Gait-based Age Estimation using a DenseNet*

Atsuya Sakata[1], Yasushi Makihara[1], Noriko Takemura[1], Daigo Muramatsu[1], and Yasushi Yagi[1]

Osaka University, Japan {sakata, makihara, muramatsu, yagi}@am.sanken.osaka-u.ac.jp, takemura@ids.osaka-u.ac.jp

Abstract. Human age is one of important attributes for many potential applications such as digital signage, customer analysis, and gait-based age estimation is promising particularly for surveillance scenarios since it can be available at a distance from a camera. We therefore proposed a method of gait-based age estimation using a deep learning framework to advance the state-of-the-art accuracy. Specifically, we employed DenseNet as one of state-of-the-art network architectures. While the previous method of gait-based age estimation using a deep learning framework was evaluated only with a small-scale gait database, we evaluated the proposed method with OULP-Age, the world's largest gait database comprising more than 60,000 subjects with age range from 2 to 90 years old. Consequently, we demonstrated that the proposed method outperform existing methods based on both conventional machine learning frameworks for gait-based age estimation and a deep learning framework for gait gait recognition.

Keywords: Gait · Age · DenseNet.

1 Introduction

Image-based human age estimation has recently become an attractive research topic in computer vision, pattern recognition, and biometrics, since there are many potential applications. For example, once a target person's age is estimated, an advertiser can change a content of a digital signage into more suitable one for the estimated age, and a shop manager may arrange goods based on customer's age statistics. It is also possible to prevent that people under age buy alcohol or cigarette based on the estimated age.

Most of the image-based human age estimation relies on facial image analysis [1–4] as we human also do so. In addition to the facial image analysis, gait video analysis for age estimation is promising, since it does not require as high image resolutions as the facial image analysis. Therefore, it can be available even at a distance from a camera without subject cooperation [5, 6], and which makes it much wider an application range of the image-based human age estimation, e.g., getting customer's age statistics from a wider area than a shop (e.g., the whole shopping mall), finding a lost child in a shopping mall, finding suspect candidates based on witness about his/her age.

In the early stage, research on gait-based age analysis started with age group classification such as classification of children and adults using representation of point light

* Supported by JST-Mirai Program JPMJMI17DH and JSPS Grant-in-aid for Scientific Research (B) 16H02848.

sources attached to some body joints [7], or classification of the younger and the elderly using minimum foot clearance from the ground [8]. Subsequently, an image analysis-based approach to classification of children, adults, and the elderly has been proposed in [9]. Thereafter, studies on gait-based age estimation has been started since 2010, by using an appearance-based gait feature such as gait energy image (GEI) [10] (a.k.a. averaged silhouette [11]), frequency-domain feature [12], or depth gradient histogram energy image (DGHEI) [13] in conjunction with machine learning techniques, for example, multi-label guided (MLG) subspace learning [14], ordinary preserving manifold learning [15], Gaussian process regression [16], ordinary preserving linear discriminant analysis (OPLDA) and ordinary preserving margin Fisher analysis (OPMFA) [17], support vector regression (SVR) [18], and age group-dependent manifold learning and regression [19].

In addition to the above-mentioned conventional machine learning-based approaches, Marin-Jimenez et al. [20] proposed a deep learning-based gait-based age estimation. More specifically, the authors design a multi-task deep model which outputs age as well as identity and gender. The model is, however, trained and tested only with small-scale gait database, i.e., TUM-GAID [21], which composed of 305 subjects almost in their twenties. Since a deep learning-based approach generally requires a huge number of training samples to reach a satisfactory accuracy, and also reasonable evaluation of gait-based age estimation requires a wide age range, we would say that the method [20] is not fully validated.

We therefore want to validate a deep learning-based approach to gait-based age estimation in this paper. Specifically, we designs a model for gait-based age estimation based on DenseNet [22], a state-of-the-art network architecture so far, and also train and test the model using the world largest gait database, OULP-Age [18], which comprising over 60,000 subjects with wide age range.

2 Proposed method

2.1 Input data

As appearance-based gait features [10–12, 23, 24] are more often used than the model-based gait features [25, 26] in the gait recognition community. In particular, silhouette-based representation is dominant in the gait recognition community, since it can avoid being affected by clothes colors and textures, unlike the person re-identification task. We therefore adopt GEI [10] as the most widely used silhouette-based gait representation as input data for our deep learning model. We set the GEI size as 88 by 128 pixels.

