. As the number of combined models increases, it appears that the fine-tuning process of ensemble models leads to a little better average accuracy compared with that of the softmax-based voting. However, it is notable that the softmax-based voter is a fast and efficient way to construct a similar accuracy model with pre-trained CNN models without further training process.

In the following section, several CNN models are briefly summarized. Section 3 introduces the majority voting, the proposed softmax-based voting schemes, and the ensemble of CNN models used in the gender prediction experiment. Section 4 describes and analyzes the gender prediction accuracy experiments for three schemes with two, three, four, and five CNN models. Summary and future works are given in the last section.

2. CNN Models used for Gender Prediction

2.1 VGG16 and VGG19

In ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) [10], VGG architecture has shown good performance. Numbers 16 and 19 after VGG indicate the number of convolutional layers [9] in the corresponding architecture. The architecture uses mainly 3×3 convolution (Conv.) filters and consists of five consecutive blocks followed by a classifier layer composed of fully-connected layers. Each block ends with a max-pooling layer [13]. The summary of VGG architecture is shown in Table 1. VGG16 and VGG19 have as "block1" commonly two sub-blocks composed of 64 3×3 Conv. filters. After "block5", there is a classifier block composed of three fully-connected layers.

Table 1. VGG architecture [9]

	VGG16	VGG19
block1	conv3-64, ×2	conv3-64, ×2
block2	conv3-126, ×2	conv3-126, ×2
block3	conv3-256, ×3	conv3-256, ×4
block4	conv3-512, ×3	conv3-512, ×4
block5	conv3-512, ×3	conv3-512, ×4

2.2 Residual Nets

Two residual nets, ResNet50 [14] and ResNet152 [14], are employed in the gender prediction experiment. As the CNN layer deepens, there happens a problem that the training error and the test error become rather large. As one way to solve this problem, they suggested residual net with skip connection as shown in Fig. 1. When such a connection is used, the vanishing gradient problem caused by making the CNN layer deep can be solved. The number after ResNet indicates the number of convolutional layers. The residual net consists of five consecutive blocks, where block1 is a 7×7 Conv. filter and the following blocks from "block2" to "block5" use 1×1 and 3×3 Conv. filters. The structure of such blocks in ResNet50 and ResNet152 can be summarized as shown in Table 2.

For example, we can see that the 2nd block is composed of three sub-blocks with 62 1×1 Conv. filters, 64 3×3 Conv. filters, and 256 1×1 Conv. filters. Following the 5th block, global average pooling [15] and fully-connected layer are usually used for classification.

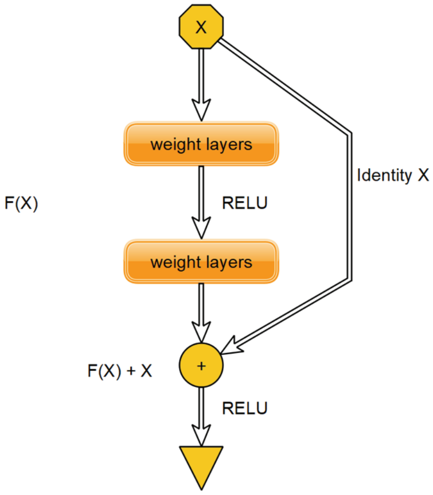


Fig. 1. The skip connection of residual net. Adapted from He et al. [14].

Table 2. The residual net architecture summary [14]

	Common	ResNet50	ResNet152
block2	[1×1(64), 3×3(64), 1×1(256)]	×3	×3
block3	[1×1(128), 3×3(128), 1×1(512)]	×4	×8
block4	[1×1(256), 3×3(256), 1×1(1024)]	×6	×36
block5	[1×1(512), 3×3(512), 1×1(2048)]	×3	×3

2.3 Densenet161

While the residual net accepts only the feature map of the previous block via skip connection, DenseNet uses dense blocks as the main component where each layer accepts all the feature maps of the previous layers [16]. Within the dense block, the *k*-th layer takes the feature maps of all preceding layers, $x_0, \dots, x_{k-1}$, as input, i.e., $x_k = F([x_0, \dots, x_{k-1}])$, where $[x_0, \dots, x_{k-1}]$ is the concatenation of all the previous feature maps [16].

