Recognition of Facial Attributes Using Multi-Task Learning of Deep Networks

Changhun Hyun and Hyeyoung Park
School of Computer Science, Kyungpook National University
Daehak-ro 80, Buk-gu, Daegu, The Republic of Korea
+82539408891, +82539507357
chhyun@knu.ac.kr, hypark@knu.ac.kr

ABSTRACT
Face recognition is one of important topics in pattern recognition field. Besides recognizing personal identity, there have been numerous studies on recognizing various facial attributes such as gender, age, race, and expression. Recently, rapid growth of deep learning techniques is leading to remarkable improvement of face recognition performances. However, facial attribute recognition is still challenging due to variety of the attributes that can be defined for human faces. As a preliminary work for efficient recognition of various facial attributes, we investigate the effect of multi-task learning of deep neural networks according to diverse combination of different attributes. Through computational experiments on recognizing six attributes by multi-task learning of convolutional neural networks, we show that the effectiveness of multi-task learning is related to the conceptual relationship among attributes, and propose a proper combination of attributes for multi-task learning of facial attribute recognition.

CCS Concepts
•Computing methodologies → neural networks.

Keywords
Facial attribute recognition; deep learning; convolutional neural networks; multi-task learning.

1. INTRODUCTION
Human facial images have been widely studied in the field of pattern recognition and computer vision. Noticeably, recent development of deep learning method achieves significant success in the problem of identity verification with facial images [1]. However, there are still many challenging issues that should be addressed in order to develop an autonomous system that can carry out natural perception of facial images just like people do when they see a picture of human face.

One of main topics that should be addressed is the recognition of various facial attributes such as expression, gender, age, and so on. The unique facial appearance of a person at a specific time is not only determined by its intrinsic position, size, and subtle angles of the facial parts (eyes, nose, and mouth), but also affected by diverse perceptual components, which can be conceptually defined such as race, hair style, skin color, and emotional status. Therefore, these facial attributes can be important factors for describing human faces.

ICMLC 2017, February 24-26, 2017, Singapore, Singapore

DOI: http://dx.doi.org/10.1145/3055635.3056618

Although there have been lots of studies on facial attribute recognition, most of conventional works have focused on a few traditional attributes: mainly age, gender, and expression [2]-[5]. Moreover, they usually took the approach to develop a well-designed individual classifier for each single specific attribute. In very recent years, however, pioneering studies have built huge facial image database with tagged labels for more than 30 attributes [6]-[7], and started to develop a deep networks for recognizing the attributes. In order to develop such a system that can recognize many attributes at the same time, multi-task learning method can play an essential role.

Multi-task learning is a machine learning technique of training multiple related pattern recognition tasks simultaneously so as to improve performance of each task [8]. Through multi-task learning techniques, it is expected to get a system capable of recognizing and analyzing various attributes at the same time more efficiently. However, the mechanism of multi-task learning has not yet been clarified, thus it is difficult to see that multi-task learning guarantees the performance improvement on all combinations of tasks. It is also not known how mutual relationship between different attributes affects the recognition performance of each task. In case of facial attribute in particular, it is essential to consider mutual effect of each attribute on the recognition performance of the other attribute, because the number of possible attributes is not limited and there could be complicated conceptual relationship among the attributes.

In the previous study [9], as a preliminary work for developing a multi-task learning system for diverse facial attribute recognition, we conducted dual-task learning of deep network for recognizing facial identity and expression, and showed the possibility of improving performance especially for expression recognition. In this paper, we further analyze the effect of multi-task learning on the recognition of six different facial attributes: identity, expression, gender, race, age, and pose. Beginning with single-task learning we compare recognition performances of multi-task learning with different combinations of the attributes. Based on the computational experiments, we try to find an appropriate learning strategy for simultaneous recognition of various facial attributes.

2. MULTI-TASK LEARNING MODEL
Multi-task learning of neural network is a performance improving technique that trains a main task and subtasks in a single network [8]. Figure 1 illustrates a structure of typical neural networks used for multi-task learning. Basically, all the tasks are conducted by a

single network. More precisely, all the nodes in input and hidden layers are commonly working for every tasks, and only output nodes are grouped and assigned to one of the tasks. Thus, in training phase, each output node just gets target value that is set for its corresponding task, and the weight update process is done by traditional error-back propagation learning algorithm. In this way, the errors for multiple tasks are reflected in the newly obtained weights in terms of improving the performance of all tasks at the same time.

