

Article

Human Gender Classification Using Transfer Learning via Pareto Frontier CNN Networks

Md. Mahbubul Islam , Nusrat Tasnim and Joong-Hwan Baek *

School of Electronics and Information Engineering, Korea Aerospace University, Goyang 10540, Korea; mahbubcse@cu.ac.bd (M.M.I.); tasnim.nishu70@kau.kr (N.T.)

* Correspondence: jhbaek@kau.ac.kr

Received: 20 February 2020; Accepted: 10 April 2020; Published: 13 April 2020



Abstract: Human gender is deemed as a prime demographic trait due to its various usage in the practical domain. Human gender classification in an unconstrained environment is a sophisticated task due to large variations in the image scenarios. Due to the multifariousness of internet images, the classification accuracy suffers from traditional machine learning methods. The aim of this research is to streamline the gender classification process using the transfer learning concept. This research proposes a framework that performs automatic gender classification in unconstrained internet images deploying Pareto frontier deep learning networks; GoogleNet, SqueezeNet, and ResNet50. We analyze the experiment with three different Pareto frontier Convolutional Neural Network (CNN) models pre-trained on ImageNet. The massive experiments demonstrate that the performance of the Pareto frontier CNN networks is remarkable in the unconstrained internet image dataset as well as in the frontal images that pave the way to developing an automatic gender classification system.

Keywords: automatic gender classification; pre-trained CNN; Pareto frontier networks; transfer learning; GoogLeNet; SqueezeNet; ResNet50

1. Introduction

Human gender classification is one of the fundamental tasks in the area of computer vision, which has recently gained a lot of traction in research communities as well as industries due to its substantial role in a notable number of real-world applications, including targeted advertisement, future of retail, forensic science, vending machines, visual surveillance, human–computer interaction systems, face-based demographical research, etc. Particularly, in social interactions, different salutations and grammar rules are used for men and women. In the targeted advertisement, the billboard’s contents can be visualized based on the demographics of pedestrians. The demographic trait gender can be used as a key characteristic to perceive the shopping nature for the future of retail. However, gender classification is still a strenuous task due to various changes in viewing angles, facial expressions, extreme poses, background, resolution variations, and face image appearance. It is more challenging in unconstrained imaging conditions.

Previous works on gender classification/recognition have focused on finding good discriminative features or ‘tailored’ feature descriptors for classification [1,2]. In recent years, attribute-based methods have gained attention where distinct features were extracted for particular attributes and used to train individual support vector machines (SVM) for each attribute. Moreover, the machine learning methods leverage by the aforementioned approaches did not fully exploit the enormous number of internet images to improve classification capabilities. A few CNN (Convolutional Neural Networks)-based methods have also been applied for learning attribute-based representations in [3,4].

Recently deep neural networks, particularly CNNs, have boosted nearly all domains of computer vision. Consequently, CNNs have been widely used for gender classification. Gil Levi et al. [5]

proposed an approach for gender classification that is so far the first CNN-based approach from unconstrained images.

In this paper, Pareto frontier transfer learning networks are employed to tackle the problem of recognizing a person's gender from an image using deep CNNs. Pareto frontier networks (e.g., GoogleNet [6], SqueezeNet [7], ResNet-50 [8]) are those pre-trained deep learning networks that are not worse than another network on both accuracy and prediction time metrics. We used a very useful dataset WIKI that contains more than 60,000 unconstrained images collected from the huge IMDB-WIKI dataset [9].

In the subsequent section, we describe the related work. Afterward, the methods will be presented from a technological perspective. Then, the experiments and their results will be discussed. Finally, we will conclude our works.

2. Background and Related Work

A satisfactory amount of literature already exists on the topic of gender classification. It is quite challenging to present all previous methods into a single ubiquitous taxonomy in the current paper. In this paper, we will provide a quick cursory overview of previous gender classification approaches.

In the earlier studies, appearance-based methods are mostly exploited for the gender classification problem where features are extracted from the face and then a classification tool is used. Few researchers extracted pixel intensity values as well and then fed these values to the classifiers [10]. The most used classifier for the automatic gender classification is the support vector machine; some other classifiers were decision trees, neural networks, and AdaBoost also applied in the following works [11–14]. Geeta et al. [15] proposed a new idea in gender classification by extracting different texture features from the face images. They evaluated their model with two different dataset FEI [16] and another self-built database, and kernel-based SVM is used for the classification. One of the leading appearance-based model Active Appearance Models (AAM) is applied independently by Xu et al. [17] and by Shih [18] for the gender recognition problem.

