DOI: https://doi.org/10.15276/aait.07.2024.6
UDC 004.93

Xception transfer learning with early stopping for facial age estimation

Marina V. Polyakova[1)]
ORCID: https://orcid.org/0000-0001-7229-7657; marinapolyakova943@gmail.com. Scopus Author ID: 57017879200
Vladyslav V. Rogachko[1)]
ORCID: https://orcid.org/0009-0005-9691-445X; rohachko.8089583@stud.op.edu.ua.
Oleksandr H. Nesteriuk[1)]
ORCID: https://orcid.org/0000-0002-0806-8259; nesteryuk@op.edu.ua. Scopus Author ID: 57207314663
Natalia A. Huliaieva[1)]
ORCID: https://orcid.org/0000-0001-6485-3094; gulyaeva.nata72@gmail.com. Scopus Author ID: 57202228675
[1)] Odessa National Polytechnic University, 1, Shevchenko Ave. Odessa, 65044, Ukraine

ABSTRACT

The rapid development of deep learning attracts more attention to the analysis of person's face images. Deep learning methods of facial age estimation are more effective compared to methods based on anthropometric models, models of active appearance, texture models, subspace of aging patterns. However, deep learning networks require more computing power to process images. Pre-trained models do not need a large training set and their training time is less. However, the parameters obtained as a result of transfer learning of the pre-training network significantly affect its efficiency. It is also necessary to take into account the properties of the processed images, in particular, the conditions under which they were obtained. Recently, the facial age estimation is implemented in applications in devices with limited resources of computing, for example, in smartphones. The memory size and power consumption of such applications are limited by the computing power of mobile devices. In addition, when photographing a person's face with a smartphone camera, it is very difficult to ensure the uniform lighting. The aim of the research is reducing the error of facial age estimation from uneven illuminated images by applying an early stopping of transfer learning of the Xception network. The proposed technique of transfer learning includes an early stopping of training, if the improvement of the results is not observed within a certain number of epochs. Then the network weights from the epoch with the lowest validation loss are saved. As a result of the proposed technique applying, the average absolute error of age estimation was about five years from unevenly illuminated test images. A number of parameters of the used in this case Xception network is less than that of other deep learning neural networks which solved the age estimation problem. Then applying of the Xception network reduces the resource consumption of devices with limited computing power. Prospects for further research are reducing the unevenness of facial image lighting to decrease the error of age estimation. Also, to reduce the computing resources, it is promising to use fast transforms in the Xception convolutional layers.

Keywords: Facial age estimation; Xception; parameter tuning; early stopping; uneven illuminated images; transfer learning

For citation: **Polyakova M. V., Rogachko V. V., Nesteriuk O. H., Huliaieva N. A. "Xception transfer learning with early stopping for facial age estimation".** ***Applied Aspects of Information Technology.*** **2024; Vol. 7 No. 1: 69–80. DOI: https://doi.org/10.15276/aait.07.2024.6**

INTRODUCTION

Content access control, targeted marketing, soft-biometrics systems, and other numerous applications with human-computer interaction employ age estimation (AE) [1, 2]. AE helps to confirm a person's age before granting them access to age-restricted content or services. At last time, more and more transactions occur online. Then in such industries as gaming, e-commerce, and alcohol/tobacco sales, AE has become increasingly important. Businesses that fail to estimate the age of their customers properly risk damaging their reputation because of violating age-restriction laws [3].

Businesses can gain insights into the preferences and behaviors of different age groups; hence AE can help businesses to understand their customer base better.

Collected age-related data can be used to improve marketing efforts and ultimately increase revenue having a higher return on marketing investment [3].

Facial biometric analysis is used to estimate person's age without requiring the user to provide an identity document (ID). AE without asking for ID facilitates access to goods or services for consumers who do not wish to provide their ID, perhaps for privacy or ethical reasons. Facial biometric analysis also allows estimating the age of the user, if he does not have an ID [4].

In addition, age plays an important role in social interaction. Very often, when addressing elders, a different vocabulary is used compared to youth [5].

