Performance Evaluation of Face Attribute Estimation Method Using DendroNet

Hiroya Kawai, Koichi Ito, and Takafumi Aoki
Graduate School of Information Sciences, Tohoku University, Japan
hiroya@aoki.ecei.tohoku.ac.jp

***Abstract*—There are many studies on face recognition, which identifies a person using distinctive features extracted from a face image. One of the problems in face recognition is that the accuracy of face recognition decreases due to environmental changes such as head pose, emotion, illumination, etc. Addressing this problem, soft biometrics, which uses attributes such as age and gender for person authentication, is expected to improve the accuracy of face recognition. This paper proposes a face attribute estimation method using the Convolutional Neural Network (CNN). The CNN architecture of the proposed method, called DendroNet, is automatically designed according to the relationships among attributes. Though experiments using the CelebA dataset, we demonstrate that the proposed method exhibits better performance than conventional methods.**

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

I. Introduction

Face recognition is one of the major topics on biometrics and has high acceptability compared with other biometrics. The accuracy of face recognition has been dramatically improved because of the recent advances in deep learning, while there are still many problems in face recognition. One of the problems in face recognition is that the accuracy of face recognition decreases due to environmental changes such as head pose, expression, illumination, etc.

One of the approaches to improve the performance of face recognition is to use soft biometrics, which uses ancillary information from the primary biometric traits such as face, fingerprint, and iris. The ancillary information used in soft biometrics is attribute such as age and gender, which can be estimated from a face image. A single attribute cannot be used for person authentication due to its low distinctiveness, while the performance of a set of attributes may be comparable with that of face recognition. So far, various attribute estimation methods from face images have been proposed, where CNN has been used in recent methods [1]–[3]. The estimation accuracy is significantly affected by the CNN architecture. Most of the CNN-based methods have employed the manually designed CNN architectures, although it is difficult to find the optimal architecture by hand, since the face attribute estimation needs the combination of many CNNs depending on the number of target attributes.

We propose a face attribute estimation method using CNN, which is automatically designed according to the relationships among attributes. The network architecture of the proposed method, called DendroNet, is based on the dendrogram, which represents the relationship among attributes. Performance evaluation using the CelebA dataset[1] demonstrates that the proposed method exhibits the efficient performance on face attribute estimation compared with conventional methods.

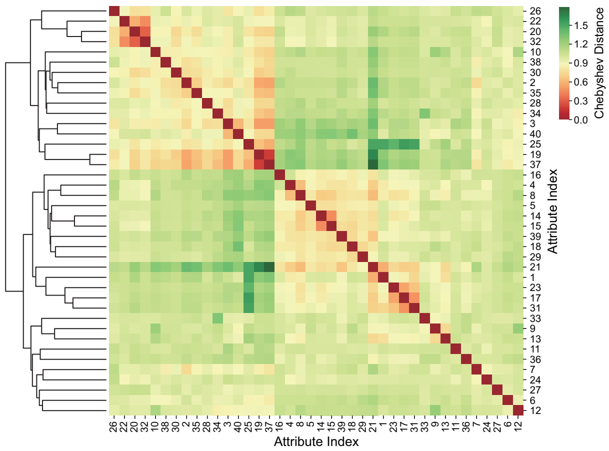

Fig. 1. Distance matrix of 40 attributes included in the CelebA dataset and its dendrogram derived by hierarchical clustering.

II. DendroNet

DendroNet is automatically designed according to the relation among facial attributes, which can be derived by attribute clustering. First, the correlation matrix is obtained for 40 face attributes in the CelebA dataset and is converted into the distance matrix according to the Chebyshev distance. 40 attributes in the CelebA dataset is shown in Table I. Next, the dendrogram is derived from the distance matrix using average linkage hierarchical clustering. The dendrogram and the heatmap of the distance matrix obtained from the CelebA dataset are shown in Fig. 1. Then, the CNN architecture is designed based on the dendrogram. The convolution blocks are located on the root, on the first to the third nodes and between the third node and leaves in the dendrogram. The parameters of each convolution block are corresponding to AlexNet [4]. Fig. 2 shows the network architecture of DendroNet for the CelebA dataset. We also design the other CNN architecture, called DendroNet-FC, where feature extractors are shared among all the attributes and fully-connected layers are branched according to the dendrogram.