DenseNet consists of the first convolution layer that takes input, followed by 4 dense blocks and the final classification layer. The transition layer composed of 1×1 Conv. layer and 2×2 average pooling layer is located between two adjacent dense blocks. Each dense block has several sub-blocks or layers with a 1×1 Conv. layer and a 3×3 Conv. layer. There are dense connections aforementioned among such sub-blocks in a dense block.

3. Voting and Ensemble Scheme for Gender Prediction

3.1 The Majority Voting of CNN Models

Gender prediction is a binary class classification problem which allows the application of the majority voter for the enhancement of accuracy based on the well-developed CNN binary classifiers mentioned in the previous section. The majority voter usually has the configuration shown in Fig. 2. As shown in Fig. 2, the voter consists of only the addition of “argmax” [17] outputs of three models followed by a comparison. The comparison is to check whether the value is greater than or equal to two. One limitation of the majority voting is that only an odd number of models can be used to avoid the tie-breaking problem.

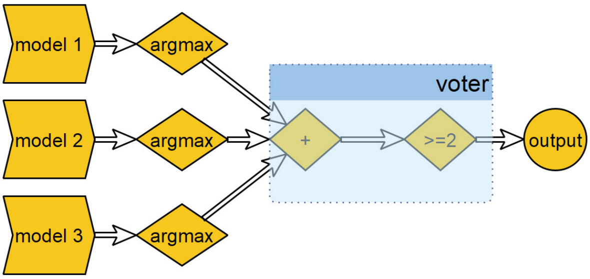


Fig. 2. The majority voting of three binary classifier models.

3.2 The Proposed Softmax-based Voting

Generally, voting schemes such as max, average, and weighted average utilize the “argmax” of the outputs of the final fully-connected layer of CNN models as inputs to the voter such as in Fig. 2 [18]. However, the “argmax” output forgets the overall tendency of outputs toward each class. The output of the final fully-connected layer retains the tendency but the value of each class output is not normalized so that it is not suitable as input to voters. In this sense, we propose a voting scheme, called softmax-based voting, using normalized outputs of CNN models and not imposing restrictions on the number of inputs of voters.

CNN classifier output can be easily converted into softmax [12] style output using softmax function or normalized exponential function. The softmax output not only indicates the class with the greatest tendency corresponding to the input but also shows the tendency toward other classes as probability values whose sum is one. These normalized probability values of the softmax output can be combined to construct a more accurate voter using outputs of CNN models. For example, Fig. 3 illustrates a classifier based on softmax-based voting of two CNN models. Unlike the majority voter, the softmax-based voter is not limited in the number of models used since it is based on the addition of softmax outputs of CNN models. In other words, any number of models can be used to build a better classifier. With experiments, the softmax-based voter is compared with the majority voter and the ensemble scheme in terms of the gender prediction accuracy while increasing the number of CNN models used.

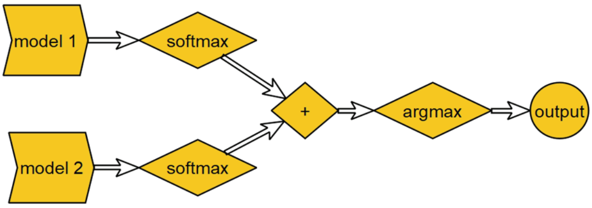
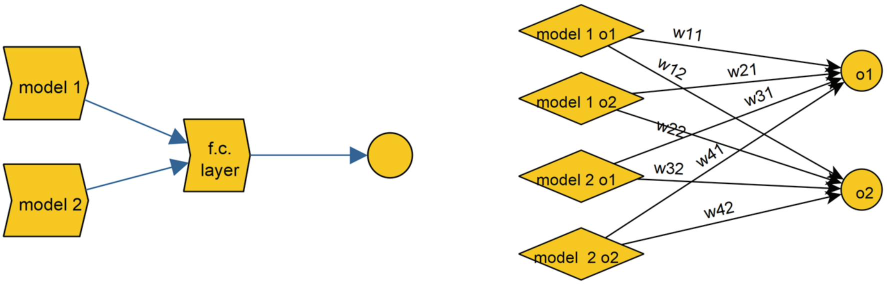