Besides the appearance-based approach, some other methods maintain a certain geometric relationship between different face parts by building a model from facial landmark information, as is known from the geometric approach for gender classification. Some of the geometric modeling approaches are presented in the following papers [19,20]. Poggio et al. [21] and Fellous [22] calculated fifteen and twenty-four facial landmark distances from human faces to recognize the gender.

Deep convolutional neural networks showed notable performance from various image recognition problems. The CNN-based methods are applied to both feature extraction as well as a classification algorithm for the automatic gender classification [5]. Some of the previous works [23,24], employed shallow CNN architecture to train the network from scratch for the gender classification where the networks are 5–6 layers deep. Compared to the aforementioned networks, the pretrained networks; GoogleNet [6], AlexNet [25] are deeper in terms of layer that produces good results mostly on the applied cases. A hybrid system for gender and age classification was presented in [26]. However, most of these methods were evaluated on the constrained imaging conditions.

Contrary to previous approaches, our work aims for a novel application of pre-trained CNN models; GoogleNet [6], SqueezeNet [7], and ResNet-50 [8] for automatic human gender classification in an unconstrained image dataset. We validate our system with one of the publicly available unconstrained image datasets IMDB-WIKI [9]. The obtained results are very interesting and confirm the effectiveness of the system for the gender classification task.

3. Methods

3.1. Convolutional Neural Networks

Convolutional neural networks are quite different from regular neural networks. In CNN, the neurons in one layer are not mandatorily connected with the neurons of the next layer. This novelty

of CNN reduces the training time and also the network complexity. The general structure of the CNN comprises three types of layers; namely, convolution, pooling, and fully connected layers.

The convolutional layer is deemed as the prime block of CNN where a convolution operation is performed to the input by some filters known as kernels to produce the neurons output. The down sampling operation is performed in the pooling layer. Max pooling and average pooling are the most used non-linear down sampling operation. With these methods, the maximum/average is taken from the evenly distributed non-overlapping areas of the output values produced by the convolution operation. Therefore, the networks preclude from overfitting degree, decrease the size of the parameters, and reduce the computational complexity. In some cases, the dropout layers are also introduced to reduce the probabilities of network overfitting. The key function of the dropout layer is to drop neurons with a precise probability [27]. The adapting activation functions of the convolutional neural network can deal with the real domain and to some extent in the complex domain [28].

All of the neurons in the fully connected layer are completely connected with the neurons of the previous layer. The fully connected layer is full of distinctive features in respect to the number of classes [29].

3.2. Transfer Learning

Due to the vastness and design complexity of deep neural network architecture, a useful technique called transfer learning can be used for a similar kind of task. In transfer learning, the deep learning model is already trained for one task and can be retrained with relatively little labeled data related to a similar task by fine-tuning the existing layers and weights. In this paper, we employed the idea of transfer learning to retrain the existing pre-trained Pareto frontier networks for gender classification problems.

3.3. Pareto Frontier Networks

Pareto frontier networks are those pre-trained deep learning networks that are not worse than another network on both metrics; accuracy and prediction time experimented on large ImageNet dataset. Consequently, GoogLeNet, SqueezeNet, ResNet networks belong to this category, whereas AlexNet does not.

3.4. Pre-Trained Deep Learning Networks

A pre-trained network has already learned to extract powerful and informative features from the natural images and the weights already fixed for the particular application. It is useful to deploy the pre-trained networks where the dataset is limited, and the application domain is related. Moreover, training CNN from scratch needs extensive computing power as well as time. Yosinski et al. [30], claimed that weights from a distant task may achieve better performance than using randomly initialized weights.

To date, a huge amount of pre-trained CNN already exist, including GoogleNet, VGGNet, AlexNet, ResNet, etc. Some of the pre-trained networks yield very good results in several applications, such as medical data analysis and disease detection. Inspired by the notable performance, the current research investigates the best configuration of some of the Pareto frontier CNN networks for gender classification. We have chosen the algorithms among the Pareto frontier standard networks considering the network simplicity and top performance in the previous years of the ILSVRC (Imagenet Large Scale Visual Recognition Challenge) competition. We also consider the time and space complexity of the networks along with the error rate shown in the ILSVRC challenge.

The pretrained networks are modified by changing the fully connected (FC) and classification layers without changing the weights of the preliminary layers. All the weights in the FC layers were initialized with random values and stochastic gradient descent with momentum (SGDM) algorithm is used for optimization, so that convergence the neural network is faster than the conventional stochastic

gradient descent optimizer. In SGDM, the updated weight Δw is combined linearly with the gradient at each iteration.

