

Article

Multi-Attribute Recognition of Facial Images Considering Exclusive and Correlated Relationship Among Attributes

Changhun Hyun ¹, Jeongin Seo ¹, Kyeong Eun Lee ² and Hyeyoung Park ^{1,*}

¹ School of Computer Science and Engineering, Kyungpook National University, Daegu 702-701, Korea; chhyun@knu.ac.kr (C.H.); lain@knu.ac.kr (J.S.)

² Department of Statistics, Kyungpook National University, Daegu 702-701, Korea; artlee@knu.ac.kr

* Correspondence: hypark@knu.ac.kr

Received: 6 April 2019; Accepted: 15 May 2019; Published: 17 May 2019



Abstract: Multi-attribute recognition is one of the main topics attaining much attention in the pattern recognition field these days. The conventional approaches to multi-attribute recognition has mainly focused on developing an individual classifier for each attribute. However, due to rapid growth of deep learning techniques, multi-attribute recognition using multi-task learning enables the simultaneous recognition of more than two relevant recognition tasks through a single network. A number of studies on multi-task learning have shown that it is effective in improving recognition performance for all tasks when related tasks are learned together. However, since there are no specific criteria for determining the relationship among attributes, it is difficult and confusing to choose a good combination of tasks that have a positive impact on recognition performance. As one way to solve this problem, we propose a multi-attribute recognition method based on the novel output representations of a deep learning network which automatically learns the exclusive and joint relationship among attribute recognition tasks. We apply our proposed method to multi-attribute recognition of facial images, and confirm the effectiveness through experiments on a benchmark database.

Keywords: multi-attribute recognition; multi-task learning; facial attributes; joint probability distribution; relationship among attributes

1. Introduction

Attribute recognition, which is the problem of finding the hidden factors composing the attributes from a given input and recognizing the pattern of attributes, is one of the main topics receiving large attention in the field of pattern recognition. Along with the development of machine learning technologies, for extracting various attributes from a single dataset, the demands for recognizing multiple attributes from one input has also increased, which is called multi-attribute recognition. Though the classic approaches to attribute recognition have been applied to develop a well-designed system for a single attribute [1–6], multi-attribute recognition requires more sophisticated machine learning methods such as deep network and multi-task learning.

Multi-task learning is one of the transfer learning methods that improves recognition performance of each task by training more than two related tasks simultaneously in a single network [7]. The concept of multi-task learning is to find common features that can be beneficial to each task. Through multi-task learning, it is expected that a system capable of recognizing and analyzing various attributes is achieved, along with performance improvement and computational cost reduction. In this regard, multi-task learning plays an important role in multi-attribute recognition. However, as mentioned in [7], multi-task learning does not always guarantee the performance improvement of all tasks, and

the tasks to be learned together should be related to each other. Both [8,9] also reported that the key of improving performance of all tasks by multi-task learning is selecting tasks that are related to each other.

Many conventional studies of multi-attribute recognition [10–13] have a limitation in that they do not take into account the mutual relationship between attributes, assuming that the attributes are independent of each other. In real world problems, however, attributes are conceptually related to each other. For instance, a person wearing a skirt is more likely to be a woman than a man. This means that a task recognizing a person's attire can positively affect a gender recognition task. Thus, through applying the relationship among attributes to multi-attribute recognition, it is expected that recognition performance will improve.

There have been several studies on multi-attribute recognition using multi-task learning, but most of them have been data-dependent. The deep learning based method for recognizing pedestrian attributes introduced by Dang et al. [14] is dependent on the portion of data and does not take into account the relationship between attributes. Hand and Chellappa [15] introduced a multi-column CNN(Convolutional Neural Network) that adopt implicit and explicit attribute relationships for facial attribute classification. Though it considers the relationship between attributes, it is only applicable to specific facial attributes since it requires prior knowledge on the relationship between input area and attributes for region-based grouping.

In this paper, we propose two multi-attribute recognition methods which use a novel output representation of a deep network based on the relationship among attributes. For considering exclusive relationship among attributes, we first compose the recognition tasks by grouping the attributes that are in a mutually exclusive relationship. For example, based on the fact that the attribute of "male" is mutually exclusive to the attribute of "female", we compose a task for "gender recognition". Similarly, using the fact that identity of each subject is also exclusive to the identities of other subjects, "identity recognition" can also be composed as single task. Through the grouping process, we obtained a number of recognition tasks such as gender recognition, expression recognition, race recognition and so on. Furthermore, in an attempt to conduct all the tasks in a single deep neural network, we exploit multi-task learning techniques with specific consideration for the mutual relationship among the individual tasks. By using our proposed output representation of the deep network, we expect the network to learn the joint probability distribution among the related tasks. The proposed method is then applied to facial attribute recognition problem to check the performance of five facial attribute recognition tasks: Identity, gender, race, age, and expression on a benchmark database.

2. Multi-Attribute Recognition Using Exclusive and Correlated Relationships

In this section, we introduce the proposed multi-attribute recognition methods in sequence. At each step, a detailed explanation of a network structure for the novel output representation and a modified cross-entropy error are introduced.

2.1. Single Task Learning for Exclusive Attributes

As the first step to consider attribute relationship in multi-attribute recognition, we took an approach of grouping attributes based on their mutually exclusive relationship. Two attributes are said to be mutually exclusive when they cannot be satisfied at the same time. For example, a facial image, cannot satisfy both male and female attributes at the same time, and these two attributes are grouped together. In this manner, all the attributes were grouped into several groups, we treated each group as a single recognition task. Accordingly, the activation function of output nodes and corresponding target needed to be redefined. In this section, we describe the learning for a single task, and then extend it to multiple tasks in the next section.