Age can be determined by the appearance of a person's face. However, compared to other facial attributes, such as race and gender, AE is influenced by underlying human factors such as living conditions and genetics. This uncertainty makes age-related face image analysis (including AE, age synthesis, and age-invariant face recognition) still an unsolved problem. In particular, AE consists in determining a specific age based on facial images. Age synthesis is related to the rendering of facial images with the effect of natural aging or rejuvenation. Age-invariant face recognition involves correct recognition of the person's face regardless of his age [6].

The paper is devoted to the problem of facial AE. The main characteristics of the methods for solving this problem are AE error and the number of parameters affecting the resource consumption [7]. It is also necessary to take into account the properties of the processed images, in particular, the conditions under which they were obtained. Recently, the facial AE is implemented in applications in devices with a limited source of computing, for example, in smartphones. The memory size and power consumption of such applications are limited by the computing power of mobile devices. In addition, when photographing a person's face with a smartphone camera, it is very difficult to ensure the uniform lighting. There is a need to tune the existing methods of facial AE taking into account these limitations.

1. ANALYSIS OF RECENT RESEARCH AND PUBLICATIONS

The literature review shows that handcrafted models and models based on deep learning are used for facial AE [1].

At the first approach, the researcher, based of his own experience, extracts aging features of facial images. This process involves the use of special filters, such as the Gabor filter or the Sobel filter, as well as histograms of oriented gradients, separatly or in combination [7]. Filters are tuned to extract features such as wrinkles, head shape, textures and contours that indicate a person's age. Methods using handcrafted features require less computing power than deep learning methods because they are not as complex. However, one of the disadvantages of these methods is usually significant error of AE [1].

The handcrafted models include anthropometric and texture models, active appearance models, subspace of aging models (AGES) [1].

Anthropometric models are based on increasing of the radius of the circle of the person's face over time [1]. They are used to quantitatively measure the size of bones, muscles and the general structure of the human body. These measurements create a geometric representation of the human body and can be useful for distinguishing between different age groups and sexes [1].

In contrast to anthropometric models, which use the distance between points on the face, texture models depend on the properties of face image textures [7, 8]. In these models, different operators are applied to analyze image textures, such as local binary patterns or biologically inspired features of skin regions, which can display spots, lines or contours indicating the presence of wrinkles.

Active appearance models are statistical models that combine anthropometric and texture facial features. Dimensionality reduction methods, such as principal component analysis, process extracted features of texture and shape. However, reducing the dimensionality of the feature space may lead to ignoring the some features of aging, such as wrinkles [9].

A subspace of aging models (AGES) was elaborated based on a set of face images taken at ages 2, 4, and 8 and sorted by increasing age. This approach helps to fill in missing age data by research the formation of a sequence of facial images of the same person as they age. However, AGES does not process effectively such aging features as wrinkles, so it is combined with texture models [1, 7].

Deep learning models automatically extract aging features from facial images. This usually allows achieving a lower AE error compared to the handcrafted models. However, deep learning models include a large number of parameters and require more computing power and time [1, 7].

Deep learning models can be elaborated from scratch, then the researcher determines the architecture and hyperparameters of the network, that is, parameters that do not tune while training [10]. After that it is necessary to evaluate network parameters by training. Therefore, design a model from scratch can be a computationally expensive and time-consuming process.

Pre-trained networks are an alternative to networks built from scratch and an effective way to reduce time and resources (Table 1) [11, 12]. To determine a person's age from a face image, networks are often pre-trained to classify images from the ImageNet dataset. When selecting such pre-trained convolutional neural network (CNN), the top-1 and top-5 classification accuracy are used [11]. It is assumed that for image classification using deep learning the prediction results with probability are obtained. Then accuracy is the number of samples that are correct predicted divided by the number of all samples. Top-1 accuracy measures the proportion of samples for which the predicted label matches the single target label. Top-5 accuracy considers a classification as correct if any of the five most probable predictions matches the target label [13]. In the last column of the Table 1, the values from work [12] are indicated in parentheses.

