Performance Evaluation of Face Attribute Estimation Method Using DendroNet

Hiroya Kawai Koichi Ito Takafumi Aoki

Graduate School of Information Sciences
Tohoku University
Sendai, Japan
hiroya@aoki.ecei.tohoku.ac.jp

***Abstract*—There are many studies on face recognition that performs personal identification based on the features extracted from a face image. identification accuracy decreases due to changes in the imaging environment, so combined use with soft biometrics that uses face attributes such as age and gender has been considered. In this paper, the design method of Convolutional Neural Network (CNN) using clustering is proposed. By applying the hierarchical relationship of attributes derived using clustering to the structure of CNN, CNN suitable for face attribute estimation is automatically generated.**

***Index Terms*—soft biometrics, face recognition, CNN**

I. Introduction

In biometric recognition using face images, its performance depends on the photographing conditions. To stabilize its performance, it is important to combine face attributes such as age and gender. Various methods for face attribute estimation are proposed so far, and in recent years, CNN based methods [1]–[3] are the mainstream. The network architecture significantly affect CNN performance, but it is difficult to manually search for a suitable architecture. Therefore, we propose automatic design method of CNN architecture using clustering. Based on the labels included in the dataset, the hierarchical relationship of face attributes is led using hierarchical clustering. By applying hierarchical relationships to the CNN architecture, the architecture suitable for face attribute estimation is automatically designed. Performance of the proposed method is evaluated using CelebA dataset [1].

II. Network Architecture Design using Hierarchical Clustering

Clustering is an analysis method that divides a set into some subsets called clusters. In the proposed method, the correlation matrix is calculated for 40 face attributes included in the CelebA dataset. It is converted to a distance matrix based on the Chebyshev distance. The list of 40 attributes is shown in Table I. The dendrogram regarding the relationship between attributes is derived from the distance matrix using the average linkage hierarchical clustering. The dendrogram and the heatmap of the distance matrix are shown in Fig. 1.

[1]http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

TABLE I
List of attributes in CelebA dataset.

Idx.	Attribute	Idx.	Attribute
1	5 O'Clock Shadow	21	Male
2	Arched Eyebrows	22	Mouth Slightly Open
3	Attractive	23	Mustache
4	Bags Under Eyes	24	Narrow Eyes
5	Bald	25	No Beard
6	Bangs	26	Oval Face
7	Big Lips	27	Pale Skin
8	Big Nose	28	Pointy Nose
9	Black Hair	29	Receding Hairline
10	Blond Hair	30	Rosy Cheeks
11	Blurry	31	Sideburns
12	Brown Hair	32	Smiling
13	Bushy Eyebrows	33	Straight Hair
14	Chubby	34	Wavy Hair
15	Double Chin	35	Wearing Earrings
16	Eyeglasses	36	Wearing Hat
17	Goatee	37	Wearing Lipstick
18	Gray Hair	38	Wearing Necklace
19	Heavy Makeup	39	Wearing Necktie
20	High Cheekbones	40	Young

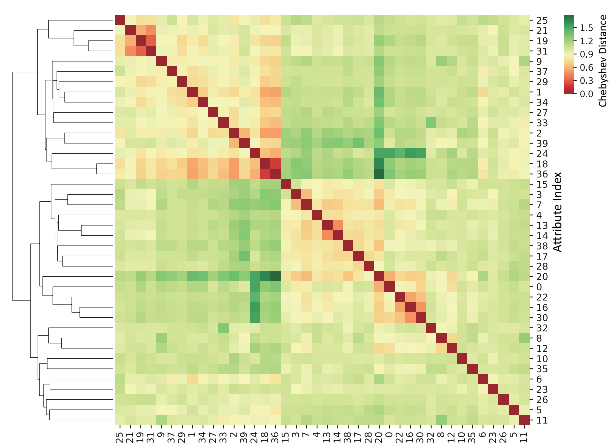

Fig. 1. The distance matrix of 40 attributes of CelebA dataset and the dendrogram derived by hierarchical clustering.