Equation (1) depicts the mathematical notation of the SGDM optimizer:

$$w \leftarrow w - \eta \nabla Q(w_i) + \alpha \Delta w, \quad (1)$$

where η is the learning rate, $Q(w_i)$ is the objective function that we want to optimize, also termed as loss function or cost function at i th data observation, w is the parameter (i.e., weights, biases, or activations), α denotes the momentum that is a temporal element for updating the neural network parameters and Δw is essentially the last changes of parameter w .

Network generalization is a major concern when neural networks are designed and trained in real-life applications. The retraining algorithms update the network weights considering the former network knowledge and the extracted knowledge of the current input. Kwok et al. [31] include constructive or pruning techniques for adaptive design of the network architecture during training and the theoretical aspects of the network generalization. In GoogLeNet [6], conventional deep plain networks (e.g., AlexNet, VGGNet) are fine-tuned by imposing 1×1 convolution filter with ReLu that help to reduce the model size by dimensionality reduction, thereby suffering less from the overfitting problem. Meanwhile, global average pooling is employed instead of fully connected layers, thus the number of weights is remarkably reduced, which can be less prone to network overfitting [32]. SqueezeNet [7], replaces 3×3 kernel with 1×1 kernel as bottleneck layer or squeeze layer to reduce the computational complexity that is $9 \times$ less parameter than the 3×3 kernels. SqueezeNet achieves $363 \times$ reduction in model size compare to AlexNet by applying the deep compression approach provided by Han et al. [33]. ResNet (Residual Network) added a skip connection to the conventional deep learning plain networks like AlexNet [25] to get rid of the vanishing gradient problem. Since the network is very deep now, the bottleneck design of ResNet-34 added 1×1 convolution layers to the start and end of the network that can reduce the number of parameters retaining the network performance even with the network turn into a 50-layer ResNet [8]. Consequently, the network is less prone to overfitting problems.

In our proposed system, we freeze the initial layers of the network so that the frozen layers will not be updated during training with the new dataset that helps to prevent the network from overfitting. Furthermore, we perform data augmentation operations including rotation, scaling, and zooming on the training dataset to get the meaningful distinct features from the image during training. In addition to that, a color filter is also applied on the training dataset as a preemptive measure to prevent the network overfitting. Finally, the dropout layer is applied to the end of the model that deletes random samples of the activations, which also helps the network from overfitting.

GoogLeNet: In ILSVRC(ImageNet Large Scale Visual Recognition Challenge) image classification competition 2014, GoogLeNet was the winner and achieved a relatively lower error rate (6.66%) compared with VGGNet and AlexNet. A 1×1 convolution filter is used in GoogLeNet as a dimension reduction module to reduce the computation. It is 22 layers deep and has almost $9 \times$ fewer parameters than AlexNet. In Figure 1, we present the customized GoogLeNet architecture designed for our experiment by replacing the learnable and classification layer and freezing the weights of the initial layers to speed up the network training as well as preventing overfitting problems.

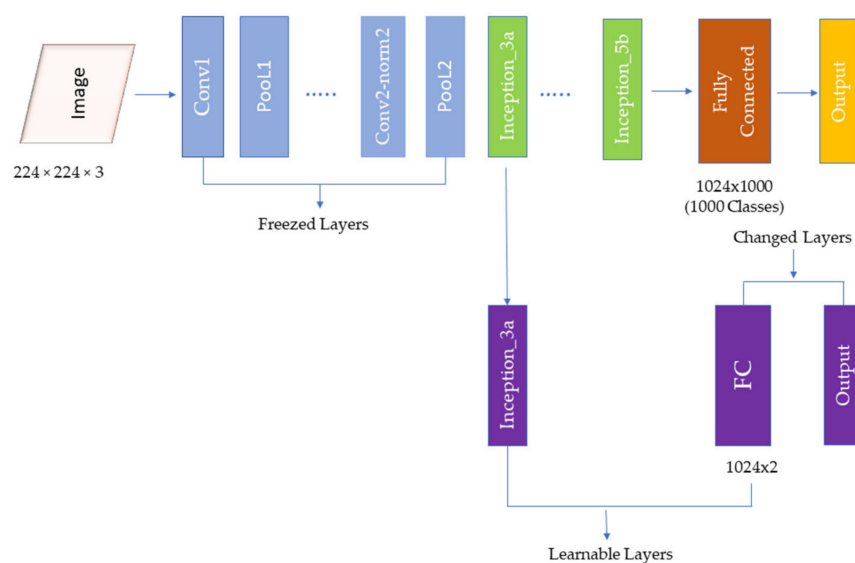


Figure 1. The customized GoogLeNet network deployed in the proposed gender classification task.

SqueezeNet: is a deep neural network designed to create a smaller network with fewer parameters maintaining the same level of accuracy with AlexNet. SqueezeNet is 18 layers deep in structure and the number of parameters is 50× fewer than AlexNet. Figure 2 shows the customized architecture of SqueezeNet.