When a group of M attributes A_1, A_2, \dots, A_M are composed by a mutually exclusive relationship, task T can be defined to assign each input to one of the attributes, which corresponds to one of M output nodes of the learning network. The network structure for the single task T using output representation

of an exclusive relationship is shown in Figure 1. Given an input data x_n the target output for m th output node $y_m^n (m = 1, \dots, M)$ needs to satisfy the conditions: $\sum_{m=1}^M y_m^n = 1$ and $y_m^n \in \{0, 1\}$. In order to design a network satisfying the conditions, the output value of m th output node $f_m(x_n, \theta)$ is defined by using softmax activation function, which can be written as

$$f_m(x_n) = \frac{e^{u_m}}{\sum_{i=1}^M e^{u_i}}, \tag{1}$$

where u_m denotes the weighted sum of input to the m th node. For training the network, we can use the conventional cross-entropy error function for the multi-class classification problem, which is written as

$$E_{ce}(\theta) = - \sum_{n=1}^N \sum_{m=1}^M y_m^n \ln f_m(x_n, \theta), \tag{2}$$

where N is the number of training data, and θ is the vector of all weight parameters in the network. Although this is the conventional setting for multi-class classification, it should be noted that the group of mutually exclusive attributes can only satisfy the underlying assumption, and thus our proposed grouping process is important in recognizing various attributes in a single network.

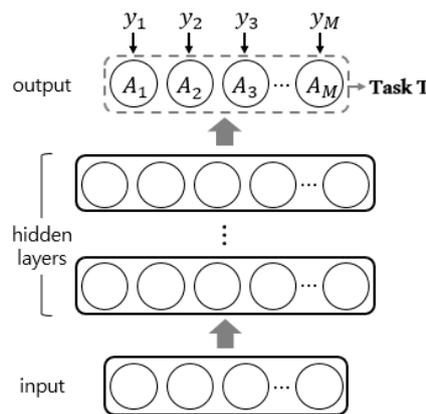


Figure 1. A network structure with single task T for recognizing M mutually exclusive attributes.

2.2. Multi-Task Learning for Independent Attributes

As the next step, in order to learn multiple tasks at the same time, the conventional cross-entropy error function for single task is extended to multiple tasks. Figure 2 illustrates the network for training two tasks at the same time. Let us assume that we have T classification tasks $T_t (t = 1 \dots T)$, and each task T_t is composed of M_t mutually exclusive attributes $A_{tm} (t = 1, \dots, T, m = 1, \dots, M_t)$. We assign one output node for each attribute so that the whole network for T tasks has $\sum_{t=1}^T M_t$ output nodes. The target value of m th output node for t th task can be denoted as y_{tm}^n and satisfies the conditions:

$$\sum_{m=1}^{M_t} y_{tm}^n = 1 (t = 1, \dots, T), y_{tm}^n \in \{0, 1\}. \tag{3}$$

In order to satisfy these conditions, the output value of the node corresponding to attribute A_{tm} is defined by using a task-wise softmax activation function such as

$$f_{tm}(x_n) = \frac{e^{u_{tm}}}{\sum_{i=1}^{M_t} e^{u_{ti}}}, \tag{4}$$

where u_{tm}^n is the weighted sum of input injected to the output node for the attribute A_{tm} when an input x_n is given. Since the softmax function is applied not to all the output nodes but to the task-wise group, we also need to modify the conventional cross-entropy error function so that summation is applied task-wisely, which can be written as

$$E_{mce}(\theta) = - \sum_{n=1}^N \sum_{t=1}^T \sum_{m=1}^{M_t} y_{tm}^n \ln f_{tm}(x_n, \theta), \tag{5}$$

where T is the number of tasks and M_t is the number of attributes in t th task.

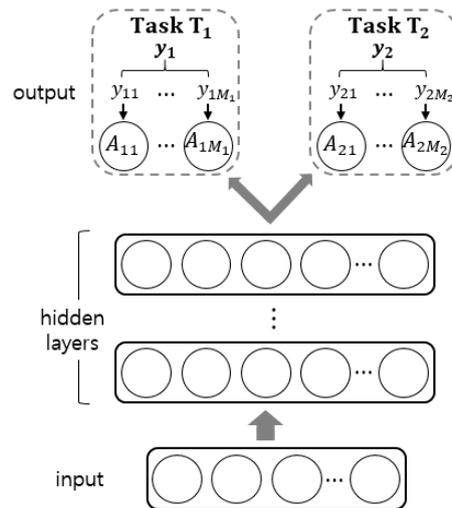


Figure 2. Extended output representation for two independent tasks T_1, T_2 .

This extension to multi-task learning from single task learning for a group of exclusive attributes is limited in the sense that it only utilizes the mutual relationship between attributes, and does not consider the relationship between tasks. The simple summation of the task-wise cross-entropy function defined by Equation (5) is derived by the assumption that the tasks are not correlated and the target random vectors $y_t = [y_{t1} \dots y_{tM_t}]$ ($t = 1 \dots T$) are mutually independent to each other. However, this strong assumption does not seem plausible in real world applications. For example, it can be easily assumed that the two random vectors of two tasks for recognizing the attributes of eye color and the attributes of hair colors are closely related note that a person with blond hair often has blue eyes. Moreover, these two recognition tasks are also associated with race recognition tasks in that Caucasians are likely to have blond hair with blue eyes. As mentioned in [7], the mutual relationship between tasks can make some influence on the performance in each task, we can expect to get better recognition performance by multi-task learning considering the relationship between tasks. Nevertheless, it is hard to define the relation in the learning model since this relationship is quite data-specific. In the next section, we propose an approach to overcome this difficulty.