Table 1. **Comparison of the characteristics of various previously trained models**

Network name	Number of layers	Number of parameters, millions	Size, Mb	Top-1 accuracy
VGG-16	16	138	515	0.706 (0.715)
VGG-19	19	144	535	0.707
ResNet-152	152	58.5	209	0.786 (0.870)
ResNet-50	50	25.6	98	0.749
Xception	71	22.9	85	0.790 (0.790)
GoogleNet	22	7	27	0.665
MobileNetv2	23	3.5	13	0.713
Inception v3	48	23.9	104	0.779 (0.782)

Source: **compiled by the [11, 12]**

Thus, a review of the literature shows that handcrafted methods tend to be less computationally expensive but less accurate in AE since aging features are extracted manually. The elaboration of such methods requires a high researcher qualification and more time. In addition, handcrafted methods are usually built for devices with limited computing resources [1].

On the other hand, methods based on deep learning provide more accurately AE since features are extracted as the CNN is training. However, deep learning requires more computing power for image processing. CNNs built from scratch allow the researcher to design the network architecture, but they require more time to train. Pre-trained CNN significantly reduces the time of the network learning and a training set size [1].

The selection of a pre-trained network determines the effectiveness of facial AE. Thus, in [7] the pre-trained VGG-16 network has a lower AE error compared to other pre-trained networks. However, the VGG-16 network requires tuning of significant number of parameters, which makes it difficult to use in mobile devices. High-quality image recognition with a fewer number of parameters compared to VGG-16 is achieved by the Xception network (Table 1) [11, 12]. Then tuning the structure and parameters of Xception is appropriate for facial AE from uneven illuminated images.

2. FORMULATION OF THE PROBLEM

Suppose that for image classification, a CNN with architecture S and parameters P is pre-synthesized, $CNN=\{S, P\}$ [14, 15]. This network has already learned earlier to extract features for solving the problem of image classification. The problem of structural tuning of the CNN for facial AE from uneven illuminated images consists in making structural changes to the top layers of the existing architecture S. Such changes should ensure the extraction of features specifically for solving a new problem for a pre-trained CNN.

To set the parameters of a pre-trained network a transfer learning is applied [16, 17]. Following this method, the weights of the network trained to solve one problem are used to solve another similar problem [7].

The transfer learning problem is formulated in terms of datasets and tasks. The dataset D consists of the feature space X and the probability distribution $P(x)$, where $x=(x_1, x_2, ..., x_n)\in X$. For a given specific dataset $D=\{X, P(x)\}$, the task consists of two components. These are the label space Y and the target predictive function f: $X\rightarrow Y$. The function $f(x)$ is used to predict the label of a new example. This task is denoted by $T=\{Y, f(x)\}$ and trained from N_T pairs of data (x_i, y_i), where $x_i\in X$, a $y_i\in Y$, $i=1, ..., N_T$ [18].

Given a base dataset D_s and a base training task T_s, a target dataset D_t and a target training task T_t, where $D_s\neq D_t$ and/or $T_s\neq T_t$, transfer learning aims to improve the loss function $L(x)$ on D_t using results of solving of T_s on D_s, i.e. $L(x_t) \leq L(x_s)$, where $x_t\in D_t$ is the optimum on the target dataset D_t, $x_s\in D_s$ is the optimum on the base dataset D_s [18].

In this paper, transfer learning is used for facial AE from unevenly lit images. The color facial image is represented as $I_0(i, j)=(I_R(i, j), I_G(i, j), I_B(i, j)\}$, where $i=1, \ldots, n$; $j=1, \ldots, m$. Here n, m are the number of rows and columns of the image, respectively, (i, j) are coordinates of the image pixel. Then each pixel of the image is described by three features $I_R(i, j)$, $I_G(i, j)$, $I_B(i, j)$ which take values from the interval [0, 255]. A lighting function $R(i, j)$ is multiplied element by element with image features. Then the unevenly lit image is represented as $I(i, j)=(I_R(i, j)R(i, j), I_G(i, j)R(i, j), I_B(i, j)R(i, j))$.

