JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 37, 1435-1448 (2021)
DOI: 10.6688/JISE.202111_37(6).0014

Age Estimation Using Correlation-Refined Features of Convolutional Neural Network

JIANN-SHU LEE[1], CHUNG-WEN CHANG[2,+], TENG-WEI KAO[1] AND JING-WEIN WANG[3]
[1]*Department of Computer Science and Information Engineering*
National University of Tainan
Tainan, 700 Taiwan
[2]*Department of Management Information Science*
Chia-Nan University of Pharmacy and Science
Tainan, 717 Taiwan
[3]*Institute of Photonics and Communication*
National Kaohsiung University of Science and Technology
Kaohsiung, 811 Taiwan
E-mail: ccwen@mail.cnu.edu.tw

Age estimation remains challenging because of its high dependence on small facial changes (based on individual propagation, increased wrinkles, and even racial or gender factors). In the past decade, some learning models of neural network based on image analysis have been rapidly developed to overcome such limitations. In this study, we developed a novel method, namely correlation-refined convolutional neural network (CR-CNN), based on some deep learning model (AlexNet). Additional to the parameters in model, the CR-CNN model considers a specific learning network, in which the neuron parameters along with refined facial features at various field-of-view levels have determined through canonical correlation analysis (CCA). Such a novel learning strategy, called low-to-middle-level-features retained transfer learning (LMLFR). Through LMLFR, the feature maps in CNN would be reorganized and join as new layer. That means the maps with high CCA values, in which neurons have high coadaptation with respect to feature-map values, are averaged and flattened; and contrarily, the maps with low CCA values are retained for the low-coadaptation neurons. All refined layers are then subjected to principal component analysis to further reduce dimensionality. At the output layer, classification is executed through support vector regressions (SVR) and Marginal Fisher Analysis (MFA) to overcome the non-Gaussian distributions of refined features on different layers. Experiments were conducted using images obtained from the well-known MORPH dataset, and the results indicated that for age estimation, the proposed model outperformed commonly used methods; the error range was approximately −5 to +5, covering approximately 80% of the learning age range. The proposed model constitutes a novel approach to feature refinement and can potentially become the basis of extensive applications.

Keywords: age estimation, CNN, transfer learning, deep learning, canonical correlation analysis

1. INTRODUCTION

Nonverbal and visual information is mostly conveyed through facial expressions. Differences in age, gender, expression, and even mental or confidence can be determined on the basis of tiny muscle changes or wrinkles using facial expression information. Estimating human age using only face images is problematic because facial features exhibit

Received February 28, 2020; revised June 10 & August 6, 2020; accepted September 10, 2020.
Communicated by Jing-Ming Guo.

extremely small variations and individual conditions vary. Age categorization is feasible for babies, young and elderly people because estimating the categories of such individuals is relatively easy [1]. However, only age ranges with specific individual factors are suitable for such categorization processes. This method for large age ranges cannot provide sufficiently accurate estimations. Recent research proposed strategies which involve multiple classification systems for improving age categorization accuracy [2]. Such strategies entail the use of sequential age patterns to represent individual facial variations, but they usually need highly complicated computing and are marred by the problem of outliers. Subsequent research has proposed an innovative strategy based on ordering and ranking concepts; in this strategy, age estimation is conducted using an ordinary sequence and is considered a double-classification problem [3, 4]. Furthermore, several strategies have been developed for deriving efficient feature maps [5-7].

In general, efficient feature map usually locates at the latter layers and can represent highly semantic image features. In addition, they perform lower co-adaptation action with others, so that those neurons achieve better balance between specificity and generality to acquire higher accuracy during model testing time [8]. The major contributions of this study can be considered as two parts. At first, the proposed CR-CNN innovatively evaluates the efficiency of feature maps in hidden layers based on the CCA values of different filters, and is accordingly determined by the proposed LMLFR strategy based on the AlexNet structure. Secondly, the new layer for feature refinement is added to minimize feature redundancy according to LMLFR result with PCA transformation to achieve efficient dimension reduction. Meanwhile, the SVR and MFA are adopted in classification to tackle non-Gaussian distribution. Accordingly, the CR-CNN successfully achieve a good balance between overfitting and underfitting problems by deriving efficient feature maps to improve model accuracy.

We proposed this paper with following structure: In Section 2, the state-of-the-art CNN-related methods for age estimation have been introduced; in Section 3, the proposed method has been described; in Section 4, the materials and experiments are discussed, and later conclusions and future works are presented in Section 5.