2.3. Multi-Task Learning for Mutually Correlated Attributes

In order to consider mutual dependency between tasks in a multi-task learning process, we propose a novel definition of network output, which can represent joint probability of multiple random vectors. By calculating the joint probability among random variables, we can figure out the proportional or inverse relationship among variables. Let us take a simple example of dual task learning. When there are two tasks, T_1 and T_2 , with exclusive attributes $\{A_{11}, \dots, A_{1M_1}\}$ and $\{A_{21}, \dots, A_{2M_2}\}$ respectively, in which the target value is given by two binary random vectors $y_1 = [y_{11} \dots y_{1M_1}]$ and $y_2 = [y_{21} \dots y_{2M_2}]$, we try to design an output layer for representing the joint probability of the random

vectors, $P(\mathbf{y}_1, \mathbf{y}_2)$. From the fact that \mathbf{y}_1 and \mathbf{y}_2 are binary vectors in which only one element can be 1 at one time, we can define a joint random vector $\mathbf{z}_{1,2}$ that represents $M_1 \times M_2$ different combination of the values, and we can assign an output node to represent each possible combination. Thus, the proposed network has $M_1 \times M_2$ output nodes, and we denote the value of each output node as $f_{m_1 m_2}(\mathbf{x}, \boldsymbol{\theta}) (m_1 = 1, \dots, M_1, m_2 = 1, \dots, M_2)$. Accordingly, the target value $z_{m_1 m_2}$ is determined by the value of \mathbf{y}_1 and \mathbf{y}_2 such as

$$z_{m_1 m_2} = \begin{cases} 1 & \text{if } y_{1m_1} = 1, y_{2m_2} = 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Note that \mathbf{z} is the $M_1 \times M_2$ dimensional random binary vector satisfying the condition: $\sum_{j=1}^{M_2} \sum_{i=1}^{M_1} z_{ij} = 1$. In order to train this target vector efficiently, the output values, $f_{m_1 m_2} (m_1 = 1 \dots M_1, m_2 = 1 \dots M_2)$ of the network need to be defined by using a softmax function such as

$$f_{m_1 m_2}(\mathbf{x}_n) = \frac{e^{u_{m_1 m_2}}}{\sum_{j=1}^{M_2} \sum_{i=1}^{M_1} e^{u_{ij}}} \quad (7)$$

where $u_{m_1 m_2}$ is the weighted sum of input injected to the corresponding output nodes. The cross-entropy error for the proposed joint representation is then defined as

$$E_{JCE}(\boldsymbol{\theta}) = - \sum_{n=1}^N \sum_{m_2=1}^{M_2} \sum_{m_1=1}^{M_1} z_{m_1 m_2}^n \ln f_{m_1 m_2}(\mathbf{x}_n, \boldsymbol{\theta}) \quad (8)$$

This can be directly extended to the case for more than two tasks $\{T_1, \dots, T_T\}$ so as to obtain

$$E_{JCE}(\boldsymbol{\theta}) = - \sum_{n=1}^N \left\{ \sum_{m_T=1}^{M_T} \dots \sum_{m_1=1}^{M_1} z_{m_1 m_2 \dots m_T}^n \ln f_{m_1 m_2 \dots m_T}(\mathbf{x}_n, \boldsymbol{\theta}) \right\} \quad (9)$$

where the random vector \mathbf{z} with $M_1 \times \dots \times M_T$ elements is defined as

$$z_{m_1 m_2 \dots m_T} = \begin{cases} 1 & \text{if } y_{im_i} = 1 \text{ for all } i = 1 \dots T \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Figure 3 shows the network model for multi-task learning with a joint random vector. Although the proposed representation of the output node needs a greater number of output nodes than the conventional multi-task learning, it can increase the representational flexibility of the network so that it enables it to learn various joint relationships between tasks. When the training of the network is completed, the classification for each task can be done by calculating marginal probability of the obtained joint probability $f_{m_1 m_2}(\mathbf{x}, \boldsymbol{\theta})$, which can be written as

$$\begin{aligned} P(y_{1m_1} = 1) &= \sum_{m_2=1}^{M_2} f_{m_1 m_2}(\mathbf{x}, \boldsymbol{\theta}), \\ P(y_{2m_2} = 1) &= \sum_{m_1=1}^{M_1} f_{m_1 m_2}(\mathbf{x}, \boldsymbol{\theta}). \end{aligned} \quad (11)$$

Then we assign the class of the current input \mathbf{x} to the node with the maximum marginal probability for each task.