Note that facial AE from an unevenly lit image is unstable problem. This is due to the fact that small variations of color features caused by uneven lighting correspond to a significant error of AE. When solving this problem with the use of CNN, the consequence of instability may be overtraining of the network. To overcome the difficult, the regularization is applied which adds some additional constraints to the training task T_t. To regularize the CNN training loss, the early stopping of training is used in this paper [19]. According to this method, the number of epochs is specified after which training stops without reducing the validation loss. Next, the network parameters corresponding to the lowest validation loss are saved. It is known from the literature that the early stopping of learning corresponds to regularization term $Q(x)\geq 0$ of the loss function, which is optimized [19]. Let M is the number of epochs after which learning stops without decreasing the validation loss. Then transfer learning with an early stopping determines the optimum point x_{tM} on the target dataset D_t, $x_{tM}\in D_t$, such that $L(x_{tM})+aQ(x_{tM}) \leq L(x_t)$, where $a\geq 0$ is the regularization constant.

The aim of the research is reducing the error of facial AE from uneven illuminated images by applying an early stopping of transfer learning of the Xception network.

3. MATERIALS AND METHODS

Structural tuning of the Xception network for facial age estimation

Following [20], a structural tuning of pre-trained models is proposed as such, which includes three stages.

At the first stage of structural tuning the architecture of a pre-trained neural network is selected for aging feature extraction from facial images. The Xception network [21, 22] has been selected because it architecture has almost the same number of parameters as the Inception v3 network and half the number of parameters compared to ResNet-152 (Table 1). But the gain in top-1 and top-5 classification accuracy on the ImageNet dataset is due to the more efficient use of Xception parameters with separable convolutions, and not to an increase in network capacity [21].

When using Xception network data first passes through an input flow, then through a middle flow that repeats eight times, and finally through an output flow. The Xception architecture has 36 convolutional layers, which are structured into 14 modules (Fig. 1). Each of the modules has linear residual connections around itself, except for the first and last modules [21, 22].

The following types of layers are used in the Xception network. Convolutional layers perform a convolution operation with filters that extract features of face images at different scales. Separable convolutions is key components of the Xception network architecture. They allow to separately model the spatial and inter-channel correlation of image features. This reduces the computing costs and the number of network parameters, as well as the error of facial AE. The convolutional and separable convolutional layers are followed by batch normalization layers, which accelerate the convergence of the network learning [21].

ReLU activation function introduces nonlinearity into the network. This helps to avoid the vanishing gradient problem, when in the process of backpropagation the values of the derivatives decrease and the learning rate vanishes with an increase in the number of layers of the network.

Max Pooling layers are used to reduce the size of the input data and increase the informativeness of the representation for further processing. Global Pooling layer returns the average value for each channel of input data. This allows reducing the amount of the data before inputting it to the fully connected layers. Flattening a tensor means to remove all of the dimensions except for one. Dense or fully connected layers obtain feature vectors from the Flatten layer and classify images based on these features using the Softmax function [22].

At the second stage of structural tuning, destructive changes were made to the architecture of the Xception network. Namely, only one fully connected layer was left. Other fully connected layers and the logistic regression layer were removed from the top of the Xception architecture (Fig. 1). This is because logistic regression takes as input the output of a linear regression function and uses a sigmoid function to estimate the probability that a sample belongs to a given class or not. Then, it is more appropriate for AE to use linear regression with continuous output from a wider range.

At the third stage of structural tuning, a constructive change was made to the architecture of the Xception network. Namely, after the Global Pooling layer only one neuron with the ReLU activation function was left on the fully connected layer. Then, this layer performs a scalar product of the input features with weights and adds a shift. As a result, the layer allows the Xception network to estimate the age using linear regression based on the trained weights and bias.

Transfer learning of Xception network with early stopping

Transfer learning means that the first layers of the network, which detect common features, remain unchanged, and the weights in the last layers adapt to the new problem. This allows to learn the network in less time to identify specific features for solving a new problem based on general features that the network has already learned to extract before. Transfer learning also helps to avoid overtraining because the base network was pre-trained on a large amount of data [16, 17].