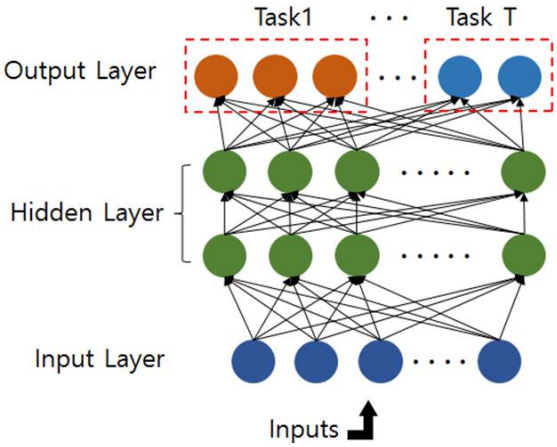


Figure 1. Network structure of multi-task learning

Ever since multi-task learning introduced, it has been applied in many fields such as video, image, and character recognition. For instance, multi-task learning was used for improving the performance of deep neural network for phoneme recognition by using suitable subtasks which are phonetically correlated: phoneme identity, subsequent acoustic states, and left and right phonetic context [10]. In the research of re-identification using multi-task learning [11], the low rank attribute embedding that learns correlation among all attributes to make use of incorrect and incomplete attributes for person re-identification was used. Both [10] and [11] mentioned that in order to improve the performance in multi-task learning, choosing appropriate subtasks is important. Based on their observations, in this paper, we try to find appropriate combination of subtasks for recognizing various facial attributes via multi-task learning. To this end, we design a deep network for recognizing multiple facial attributes and investigate the change of recognition performances according to the various combinations of the attributes.

As shown in Figure 2, we design a convolutional neural network composed of two convolutional and max pooling layers followed by fully connected multilayer perceptron (MLP) that has two hidden layers, and an output layer. The number of filter maps in convolution layer 1 and 2 are set to 64 and 32 respectively. The number of input nodes depend on input image, hidden nodes are designed as 300 that provide stable performance over all attributes through single-task learning, and the number of output node is changeable according to the task. The ReLU function is used in convolutional layers, sigmoid activation in hidden layers, and the softmax function with cross entropy error function in the output [12].

Since the proposed network is designed for learning multiple tasks at the same time, the conventional cross entropy error function for single tasks is extended for applying multiple tasks so as to obtain

$$E_{mce} = \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M_t} y_{ntm} \ln f_{tm}(x_n, \theta) \qquad (1)$$

where N is the number of training data, T is the number of tasks, and M_t is the number of classes in tth task. The f_{tm} is the value of output node corresponding mth class of tth task with an input x_n, and y_{ntm} is the target value of the node. Note that the target value is binary, and satisfy the condition:

$$\sum_{m=1}^{M_t} y_{ntm} = 1 \qquad (2)$$

for any $t = 1 \ldots T$ and $n = 1 \ldots N$. In order to make the value of network output node f_{tm} to satisfy this condition, the softmax function is applied to each group of nodes associated with tth task (t=1, … T), which can be defined as

$$f_{tm}(x_n) = \frac{e^{u_{tm}}}{\sum_{i=1}^{M_t} e^{u_{ti}}} \qquad (3)$$

where u_{tm} is the weighted sum of inputs to the output node for mth class of tth task with an input x_n.