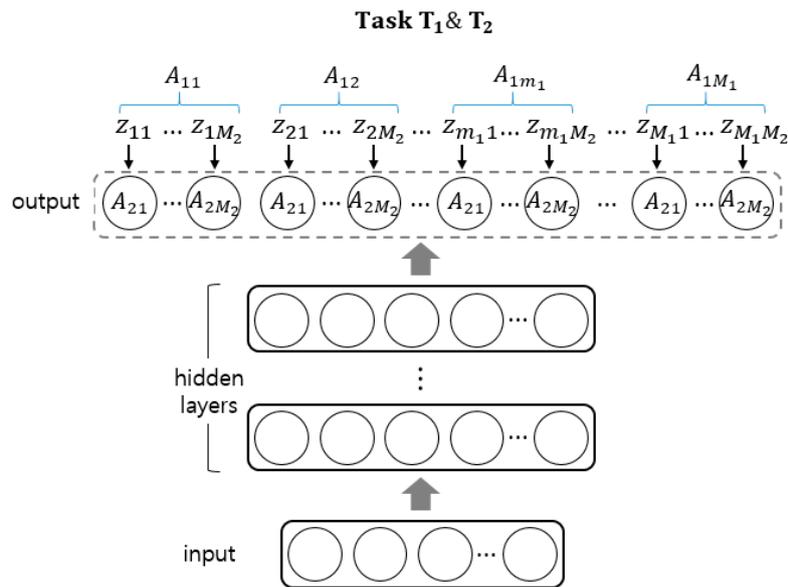


Figure 3. Network for multi-task learning for joint random vector output.

3. Multi-Attribute Recognition of Facial Images

We applied the proposed multi-attribute recognition method to the facial attribute recognition problem. We considered five facial attributes: Identity, expression, gender, race, and age, which will be explained in detail later. Figure 4 shows two network structures in case of multi-task learning of race and gender: (a) For mutually independent tasks and (b) for mutually correlated tasks.

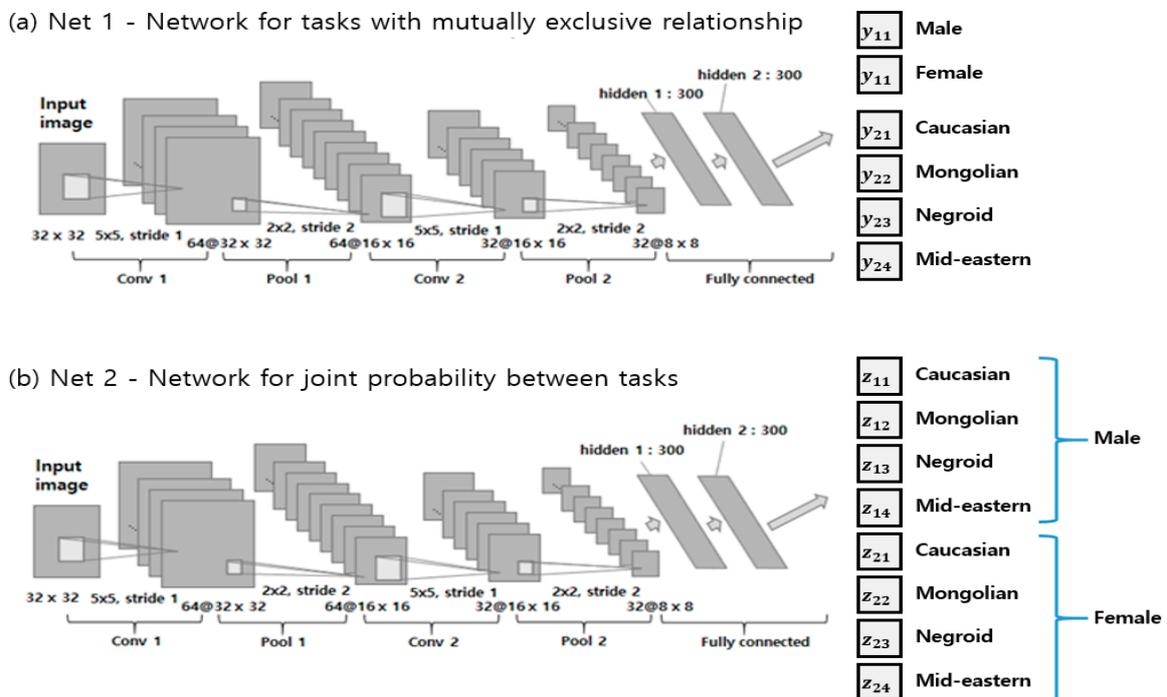


Figure 4. Overview of network structure for multi-attribute recognition with two tasks: (a) Net 1—Network for tasks with a mutually exclusive relationship, (b) Net 2—Network for the joint probability between tasks.

As shown in Figure 4, since gender has two attributes (male and female) and race has four attributes (Caucasian, Mongolian, Negroid, and Middle-eastern), in this example, the number of required output nodes are six for the network (a), which is designed for the mutually independent tasks. On the other hand, the network (b) designed for the mutually correlated tasks has eight (2×4) output nodes. For the sake of understanding the output representation for the learning of more than two tasks, Figure 5 shows the output representation for the case of three recognition tasks: Gender, race, and age. Since gender has two, race has four, and age has five attributes, 40 output nodes were required. Likewise, the number of output nodes required for learning depended on the number of tasks to be learned together.