Fig. 3. Softmax-based voting of two models.

3.3 The Ensemble of CNN Models

The ensemble of CNN models usually combines already trained models with added fully-connected layers as shown in Fig. 4(a). Each model has two outputs and each output is connected to the final output o1 and o2 with weighted connections. The weights are determined with further training of the ensemble model. Weights of each model are fixed during the additional training. Just like softmax-based voting, the ensemble scheme does not restrict the number of models. During experiments, the numbers of models combined for the ensemble are 2, 3, 4, and 5. For example, the ensemble of two models with a final fully-connected layer is illustrated in Fig 4(a). The internal structure of the fully-connected layer is shown in Fig. 4(b).



(a) (b)

Fig. 4. Ensemble of models with a fully-connected layer. (a) The ensemble of two CNN models. (b) A fully-connected layer structure.

4. Experiments

Adience dataset contains face photos and two kinds of labels, i.e., age group and gender, corresponding to each photo. For the gender prediction experiment, only gender labels are used. The statistics of Adience data set for gender prediction are shown in Table 3. Normally, the 5-fold cross-validation technique [19,20] is used to measure the prediction accuracy as an average value for five cases. However, the experiment aims to test or to show whether the voter or the ensemble method improves the prediction accuracy, rather than to make an accurate comparison with the other methods. Therefore, only one case is used in this experiment as shown in Table 3.

Table 3. The statistics of Adience dataset used in the experiment

	Female	Male	Total
Train	7,424	6,073	13,497
Test	1,948	2,047	3,995
Total	9,372	8,120	17,492

As mentioned before, ResNet50, ResNet152, DenseNet, VGG16, VGG19 are used as CNN models for the experiment. As a deep learning platform, PyTorch is used since it provides the aforementioned CNN models and their pre-trained weights for ImageNet challenge. Stochastic gradient descent optimizer is employed for model training. The initial learning rate and the momentum are 0.01 and 0.9, respectively. In every 7 epochs, the learning rate is decayed by multiplying 0.1. CNN models are initialized with pre-trained weights. CNN models used in the experiment are obtained by training for 150 epochs. During training, data augmentation techniques are applied such as random resized crop, random horizontal flip and color jitter of hue, saturation, brightness, and contrast within 20%.

Accuracies of five CNN models are shown in Table 4. They are obtained with averaging the accuracies of 20 repetitions of test data evaluation. The numbers are standard deviations. The two best CNN models are "vgg19" and "res152".

In Table 4, the 2st column contains the model name and its corresponding number. Voter or ensemble model is named with voter/ensemble type and model numbers composing the voter or ensemble model. For example, "v012" means the majority voter composed of three models "0" (dense), "1" (res50), and "2" (vgg19).

Table 4 also shows accuracies and standard deviations of majority voters. Given five CNN models, the experiment is performed on 10 three-model majority voters and one five-model majority voter. Among three-model voters, "v024" and "v124" show accuracies greater than 95%. The five-model voter "v01234" show similar accuracy to "v024".