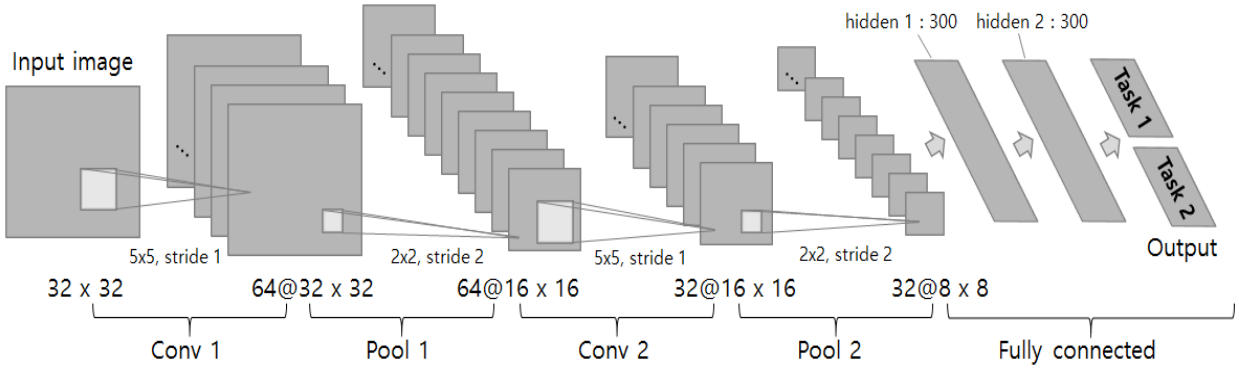


Figure 2. Structure of proposed convolutional neural network

3. EXPERIMENTAL SETUP

In order to investigate the performance of multi-task learning, we use the CMU Multi-PIE Face database [13] to train CNN. Multi-PIE database originally contains over 750,000 images of 337 subjects. The size of the image varies and each person has six facial expressions with some variations in pose, flash, and time (session). In this study, we chose 30 subjects that have all variations and normalize the size of image as 32x32. For training CNN, we added gender, race, and age labels for each image data manually (See Figure 3). Total number of data is 23,863, of which

5,086(20%) are used for training, and the remaining 18,777(80%) are used for test. Table 1 shows the componential ratio of each attribute.

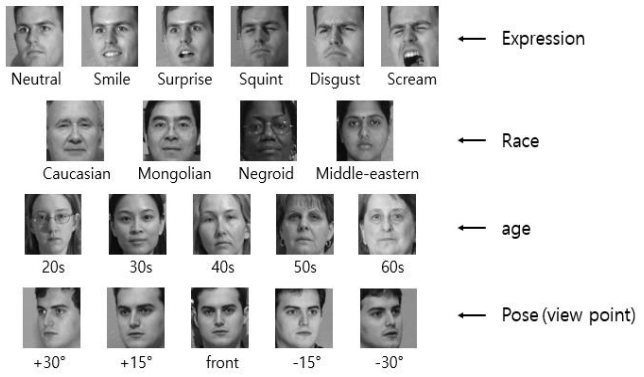
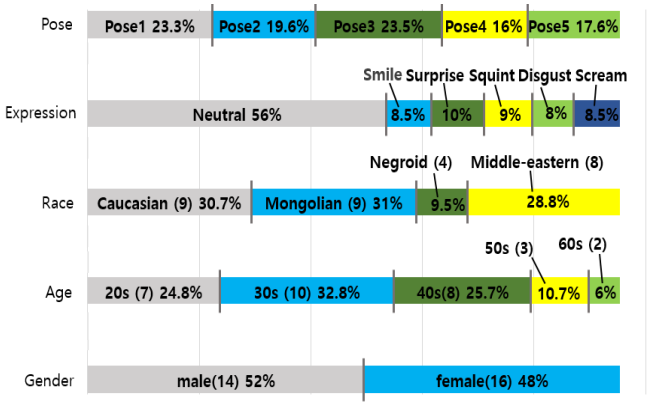


Figure 3. Facial attributes used for multi-task learning

In all experiments, we set the learning rate as 1.5, batch size of mini-batch mode as 500, 30000 epoch training, 50% of dropout, and use zero padding as default. During each training phase, we evaluated the test error at every epoch, and picked minimum test error for the performance comparison of the task. We conducted all experiments three times with random parameter initialization to obtain average results, and all the performance are measured as the misclassification ratio.

Table 1. Experimental data configuration

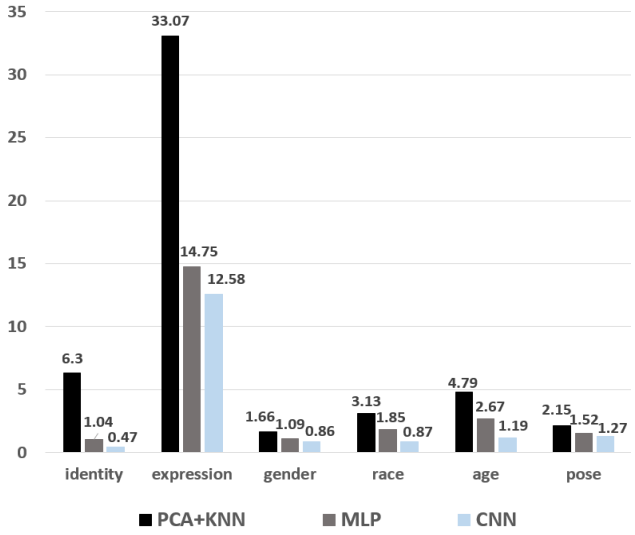


4. EXPERIMENTAL RESULTS

4.1 Single-Task Learning

Ahead of multi-task learning, we first evaluated the performance of K-Nearest Neighbor classifier with PCA features as well as the single-task learning of multi-layer perceptron (MLP) for each attribute. Table 2 shows compared the performance of single task learning with the proposed CNN with those of K-NN followed by feature extraction using PCA (316 dimension) and single-task learning with MLP. Form the table, we can see that the performance of the proposed deep network is superior to the conventional classifiers even in the case of single-task learning. This result is used as a baseline for the following multi-task learning experiments.