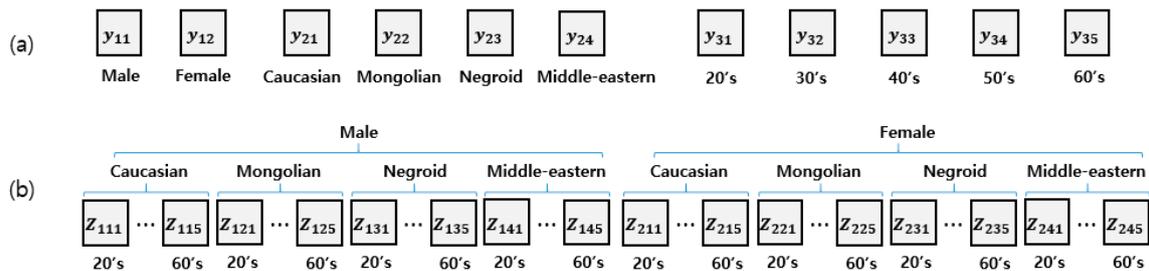


Figure 5. Output representation for three attribute recognition tasks (gender, race, and age): (a) Representation for independent tasks and (b) for correlated tasks.

We designed a convolutional neural network composed of two convolutional and max pooling layers followed by a fully connected multilayer perceptron (MLP) that had two hidden layers, and an output layer. The number of filter maps in convolution layers 1 and 2 were set to 64 and 32 respectively. The number of input nodes depended on the size of input image, hidden nodes in fully connected layers were designed as 300 which provided stable performance over all attributes through single task learning for each attribute, and the number of output node was changeable according to the task as described in detail in the following paragraph. The ReLU function was used in convolutional layers, sigmoid activation in hidden layers, and the softmax function with cross-entropy error function in the output layer [16].

4. Experimental Results

4.1. Multi-Attribute Recognition on the CMU Multi-PIE Dataset

As a benchmark database, we took CMU (Carnegie Mellon University) Multi-PIE (Pose, Illumination, and Expressions) which is a well-known dataset with facial attributes [17]. From the original data with more than 750,000 images of 337 subjects with some variations in pose, flash, and time, we selected images about 30 subjects for experiment. Other than pre-labeled identity and expression class, we manually labeled three facial attributes: Gender, race, and age for attribute recognition tasks. The total number of labeled data was 23,863, of which 5086 (20%) were used for training, and the remaining 18,777 (80%) were used for the test. The data was divided in a way that kept the composition ratio of the attributes. The table of experimental data configuration is shown in our previous work [18]. We composed five data settings for cross validation, and conducted all the experiments on the same 30,000 epochs training for three random initializations, and obtained average results. Since we used images for 30 individuals, we treated 30 identities as 30 different attributes for identity recognition. Obviously, these identities were in a mutually exclusive relationship. Similarly, we treated the six variations in facial expression as six distinct attributes. Moreover, we manually labeled two attributes concerning gender, four different attributes concerning race (Caucasian, Mongolian, Negroid, and Middle-eastern) according to the race category used in [19], as well as five attributes concerning age groups (20s, 30s, 40s, 50s, 60s). Since the size of the image in the dataset varied, we resized all data to 32×32 . Figure 6 shows the example of the Multi-PIE data used in the experiment.

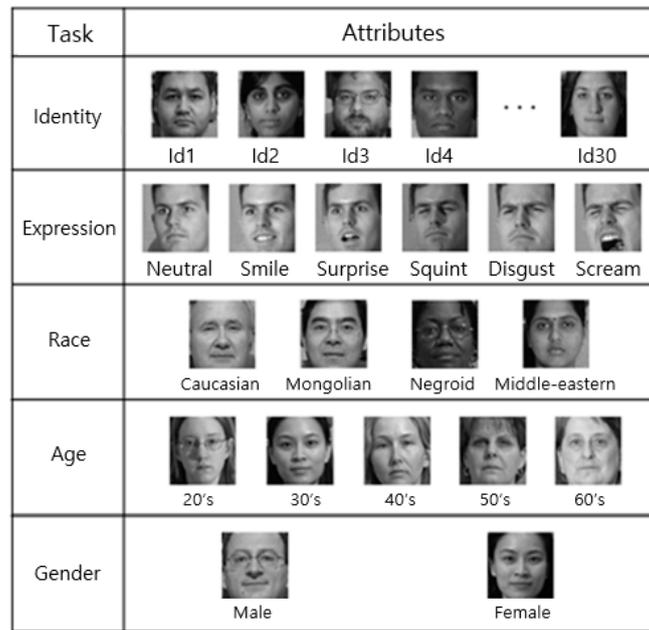


Figure 6. Examples of CMU Multi-PIE data.

We first grouped the mutually exclusive attributes into five tasks (identity, gender, race, age, and expression recognition), and conducted the single task learning for each task, so as to compare the results with the proposed multi-task learning method. We applied the proposed output representation methods for independent and correlated relationships to various combinations of two or more tasks.

For the basis of the experiment, we conducted dual task learning for all possible combinations of two recognition tasks. For each experiment, we implemented two different settings: (a) The output representation for independent tasks, and (b) for correlated tasks, of which the results are shown in the Table 1. The diagonal elements of the tables show the performance of single task learning, and the value in the *i*th row and *j*th column represents the classification error for *i*th task in the dual-combination learning of the *i*th task and *j*th task. For example, the first row indicated the error of the identity classification when the identity task was combined with other tasks. The underlined value in each row corresponds to the minimum misclassification rate for each task, and the shaded cells show the cases in which improved performance was obtained by applying multi-task learning. The values in bold imply improved performance in the comparison of two different settings (a) and (b).

Table 1. Recognition error (%) of dual task learning with independent tasks (a), and correlated tasks (b).