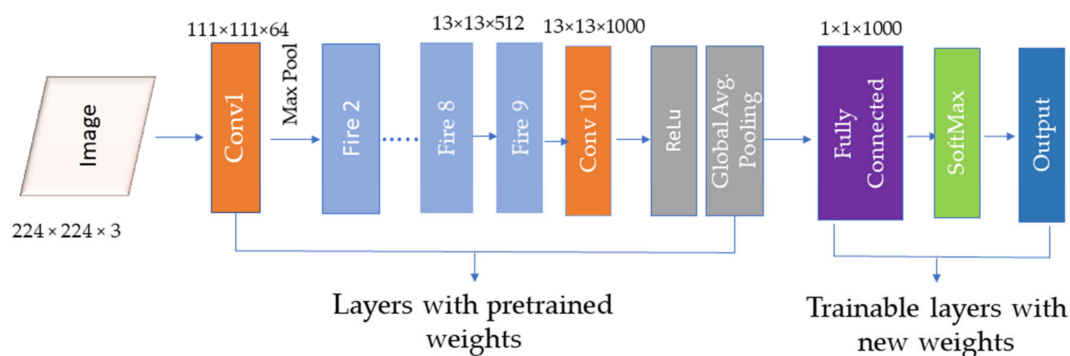


Figure 2. The customized SqueezeNet architecture deployed in the proposed gender classification task.

ResNet-50: ResNet-50 is a 50-layers deep convolutional neural network that is already trained on more than a million images. In 2015, ResNet-50 won 1st place in the ILSVRC classification competition with a top-5 error rate of 3.57%. The retrained architecture of ResNet50 shown in Figure 3.

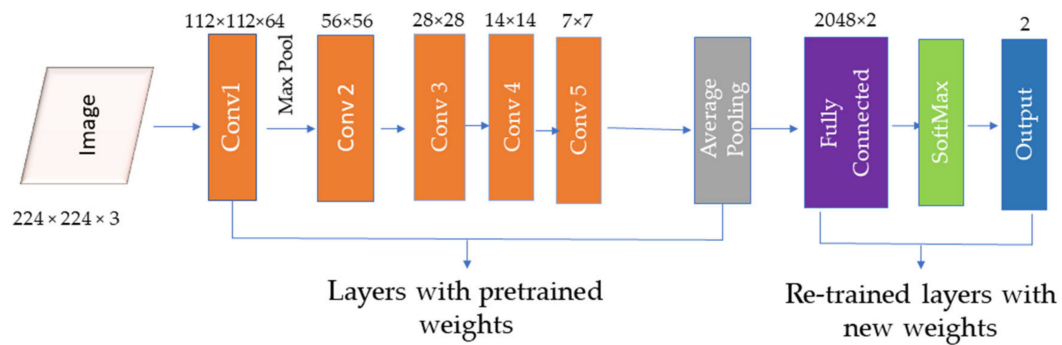


Figure 3. The customized ResNet50 architecture deployed in the proposed gender classification task.

In this research, the aforementioned pre-trained Pareto frontier networks are retrained to classify human gender by replacing the fully connected layer and the classification-output layer with two classes namely, male and female. The overall schematic diagram of our work is shown in Figure 4.

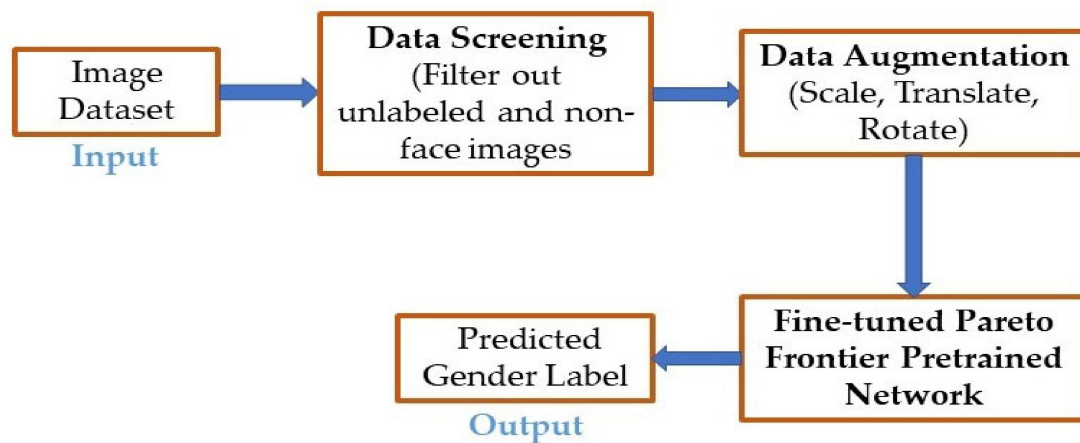


Figure 4. The schematic diagram of the proposed system for gender classification.