2. RELATED WORK

Age estimation procedures can be divided into two parts: age feature extraction and feature-based age estimation. Regarding age feature extraction, Kwon and Lobo have applied craniofacial development theory and skin wrinkle analysis to estimate human age [5, 9]. However, these methods cannot provide sufficient accuracy, and their applications are limited. Guo investigated biologically inspired features using a pyramid-type Gabor filters to obtain a wider range of features in order to partially overcome personal aging variations in cross-race age estimation [6].

To overcome age estimation problems, several CNN-related age estimation methods have been developed. Geng developed an innovative approach to model aging patterns through a time-order sequence of face images of individuals, and called as aging pattern subspace (AGSE) [2]. Another study considered a multitask learning problem and accordingly developed a general age estimation model called multi-task warped Gaussian process (MTWGP) [10]. In this model, a kernel function and Gaussian mixture model are employ-

ed to merge the properties of individually learned parameters. Additionally, because age data has sparsely distribution and the its use in training would affect estimation accuracy, some continuous-type data must be extracted (such as ranking or cumulative data). Chang [4] adopted relative ordering information (rather than age data) with respect to labeled data in a dataset and divided an ordered hyperplane into two groups to simplify the multiple-classification problem (OHRANK).

In the last decade, researchers have developed age estimation strategies after considering various aspects [11, 12]. Studies have explored approaches for determining optimal network structures and using dataset characteristics as prior knowledge (*i.e.*, using pre-training parameters) in order to develop innovative networks. However, optimal structures that can change hidden activities in networks and the mechanisms underlying the function of such structures have yet to be resolved. Liu [11] adopted a compound structure of deep CNNs with end-to-end approach, called ageNet, in which the parameters should be pre-trained initially using large-scale datasets. In ageNet, the classification and Gaussian-based regression are combined to solve estimation problem. The general-to-specific deep transfer leaning plays a crucial role in decreasing overfitting, and thus increases estimation accuracy. In experiments, ageNet outperformed several famous strategies on benchmark databases. Furthermore, Yosinski discussed the benefits of transfer learning by conducting some representative experiments [8] and determined that coadaptation between neurons can be reduced through such learning. He reported the necessity of reducing overfitting through the use of strategies such as dropout approaches, transfer learning, and decorrelation neural networks [8, 12, 13]. Accordingly, this study developed a CR-CNN system in which pre-training parameters are determined using a general-purpose dataset for transfer learning. The study subsequently investigated the coadaptation of feature maps using correlation information. The results indicated that the proposed method outperformed some famous methods.

3. METHODS

3.1 CNN Models

CNNs are commonly used in many fields because of their high efficiency and accuracy in modeling complex classification and segmentation problems. In general, CNN contains convolution and full-connected layers with interlacing positions, with applying forward and back propagations by stochastic gradient descent optimization. Recently, in deep learning strategy, in order to overcome highly variant features in facial images, scholars have recommended the use of additional layers which could extract geometry-invariant and location-invariant properties of features at different levels. Research thus applied some benchmark structures, such as AlexNet, to explore strategies and reported inspiring results [14].

Basically, there are three stages in a basic CNN, namely a convolution layer, a pooling layer, and active function transformation. The convolution layer plays a crucial role in the integration of neighboring data filtered using specific weighting masks (called kernels). Different kernels are used to apply to the neuros of current layer and would conduct various feature maps on subsequent layers. Thus, image characteristics can be efficiently extracted; for example, low-level visual features such as edges or shapes, and high-level visual fea-

tures such as regions or structures can be extracted. The pooling layer plays a role in subsampling for dimensional reduction. In our proposed CNN, a rectified linear unit, which can provide faster network convergence, is applied as the activation function [15]; a max polling strategy is also used in the CNN. Thus, in a fully connected output layer, a softmax function is applied to restrict node weightings to a probability value ranging from 0 to 1. A back-propagation algorithm computes changes in node weightings (θ) to minimize the cost function ($J(\theta)$).

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_i^T x(i)}}{\sum_{l=1}^{k} e^{\theta_i^T x^{(i)}}}\right]$$

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[x^{(i)}(1\{y^{(i)} = j\} - p(y^{(i)} = j \mid x^{(i)}, \theta))]$$

$$\theta_{t+1} = \theta_t - \gamma \nabla J(\theta_t)$$

where m is the number of data, and γ is the learning rate.

3.2 Network Structure

The proposed network structure is based on AlexNet, which has been reported to exhibit high performance on several benchmark databases. The proposed network comprises an input layer (with a 227×227 RGB image being the input data), seven convolution layers, and three fully connected layers including the output layer (Fig. 1). In the first convolution layer, a kernel with an 11×11 convolution mask of stride 4 and 96 specified filters with various weighted kernels are applied. Thus, 96 feature-map images measuring 55×55 are obtained. The second layer reduces the feature-map size to 27×27 using the max pooling strategy. Moreover, the convolution layer and the pooling layer have been arranged as interpolated positions from three to seven layers. Subsequently, the feature map is then flattened as the fully connected layer, and the total number of neurons is 4096. Finally, the third fully connected layer exhibits 80 neurons, representing activations for ages ranging from 1 to 80 years.