(a)	Identity	Expression	Gender	Race	Age
Identity	0.48	<u>0.40</u>	0.43	0.52	0.46
Expression	9.68	11.48	10.86	10.97	10.42
Gender	0.16	0.61	0.77	0.33	0.24
Race	0.27	0.67	0.48	0.78	0.34
Age	0.46	1.30	0.67	0.73	1.12
(b)	Identity	Expression	Gender	Race	Age
Identity	0.48	0.98	0.50	0.52	0.51
Expression	9.29	11.48	10.88	10.94	10.50
Gender	0.13	0.61	0.77	0.28	0.23
Race	0.23	0.83	0.47	0.78	0.36
Age	0.40	1.68	0.66	0.64	1.12

As shown in Table 1, it is difficult to say that multi-task learning can always improve the performance in all attribute recognition, which corresponds with the arguments in [7]. However, we can still see that dual task learning can improve the performance in many cases (the shaded cells), which supports the empirical efficiency of multi-task learning reported in many practical applications [8,9]. Further, we can also see that the different tendency in the performances of settings (a) and (b). The setting (a) generally gave slight improvement compared to single task learning, whereas the setting (b) showed more apparent discrepancy in the performance gains and losses. This tendency can be expected from the theoretical property of the joint representation setting, in which the learning model is free from the strong independency assumption. In the experiment of expression and other tasks, expression itself was always improved in performance from both methods, but it did not help the performance of other recognition tasks when learning a joint relationship. Thus, expression is likely to improve the performance of other tasks when learning as an independent task rather than correlated task. In the experiment of identity and other tasks, identity always helped to improve the performance of other tasks, whereas identity itself did not get a positive effect by other tasks when learning a joint relationship. This seems to be due to a lack of data compared to the increasing number of nodes required. Therefore, it seems appropriate to consider identity as an independent task in order to get the performance improvement on all tasks. In particular, we found that gender, race, and age were complementarily related to each other which is marked with a red box. Thus, we further examined how expression and identity affected the learning of multiple tasks by sequentially adding them to the multi-attribute recognition problem with three (gender, race, and age) recognition tasks so as to find an optimal combination for the multi-attribute recognition of faces.

Table 2 shows the recognition performances of five facial attributes under several output representation settings. G, R, A, E and I are simplified representations of the tasks: Gender, race, age, expression, and identity, respectively. The symbol '+' means that tasks are combined with the assumption that tasks are independent of each other. Another symbol '*' represents the proposed joint representation method for mutually correlated tasks. Bold numbers with underline denote the best performance for the specific tasks. The values in gray cells indicate the best recognition performance among the multi-attribute recognition experiments with the same type and number of tasks.

Table 2. Recognition error (%) of five facial attributes under different representation methods.

Representation Method	Identity	Expression	Gender	Race	Age
Single task learning	0.48	11.48	0.77	0.78	1.12
G+R+A	-	-	0.17	0.26	0.47
G*R*A	-	-	0.16	0.25	0.47
E+G+R+A	-	10.30	0.24	0.24	0.50
E+G*R*A	-	<u>9.51</u>	0.14	0.21	0.36
E*G*R*A	-	9.52	0.30	0.46	0.79
I+G+R+A	0.49	-	0.15	0.19	0.41
I+G*R*A	0.45	-	0.13	0.20	0.37
I+E+G+R+A	0.46	10.35	0.14	0.19	0.39
I+E+G*R*A	<u>0.36</u>	9.56	<u>0.10</u>	<u>0.17</u>	<u>0.29</u>
I+E*G*R*A	0.47	9.65	0.18	0.26	0.44

From Table 2, multi-attribute recognition using proposed output representation methods gave better performance for most of the attributes compared to those of single task learning. When training complementarily related three attributes (gender, race, and age) at the same time, the performance of all tasks was much improved compared to the result of dual task learning, and learning with a correlated relationship showed better performance than an independent relationship. In the experiment of learning expression, gender, race, and age together, the best performance was obtained when regarding

expression as an independent task and the other three attributes were correlated (E+G*R*A in Table 2), and this is consistent with what was revealed in the previous dual task (Table 1) and three task experiments. For the next step, we combined identity with gender, race, and age recognition. As with the results from Table 2, we could confirm that identity, as an independent task, helped overall recognition performance of three other recognition tasks, which is also consistent with the results in the previous experiments. The reason why we did not conduct the joint relationship learning with identity and other tasks is that the number of output nodes becomes too large when a novel output node is created for the joint representation which may lead to overfitting the problem. Lastly, we implemented multi-task learning with five tasks all together by applying the appropriate combination of the output representation methods, and confirmed that we can get a considerable performance improvement on all attributes except for expression. Through the whole experiment, we found the best output representation for multi-attribute recognition of facial images, and confirmed our proposed output representation methods can improve the performance of all recognition tasks when properly combining them.