Table 4. Experimental result of CNN models and majority voters

		Accuracy (%)	Standard deviation
Model	dense(0)	93.96	0.24
	res50(1)	93.85	0.16
	vgg19(2)	94.28	0.10
	vgg16(3)	93.18	0.11
	res152(4)	94.11	0.20
Majority voters	v012	94.93	0.16
	v013	94.61	0.13
	v014	94.80	0.15
	v023	94.55	0.11
	v024	95.14	0.17
	v034	94.88	0.15
	v123	94.69	0.12
	v124	95.05	0.14
	v134	94.80	0.14
	v234	94.75	0.11
	v01234	95.14	0.15

The experimental result with softmax-based voters is shown in Table 5. In the case of softmax-based voters, the experiment is performed on 10 two-model voters, 10 three-model voters, 5 four-model voters, and one five-model voter. In the case of two-model voters, four models such as "sv02", "sv04", "sv24", and "sv34" show accuracies greater than 95%. The best of them is "sv24", e.g. the softmax-based voter with two best accuracy CNN models "vgg16" and "res152". Among 10 three-model voters, only three models "sv013", "sv023", "sv123" show accuracies less than 95%. The model "3" seems to contribute to the reduction of accuracy. As shown in Table 4, the CNN model "3", i.e., "vgg16", has the least accuracy among five CNN models. The best two softmax-based voters are "sv024" and "sv124" without the CNN model "3" and with the best two CNN models "2" and "4". All the four-model voters have accuracies greater than 95%. The best one is "sv0124". The unique five-model voter (sv01234) accuracy is slightly less than the accuracy of the best four-model voter, i.e., "sv0124". Also, in this case, the CNN model "3" appears to be the source of accuracy decrement.

Table 5. Experimental result of softmax-based voters

	Accuracy (%)	Standard deviation
Two-model voters		
sv01	94.65	0.18
sv02	95.13	0.07
sv03	94.74	0.13
sv04	94.92	0.13
sv12	94.98	0.10
sv13	94.48	0.12
sv14	94.81	0.19
sv23	94.22	0.08
sv24	95.40	0.14
sv34	95.13	0.15
Average	94.85	0.13
Three-model voters		
sv012	95.17	0.13
sv013	94.86	0.14
sv014	95.07	0.17
sv023	94.72	0.10
sv024	95.38	0.13
sv034	95.08	0.13
sv123	94.70	0.10
sv124	95.33	0.11
sv134	95.11	0.11
sv234	94.89	0.14
Average	95.03	0.12
Four- and five-model voters		
sv0123	95.18	0.13
sv0124	95.43	0.14
sv0134	95.19	0.11
sv0234	95.29	0.10
sv1234	95.28	0.10
sv01234	95.35	0.13
Average	95.29	0.12

Table 6. Experimental result of ensemble voters

	Accuracy (%)	Standard deviation
Two-model voters		
fc01	94.69	0.10
fc02	95.13	0.12
fc03	94.81	0.10
fc04	95.14	0.13
fc12	94.92	0.09
fc13	94.41	0.08
fc14	94.73	0.12
fc23	93.99	0.07
fc24	95.23	0.12
fc34	95.13	0.12
Average	94.82	0.10
Three-model voters		
fc012	95.28	0.08
fc013	94.86	0.15
fc014	95.07	0.15
fc023	95.11	0.08
fc024	95.44	0.13
fc034	95.24	0.13
fc123	94.79	0.07
fc124	95.30	0.11
fc134	95.14	0.17
fc234	95.27	0.15
Average	95.15	0.12
Four- and five-model voters		
fc0123	95.17	0.11
fc0124	95.47	0.16
fc0134	95.41	0.10
fc0234	95.51	0.10
fc1234	95.42	0.11
fc01234	95.45	0.09
Average	95.41	0.11

Ensemble models require further training to get better accuracy. The experiment is performed to find out how many epochs are appropriate for the fine-tuning process. It is observed that 20 epochs of fine-tuning are enough to get better accuracy and 40 or 100 epochs often lead to worse accuracies.