3.3 Training Strategy

The general features are perceived in a small field of view (FOV) in human visual system. The features fit low-level visual perception and thus generally appear in wide-range images, and called the low-level features (*i.e.*, that indicate simple edges or points in image). On the other hand, the specific features perform in a large FOV and only appear in specific-domain images, called high-level features (*i.e.*, that contain compound structural features of image).

In order to achieve accurate estimation, the applied training strategy must take into accounts both underfitting and overfitting problems. For avoiding overfitting problem, more data with general features, as training data is necessary; but regarding to the underfitting problem, specific features have been assigned. During training model, these two problems would be determined as from general features in the shallow layers to specific features learnings in the deeper layer. Thus, to select proper combination of layers so as to make good balance between two problems is crucial in model structure. In a CNN, low-

level features are learned on shallow network layers and high-level features are learned on deep layers [16]. Furthermore, due to high sparsity of the aging data, few training data for some age ranges may conduct specific features on layers to aggravate overfitting problem. Thus, we propose transfer learning for model training.

As we known, skin color, wrinkle, beard, and musculature features vary in aging humans; and small changes in low-level features are thus considerably less discriminative. We attempt to increase the discrimination by using the pre-trained parameters on larger image dataset (called the transfer learning) in from low-level to mid-level feature maps. The proposed model is based on an 8-layer AlexNet framework, whose parameters are acquired from the ImageNet dataset with approximately 15,000,000 labeled images over 22,000 image classes [17]; therefore, the parameters are highly suitable for general images.

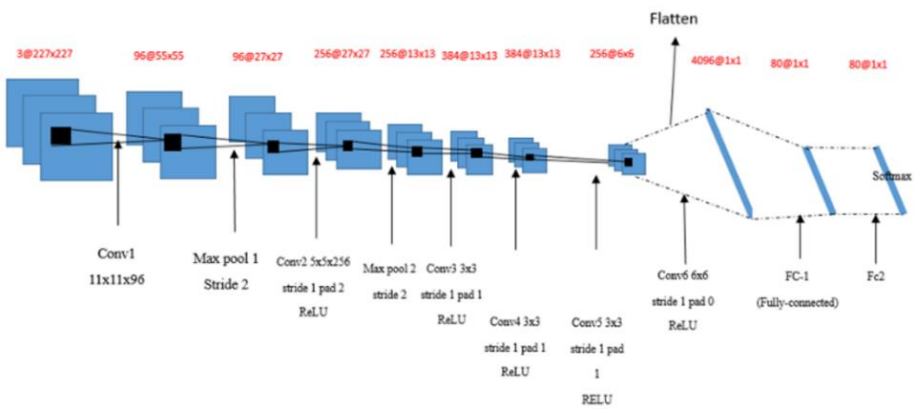


Fig. 1. Proposed CR-CNN based on an AlexNet structure.

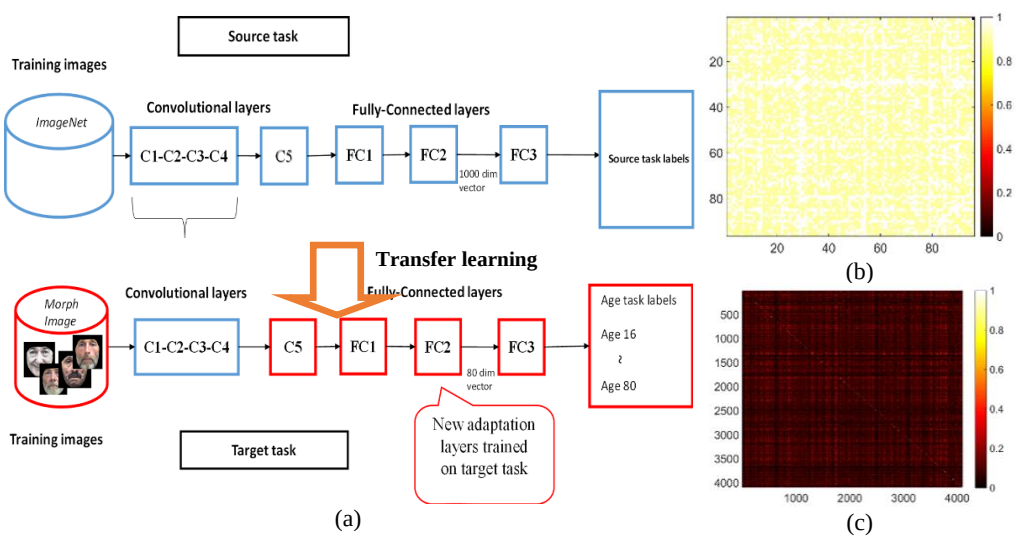


Fig. 2. (a) LMLFR processing; In the proposed CR-CNN, layers C1-C4 are used for training through transfer learning using pre-trained parameters from other dataset; while the C5, FC1-FC3 are new

trained layers; (b) CCA hot map of layer C1 and (c) CCA hot map of layer FC1.