4.2. Analysis of Toy Problem

The method of learning independent relationships was based on the assumption that tasks were independent of each other, and the method of learning joint relationships uses the strong assumption that the tasks are related to each other. For further analysis on the difference between independent representation and joint representation, we conducted a simple experiment using the MNIST (Modified National Institute of Standards and Technology) dataset. For a task that uses binary classifying numbers, we chose two numbers (five and nine) among ten digits of MNIST data, and gave binary labels 1 and 0 respectively. For another task of noise recognition, we added some gaussian noise with $\sigma = 0.2$ to original images, and assigned binary label 1 if an image was noisy, and 0 otherwise. Figure 7 shows the example images of MNIST data. In Figure 7, the images from the top and second row indicate that 50% of images are noisy for each digit, we call this dataset p50. The third and fourth rows indicate that 10% of "5" images and 90% of "9" images are noisy, which is denoted as p10. Likewise, we made 10 different datasets by changing the portion of noise image in the "5" digit class from 10% to 90%. Note that, the portion of noise image in the "9" class changes from 90% to 10%, and thus, the total ratio of noisy data is always 50% of total data. The number of training data is 10,000, and 1000 for test data for each set. Experiments were conducted on the nine datasets and compared to change of performance. The network structure consisted of two convolutional and max pooling layer followed by one fully connected layer with 50 hidden nodes. For each experiment, we set the learning rate as 0.01, batch size as 1000, and used 50 epochs training. For each dataset, the learning was done for 20 random initializations to get average performance.

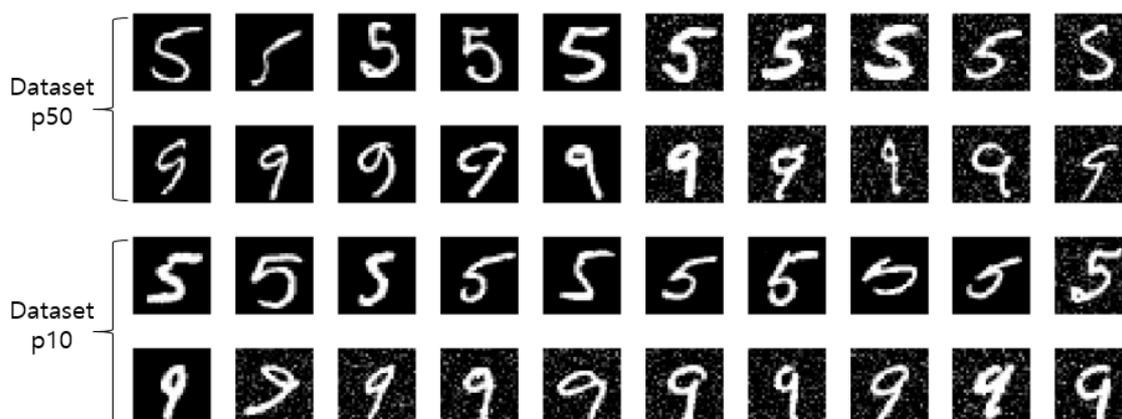


Figure 7. Examples of noisy MNIST data with addition of Gaussian noise.

Figure 8 shows the change of classification rate of digit recognition tasks. When the ratio of noisy data to original data was even (p50), which means two tasks are independent, our proposed method for correlated tasks was worse than those for independent tasks. On the contrary, when the discrepancy of noisy data portion between two digit classes became large, the joint representation showed better performance than the independent representation. We can also observe that the performance improvement in the p10 and p90 sets was less than that in the p20 and p80 sets. This phenomena may be due to the limited size of data in a specific joint class, such as “digit 5” and “noisy” class in the p10 set and “digit 9” and “noisy” class in the p90 set.

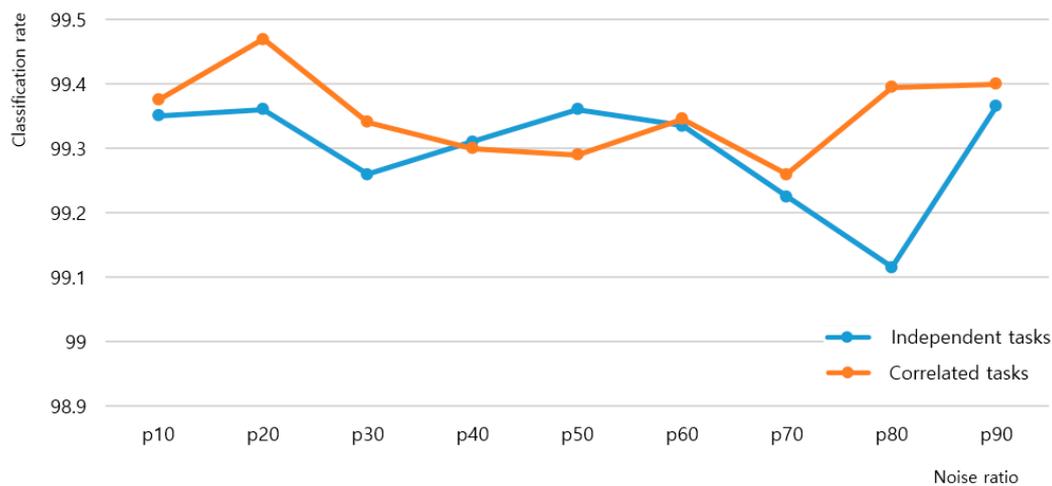


Figure 8. Classification rate (%) for digit recognition.

5. Conclusions

In this paper, we propose an overall design for various attribute recognition using multi-task learning of deep networks. Whereas conventional multi-attribute recognition method does not consider a mutual relationship between attributes and regards each of them as independent random values, we proposed to design an output representation of the learning network considering their relationships. For a given set of various attributes, we first used their exclusive relationship, and made groups for all the attributes into several tasks which was a simple extension of the conventional attribute recognition method. In addition, with an assumption that tasks were in dependent relationships, we considered joint relationships among tasks, so that the network could learn dependency among recognition tasks. Based on the results of applying two proposed methods to facial attribute recognition, we verified that the proper combination of these two methods can bring considerable improvement on the recognition performance of multi-attribute recognition.