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

978-1-7281-3575-5/19/$31.00 ©2019 IEEE

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

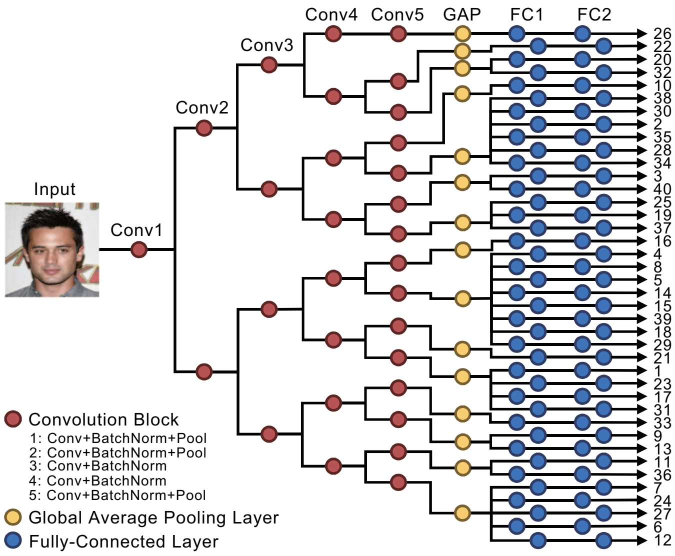

Fig. 2. Network architecture of DendroNet for the CelebA dataset.

III. Experiments and Discussion

We evaluate the performance of the proposed method using the CelebA dataset [1]. CelebA dataset consists of 202,599 face images and 40 face attribute labels for each image. The dataset is divided into 182,637 for training and 19,962 for test followed by the official partition. 10% of training data is separated as validation data, which is used to check overfitting. Cross-entropy loss is used for the 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. The size of the input image is set to 227×227 pixels, and standardization of pixel values is applied to each image. We compare the estimation accuracy for LNets+ANet [1], FaceNet+SVM [2], MCNN-AUX [3], and two proposed methods.

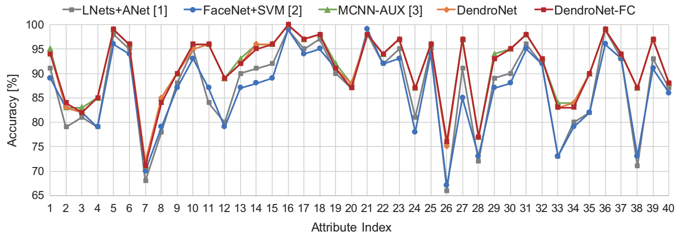

Fig. 3. Accuracy of face attribute estimation for each attribute.

TABLE II
Average estimation accuracy of 40 attributes.

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

The Fig. 3 shows the estimation accuracy for each attribute and Table II shows the average estimation accuracy of 40 attributes for each method. The estimation accuracy of the two proposed methods is higher than that of the conventional methods using Support Vector Machine (SVM) [1], [2] and is comparable with that of MCNN-AUX [3]. The manual design of network architectures such as MCNN-AUX is time-consuming. while DendroNet can be automatically designed within 10 seconds on a common laptop PC. As observed above, the proposed method exhibits efficient performance on face attribute estimation compared with conventional methods.

IV. Conclusion

This paper proposed a face attribute estimation method using DendroNet, which is automatically designed according to the relation among face attributes. Performance evaluation using the CelebA dataset demonstrated that the proposed method exhibits the efficient performance on face attribute estimation compared with conventional methods. We will develop a compact and flexible architecture for face estimation in our future work.

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.