2.2 Network structure

We employ DenseNet [22] as a state-of-the-art network structure for our gait-based age estimation task. While the residual network (ResNet) [27] exploits a skip connection from a single previous layer, DenseNet exploits skip connections from all the preceding layers as shown in Fig. 1. DenseNet achieves the state-of-the-art accuracies with less

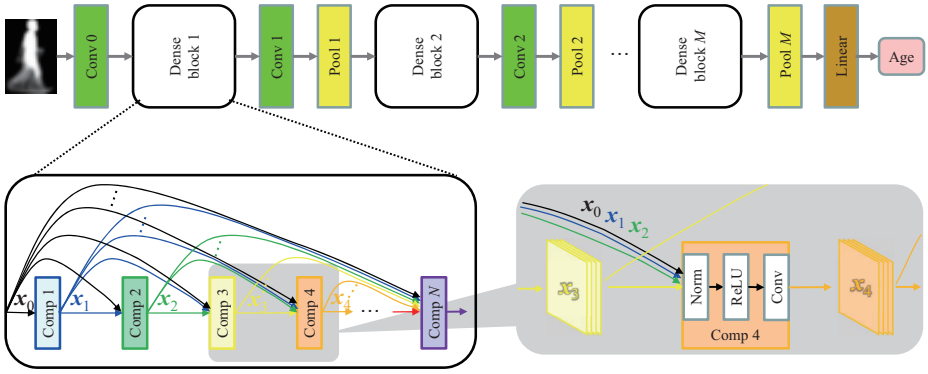


Fig. 1. An illustration of network architecture with M dense blocks, M composite layers, and growth rate $k = 4$.

parameters thanks to the dense connectivity, on object classification tasks as well as other tasks such as optical flow estimation [28] and saliency map estimation [29]. We briefly describe the network structure below and may refer the readers to [22] for more details.

DenseNet is mainly composed of multiple dense blocks and in-between transition layers including a pair of 1×1 convolution layer and a 2×2 average pooling layer with stride 2. Each dense block is then composed of multiple composite layers with dense connection, where each composite layer comprises a triplet of a 1×1 batch normalization layer, a rectified linear unit (ReLU) [30], and a 3×3 convolution layer. Here, we consider M dense blocks, and each dense block is composed of N composite layers. Given an input $\boldsymbol{x}_0$ to a dense block whose number of feature maps is k_0, it is fed into the following composite layers (i.e., Comp 1, Comp 2, ..., Comp N) as shown in Fig. 1. The l-th composite layer receives feature maps from all the preceding composite layers, i.e., $[\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{l-1}]$, and then output $\boldsymbol{x}_l$ whoes number of feature maps is k, which is so-called growth rate. In summary, the number of feature maps which the l-th composite layer receives, sums up to $k_0 + k(l-1)$, and hence each composite layer enjoys a sort of collective knowledge from the preceding layers. Specifically, we experimentally used $M = 5$ dense blocks and each dense blocks has $N = 5$ composite layers with $k = 12$ growth rate.

In addition to the above-mentioned main part, we insert a 7×7 convolution layer with stride 2 (Conv 0) before the first dense block, and also a 7×7 global average pooling layer followed by a one-dimensional full connection layer to output an age.

3 Experiments

3.1 Data set

We employed OULP-Age [18] as the world's largest gait database for gait-based age estimation. OULP-Age was collected in conjunction with long-run exhibition of experience-

based gait analysis demo in a science museum for an approx one year [31]. It consists of 63,846 (31,093 males and 32,753 females) with age ranging from 2 to 90 years old. The database is divided into a training set composed of 31,923 subjects (15,596 males and 16,327 females) and a test set composed of 31,923 subjects (15,497 males and 16,426 females).

3.2 Training

The network was trained using stochastic gradient descent (SGD) [32] with batch size 64 and 100 epochs. An initial learning rate was set to 0.1 and divided by 10 at epoch 50 and 75. We use a Nesterof momentum [33] of 0.9 without dampening. We adopt the weight initialization introduced by He et al. [34]. A loss function was set to sum of absolute difference between an output (i.e., estimated) age and the ground truth age.