On the other hand, we need to note that the necessary number of output nodes for the proposed joint representation increases rapidly as the number of attributes to be recognized increases. Moreover, learning exclusive and joint relationship can adversely affect performance if tasks are uncorrelated, and there is also a risk of overfitting when data is insufficient. Therefore, although our proposed method should be solved by designing a learning model using prior knowledge among attributes and tasks, there is no standard for how to properly combine two methods, which will remain our future work. Finally, though we have focused on the multi-attribute recognition for facial images, the proposed method can be applied to general multi-attribute recognition problems such as attribute recognition of pedestrians, cars, and so on as shown in the simple experiment for digit images.

Author Contributions: Conceptualization, K.E.L. and H.P.; data curation, C.H. and J.S.; formal analysis, C.H. and H.P.; methodology, C.H., J.S. and H.P.; software, C.H. and J.S.; validation, C.H. and H.P.; visualization, C.H.; writing—original draft, C.H. and H.P.; writing—review and editing, C.H. and H.P.

Funding: This work was supported by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2016-0-00145, Smart Summary Report Generation from Big Data Related to a Topic). This work was supported by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding).

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

References

1. Li, W.; Li, M.; Su, Z.; Zhu, Z. A Deep-Learning Approach to Facial Expression Recognition with Candid Images. In Proceedings of the 2015 14th IAPR International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 18–22 May 2015; pp. 279–282.
2. Karthigayani, P.; Sridhar, S. Decision tree based occlusion detection in face recognition and estimation of human age using back propagation neural network. *J. Comput. Sci.* **2014**, *10*, 115–127. [[CrossRef](#)]
3. Dehshibi, M.M.; Bastanfard, A. A new algorithm for age recognition from facial images. *Signal Process.* **2010**, *90*, 2431–2444. [[CrossRef](#)]
4. Ramesha, K.; Raja, K.B.; Venugopal, K.R.; Patnaik, L.M. Feature Extraction based Face Recognition, Gender and Age Classification. *IJCSE* **2010**, *2*, 14–23.
5. Dehghan, A.; Ortiz, E.G.; Shu, G.; Masood, S.Z. DAGER: Deep age, gender and emotion recognition using convolutional neural network. *arXiv*, 2017; arXiv:1702.04280.
6. Günther, M.; Rozsa, A.; Boulton, T.E. AFFACT: Alignment-free facial attribute classification technique. In Proceedings of the IEEE International Joint Conference on Biometrics, Denver, CO, USA, 1–4 October 2017; pp. 90–99.
7. Caruana, R. Multitask Learning. *Mach. Learn.* **1997**, *28*, 41–75. [[CrossRef](#)]
8. Seltzer, M.L.; Droppo, J. Multi-task learning in deep neural networks for improved phoneme recognition. In Proceedings of the 2013 IEEE International Conference on Acoustics, Vancouver, BC, Canada, 26–31 May 2013; pp. 6965–6969.
9. Su, C.; Yang, F.; Zhang, S.; Tian, Q.; Davis, L.S.; Gao, W. Multi-task learning with low rank attribute embedding for person re-identification. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 3739–3747.
10. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep Learning Face Attributes in the Wild. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 3730–3738.
11. Zhang, N.; Paluri, M.; Ranzato, M.A.; Darrell, T.; Bourdev, L. Panda: Pose aligned networks for deep attribute modeling. In Proceedings of the IEEE International Conference on Computer Vision and pattern recognition, Washington, DC, USA, 23–28 June 2014; pp. 1637–1644.
12. Zhong, Y.; Sullivan, J.; Li, H. Face attribute prediction using off-the-shelf cnn features. In Proceedings of the 2016 International Conference on Biometrics, Halmstad, Sweden, 13–16 June 2016; pp. 1–7.
13. Kang, S.; Lee, D.; Yoo, C.D. Face attribute classification using attribute-aware correlation map and gated convolutional neural networks. In Proceedings of the 2015 IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27–30 September 2015; pp. 4922–4926.
14. Li, D.; Chen, X.; Huang, K. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios. In Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition, Kuala Lumpur, Malaysia, 3–6 November 2015; pp. 111–115.
15. Hand, E.M.; Chellappa, R. Attributes for improved attributes: A multi-task network utilizing implicit and explicit relationships for facial attribute classification. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–10 February 2017; pp. 4068–4074.
16. Dunne, R.A.; Campbell, N.A. On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function. In Proceedings of the 8th Australian Conference on the Neural Networks, Melbourne, Australia, 11–13 June 1997; pp. 181–185.
17. Gross, R.; Matthews, I.; Cohn, J.F.; Kanade, T.; Baker, S. Multi-PIE. *Image Vis. Comput.* **2010**, *28*, 807–813. [[CrossRef](#)] [[PubMed](#)]

18. Changhun, H.; Hyeyoung, P. Recognition of Facial Attributes Using Multi-Task Learning of Deep Networks. In Proceedings of the 9th International Conference on Machine Learning and Computing, Singapore, 24–26 February 2017; pp. 284–288.
19. Kelly, D.J.; Quinn, P.C.; Slater, A.M.; Lee, K.; Gibson, A.; Smith, M.; Ge, L.; Pascalis, O. Three-month-olds, but not newborns, prefer own-race faces. *Dev. Sci.* **2005**, *8*, F31–F36. [[CrossRef](#)] [[PubMed](#)]



© 2019 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/>).