4. Experiments and Results

The experiments are conducted to appraise the performance of the three different CNN in the application of human gender classification on the unconstrained WIKI images from the huge IMDB-WIKI dataset.

The simulations are performed using MATLAB 2019 software. The networks were trained in a standalone system with an Intel Core i7-7700 CPU @3.60 GHz, 8 core processor. The memory of the system is 16 GB and GeForce GTX 1080 Ti version of CUDA enabled GPU is used as a parallel computing toolbox.

4.1. Training Dataset

In this work, all used CNNs have been trained on the WIKI_Cleaned dataset, which is a subgroup of the public IMDB-WIKI dataset [9]. The WIKI dataset includes images of 62,328 celebrities from different sectors including sports, politics, and the film industry. The original dataset endures from a huge number of incorrect gender annotations and non-face images. We filter out those problematic images that make our dataset fit for the experiment. Finally, the WIKI_Cleaned dataset becomes a set of about 43,000 images, which is 30% less in size than the original WIKI dataset.

4.2. Experiments

For gender classification, we validate our proposed system by randomly taking 30% (13,000 images) of the WIKI_Cleaned dataset as a testing dataset as well as the Caltech_Faces [34] dataset, which contains 453 frontal images of 27 subjects captured under different lighting conditions, expressions, and backgrounds.

We fine-tune the training options like mini-batch size, learning rate, and number of epoch for the training. We set mini-batch size to 64 for faster processing because the training data set is more than 30,000. The networks are explored with two different learning rate to evaluate the most appropriate setting, i.e., [0.0001, 0.0003]. We choose these nominal learning rate to learn faster in the new layers than transferred layers. We perform some data augmentation operation on the training images to prevent the network from overfitting. The operations include image resizing according to network input where the image sizes are varying in the dataset, color preprocessing to make distinct image channels, randomly flipping in the y-direction, translate up to 30 pixels, and 10% scaling in both directions.

4.3. Results

We have reported the classification result with the metric, accuracy (Cacc), defined as the percentage of images correctly matched with their ground truth label. In Tables 1 and 2, we summarize the evaluations of the GoogLeNet, SqueezeNet, and ResNet50 deep learning networks in the task of gender classification on the WIKI images and Caltech_Faces images. The performance of the deployed networks on WIKI_Cleaned and Caltech_Faces images are graphically presented in Figures 5 and 6. Table 3 presents the run times consumes for the experiments by the fine-tuned Pareto frontier networks. In our experiment, we observed that after screening the WIKI dataset classification, accuracy increases from 84.01% to 92.57% using GoogLeNet architecture.

Table 1. Performance of the fine-tuned Pareto Frontier networks on the WIKI_Cleaned dataset with two different learning rates (LR).

Network	Accuracy (%)	
	LR = 0.0001	LR= 0.0003
GoogLeNet	90.97	92.57
SqueezeNet	90.50	90.76
ResNet50	92.51	92.22

Table 2. Performance of the fine-tuned Pareto Frontier networks on the Caltech_faces dataset with two different learning rates (LR).

Network	Accuracy (%)	
	LR = 0.0001	LR= 0.0003
GoogLeNet	88.89	82.96
SqueezeNet	80	73.33
ResNet50	70.37	88.15

Table 3. Runtime consumption for the learning process of the deployed networks on the WIKI dataset.

Network	Runtime (Minutes)	
	LR = 0.0001	LR = 0.0003
GoogLeNet	39:15	40:12
SqueezeNet	25:48	46:58
ResNet50	89:11	158:47

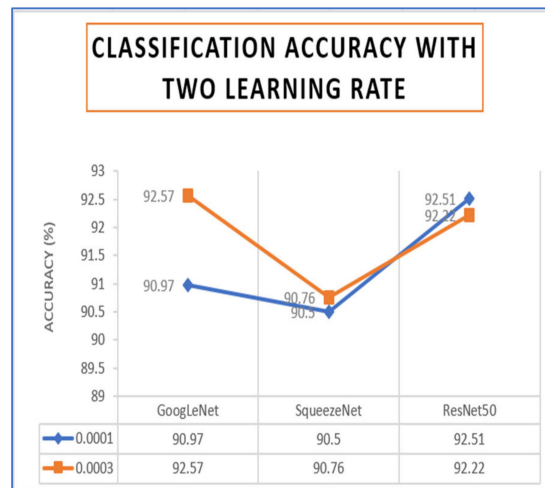


Figure 5. Overall performance of the implemented networks on the WIKI_Cleaned dataset.