Accuracies and standard deviations of ensemble models are shown in Table 6. Just like softmax-based voters, the ensemble method has 10 models composed of two CNN models, another 10 models composed of three CNN models, five models composed of four CNN models, and one model composed of five CNN models. In the ensemble method, it is noticed a similar tendency as in softmax-based voting. Ensemble models composed of two CNN models with greater accuracy than 95% are “fc02”, “fc04”, “fc24”, and “fc34”. The best of them is “fc24”, but its accuracy is slightly less than ‘sv24’. Among the three-model ensemble, only “fc013” and “fc123” show accuracies less than 95%. The best accuracy is observed in “fc024” among three-model ensembles. Just like as in the softmax-based voter, all four-model ensembles show accuracy greater than 95%. The best of them is “fc0234”. Though “fc0234” contains the worst accuracy CNN model “3”, it seems to reach the best accuracy with the fine-tuning training process for 20 epochs. Besides, it is notable that the model “fc01234” has the second-best accuracy though it contains the worst CNN model “3”.

It is necessary to compare the softmax-based voting and the ensemble scheme since each case has the corresponding case in the other scheme. The experimental results of models composed of two CNN models are shown in Tables 5 and 6. In this two-model case, we cannot conclude the ensemble scheme is better than the softmax-based voting since seven cases “02”, “12”, “13”, “14”, “23”, “24”, and “34” among ten cases the softmax-based voting shows the same or better accuracy than the ensemble scheme. It is notable that the best two-model softmax-based voter “sv24” shows similar accuracy to those of the best ensemble models “fc024”, “fc0124”, “fc0134”, and “fc01234”. The experimental results of the three-model case are shown Tables 5 and 6. In this case, the ensemble scheme shows similar or better accuracy than the softmax-based voting. In Tables 5 and 6, it is observed that the ensemble model combined with four or five models shows similar or better accuracy than the corresponding softmax-based voter. The best accuracy model is “fc0234” whose accuracy is 95.51. Though “fc0234” has the CNN model “3”, i.e., “vgg16”, it reaches to the top accuracy through the fine-tuning process.

The average accuracy of models consisting of two, three, four, and five CNN models are shown in Tables 5 and 6. Though the softmax-based voting is somewhat better in the two-model case, the ensemble scheme is better in three or more model cases. It appears that as the number of models combined increases, the accuracy improvement by fine-tuning becomes evident in the ensemble scheme. However, the softmax-based voter may be a fast and effective way to obtain a better accuracy by selecting and combing easily the best accuracy two or more CNN models. As shown in Table 5, the best accuracy model composed of two, three, and four CNN models are “fc24”, “fc024”, “fc0124”, and their accuracies are over 95%. Notably, the model “fc01234” made with the addition of the model “3” with the worst accuracy to the model “fc0124” with accuracy 95.43% has a slightly lower accuracy 95.35% than the model “fc0124”. That is, the addition of inferior models may lead to inferior accuracy than before.

With the observation of experimental results, it seems to be worthy to suggest the rule of constructing better softmax-based voters as follows.

(1) Construct a softmax-based voter SV_2 composed of two CNN models with the addition of the second-best CNN models to the best accuracy CNN model SV_1.

(2) Add the third-best CNN model to SV_2 to construct the best accuracy softmax-based voter SV_3 composed of three models.

(3) Continue to add the next best CNN models to SV_{k-1} to construct the best accuracy softmax-based voter SV_k composed of k CNN models.

(4) In the end, we can select the best accuracy softmax-based voter among those models SV_2, SV_3, ..., SV_k.

In the case of the ensemble scheme, the above rule can be a guideline to get a better model. However, the fine-tuning can make the model combined with a somewhat inferior CNN model have better accuracy than the model combined with superior accuracy CNN model. Note the accuracies and the standard deviations of two ensemble models “fc0124” and “fc0234” illustrated in Table 4. The model “fc0234” has better accuracy and standard deviation than the model “fc0124” even though the accuracy of the model “1” is better than that of model “3” as shown in Table 4. Therefore, it is necessary to evaluate alternative models one by one to construct the best accuracy model composed of (k+1) CNN models from the best accuracy model composed of k CNN models by the addition of one CNN model.