3.3 Evaluation measure

We evaluated the accuracy of gait-based age estimation using a mean absolute error (MAE). MAE is computed by comparing the estimated age $\hat{a}_i$ for the i-th test sample with its corresponding ground truth age a_i as

$$M = \frac{1}{n}\sum_{i=1}^{n}|\hat{a}_i - a_i|, \tag{1}$$

where n is the number of the test samples. In addition, we employed a cumulative score (CS) for evaluating gait-based age estimation. Specifically, we define the number of test samples whose absolute difference between an estimated age and the ground truth age is less than or equal to y as $n(y)$, and then CS of the y-year absolute error as

$$\mathrm{CS}(y) = \frac{n(y)}{n}. \tag{2}$$

3.4 Sensitivity analysis

Since the number of blocks M and the number of composite layers N are two key hyper-parameters for DenseNet, we analyze the sensitivity of these two hyper-parameters on gait-based age estimation accuracy. Specifically, we set $M \in \{2, 3, 4, 5, 6, 7\}$ and $N \in \{4, 5, 6\}$ and then evaluate an MAE for each parameter combination as shown in Fig. 2. As result, we can see that the number of dense blocks has an impact on the accuracy, namely, the accuracy significantly drop if the number of dense blocks is less than 5. On the other hand, the number of composite layers does not have much impact on the accuracy, at least within the range from 4 to 6.

3.5 Comparison with state-of-the-arts

Finally, we compared the proposed method with benchmarks. As a baseline algorithm, a method using GPR with radial basis function [16] was adopted. More specifically, it

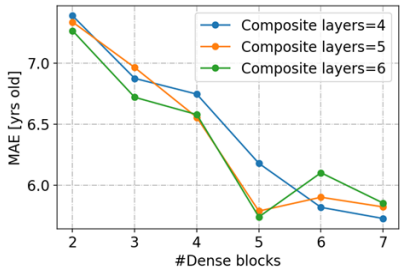


Fig. 2. Sensitivity analysis of the number of dense blocks and composite layers on MAE.

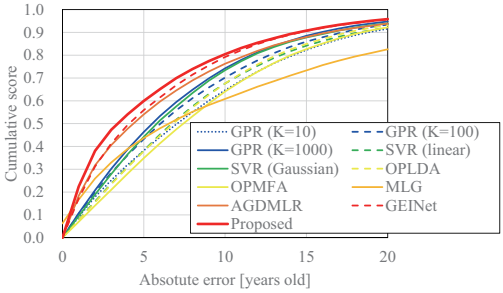


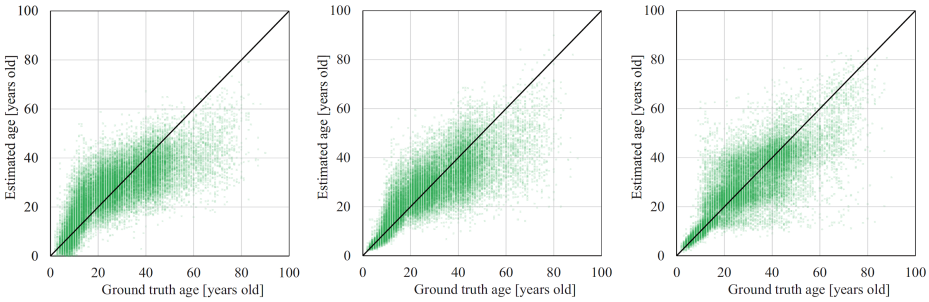
Fig. 3. Cumulative score.

Table 1. MAE [%] and cumulative score at 1, 5, and 10 years absolute errors. Bold and italic bold indicate the best and the second best performances, respectively.

Method	MAE	CS(1) [%]	CS(5) [%]	CS(10) [%]
MLG [14]	10.98	16.7	43.4	60.8
GPR ($k = 10$) [16]	8.83	9.1	38.5	64.7
GPR ($k = 100$) [16]	7.94	10.5	43.3	70.2
GPR ($k = 1000$) [16]	7.30	10.7	46.3	74.2
SVR (linear)	8.73	7.9	38.2	67.6
SVR (Gaussian)	7.66	9.4	44.2	73.4
OPLDA [17]	8.45	7.7	37.9	67.6
OPMFA [17]	9.08	7.0	34.9	64.1
ADGMLR [19]	6.78	***18.4***	54.0	76.2
GEINet [35]	***6.22***	17.2	***55.9***	***79.2***
Proposed method	**5.79**	**22.5**	**55.9**	**80.4**

requires huge computation on a gram matrix and its inverse matrix if all the training samples are used, we employ an active set method in the same way as in [31], where k nearest neighbors for each test sample are used for GPR and $k = 10, 100, 1000$ were evaluated. We also tested support vector regression (SVR) with linear and Gaussian kernels, denoted as SVR (linear) and SVR (Gaussian). We also evaluated existing methods of gait-based age estimation using conventional machine learning techniques such as MLG [14], OPLDA [17], OPMFA [17], and AGDMLR [19]. In addition, we employed a slightly modified version of GEINet [35] as a deep learning-based approach to age estimation. More specifically, the original GEINet outputs class (i.e., subject) likelihoods and hence the number of nodes at the last layer is equal to the number of subjects. On the other hand, the modified version of GEINet outputs an age and hence the last layer has just a single node.