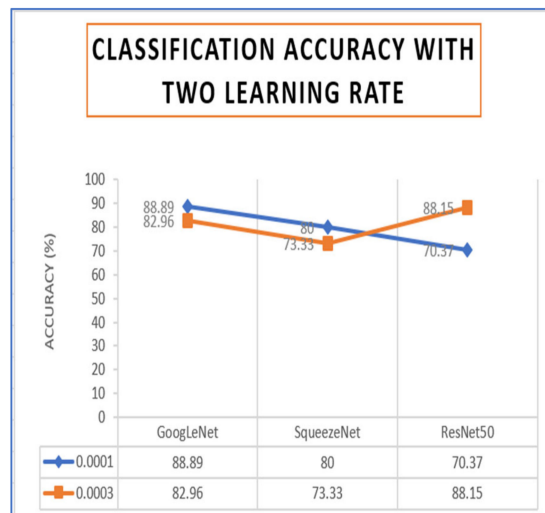


Figure 6. The overall performance of the implemented networks on the Caltech_Faces dataset.

In Figure 7, we visualize some misclassified example images where most of the mistakes were caused by blur and low-resolution images. It is also perceivable that our system wrongly classified due to extremely challenging viewing conditions of the internet images. Figures 8 and 9 visualize the performance progress during training for the best performing CNN networks.



Figure 7. Some examples of wrongly classified images from the experimented image datasets.

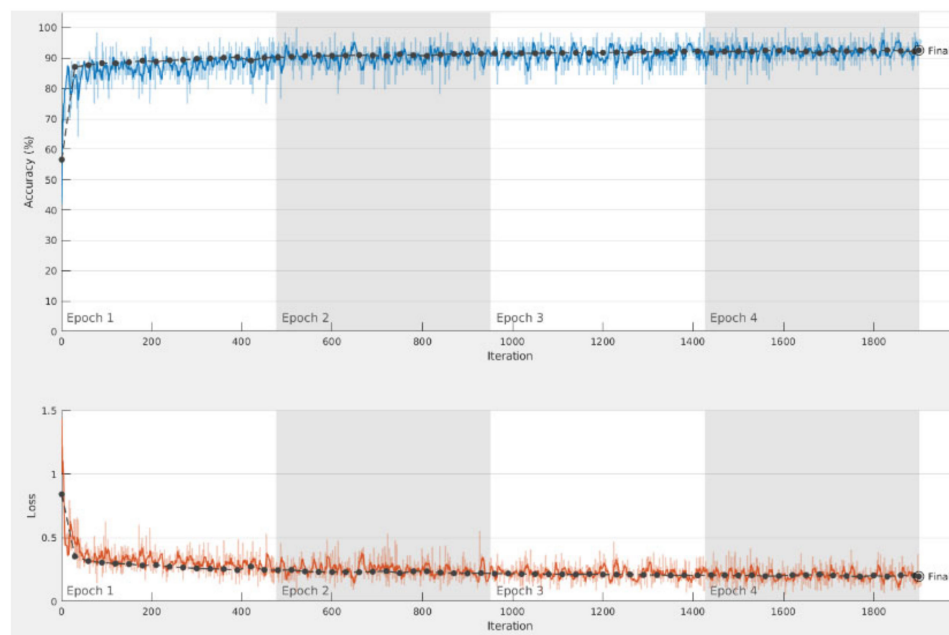


Figure 8. GoogLeNet training progress over WIKI_Cleaned dataset with learning rate 0.0003 and number of epoch 4. In the figure, upper part: training accuracy (blue line), validation accuracy (black line) lower part: training loss (red line), validation loss (black line).