Specifically, the parameters acquired from the training results of ImageNet can be applied to all convolution layers to determine the number of layers that should be used as transfer learning. Accordingly, we conducted experiments to determine the suitable number of layers with transfer learning; and the results revealed that the C1-C4 layers could be used by transfer learning because such network structure engendered the lowest model error (Fig. 2).

3.4 Feature Refinement

In CNN models, the last two layers are mostly extracted as feature mapping layers. This approach may not be applicable to all physical situations. As reported by Zeiler [16], with image processing aspect, some individual-layer neurons in CNN would perform especially well for representing specific high-level or low-level visual features. That means it is potentially improve model performance by evaluating the performance of different-layer feature mappings and then to refine the resultant mapping diagrams. This is also the key inspiration of the proposed method. Furthermore, Yosinski [8] figured out the coadaptation in hidden-layers activities can affect the estimation accuracy. Hence, some decorrelation in loss function have been adopted later [14]. Accordingly, we propose our innovative system to address coadaptation problems and thus refine features in layers.

Because featured-mapping diagrams would be produced in the consequence of convolution layers, the first stage of refinement, as shown in Fig. 2, is to evaluate the correlation and dependency between feature maps. Accordingly, a study proposed a canonical correlation analysis (CCA) approach, which has been successfully used to evaluate affine-invariant features in image processing [17]. In one layer, to extract low-level visual features (*i.e.*, edges or textures), various filters can be applied in specific directions to derive feature maps. As an image object is usually composed of many simple structural edges or textures, they exhibit affine invariant and be evaluated by CCA (rather than normal correlation analysis). Thus, neurons associated with such low-level features (extracted by different kernels) exhibit high coadaptation because they all tend to exhibit affine invariant properties of the same image object, with respect to CCA. Notably, the CCA approach does not apply to high-level visual features because they exhibit no properties of direction or affine invariants.

Feature refinement is typically executed to determine feature map diagram with minimal interneuron feature redundancy. Therefore, we can compute CCA values to evaluate the coadaptation degree between low-level feature maps of different filters in each layer. The computation of CCA values is described as follows. For paired data sets X (with p dimension) and Y (with q dimension), which are also called random variables in probability (*i.e.*, two feature maps of different filters in the proposed network), we may find two coefficient vectors A and B, such that the linear-mapped random variables A^TX and B^TY achieve maximum correlation (ρ). In equations, $Var(\cdot)$ and $Cov(\cdot)$ represent the variance and the covariance matrices respectively. With new notations $X^* = A^TX$ and $Y^* = B^TY$ applied, the found coefficients must satisfy the conditions $Var(X^*) = A'\sum_{XX}A = 1$ and $Var(Y^*) = B'\sum_{YY}B = 1$ (where the $\sum_{MN}$ indicates the covariance of M and N, with $M, N \in \{X, Y\}$). Then the CCA value is computed and denoted as $\rho(X^*, Y^*)$.

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \; Y = \begin{pmatrix} y_1 \\ \vdots \\ y_q \end{pmatrix} \; A = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} \; B = \begin{pmatrix} b_1 \\ \vdots \\ b_q \end{pmatrix}$$

$$X^* = a_1 * x_1 + \ldots + a_p * x_p$$

$$Y^* = b_1 * x_1 + \ldots + b_q * y_q$$

$$\rho(X^*, Y^*) = \frac{Cov(X^*, Y^*)}{\sqrt{Var(X^*) \times Var(Y^*)}} = \frac{A' \sum_{XY} B}{\sqrt{A' \sum_{XX} AB' \sum_{YY} B}}$$

$$(A, B) = \arg\max_{A,B} \rho(X^*, Y^*)$$

CCA can be used to determine the correlation between sets of high-dimensional data. However, to correctly acquire such two vectors, as reported previously [17], the eigenvectors of the covariance of the two data sets X and Y must be used. Figs. 2 (b) and (c) illustrate a hot map of CCA values in two feature mapping diagrams for a specified network layer. For example, using 96 kernels can result in 96 maps in the C1 layer; thus, the CCA hot map exhibits 96 × 96 values. In the proposed network structure, C1-C5 are shown to exhibit high CCA values. By contrast, the CCA values derived for the fully connected layers are shown to be low in Fig. 2 (c).

