# SIMILARITY-AWARE DEEP ADVERSARIAL LEARNING FOR FACIAL AGE ESTIMATION

*Penghui Sun*[1], *Hao Liu*[1,2,*] , *Xing Wang*[1], *Zhenhua Yu*[1,2], *Suping Wu*[1,2]

[1] School of Information Engineering, Ningxia University, Yinchuan, 750021, China
[2] Collaborative Innovation Center for Cloud Computing and Big Data Applications of Ningxia, Yinchuan, 750021, China

## ABSTRACT

In this paper, we propose a similarity-aware deep adversarial learning (SADAL) approach for facial age estimation. Instead of making access to limited training samples which likely leads to sub-optima, our SADAL seeks sets of unobserved and plausible hard-examples based on existing training samples, which typically reinforces the discriminativeness of the learned feature descriptor for ages. Motivated by the fact that age labels are usually correlated in the real-world applications, we carefully develop a similarity-aware function in our approach, which dynamically measures each face pair with different weights based on different age value gaps. During the learning process, we jointly optimize both procedures of generating hard-examples and learning age estimator via a sequence of adversarial-game iterations. As a result, the smoothing aging pattern is exploited in the reconstructed hard-example space for robust age estimation. Experimental results on two standard benchmarking datasets show that our approach achieves superior performance compared with most state-of-the-art age estimation methods.

***Index Terms***— Facial age estimation, deep learning, generative adversarial network, biometrics.

## 1. INTRODUCTION

The main purpose of facial age estimation targets on estimating the precise biologicals ages or age ranks based on the given facial image, which incorporates with a set of popular and potential applications such as the facial attributes detection [12, 16], video content analysis [1, 24] and human-computer interaction [10]. Despite numerous facial age estimation approaches have been proposed recently [3, 9, 11, 16], the performance still remains unsatisfied in practice due to variations including different expressions, facial accessories, poses, races, low resolution or extreme illumination.

* Corresponding author (liuhao@nxu.edu.cn). This work was supported in part by the Natural Science Foundation of Ningxia under Grant 2018AAC03035, in part by the National Natural Science Foundation of China under Grant 61806104, Grant 61662059 and Grant 61862050, in part by the Scientific Research Projects of Colleges and Universities of Ningxia under Grant NGY2018050, and in part by the Youth Science and Technology Talents Enrollment Projects of Ningxia under Grant TJGC2018028.

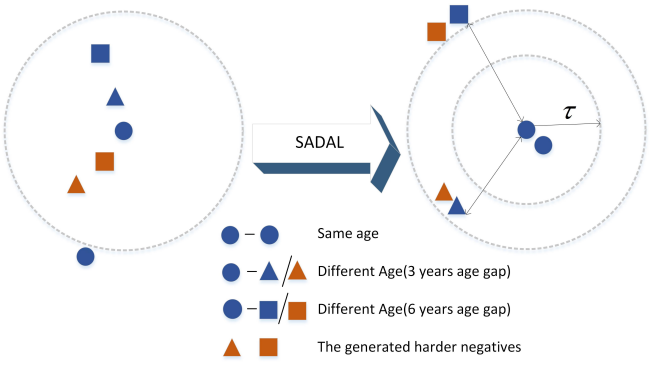


**Fig. 1**. Insight of our SADAL approach. Our basic idea aims to utilize existing feature points to reconstruct the hard-negative samples, so that these generated hard-negatives provide essential complementary information for effective feature representation or metric learning. Suppose we have the existing samples (blue) and the generated hard-negatives (orange), we enforce two-fold criterions in our SADAL: 1) The distance between each pair from different classes with a small age gap (circle and triangle) is smaller than that from a negative pair with a large age gap (circle and square). 2) The distances of the pairs with same ages should become as smaller as possible. This figure is best viewed in color pdf.