Table 2. Classification error on single-task classifiers (%)



4.2 Dual-Task Learning

In dual-task learning scenario, we conducted learning for all possible combination of two tasks (attributes). Table 3 present the results compared with those of single-task learning. The diagonal components of the table are the result of single-task learning, and the value in *i*th row and *j*th column represents the performance for the attribute in *i*th row in the learning of the dual-combination of the attributes in *i*th row and *j*th column. In other words, the first row indicates the performance of the identity classification when the identity attribute and the other attributes are combined. The underlined values correspond to the minimum misclassification rate for each attribute, and the shaded cells show the case of obtaining improved performance via dual-task learning.

Table 3. Classification error on dual-task learning with CNN (%)

	identity	expression	gender	race	age	pose
identity	0.47	0.48	0.50	0.60	0.45	0.51
expression	10.58	12.58	11.89	11.93	11.48	14.11
gender	0.25	0.68	0.86	0.40	0.23	0.44
race	0.30	0.77	0.54	0.87	0.37	0.66
age	0.46	1.63	0.78	1.01	1.19	1.75
pose	1.05	1.40	1.12	1.24	1.19	1.27

Through this experimental investigation, we confirmed that dual-task learning does not always give performance improvement. Furthermore, we could find some tendencies among attributes and marked it with red box. Combining identity with other attributes degrades performance of identity while performances of other attributes are improved. Gender, race, and age are always mutually synergistic in dual-task learning, not only in the three individual trials but also in the mean value table.

4.3 Multi-Task Learning

Based on the investigation results in dual-task learning, in multi-task learning, we first investigate the mutual performance improvement by combining gender, race, and age, which clearly show learning synergies in dual-task learning. Furthermore, we added facial expression and identity to gender, race, and age to find out the effect on performance when learning four attributes together. In order to investigate the performance of multi-task learning with all attributes, we conduct the 5 and 6-task learning.

Table 4. Comparison of classification error on single, multi-task learning (%)