In common, the high coadaptation neuros in neural network would reveal much redundant parameters to reduce the model training and testing performance. That means too much highly-correlated neurons would strongly force model into over-fitting performance. Thus, some solutions of dropping node have been developed. Aruni conducted a study on a CNN model [18] and showed that the duplication reduction would efficiently improve the performance for CNN models.

To achieve efficient feature refinement in CNN models, the diversity of feature maps should be ensured. Accordingly, in the proposed method, feature maps are re-arranged according to their CCA values. One diagram of averaging high-CCA featured maps would be acquired, because it can decrease redundant information to alleviate overfitting. That means only one 55×55 featured-mapping diagram is acquired from averaging the original 96 diagrams (55×55) in the proposed system. We call that as the averaged feature map (AFM). Moreover, some degree of dimensional reduction is achieved. By contrast, for fully connected layers, future maps must be preserved because they have low CCA values. Such maps can be cascaded to produce a cascaded feature map called designated concatenated feature (CF). Consequently, eight (C1-C5 AFMs and FC1-FC3 CFs) rearranged feature maps are obtained from the inner layers of the CNN model; for each map, the principal component analysis (PCA) can be conducted for dimensional reduction. Fig. 3 presents vectors with maximum eigenvalues, which are representative feature maps of individual AFMs and CFs. All representative maps can be cascaded and flattened to form a feature refinement layer. Considering the non-Gaussian distribution of such cascaded feature data, the marginal Fisher analysis (MFA) can be applied for efficient dimensional reduction, resulting in data called MFA features [19]. Finally, support vector regression (SVR) can

be applied to map the refinement layer to the output layer as the age data, as illustrated in Fig. 3 (c).

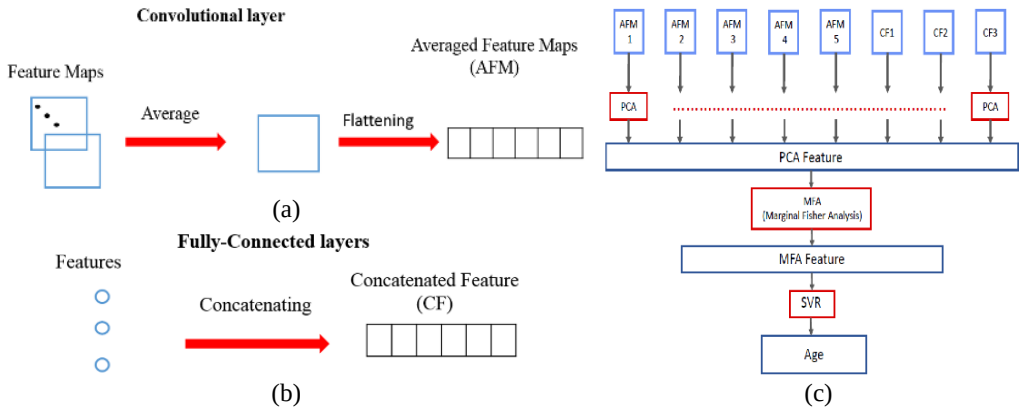


Fig. 3. (a) Refinements for low-CCA layer (map to AFM) and (b) high-CCA layer (map to CS); (c) Contents of feature refinement layer (MFA feature) and associated methods.

4. EXPERIMENTS AND RESULTS

To demonstrate the performance of our method, we conducted experiments to assess age estimation accuracy by using mean absolute error (MAE) and cumulative score (CS), two commonly used measures. MAE is based on the absolute error between actual and estimated ages. CS represents the ratio of data inside a target range; that is, in a selected error range, CS represents the ratio of estimated ages for which the errors are lower than the established tolerance level. These measures can be derived as follows:

$$MAE = \frac{1}{N}\sum_{n=1}^{N} | g_n - e_n | \qquad CS_a = \frac{N'_a}{N} \times 100\%$$

where g_n and e_n denote actual and estimated ages, respectively, N denotes the number of data sets, and N'_a denotes the number of estimated data sets within the error bound of $\pm a$ (called the tolerance-error range).

4.1 MORPH Dataset

We used the MORPH dataset as the benchmark in our experiments [9]. This dataset contains 55132 face images of 13618 individuals aged 16-77 years. Four images are available for each person, and the dataset includes various sexes and races.

4.2 Data Description

For fair comparisons, we used the same data setting as did previous works (Table 3). We selected images of white people only from the MORPH dataset, collecting 5475 face images of 2648 people. We randomly selected 80% of the images as training data, with the remaining 20% serving as testing data. The training and testing data exhibited no overlap,

and about 80% data are selected from the range of 20-40 ages. Moreover, only few of the selected people were aged > 50; this could thus affect our model's accuracy in estimating their ages.