Conventional facial age estimation methods can be roughly classified into two categories: facial feature representation [19, 21] and age estimation [3, 11]. Representative facial feature representation approaches include holistic subspace features, active appearance model (AAM) [9, 10, 16], local binary patterns (LBP) [1, 21], and bio-inspired features (BIF) [12]. Having obtained the feature descriptors for each facial image, we feed the feature vector to a pre-trained age estimator to predict the chronological age value. However, the employed feature descriptors in those methods are hand-crafted, which usually requires strong expert knowledge by hand-design and consumes much effort. To address this, learning-based feature representation has been developed in the recent literature [8, 21], which aim to learn discriminative features directly from the raw image pixels. While these data-

driven methods achieve promising performance, they ignore the complex and non-linear relationship between facial images and age labels especially when faces are captured in wild condition. To circumvent this, deep learning-based methods [19, 22, 26] have been leveraged to exploit the relationship between face features and aging process via a sequence of nonlinear transformations. While promising performance has been achieved, these methods undergo unbalanced training samples during model learning which likely leads to bias prediction. Moreover, they cannot explicitly exploit the plausible samples in the unobserved space for robust age estimation.

Besides from developing both types of discriminative methods on feature representation and age estimator, *learning to ranking* has also been applied to be effective on improving the performance in recent works [5, 17, 18, 20, 22]. These methods typically project face samples to a latent common space, where the similarity of face pairs is equivalently isotonic to the age difference in a ranking-preserved manner. For example, Chen *et al.* [5] proposed a ranking-based CNN for age estimation, which was trained with ordinal age labels represented by a series of aggregated binary outputs. Liu *et al.* [19] proposed a label-sensitive deep metric learning (LSDML) approach, which aims to mine hard meaningful samples and learn a similarity distance function via the deep residual network. However, these methods are limited to be learned by existing samples, which cannot fully describe the distribution of hard-negative samples [6]. Since a facial image with a specific age undergoes the facial variations due to facial expressions, lighting or any other external factors, this may lead to a neighboring age category as the final bias prediction.

To address the above-mentioned challenges, we propose a similarity-aware deep adversarial learning (SADAL) approach for facial age estimation, which attempts to generate potential hard-examples and learning discriminative metric in a ranking-preserving manner. Fig. 1 illustrates the basic idea of the proposed approach. Different from traditional age estimation methods, our approach aims to reconstruct negative samples, which provides complementary information for the existing training samples. To achieve this, we propose a hard-example generation strategy, where the abundant observed easy negatives can be used to generate potential hard-negatives to reinforce our model. To further efficiently exploit the age-difference information in our approach, we carefully design a similarity-aware function to dynamically measure face pairs with different age value gaps between adjacent ages. By doing this, the distance between the pair from different classes with a small age gap is smaller than that from a negative pair with a large age gap, and the distances of the pairs with same ages become smaller. During the training process, we globally optimize both tasks of learning similarity-aware distance metric and mining hard-negatives with a sequence of dynamic game iterations. When the game process reaches Nash equilibrium, the generator restores the true distribution of the training data and utilizes the observed data as input to generate hard samples. As a result, the generated hard-negatives enhance the learned distance metric space, where the correlation of age labels is exploited simultaneously for robust age estimation. Experimental results on two widely-evaluated face aging datasets highlight the effectiveness of the proposed method.

## 2. SIMILARITY-AWARE DEEP ADVERSARIAL LEARNING

As illustrated in Fig. 2, our SADAL mainly learns the generator and the discriminator simultaneously in an adversarial manner. Specifically, the generator aims to generate hard-negatives on which the learned metric would mis-classify. Accordingly, the discriminator optimizes a discriminative distance metric, where the inter-class separability, intra-class compactness and label correlation of age classes are exploited to characterize the feature similarity simultaneously.

### 2.1. Hard-Example Generation

Let $X = [x_1, ..., x_i, ..., x_N] \in R^{d \times N}$ be the training set and $Y = [y_1, ..., y_i, ...y_N] \in R^{1 \times N}$ be the corresponding age labels, where $x_i \in R^d$ $(1 \leq i \leq \mathrm{N})$ denotes the $i$-th face image of $d$ pixels and $y_i \in R^1$ specifies the chronological age value, respectively. We feed each batch of facial images into the designed deep convolutional neural networks (CNN) and then compute the feature descriptors. Specifically, the deep feature representation is performed by a sequence of operations as follows:

$$f(x_i) = \mathrm{pool}(\mathrm{ReLU}(W \otimes x_i + b)), \tag{1}$$

where $\otimes$ denotes the convolution operation, $\mathrm{ReLU}(\cdot)$ is the rectifier nonlinear function, $\mathrm{pool}(\cdot)$ specifies the max-pooling operation. In general, the main goal of traditional metric learning is to measure the similarity between samples, which aims to maximize the inter-class variations and minimize the intra-class variations, simultaneously. Based on this, we compute the similarity for each sampled face pair $f(x_i)$ and $f(x_j)$ as follows:

$$d_f\,(x_i, x_j) = \| \, f(x_i) - f(x_j)\|_2, \tag{2}$$

where $\| \cdot \|_2$ denotes the squared Euclidean distance between each face pairs in the learned feature space.