CSs for the benchmarks and the proposed method are shown in Fig. 3. Also, MAEs and CSs for 1, 5, and 10 years tolerance are summarized in Table 1. As a result, deep learning-based methods (i.e., GEINet [35] and the proposed method) significantly out-



(a) GPR ($k = 1000$) [16] (b) GEINet [35] (c) Proposed method

Fig. 4. Scatter plot between ground truth age and estimated age. Diagonal lines show equal lines of ground truth age and estimated age.

perform the other methods without deep learning. In addition, the proposed method outperforms GEINet. This is because the proposed method exploit more collective information by dense connectivity while a basic structure of GEINet is derived from AlexNet [36] without such a dense connectivity.

Moreover, scatter plots between the ground truth age and the estimated ages are shown for the baseline GPR ($k = 100$) as a non-deep learning-based method [16], GEINet [35] as a deep learning-based method, and the proposed method in Fig. 4. While the estimated ages for the baseline algorithm (Fig. 4 (a)) is largely deviated for all the age range, GEINet (Fig. 4 (b)) suppressed such a deviation, particularly, for the children under 15 years old. The proposed method further suppress the deviation for the children and then yielded the best accuracy as a result.

4 Conclusions

In this paper, we proposed a method of gait-based age estimation using a deep learning framework. Specifically, we employed DenseNet as a state-of-the-art network architecture for gait-based age estimation, and demonstrated its effectiveness with OULP-Age, the world's largest gait database comprising more than 60,000 subjects with age range from 2 to 90 years old.

One of future research avenues is further investigating suitable network architectures for gait-based age estimation, e.g., ResNet [27], PyramidNet [37], and ShakeNet [38]. In addition, it is also worth investigating multi-task learning framework, e.g., [20].

References

1. Kwon, Y.H., Lobo, N.V.: Age classification from facial images. Computer Vision and Image Understanding **74** (1999) 1–21
2. Zhang, Y., Yeung, D.Y.: Multi-task warped gaussian process for personalized age estimation. In: Proc. of IEEE computer society conferene on Computer Vision and Pattern Recognition 2010, San Francisco, CA, USA (2010) 1–8

3. Lanitis, A., Draganova, C., Christodoulou, C.: Comparing different classifiers for automatic age estimation. IEEE Trans. on System, Man, and Cybernetics, Part B **34** (2004) 621–628
4. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Trans. on Pattern Analysis and Machine Intelligence **24** (2002) 442–455
5. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recognition Letters **31** (2010) 2052–2060
6. Martin-Felez, R., Xiang, T.: Uncooperative gait recognition by learning to rank. Pattern Recognition **47** (2014) 3793 – 3806
7. Davis, J.: Visual categorization of children and adult walking styles. In: Proc. of the Int. Conf. on Audio- and Video-based Biometric Person Authentication. (2001) 295–300
8. Begg, R.: Support vector machines for automated gait classification. IEEE Trans. on Biomedical Engineering **52** (2005) 828–838
9. Mannami, H., Makihara, Y., Yagi, Y.: Gait analysis of gender and age using a large-scale multi-view gait database. In: Proc. of the 10th Asian Conf. on Computer Vision, Queenstown, New Zealand (2010) 975–986
10. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. on Pattern Analysis and Machine Intelligence **28** (2006) 316– 322
11. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: Averaged silhouette. In: Proc. of the 17th International Conference on Pattern Recognition. Volume 1. (2004) 211–214
12. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: Proc. of the 9th European Conference on Computer Vision, Graz, Austria (2006) 151–163
13. Hofmann, M., Bachmann, S., Rigoll, G.: 2.5d gait biometrics using the depth gradient histogram energy image. In: Proc. of the 5th IEEE Int. Conf. on Biometrics Theory, Applications, and Systems, Washington D.C., USA (2012) Paper ID 172, 1–5
14. Lu, J., Tan, Y.P.: Gait-based human age estimation. IEEE Trans. on Information Forensics and Security **5** (2010) 761–770
15. Lu, J., Tan, Y.P.: Ordinary preserving manifold analysis for human age estimation. In: IEEE Computer Society and IEEE Biometrics Council Workshop on Biometrics 2010, San Francisco, CA, USA (2010) 1–6
16. Makihara, Y., Okumura, M., Iwama, H., Yagi, Y.: Gait-based age estimation using a whole-generation gait database. In: Proc. of the Int. Joint Conf. on Biometrics (IJCB2011), Washington D.C., USA (2011) 1–6
17. Lu, J., Tan, Y.P.: Ordinary preserving manifold analysis for human age and head pose estimation. IEEE Transactions o Human-Machine Systems **43** (2013) 249–258
18. Xu, C., Makihara, Y., Ogi, G., Li, X., Yagi, Y., Lu, J.: The ou-isir gait database comprising the large population dataset with age and performance evaluation of age estimation. IPSJ Transactions on Computer Vision and Applications **9** (2017) 24
19. Li, X., Makihara, Y., Xu, C., Yagi, Y., Ren, M.: Gait-based human age estimation using age group-dependent manifold learning and regression. Multimedia Tools and Applications **77** (2018) 28333–28354
20. Marin-Jimenez, M., Castro, F., de la Torre, N.G.F., Medina-Carnicer, R.: Deep multi-task learning for gait-based biometrics. In: Proc. of 2017 IEEE International Conference on Image Processing. (2017) 106–110
21. Hofmann, M., Geiger, J., Bachmann, S., Schuller, B., Rigoll, G.: The tum gait from audio, image and depth (gaid) database: Multimodal recognition of subjects and traits. J. Vis. Comun. Image Represent. **25** (2014) 195–206
22. Huang, G., Liu, Z., v. d. Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017) 2261–2269