4.3 Preprocessing

Preprocessing was conducted to determine regions of interest (ROIs) in face images. We employed the active appearance model (AAM) algorithm (implemented using Open CV Library) and then developed a convex hull as an ROI based on AAM landmarks [20]. Subsequently, face images were preprocesses for scaling and rotating to obtain a patch of normalized data.

4.4 Instruments

The hardware comprised a Windows 7 personal computer with a Core i7-3770 3.40 Hz CPU with a GPU of GTX970 4 GB. Matlab and C++ programs constituted the software. In the proposed CR-CNN, each model spent > 20 min, on average, in the training stage.

4.5 Experimental Results

The proposed CR-CNN includes several layers for transfer learning. For training, pre-training parameters for low-level and mid-level image features (hereafter designated as LMLFR) were used. However, we had to determine the optimal layers for use in transfer learning. Accordingly, we applied various layers (none, C1 to C2, C1-C3, C1-C5, and additionally FC1) along with their relevant parameters and evaluated the associated learning time and MAEs. The experimental results indicated that the C1-C4 layers were associated with the lowest MAE in the MORPH dataset (MAE = ~4.49); therefore, these layers were used for transfer learning.

Table 1. Evaluations for the determination transfer-learning layer.

Transfer-learning layers	All	None	C1, C2	C1-C3	C1-C4	C1-C5,	C1-C5, FC1
Random weightings	None	All	C3-C5, FC1-FC3	C4-C5, FC1-FC3	C5, FC1-FC3	FC1-FC3	FC2-FC3
Adjust transfer learning weightings	None	None	None	None	None	None	None
Adjust random weightings	None	Yes	Yes	Yes	Yes	Yes	Yes
MAE	14.75	6.8	5.46	4.77	**4.49**	4.99	5.34
Learning time	0	180	90	60	20	5	4 (mins)

Immediately after determining the layers for LMLFR, the experimental data have been designed. We divided all images randomly into five parts, thus obtaining five data sets for model training. In each training trial, four of the five data sets were considered training data and remaining one was considered testing data. Thus, five training trials were conducted by varying the data sets that served as training or testing data. Model accuracy was evaluated using MAE and CS. The mean, standard deviation (std), and std of error

mean obtained are presented in Table 2. We also compared the proposed CNN without CR (f_error) and with CR (f_CR_error). Table 2 presents the CS results. Subsequently, to validate the significance of refinement, the paired t test was performed for the five trials, as presented in the last two columns of Table 2. For example, the *t* value was 8.639 for trial 1, which was significant ($p < 0.05$), indicating that the difference between f1_error and f1_CR_error was significant (two-tailed significance and 95% confidence interval).

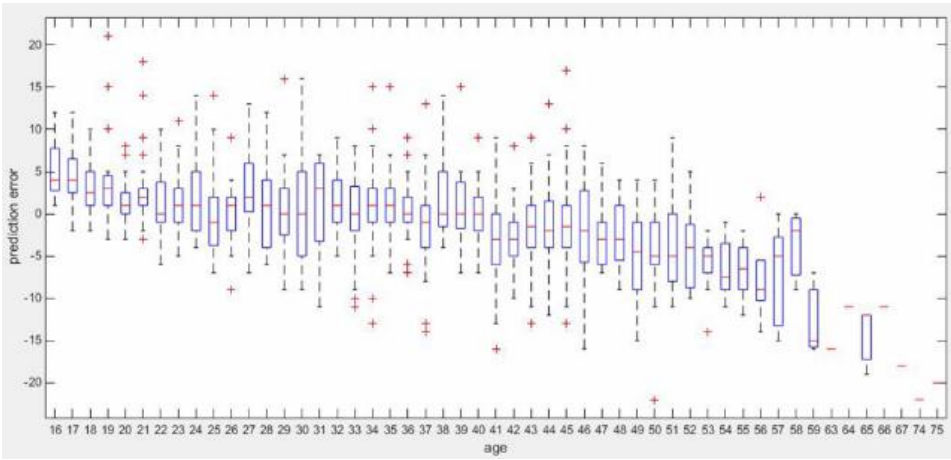


Fig. 4. Error distribution observed for proposed method.

Fig. 4 illustrates box plots of the error distribution for the proposed method, indicating mean, third quartile, first-quartile, minimum, and maximum values. The MAE distributions for all ages ranged from +5 to −15. On average, approximate 80% ranges of the whole data performed good enough results with the MAE ranged from −5 to +5. In addition, the MAE values were large for people aged >55 years because of the imbalance in age data used for training, thus inducing overfitting, which prevented thorough model training. The problem of overfitting could be overcome by using an adequately large data set, and then the MAE may decrease as the number of data increased expectedly.