**Similarity-Aware Function:** It is worth noting that facial age estimation especially in unconstrained environments is a similarity-aware problem, where the similarity of each face pair with different age values is sensitive to the age smoothness in real-world applications. To achieve the encouraging performance of facial age estimation, we should pay more attention to the relationship of face pairs based on the similarity between age differences. Motivated by this, we propose a

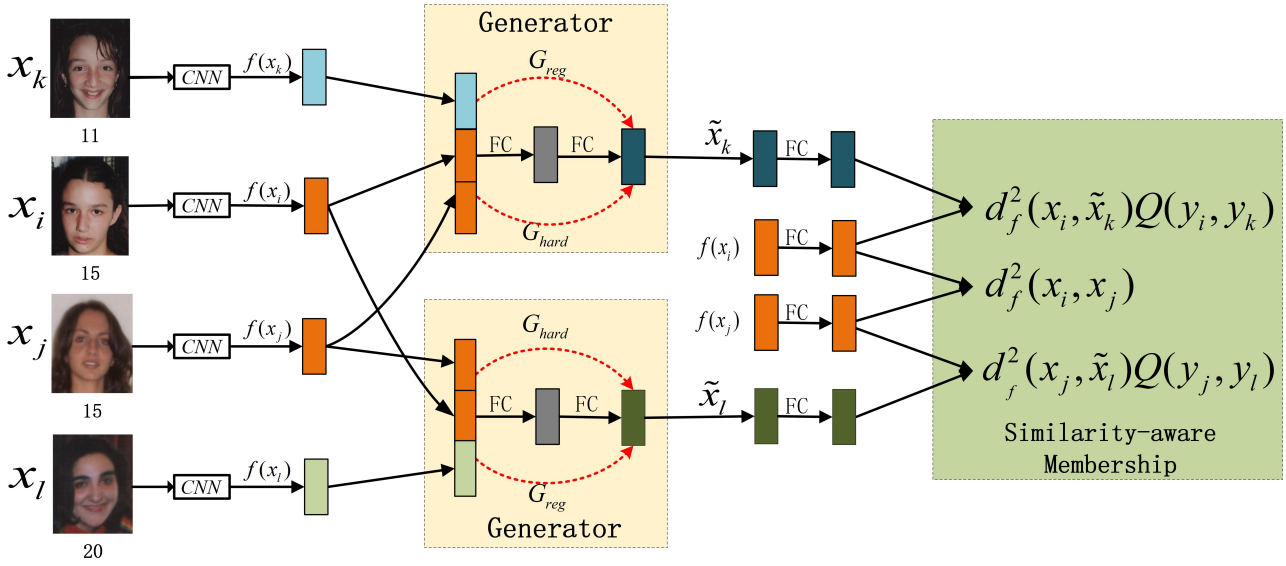


**Fig. 2**. Framework of our proposed SADAL. Our SADAL basically consists of the generator and the discriminator. The generator takes the input with extracted features from a deep CNN. Then it generates discriminative hard-examples to complement the training space for robust feature learning. Accordingly, the discriminator is equipped with a similarity-aware membership function, which aims to identify whether the generated sample is fake or real via a pre-trained age ranker. During the training process, we jointly optimize the generator and the discriminator in an adversarial manner for robust training convergence.