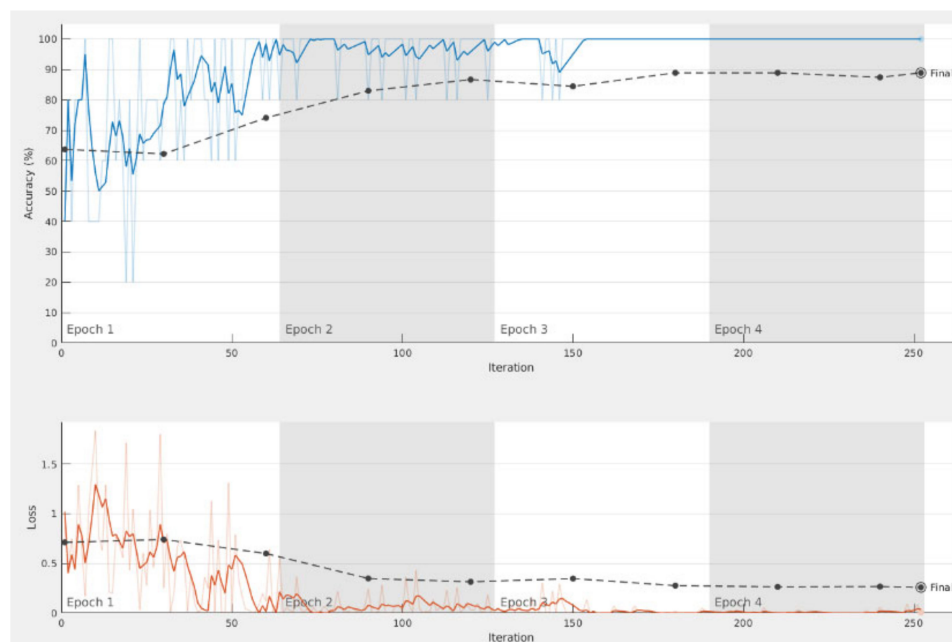


Figure 9. GoogLeNet training progress over Caltech_Faces dataset with learning rate 0.0003 and number of epoch 4. In the figure, upper part: training accuracy (blue line), validation accuracy (black line) lower part: training loss (red line), validation loss (black line).

From Tables 1 and 2, it is obvious that the performance of the networks is harmonic with the learning rates 10^{-4} and 3×10^{-4} . It is also noticeable that GoogleNet and ResNet50 networks perform better than the SqueezeNet model, whereas its runtime cost is less among the networks. There is a performance trade-off between the GoogLeNet and ResNet50 networks, where one ensures higher classification accuracy with learning rate 0.0001 and another with learning rate 0.0003 on both datasets.

5. Discussion

This work aimed to classify one of the most important human demographics (i.e., gender) deploying three different pretrained CNN architectures by following the transfer learning concept. The comparative results of these three networks are summarized in Tables 1 and 2. The last layers of GoogLeNet, SqueezeNet, and ResNet50 deep learning networks provide the necessary information to calculate the validation accuracy and losses.

The data preprocessing steps up the result almost 9% with the same parameter settings compared to the raw WIKI dataset. The details of the preprocessing are discussed in Section 4.1. Based on Table 3, the ResNet50 model takes training time more than twice compared to GoogLeNet, whereas the performance is almost the same and SqueezeNet takes less time, but performance is also poor than the other networks. This training time consumption may be caused by more layers in the ResNet50 architecture.

It is now evident from the results that all of the employed networks perform interchangeably with the two learning rates 0.0001 and 0.0003. We also found that none of them provide satisfactory results with a high learning rate 0.01, so that it is not mentioned in the Section 4.3. We observed that in case of both dataset WIKI and Caltech Faces, GoogLeNet shows the highest classification accuracy, 92.57% (WIKI _cleaned) and 88.89% (Caltech Faces) irrespective to the learning rate. We observed that the classification accuracy can be improved further by setting the network parameters, by changing the network structure, and by using more sophisticated data augmentation process in the future.

6. Conclusions

In this paper, we propose a gender classification framework deploying Pareto frontier pre-trained CNN networks with the concept of transfer learning. The novelty of this research lies in demonstrating the use of Pareto frontier pre-trained deep learning models for gender classification in the unconstrained internet image dataset and prove the concept of Pareto efficiency by their experimental results.

The experimental results observed by the deployed CNN models demonstrated their potentials in the automated analysis of face images and strengthen their use in a similar kind of classification task. Despite the heterogeneity of the WIKI images, the Pareto frontier pre-trained CNN networks, GoogLeNet, SqueezeNet, and ResNet50 demonstrated an impressive classification rate that is more than 90% with the best combination of the network parameters. We observed an unsteady classification rate (i.e., >80%) in the case of Caltech_Faces dataset due to the minimum number of labeled data.

Furthermore, this work is perhaps the maiden attempt to use the tailored Pareto frontier pre-trained CNN models for the task of gender classification in the unconstrained WIKI dataset.

Author Contributions: Conceptualization, methodology, manuscript preparation, and experiments, M.M.I.; data curation, writing—review and editing, M.M.I. and N.T.; supervision, J.-H.B.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the GRRC program of Gyeonggi province, grant number (GRRC Aviation, 2017-B04).

Acknowledgments: We would like to acknowledge Korea Aerospace University with much appreciation for its ongoing support to our research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Eiding, E.; Enbar, R.; Hassner, T. Age and gender estimation of unfiltered faces. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2170–2179. [[CrossRef](#)]
2. Liu, C.; Wechsler, H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Process.* **2002**, *11*, 467–476. [[PubMed](#)]
3. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3730–3738.