5. DISCUSSION AND CONCLUSION

As revealed by the experimental results, adding feature refinement layers in the proposed CNN models enhances model performance. For the five data sets used in our experimental trials, the CR-CNN always outperformed the CNN without CR. In all trials, the proposed CR-CNN exhibited smaller MAE, std, and mean error std values than did the CNN without CR. The lower std indicates the higher convergence of the model. Additionally, the lower mean error std signifies higher model performance. This finding indicates that most cases presented near-zero mean error, signifying high model stability. Compared with other commonly-used models, the MAE value of the proposed method was only 3.92 (Table 2).

Table 2. MAE values for five trials data and (two columns on right side) paired t test conducted to compare performance of proposed CNN with and without refinement.

		Mean	m	Std.	Std. Error Mean	Mean of difference	Significance $t(p)$(2-tails)
Trial1	f1_error	4.4927	1094	4.03659	.12204	.58501	8.6 (< 0.05)
	f1_CR_error	**3.9077**	1094	3.55478	.10747		
Trial2	f2_ error	4.4779	1084	3.86457	.11738	.54982	7.7 (< 0.05)
	f2_CR_error	**3.9280**	1084	3.41719	.10379		
Trial3	f3_ error	4.4687	1118	3.98435	.11916	.52236	6.1 (< 0.05)
	f3_CR_error	**3.9463**	1118	3.62286	.10835		
Trial4	f4_error	4.5262	1108	3.99890	.12014	.63448	6.7 (< 0.05)
	f4_CR_error	**3.8917**	1108	3.41221	.10251		
Trial5	f5_error	4.5000	1062	4.01311	.12315	.52448	6.5 (< 0.05)
	f5_CR_error	**3.9755**	1062	3.44026	.10557		

Table 3. Comparison with other methods.

Method	MORPH (measured in MAE)
AGES.	8.83
SVR	5.77
MTWGP	6.28
OHRank	5.69
CA-SVR	5.88
DLA	4.77
Proposed	**3.92**

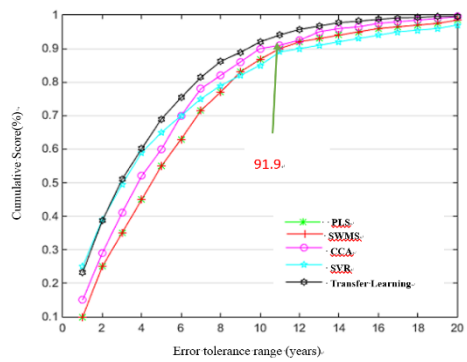


Fig. 5, Ratio of acceptable estimation errors (*i.e.*, error not larger than the tolerance-error range) obtained by setting specified CS.

CS can be used to determine ratios of acceptable estimation errors (error must be within the tolerance-error range), as illustrated in Fig. 5. The proposed CR-CNN had the highest performance in the various ranges of estimation errors compared with partial least squares (PLS), support vector machine (SVM), CCA, and support vector regression (SVR) methods. For example, when the target range of CS was 10 (estimation error < 10), the ratio of acceptable estimation results was 91.9% for the selected MORPH dataset.

This study developed a novel deep-learning structure (CR-CNN) for age estimation based on AlexNet and determined that this method outperformed other famous methods. The proposed CR-CNN includes a new strategy of transfer learning (LMLFR), which is used to achieve balance between overfitting and underfitting. To extract efficient (less coadaptation) and low-dimensional feature maps, the diversity of features is determined by evaluating CCA values in different layers. For each average and cascaded maps, the PCA is applied for first-stage dimension reduction. Then, dimension of the final feature layer (obtained by combining feature maps to achieve a non-Gaussian distribution) is also further reduced using MFA. The subsequent output is estimated using SVR. Experiments validated the accuracy and efficiency of the proposed CR-CNN.

Because the data used in this study were obtained from the MORPH dataset, the CR-CNN model was determined to be suitable for white people and may not applicable in age estimation for people of other races because the age properties of faces can be different from those in the training data. Thus, retraining the model using data for people of different races is necessary to acquire an accurate age estimation model. In addition, the robustness of the proposed CR-CNN can be validated using other datasets in the future. Thus, the proposed CR-CNN model may provide a feasible approach for age estimation in many fields.

ACKNOWLEDGMENT

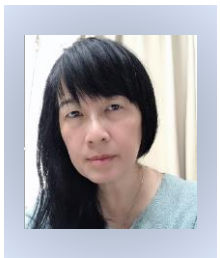
This work was supported by the Ministry of Science and Technology under grant number MOST 108-2221-E-024-011-MY3.