The dendrogram is searched in the direction from the root to the leaves, and the convolution blocks are respectively arranged at the root, up to the third nodes, and between the third nodes and the leaves. The parameters of each convolution

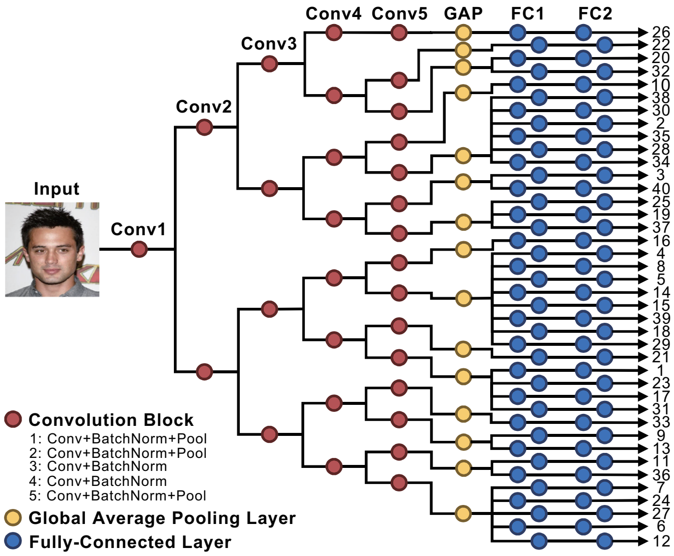

Fig. 2. The network architecture of DendroNet designed for CelebA dataset.

block are based on AlexNet [4]. The network architecture designed in this method is named DendroNet. The network architecture of DendroNet designed for CelebA dataset is shown in Fig. 2. We also design another CNN architecture in which feature extractors are shared among all attributes and fully-connected layers are branched based on the dendrogram. This network architecture is named DendroNet-FC.

III. Experiments and Discussion

The performance of DendroNet and DendroNet-FC is evaluated using CelebA dataset [1]. CelebA dataset includes 202,599 face images and 40 face attribute labels. In this experiment the dataset is devided into 182,637 and 19,962 according to the official partition, and each is used for training and test. 10% of training data is separated as validation data, and it is used to confirm overfitting. Cross-entropy loss is used for loss function, and Nesterov Accelerated Gradient (NAG) is used for optimization. The initial learning rate is set to 0.001, and the learning rate and the number of epochs are controlled according to the validation loss. Also, the size of the input image is set to 227×227 pixels, and standardization is applied to each image. We compare the estimation accuracy for test data of LNets+ANet [1], FaceNet+SVM [2], MCNN-AUX [3], and two proposed methods.

The Fig. 3 shows the estimation accuracy for each attribute. Also, the average accuracy for 40 attributes is shown in Table II. The two proposed methods show higher accuracy than the method using Support Vector Machine (SVM) as a classifier [1], [2]. Also, the accuracy of the proposed methods are as higher as MCNN-AUX [3]. In contrast to the conventional method of manually designing the network over time, the proposed method automatically design network within 10 seconds using a common laptop. The proposed method is effective for task such as face attribute estimation required complex network design.

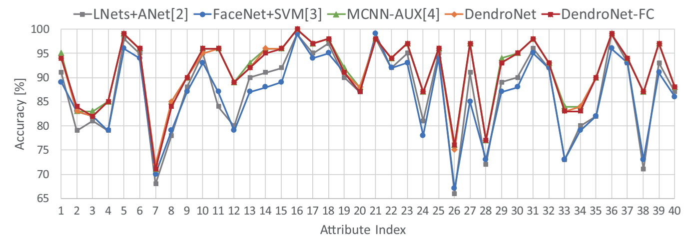

Fig. 3. The accuracy of face attribute estimation for CelebA dataset.

TABLE II
The average accuracy for 40 attributes of CelebA dataset.

Method	Network/Classification	Acc.[%]
Liu et al. [1]	5conv+4conv/SVM	87.30
Zhong et al. [2]	FaceNet [5]/SVM	86.60
Hand et al. [3]	3conv(multi)+2fc/softmax	91.28
DendroNet	5conv(multi)+2fc/softmax	91.15
DendroNet-FC	5conv+2fc(multi)/softmax	91.10

IV. Conclusion

In this paper, we proposed an automatic design method of CNN for face attribute estimation. In the performance evaluation using CelebA dataset, the CNN designed by the proposed method showed the same accuracy as the conventional CNN designed manually. In the future, we plan to study regarding other automatic design methods of CNN for face attribute estimation.

References

[1] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the Wild," *Proc. Int'l Conf. Computer Vision*, pp. 3730–3738, Dec. 2015.
[2] Y. Zhong, J. Sullivan, and H. Li, "Face attribute prediction using off-the-shelf CNN features," *Proc. Int'l Conf. Biometrics*, pp. 1–7, Jun. 2016.
[3] E. Hand and R. Chellappa, "Attributes for improved attributes: A multi-task network utilizing implicit and explicit relationships for facial attribute classification," *Proc. the Thirty-First AAAI Conf. Artificial Intelligence*, pp. 4068–4074, Feb. 2017.
[4] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Proc. Advances in Neural Information Processing Systems*, pp. 1097–1105, Dec. 2012.
[5] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 815–823, Dec. 2015.