similarity-aware function $\mathrm{Q}(y_m, y_n)$ to dynamically measure the degree of variation face pairs with different age values. With this function, the face pair with a larger age gap has a higher weight than that with a smaller age gap. At the same time, we amplify the differences between the pairs with interval $L$ in the transformed feature representation space by following the Wing-Loss pattern [7]. Fig. 3 illustrates the similarity-aware function to show how they exploit different relations for neighboring age classes. Based on the above senses, we compute the similarity-aware membership for the given face pair as follows:

$$\mathrm{Q}(y_m, y_n) = \begin{cases} \frac{1}{L}\ln(1+\frac{|y_n - y_m|}{\varepsilon}), & |y_m - y_n| \le L, \\ \frac{|y_n - y_m|}{L} - C, & \text{otherwise}, \end{cases} \tag{3}$$

where the non-negative $L$ is a constant parameter that describes the tolerance level of varying age relationship, $\varepsilon$ denotes the curvature of the nonlinear-region $(-L, L)$ and $C = 1 - \frac{1}{L}\ln(1+\frac{|y_n - y_m|}{\varepsilon})$ is a constant that smoothly links the piecewise-defined linear and nonlinear parts, respectively. Obviously, we specify $\mathrm{Q}(y_m, y_n) \ge 0$ by the value 0 only if $y_i$ and $y_k$ are belong to the same age category.

**Objective Formulation:** Traditional hard-negative mining methods such as [19] only focus on seeking hard-negatives from existing samples within the training set, which is limited to describe the full distribution of the hard-negative space. Moreover, it is very challenging to collect images of people of all age progression in the personalized real life. This probably leads to unbalanced data for training, and the potential hard-negatives from the unobserved space would be misclassified which probably gives rise to bias prediction. In our approach, we propose a hard-example generation method, which leverages the observed training samples to produce plausible hard-negatives and provide the complementary information for existing training samples. Each time, we sample a quadruplet $(i, j, k, l)$ of facial images as the input of the CNN, where the quadruplet composes of two positive point $x_j$ and $x_i$ with the same label $y_i = y_j$, and two negative points $x_k$ and $x_l$ under the label-inequality $y_i \neq y_k \neq y_l$. After that, the generator directly concatenates the features $(i, j, k)$ and $(i, j, l)$ as the input to the generator, and outputs $\tilde{x}_k$ and $\tilde{x}_l$ as the generated hard-example as the input to the discriminator. Overall, we formulate our objective function of the hard-example generator as follows:

$$\min_{\theta_g} G_{(i,j,k,l)} = G_{hard} + G_{reg} + G_{adv}, \tag{4}$$

subject to

$$G_{(i,j,k,l)} = \begin{cases} \sum\limits_{(i,j,k)} [\| \tilde{x}_k - x_i \|_2^2 + \| \tilde{x}_k - x_k \|_2^2 \\ \quad + \max(0, d_f(x_i, \tilde{x}_k)[1 - \mathrm{Q}(y_i, y_k)] - \tau)^2], \\ \sum\limits_{(i,j,l)} [\| \tilde{x}_l - x_j \|_2^2 + \| \tilde{x}_l - x_l \|_2^2 \\ \quad + \max(0, d_f(x_j, \tilde{x}_l)[1 - \mathrm{Q}(y_j, y_l)] - \tau)^2], \end{cases}$$

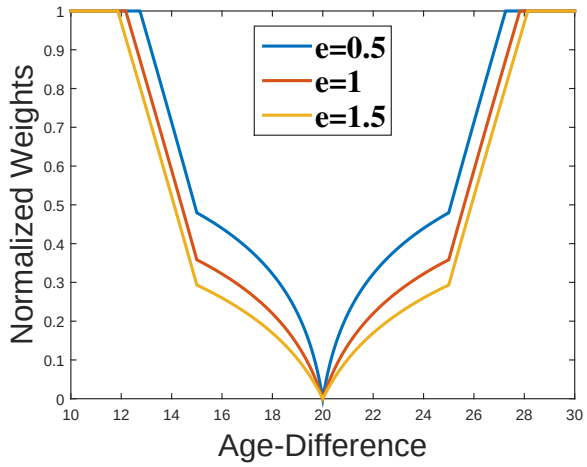


**Fig. 3**. Illustration of the similarity-aware function, which typically exploits the age-difference relations for neighboring age classes. In this example, $y_m$ is set as 20 and $L$ is set as 5. It can be figured out that the larger weight $\mathrm{Q}(y_m, y_n)$ is assigned by the higher difference between $y_n$ and $y_m$. By doing this, the smoothness of age differences is exploited in the transformed subspace.