REFERENCES

1. H. K. Young and N. da V. Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, Vol. 74, 1999, pp. 1-21.
2. X. Geng, Z. H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, 2007, pp. 2234-2240.
3. K. Chen and S. Gong, "Cumulative attribute space for age and crowd density estimation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2467-2474.
4. K. Y. Chang, C. S. Chen, and Y. P. Hu, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 585-592.
5. N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, 2006, pp. 387-394.
6. G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 112-119.
7. X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *Proceeding of IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 534-541.
8. J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, Vol. 27, 2014, pp. 3320-3328.
9. "MORPH non-commercial release whitepaper," http://www.faceaginggroup.com, 2007.
10. Z. Kuang, C. Huang, and W. Zhang, "Deeply learned rich coding for cross-dataset facial age estimation," in *Proceedings of IEEE International Conference on Computer Vision Workshops*, 2015, pp. 96-101.

11. X. Liu, S. Li, M. Kan, J. Zhang, S. Wu, W. Liu, H. Han, and S. Shan, "AgeNet: Deeply learned regressor and classifier for robust apparent age estimation," in *Proceedings of IEEE International Conference on Computer Vision Workshop*, 2015, pp. 16-24.
12. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
13. F. Dornaika, I. Arganda-Carreras, and C. Belver, "Age estimation in facial images through transfer learning," *Machine Vision & Applications*, Vol. 30, 2019, pp. 177-187.
14. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, Vol. 60, 2012, pp. 84-90.
15. K. Hara, D. Saito, and H. Shouno, "Analysis of function of rectified linear unit used in deep learning," in *Proceedings of International Joint Conference on Neural Networks*, 2015, pp. 1-8.
16. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional neural networks," in *Proceeding of the 13th European Conference on Computer Vision*, 2014, pp. 818-833.
17. Y. Horikawa, "Use of autocorrelation kernels in kernel canonical correlation analysis for texture classification," in *Proceedings of International Conference on Neural Information Processing*, LNCS, Vol. 3316, 2004, pp. 1235-1240.
18. R. C. Aruni, P. Sharma, and E. Learned-Miller, "Reducing duplicate filters in deep neural networks," *NIPS Workshop on Deep Learning: Bridging Theory and Practice*, Vol. 1, 2017.
19. D. Xu, S. Yan, D. Tao, S. Lin, and H. J. Zhang, "Marginal fisher analysis-based feature extraction for identification of drug and explosive concealed by body packing," *IEEE Transactions on Image Processing*, Vol. 11, 2007, pp. 2811-2821.
20. A. U. Batur and M. H. Hayes, "Adaptive active appearance models," *IEEE Transactions on Image Processing*, Vol. 14, 2005, pp. 1707-1721.

Jiann-Shu Lee received his Ph.D. degree in Electrical Engineering from National Cheng Kung University. He is specialized in multimedia signal analysis, image processing, medical image processing, and computational intelligence and he is also the author of many academic articles, including publications in journals such as IEEE Transactions on Image Processing, IEEE Transactions on Information Technology in Biomedicine and Pattern Recognition. He has served as Session Chairs and Program Committees in many domestic and international conferences. He has also received numerous awards and certificates from organizations such as Cisco, Acer, IAENG, IIHMSP, and Xerox. Currently, he is a Professor with the Department of Computer Science and Information Engineering, National University of Tainan, where he also serves as the Director of Computer Center.



Chung-Wen Chang received her Ph.D. degree in Computer Science and Information Engineering from Cheng Kung University. She is specialized in image processing, medical image processing, and some cross-domain topics, including color science and hand function analysis in rehabilitation. Currently, she is an Associated Professor with the Department of Management Information Science in Chia Nan University of Pharmacy and Science. Her current research interests include the IoT with data analysis, AI and machine learning with their applications.

Teng-Wei Kao is specialized in image processing and computational intelligence. Currently, he is a graduate student with the Department of Computer Science and Information Engineering, National University of Tainan.

Jing-Wein Wang received B.S. and M.S. degrees in Electrical Engineering from the National Taiwan University of Science and Technology, in 1986 and 1988, respectively. He received a Ph.D. degree in Electrical Engineering from National Cheng Kung University, Taiwan, in 1998. He was a chief project leader at the Equipment Design Center of PHILIPS, Taiwan, from 1992 to 2000. In 2000, he joined the faculty of the National Kaohsiung University of Science and Technology and served as the Dean of the College of Electrical Engineering and Information Science from August 2017 to July 2019. Currently, he is a Distinguished Professor at the Institute of Photonics Engineering. His current research interests include automated optical inspection, pattern recognition, deep learning with their applications.