23. Wang, C., Zhang, J., Wang, L., Pu, J., Yuan, X.: Human identification using temporal information preserving gait template. IEEE Trans. on Pattern Analysis and Machine Intelligence **34** (2012) 2164 –2176
24. Lam, T.H.W., Cheung, K.H., Liu, J.N.K.: Gait flow image: A silhouette-based gait representation for human identification. Pattern Recognition **44** (2011) 973–987
25. Johnson, A., Bobick, A.: A multi-view method for gait recognition using static body parameters. In: Proc. of the 3rd Int. Conf. on Audio- and Video Based Biometric Person Authentication, Halmstad, Sweden (2001) 301–311
26. Ariyanto, G., Nixon, M.: Marionette mass-spring model for 3d gait biometrics. In: Proc. of the 5th IAPR International Conference on Biometrics. (2012) 354–359
27. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016) 770–778
28. Zhu, Y., Newsam, S.D.: Densenet for dense flow. In: 2017 IEEE International Conference on Image Processing (ICIP). (2017) 790–794
29. Oyama, T., Yamanaka, T.: Influence of image classification accuracy on saliency map estimation. CAAI Transactions on Intelligence Technology **3** (2018) 140–152(12)
30. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In Furnkranz, J., Joachims, T., eds.: Proceedings of the 27th International Conference on Machine Learning (ICML-10), Omnipress (2010) 807–814
31. Makihara, Y., Kimura, T., Okura, F., Mitsugami, I., Niwa, M., Aoki, C., Suzuki, A., Muramatsu, D., Yagi, Y.: Gait collector: An automatic gait data collection system in conjunction with an experience-based long-run exhibition. In: Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016). Number O17, Halmstad, Sweden (2016) 1–8
32. Bousquet, O., Bottou, L.: The tradeoffs of large scale learning. In Platt, J., Koller, D., Singer, Y., Roweis, S., eds.: Advances in Neural Information Processing Systems 20, Curran Associates, Inc. (2008) 161–168
33. Sutskever, I., Martens, J., Dahl, G., Hinton, G.: On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28. ICML'13, JMLR.org (2013) III–1139–III–1147
34. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: 2015 IEEE International Conference on Computer Vision (ICCV). (2015) 1026–1034
35. Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., Yagi, Y.: Geinet: View-invariant gait recognition using a convolutional neural network. In: Proc. of the 8th IAPR Int. Conf. on Biometrics (ICB 2016). Number O19, Halmstad, Sweden (2016) 1–8
36. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L., Weinberger, K., eds.: Advances in Neural Information Processing Systems 25, Curran Associates, Inc. (2012) 1097–1105
37. Han, D., Kim, J., Kim, J.: Deep pyramidal residual networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017) 6307–6315
38. Gastaldi, X.: Shake-shake regularization of 3-branch residual networks. In: 5th International Conference on Learning Representations, Workshop track. (2017) 1–5