There are three objectives for (4) accordingly: The term $G_{hard}$ ensures that the generated the samples which are close to the positive sample, leading to adversarial hard-negatives. The term $G_{reg}$ minimizes reconstruction error between the generated samples and the original ones, which exploits the original structure in the transformed subspace. The term $G_{adv}$ enforces our training goal that the similarity distance between the generated negative and the original positive should be smaller than a pre-specified threshold $\tau$ (assigned to 1 in our experiments). Moreover, the function of $[1 - Q(\cdot, \cdot)]$ in $G_{adv}$ is used to exploit the age-difference information smoothly. In this way, the generator captures the full distribution of training data via an adversarial manner. These generate hard-examples likely mislead bias prediction for age estimation. For robust model training, we introduce an adversarial learning method to optimize the generator parameters by playing a dynamic-game with the discriminator, *i.e.*, an age estimator or a discriminative metric subspace learning across iterations.

### 2.2. Deep Adversarial Learning

We formulate the objective function of the discriminator by minimizing the following optimization problem:

$$\min_{\theta_d} D_{(i,j,k,l)} = D_{(i,k)} + D_{(j,l)} + D_{(i,j)} \quad (5)$$

$$= \sum_{(i,j,k,l)} [ \max_{(i,k)\in\widehat{N}} (0, \tau - d_f(x_i, \tilde{x}_k) Q(y_i, y_k))^2$$

$$+ \max_{(j,l)\in\widehat{N}} (0, \tau - d_f(x_j, \tilde{x}_l) Q(y_j, y_l))^2$$

$$+ \max_{(i,j)\in\widehat{P}} (0, d_f(x_i, x_j))^2].$$

There are three objectives for (5): 1) $D_{(i,k)}$ and $D_{(j,l)}$ in (5) denote the similarity-aware membership score between the learned feature vectors of face pairs from different classes. The membership score is enforced larger than the pre-defined threshold $\tau$, and the goal of $D_{(i,j)}$ is to ensure that the similarity of positive face pairs are minimized. Hence, the margin between positive-positive pair is minimized and meanwhile the positive-negative pairs is maximized in the learned subspace. 2) The designed similarity-aware function $\mathrm{Q}(\cdot, \cdot)$ is applied to measure the age-difference information in a ranking-preserving manner. The dissimilarity between negative pairs with a small age gap is smaller than that from a negative pair with a large age gap. Therefore, different weights should be smoothly assigned to different negative pairs according to age-difference values. 3) The training procedure jointly optimizes both tasks of hard-example generator and the discriminator for age classification in an adversarial manner. By employing these objectives in our discriminator, our SADAL achieves a robust training convergence via the widely-used dynamic-game iterations.

## 3. EXPERIMENTS

### 3.1. Experimental Settings and Implementation Details

We evaluated the performance of the proposed SADAL method on two widely-evaluated databases including FG-NET [16] and MORPH (Album2) [14]. For the evaluation metric, we employed the mean absolute error (MAE) [21] to measure the error between the predicted age and the ground-truth. For the employed network, we performed the initialization with VGG Face Net [23] and appended the fully connected layer. We optimized the new layer with ten times learning rate compared with other layers for fast convergence. We also used a 2-layer fully connected network as the generator by concatenating the features as the input. For the parameters of the designed network, we specified the values of the weight decay, moment and learning rate empirically to 0.0001, 0.9 and 0.001, respectively. For the parameters of the similarity-aware function, we empirically specified the values of $L$ and $\varepsilon$ as 5 and 1 to exploit the age-difference information on neighboring age labels, respectively. Besides, we fixed the maximum training iteration to 20,000 and set the batch-size as 64 for the quadruplet input.

### 3.2. Experiments on the MORPH dataset

The MORPH (Album 2) dataset [14] consists of 55608 face images, whose face images were collected from about 13000 people with different races. The age range is from 16 to

**Table 1**. Comparisons of MAEs with different state-of-the-art approaches on the MORPH dataset. The MAE (in years) is reported for each method.