A technique of the tuning the parameters of pre-trained networks is proposed. It applies an early stopping of learning and saves weights of the network with the lowest validation loss.

This technique consists of the following stages:

1) determining the base dataset D_s and the base training task T_s for the considered neural network;

2) initialization of weights of the neural network with values obtained as a result of solving the base training task T_s ;

3) determining the target dataset D_t and the target training task T_t for the neural network pre-trained on stages 1) and 2);

4) choosing the loss function $L(x)$, optimization algorithm, number of epochs and other parameters of the network training;

5) the transfer stage of network training, when all pre-trained parameters are frozen, and only the weights on the layers added for AE are updated;

6) the fine-tuning stage when all pre-trained parametes are unfrozen and updated together with the weights of the layers added for AE;

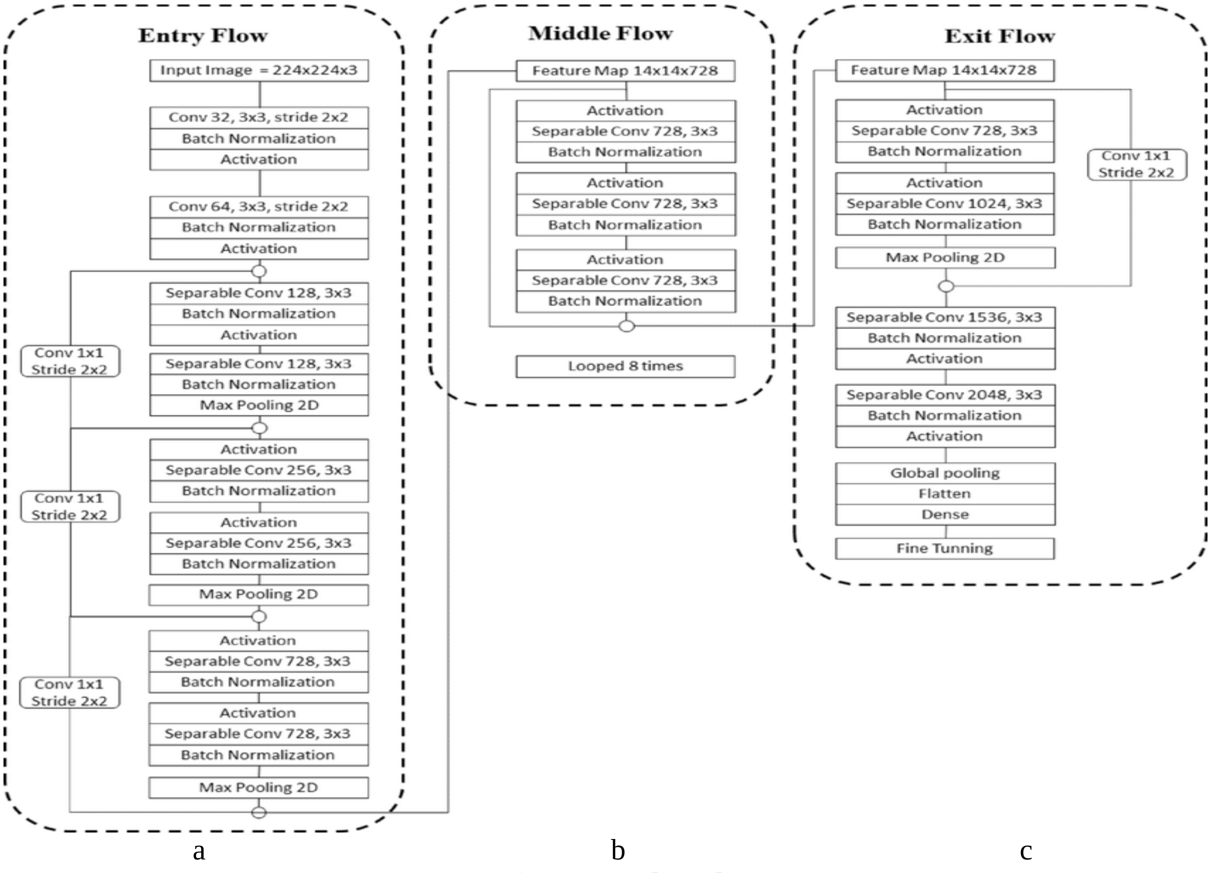

Fig. 1. **Xception network architecture:**
a – input flow; b – middle flow; c – output flow
Source: **compiled by the [22]**

7) early stopping of training if the improvement of the results is not observed during a certain number of epochs;

8) saving the network parameters, which correspond to the lowest validation loss, in order to further apply the network with these parameters for facial AE.