4. Zhang, N.; Paluri, M.; Ranzato, M.; Darrell, T.; Bourdev, L. PANDA: Pose aligned networks for deep attribute modeling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 1637–1644.
5. Levi, G.; Hassner, T. Age and Gender Classification using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 34–42.
6. Christian, S.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
7. Iandola, N.; Forrest, S.; Han, M.; Moskewicz, W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
9. Rothe, R.; Timofte, R.; Gool, L.V. DEX: Deep EXpectation of apparent age from a single image. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), Santiago, Chile, 7–13 December 2015.
10. Moghaddam, B.; Yang, M.H. Learning gender with support faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 707–711. [[CrossRef](#)]
11. Makinen, E.; Raisamo, R. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 541–547. [[CrossRef](#)] [[PubMed](#)]
12. Yu, S.; Tan, T.; Huang, K.; Jia, K.; Wu, X. A study on gait-based gender classification. *IEEE Trans. Image Process.* **2009**, *18*, 1905–1910. [[PubMed](#)]
13. Golomb, B.A.; Lawrence, D.T.; Sejnowski, T.J. Sexnet: A neural network identifies sex from human faces. *Adv. Neural Inf. Process. Syst.* **1990**, *3*, 572–577.
14. Baluja, S.; Rowley, H.A. Boosting sex identification performance. *Int. J. Comput. Vis.* **2006**, *71*, 111–119. [[CrossRef](#)]
15. Geetha, A.; Sundaram, M.; Vijayakumari, B. Gender classification from face images by mixing the classifier outcome of prime, distinct descriptors. *Soft Comput.* **2019**, *23*, 2525–2535. [[CrossRef](#)]
16. Centro Universitario da FEI. FEI Face Database. Available online: <http://www.fei.edu.br/~{cet}/facedatabase.Html> (accessed on 25 January 2020).
17. Xu, Z.; Lu, L.; Shi, P. A hybrid approach to gender classification from face images. In Proceedings of the 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008.
18. Shih, H.C. Robust gender classification using a precise patch histogram. *Pattern Recognit.* **2013**, *46*, 519–528. [[CrossRef](#)]
19. Matthias, D.; Juergen, G.; Gabriele, F.; Luc, V.G. Real-time facial feature detection using conditional regression forests. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2578–2585.
20. Rajeev, R.; Vishal, M.P.; Rama, C. HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 121–135.
21. Poggio, B.; Brunelli, R.; Poggio, T. HyberBF Networks for Gender Classification. 1992. Available online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.54.2814> (accessed on 13 April 2020).
22. Fellous, J.M. Gender discrimination and prediction on the basis of facial metric information. *Vis. Res.* **1997**, *37*, 1961–1973. [[CrossRef](#)]
23. Castrillón-Santana, M.; Lorenzo-Navarro, J.; Ramón-Balmaseda, E. Descriptors and regions of interest fusion for gender classification in the wild. comparison and combination with cnns. *CVPR* **2016**. [[CrossRef](#)]
24. Antipov, G.; Berrani, S.A.; Dugelay, J.L. Minimalistic cnn-based ensemble model for gender prediction from face images. *Pattern Recognit. Lett.* **2016**, *70*, 59–65. [[CrossRef](#)]
25. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 1, pp. 1097–1105. Available online: <https://dl.acm.org/doi/10.5555/2999134.2999257> (accessed on 13 April 2020).

26. Duan, M.; Li, K.; Yang, C. A hybrid deep learning CNN-ELM for age and gender classification. *Neurocomputing* **2018**, *275*, 448–461. [[CrossRef](#)]
27. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
28. Dosenovic, T.; Kopellaar, H.; Radenovic, S. On some known fixed point results in the complex domain: Survey. *Mil. Techn. Cour.* **2018**, *66*, 563–579.
29. Fan, S.; Xu, L.; Fan, Y.; Wei, K.; Li, L. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. *Phys. Med. Biol.* **2018**, *63*, 165001. [[CrossRef](#)]
30. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How Transferable are Features in Deep Neural Networks. *Adv. Neural Inf. Process. Syst.* **2014**, 3320–3328.
31. Kwok, T.-Y.; Yeung, D.-Y. Constructive Feedforward Neural Networks for Regression Problems: A Survey. *HKUST-CS95* **1995**, 1–29.
32. Lin, M.; Chen, Q.; Yan, S. Network in Network. In Proceedings of the 2nd International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
33. Han, S.; Mao, H.; William, J.D. Deep Compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *arXiv* **2015**, arXiv:1510.00149.
34. Computational Vision. Available online: <http://www.vision.caltech.edu/html-files/archive.html> (accessed on 2 January 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).