| Method | MAE | Year |
|---|---|---|
| BIF+KNN | 9.64 | - |
| Raw+OHRanker [3] | 7.34 | 2011 |
| LBP+OHRanker [3] | 6.88 | 2011 |
| BIF+OHRanker [3] | 6.49 | 2011 |
| LDL [9] | 5.69 | 2013 |
| CPNN [9] | 5.67 | 2013 |
| CA-SVR [4] | 4.87 | 2013 |
| CS-LBFL [21] | 4.52 | 2015 |
| CS-LBMFL [21] | 4.37 | 2015 |
| CSOHR [2] | 3.74 | 2015 |
| DeepRank [25] | 3.57 | 2015 |
| DeepRank+ [25] | 3.49 | 2015 |
| OR-CNN [22] | 3.27 | 2016 |
| ODFL [20] | 3.12 | 2017 |
| LSDML [19] | 3.08 | 2018 |
| M-LSDML [19] | 2.89 | 2018 |
| **SADAL** | **2.75** | **-** |

**Table 2**. Comparisons of MAEs with different deep learning approaches on the MORPH dataset.

| Method | MAE |
|---|---|
| unsupervised VGG + KNN | 7.21 |
| unsupervised VGG + OHRanker | 4.58 |
| VGG + Single Label | 3.63 |
| VGG + Gaussian Label | 3.44 |
| ODFL [20] | 3.12 |
| **SADAL** | **2.75** |

77 years old and everyone has about 6-15 face images that vary slightly by the factors of facial expressions, background, resolution and illumination. Each face image is labeled with its chronological age value of the corresponding person. For evaluation on the MORPH dataset, we performed 10-folds cross-validation for evaluation by following the settings in [21]. Specifically, we divided the whole dataset into ten folds and each fold has the nearly equal size. We used nine folds as the training set, and the remaining one was used for the testing set. We repeated this procedure 10 times and computed the average results as the final age estimation performance.

**1) Comparisons with the State-of-the-arts:** Table 1 tabulates the MAEs of our SADAL on the MORPH dataset, compared with state-of-the-arts facial age estimation methods. Results of the existing state-of-the-art methods are directly cropped from the related papers. As can be seen, our SADAL achieves competitive performance with the different facial age estimation methods and even obtains better performance than that of the deep learning methods such as DeepRank [25] and OR-CNN [22].

**2) Comparisons with Different Deep Learning Methods:** We also compared our SADAL with different deep learning methods. To be specific, we first employed the pre-trained VGG Face Net [23] without the fine-tuning training as the feature extractors. We created a baseline method with the unsupervised VGG features and KNN. Then, we deployed the softmax loss [15] as the single label method, and the deep label distribution learning [26] as the Gaussian label methods at the top of the VGG Face Net and fine-tuned the network. Table 2 tabulates the performance of different deep learning methods. We see that our model obtains the best performance, which is because the structural ordinal relation is exploited by our model in the learned face feature representation, which takes advantages of full orders in quadruplet comparisons.

**3) Computational Time**: Our proposed approach was implemented under the open-source CAFFE [13] for accelerated deep learning architecture. We trained our model with a speed-up parallel computing technique by using a single GPU with the NVIDIA Titan V. Our proposed SADAL converges at 20000 iterations, which typically consumes around 5 hours for the training convergence.

**Table 3**. Comparisons of MAEs compared with state-of-the-art approaches on the FG-NET dataset. The MAE (in years) is reported for each method.

| Method | MAE | Year |
|---|---|---|
| BIF+KNN | 8.24 | - |
| Raw+OHRanker [3] | 6.25 | 2011 |
| LBP+OHRanker [3] | 4.92 | 2011 |
| BIF+OHRanker [3] | 4.48 | 2011 |
| LDL [9] | 5.77 | 2013 |
| CPNN [9] | 4.76 | 2013 |
| CSOHR [2] | 4.70 | 2015 |
| CS-LBFL [21] | 4.43 | 2015 |
| CS-LBMFL [21] | 4.36 | 2015 |
| ODFL [20] | 3.89 | 2017 |
| LSDML [19] | 3.92 | 2018 |
| M-LSDML [19] | 3.74 | 2018 |
| **SADAL** | **3.67** | **-** |

### 3.3. Experiments on the FG-NET dataset

The FG-NET dataset [16] contains 1002 face images from 82 persons. The age range of this dataset covers from 0 to 69 years old, and every identity has about 12 face samples. Each face image is annotated with its accurate age value of the corresponding person. Since the FG-NET dataset is small, we directly set the learning rate of the 8-10 convolution layer of VGG Face Net to 0 to prevent over-fitting. Note that

we applied the leave-one-person-out protocol to conduct the age estimation experiments. Specifically, we selected the face image of one person as the test set, and the remaining were used for training. After 82 folds, each person has been used as test set once, and the final age estimation performance are computed the average results from all the estimates. Table 3 shows the MAEs of our SADAL compared with the state-of-the-arts. From the results, we see that our SADAL outperforms most state-of-the-arts even with limited training samples, which benefits from the generated hard-examples.