Next, the implementation of the stages of tuning of the Xception network parameters for the facial AE is considered.

At the first stage, the ImageNet dataset was used as a base dataset D_s [23, 24]. It consists of training data (1.2 million images), validation data (50,000 images), test data (100,000 images), and image labels (1000 categories). The image classification by categories was performed by the Xception network. We consider this task as the base training task T_s.

At the second stage, the parameters of the Xception network for AE were initialized with the weights of the same network, pre-trained on the ImageNet dataset to classify images by categories.

At the third stage, the formulation of the target learning task T_t depends on how the AE is performed [1]. If the person's age is estimated on interval, then the problem of multiclass classification is solved, where the classes correspond to age ranges. If a point estimation of age is performed, the target learning task T_t is formulated as a regression task, as in this paper. Uneven illuminated images from UTKFace dataset [25] will chose as the target dataset D_t in the experimental research section.

At the fourth stage, the Nadam method, which combines the Nesterov Accelerated Gradient and Adam optimization methods [26], was applied to optimize the loss function of the neural network.

Nadam uses gradient descent with momentum and is an improved version of the Adam method with the following advantages:

1) accelerating the convergence of the optimization and overcome local minima, especially in the case of significant loss function gradient or its significant variations;

2) automatic adaptation to the dimension of the gradients and to the dynamic range of their values. This allows to effectively processing data of different complexity and dimension;

3) adding a momentum term improves the robustness of the gradient optimization;

4) the optimization rate is automatically tuned for each network parameter using a moving average of the gradient values on previous iterations.

At the fifth and sixth stages the Xception network weights are trained for the AE on UTKFace dataset facial images [27]. Note that at these stages validation dataset is used to unbiased estimate the CNN performance after each epoch. But the network does not learn based on the validation data, its parameters are not tuned then the learning process are controlled based on the validation set after each epoch. The validation set influence on the model selection and stopping rules. It allows to tune the model’s hyperparameters and adjust the complexity of the model while training. The validation set is also separated from the training data and the test data to prevent overfitting. The last occurs when a model is learned the training data too well and performs poorly on unseen data. By monitoring the model’s performance on the validation set, we can stop the training when the model starts to overfit.

At the seventh and eighth stages the early stopping is used. It is a method that stops training once the model performance stops improving on a validation dataset. The parameters of the network corresponding to the lowest validation loss are saved, in order to further use the network with these parameters for facial AE.

4. EXPERIMENTAL EVALUATION OF ERROR OF FACIAL AGE ESTIMATION BY XCEPTION NETWORK

In the experimental research the error of the facial AE by the Xception neural network for images from the UTKFace and FG-NET datasets [27, 28] was evaluated.

Choosing a dataset for network training faces the problem of uneven distribution of samples by age groups [1]. The UTKFace dataset consist of 23,708 human face images ranging in age from 0 to 116 years, with an average age of 33.2 years. This dataset differs in diversity, volume and representativeness of different age and ethnic groups. Images from the UTKFace dataset are characterized by different distance to the face, illumination, face rotation angle, and resolution.

Unevenly illuminated images from the UTKFace dataset were selected using the software presented in [27]. This software is designed to implement two aesthetic and technical image quality models based on neural image assessment (NIMA). The NIMA's models are trained via transfer learning, where ImageNet pre-trained CNNs are used and fine-tuned for the image classification. In this paper NIMA was used to detect uneven lighting on facial images from UTKFace. In this way, 14,544 unevenly lit images from UTKFace were selected. Since the UTKFace dataset contains only facial images, and not full-length photos, the pre-processing of the selected images is minimized.