5. Summary and Future Works

This paper deals with the accuracy enhancement of photo-based gender prediction utilizing pre-trained CNN models. There have been presented and compared three approaches such as majority voting, softmax-based voting, and ensemble scheme. Models of each approach are constructed on pre-trained CNN models. The first two approaches do not require further training while the last approach requires fine-tuning. With the experiment, it is shown that three approaches helped to achieve better accuracy than the accuracies of the constituent CNN models. Also, it is notable that the softmax-based voter always shows better accuracy than the majority voter does though they consist of the same CNN models. The ensemble scheme shows similar or slightly better accuracy than the softmax-based voting. The softmax-based voting appears to be a fast and efficient way to obtain a better accuracy model without further training though overall it shows a little inferior performance compared with the ensemble scheme. Besides, the performance improvement by the softmax-based voting just like other voters or ensemble methods is limited within a certain range, that is, about 1%–2% in this paper.

The softmax-based voting together with the ensemble scheme needs to be applied to multi-class classification problems such as age group predictions to see the same or similar effects as in binary classification problems.

Acknowledgement

This work was supported by the research fund of Chungnam National University in Daejeon, Korea.

References

[1] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170-2179, 2014.

[2] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston, MA, 2015, pp. 34-42.

[3] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood, "DAGER: deep age, gender and emotion recognition using convolutional neural network," 2017 [Online]. https://arxiv.org/abs/1702.04280.

[4] S. Lapuschkin, A. Binder, K. R. Muller, and W. Samek, "Understanding and comparing deep neural networks for age and gender classification," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, Venice, Italy, 2017, pp. 1629-1638.

[5] K. Zhang, C. Gao, L. Guo, M. Sun, X. Yuan, T. X. Han, Z. Zhao, and B. Li, "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, pp. 22492-22503, 2017.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.

[7] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrel, "Caffe: convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM International Conference on Multimedia*, Orlando, FL, 2014, pp. 675-678.

[8] 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*, Boston, MA, 2015, pp. 1-9.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015 [Online]. Available: https://arxiv.org/abs/1409.1556.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, pp. 211-252, 2015.

[11] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision*, vol. 126, pp. 144-157, 2018.

[12] Wikipedia, "softmas function," 2020 [Online]. Available: https://en.wikipedia.org/wiki/Softmax_function.

[13] Wikipedia, "Convolutional neural network,", 2020 [Online]. Available: https://en.wikipedia.org/wiki/Convolutional_neural_network.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015 [Online]. Available: https://arxiv.org/abs/1512.03385.

[15] A. Thomas, "An introduction to Global Average Pooling in convolutional neural networks," 2019 [Online]. Available: https://adventuresinmachinelearning.com/global-average-pooling-convolutional-neural-networks/.

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, pp. 4700-4708.

[17] Wikipedia, "Arg max," 2020 [Online]. Available: https://en.wikipedia.org/wiki/Arg_max.

[18] A. Singh, "A comprehensive guide to ensemble learning (with Python codes)," 2018 [Online]. Available: https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/.

[19] R. R. Picard and R. D. Cook, "Cross-validation of regression models," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 575-583, 1984.

[20] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics Surveys*, vol. 4, pp. 40-79, 2010.

Kyoungson Jhang **https://orcid.org/0000-0001-5659-0503**

He received B.S., M.S., and Ph.D. degrees in Department of Computer Engineering from Seoul National University in 1986, 1988, and 1995, respectively. From 1996 to 2001, he has been a faculty of the computer engineering department at Hannam University. Since September 2001, he has been working as a professor for the Department of Computer Science and Engineering at Chungnam National University, Daejeon, Korea. His research focuses on image processing and computer vision to ease human-computer interactions.