## 4. CONCLUSIONS AND FUTURE WORK

We have proposed a similarity-aware deep adversarial learning (SADAL) for accurate facial age estimation. Experimental results on two datasets show the effectiveness of the proposed approach. It is desirable to address facial age estimation with dual-cycle networks to further exploit the aging pattern specific for personalized aging pattern in future works.

## 5. REFERENCES


[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *TPAMI*, 28(12):2037–2041, 2006.

[2] K. Chang and C. Chen. A learning framework for age rank estimation based on face images with scattering transform. *TIP*, 24(3):785–798, 2015.

[3] K. Chang, C. Chen, and Y. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *CVPR*, pages 585–592, 2011.

[4] K. Chen, S. Gong, T. Xiang, and C. C. Loy. Cumulative attribute space for age and crowd density estimation. In *CVPR*, pages 2467–2474, 2013.

[5] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao. Using ranking-cnn for age estimation. In *CVPR*, July 2017.

[6] Y. Duan, W. Zheng, X. Lin, J. Lu, and J. Zhou. Deep adversarial metric learning. In *CVPR*, pages 2780–2789, 2018.

[7] Z.-H. Feng, J. Kittler, M. Awais, P. Huber, and X.-J. Wu. Wing loss for robust facial landmark localisation with convolutional neural networks. In *CVPR*, June 2018.

[8] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *TPAMI*, 32(11):1955–1976, 2010.

[9] X. Geng, C. Yin, and Z. Zhou. Facial age estimation by learning from label distributions. *TPAMI*, 35(10):2401–2412, 2013.

[10] X. Geng, Z. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *TPAMI*, 29(12):2234–2240, 2007.

[11] G. Guo and G. Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *CVPR*, pages 657–664, 2011.

[12] G. Guo, G. Mu, Y. Fu, and T. S. Huang. Human age estimation using bio-inspired features. In *CVPR*, pages 112–119, 2009.

[13] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[14] K. R. Jr. and T. Tesafaye. MORPH: A longitudinal image database of normal adult age-progression. In *FG*, pages 341–345, 2006.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.

[16] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *TPAMI*, 24(4):442–455, 2002.

[17] K. Li, J. Xing, C. Su, W. Hu, Y. Zhang, and S. Maybank. Deep cost-sensitive and order-preserving feature learning for cross-population age estimation. In *CVPR*, pages 399–408, 2018.

[18] H. Liu, J. Lu, J. Feng, and J. Zhou. Group-aware deep feature learning for facial age estimation. *Pattern Recognition*, 66:82–94, 2017.

[19] H. Liu, J. Lu, J. Feng, and J. Zhou. Label-sensitive deep metric learning for facial age estimation. *TIFS*, 13(2):292–305, 2018.

[20] H. Liu, J. Lu, J. Feng, and J. Zhou. Ordinal deep learning for facial age estimation. *TCSVT*, 29(2):486–501, 2019.

[21] J. Lu, V. E. Liong, and J. Zhou. Cost-sensitive local binary feature learning for facial age estimation. *TIP*, 24(12):5356–5368, 2015.

[22] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua. Ordinal regression with multiple output cnn for age estimation. In *CVPR*, pages 4920–4928, 2016.

[23] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *BMVC*, 2015.

[24] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan. Personalized age progression with aging dictionary. In *ICCV*, pages 3970–3978, 2015.

[25] H.-F. Yang, B.-Y. Lin, K.-Y. Chang, and C.-S. Chen. Automatic age estimation from face images via deep ranking. In *BMVC*, 2015.

[26] X. Yang, B. Gao, C. Xing, Z. Huo, X. Wei, Y. Zhou, J. Wu, and X. Geng. Deep label distribution learning for apparent age estimation. In *ICCVW*, pages 344–350, 2015.