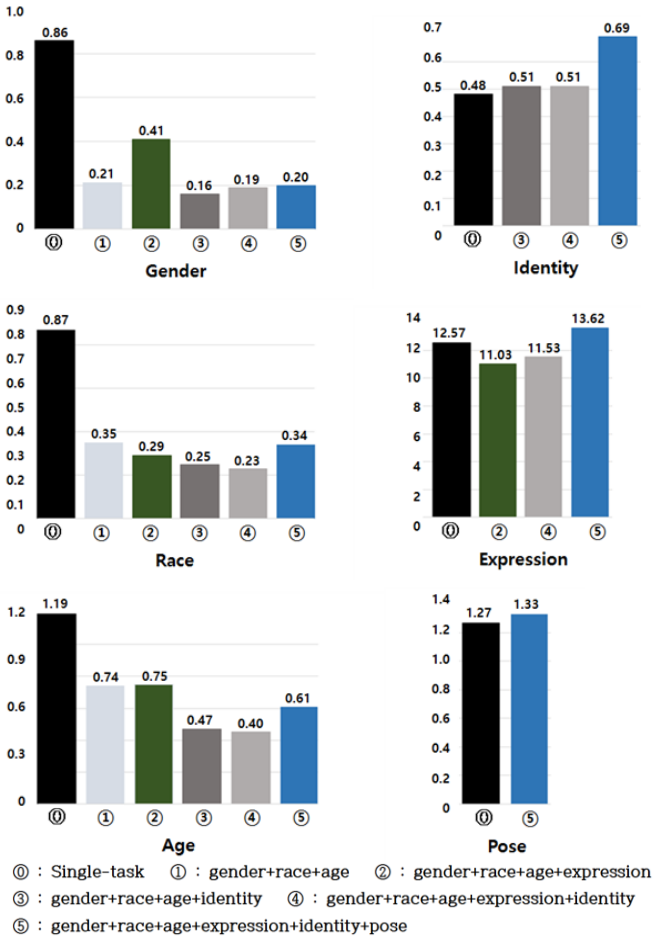


Table 4 shows the performance of the multi-task learning classifier for five different kinds of combinations, compared with the single-task learning case. In summary, the best performing combination of attributes is gender, race, age, expression and identity (experiment ④). Experiment ④ shows that overall recognition performance of all attributes is improved compared to other combinations. Except for identity and pose, the performance improvement is obtained by the multi-task learning in general. In case of gender, race, and age, not only on single-task learning, but also on dual-task, the performances are always improved by multi-task learning. In experiment ③, by training identity with gender, race, and age, performances are improved except for identity, and it accords with the tendency we found in dual-task experiment. As a result, the tendencies found in dual-task learning also correspond with multi-task learning, and it seems to be related to the conceptual relationship among attributes.

5. CONCLUSION & DISCUSSION

In this paper, we investigate the mutual effect of combination of various facial attributes on the recognition performance in multi-task learning. In our experiments, we found some tendencies: gender, race, and age have mutual synergy in multi-task learning; identity gives positive effects to other attributes while identity itself has a negative effect. Through computational experiments, we show that multi-task learning does not always result in improving performance for all attributes, and it is important to choose correlated tasks appropriately. In order to obtain positive effects through multi-task learning, it is necessary to train the neural network considering the conceptual relation among attributes. For future work, we plan to improve the performance of multi-task learning by using more various facial attributes [14], [15].

6. ACKNOWLEDGMENT

This study was partially supported by the BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005). This research was partially supported by Basic Science Research Program through the National Research Foundation of Kore (NRF) funded by the Ministry of Education (2013R1A1A2061831). This work was supported by ICT R&D program of MSIP/IITP. [R7124-16-0004, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding].

7. REFERENCES

[1] Taigman, Y., Yang, M., Ranzato, M. A., and Wolf, L. *Deepface: Closing the gap to human-level performance in face verification*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 141,2016

[2] W. Li, M Li, Z Su, Z. Zhu, *A deep-learning approach to facial expression recognition with candid images*, International Conference on Machine Vision Applications (MVA), pp. 279-282, 2015

[3] P. *Karthigayani*, S. Sridhar, *Decision tree based occlusion detection in face recognition and estimation of human age using back propagation neural network*, Journal of Computer Science, Vol. 10(1) pp. 115-127, 2014

[4] *Dehshibi*, M. M., and Bastanfard, A. *A new algorithm for age recognition from facial images*. Signal Processing, 90(8), pp. 2431-2444, 2010

[5] *Ramesha* K et al. *Feature extraction based Face Recognition, Gender and Age Classification*. International Journal on Computer Science and Engineering, Vol. 02, No.01S, pp.14-23, 2010

[6] Zhong, Y., Sullivan, J., and Li, H., *Leveraging mid-level deep representations for predicting face attributes in the wild*. Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, pp. 3239-3243

[7] Ranjan, R., Patel, V. M., and Chellappa, R., *Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition.* arXiv preprint arXiv:1603.01249, 2016

[8] R. *Caruana*, *Multitask Learning*, Machine Learning, Vol. 28(1) pp. 45-75, 1997

[9] Seo, J., Hyun, C., and Park, H., *Improving Performance of Facial Expression Recognition using Multi-task Learning of*

Neural Networks, Proceedings of the 3rd International Conference on Human-Agent Interaction, ACM, pp. 327-328, 2015

[10] Seltzer, M. L., and Droppo, J., *Multi-task learning in deep neural networks for improved phoneme recognition*. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6965-6969, 2013

[11] Su, Chi, et al. *Multi-task learning with low rank attribute embedding for person re-identification*, Proceedings of the IEEE International Conference on Computer Vision. pp. 3739-3747, 2015

[12] Dunne, R. A., and Campbell, N. A. *On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function*. Proc. 8th Aust. Conf. on the Neural Networks Melbourne, 181. Vol. 185, 1997

[13] R. *Gross*, I. Matthews, J. Cohn, T. Kanade, S. Baker, *Multi-PIE*, Image Vision Computing, Vol. 28(5) pp. 807-813, 2010

[14] Wolf, L., Hassner, T., and Taigman, Y. *Effective unconstrained face recognition by combining multiple descriptors and learned background statistics*, IEEE transactions on pattern analysis and machine intelligence, 33(10), 1978-1990, 2011

[15] *Ehrlich*, M., Shields, T. J., Almaev, T., and Amer, M. R. *Facial attributes classification using multi-task representation learning.* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 47-55, 2016