Specifically, each image must be reduced to the size that the neural network input assumes.

An annotation with an age category corresponding to each image allows using the data for neural network training [25]. Before training the network, the dataset was divided into three subsets, namely, training set (70 % of images of the entire dataset), validation set (10 %), and test set (20 %).

The Nadam method with an initial learning rate of 0.0001 was applied to train the Xception neural network. The mean squared error (MSE) was used as a loss function. To calculate the MSE, the squared differences between the annotated age values and the age values estimated from the face images are determined. These differences are averaged over the images of the training and validation datasets.

The training phase of the Xception neural network, where all pre-trained parameters are frozen and only the weights on the layer added to determine the age are updated, took 15 epochs. Training loss after each epoch determines the updating of network weights. Validation loss after each epoch affects the ability of the network to generalize data. The training loss and validation loss should be similar and they are presented in Fig. 2. The blue line shows the training loss, the red line shows the validation loss. Although at the beginning of learning validation loss is less than training loss.

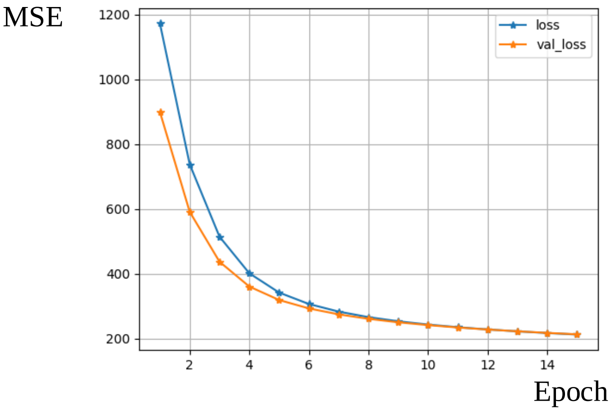

Fig. 2. **Training and validation losses at transfer stage**

***Source:* compiled by the authors**

Next, while fine-tuning the other Xception weights are unfrozen, allowing them to adapt to the UTKFace dataset while training. For the first epochs the validation loss and training loss are approximately the same, and then the validation loss exceeds the training loss (Fig. 3). The blue line corresponds the training loss, the red line corresponds the validation loss.

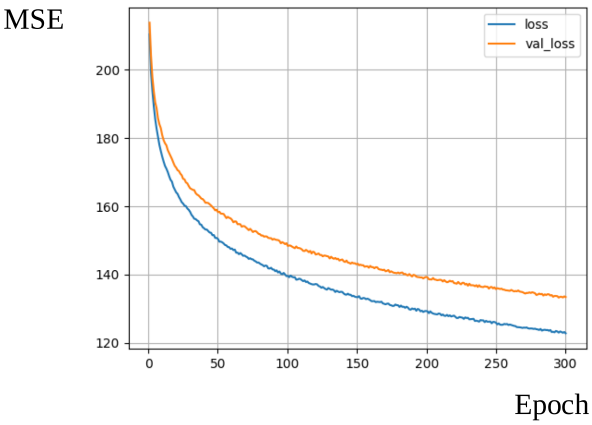

Fig. 3. **Training and validation losses at the first 300 epochs of fine-tuning**

***Source:* compiled by the authors**

Training stopped if the values of the validation loss did not decrease for 50 consecutive epochs. The weights of the network from the epoch with the best value of the validation loss were saved for facial AE on unevenly illuminated images. 1000 epochs are chosen as the maximum number of iterations. But learning was then stopped after 732 epochs with the applying of early stopping. The mean absolute error (MAE) of AE with the Xception network is 5.17 years from unevenly illuminated images. The number of training epochs affects the reducing the MAE of AE with the application of an early stopping of learning in comparison with the absence of an early stopping. For example, after the first 300 iterations of training on unevenly lit UTKFace images, the MAE of AE was 10.59 years (Fig. 4), which together with Fig. 3 indicates a clear undertraining of the network.

The trained neural network processed images from the UTKFace dataset in 22 ms. The research was performed using an AMD Ryzen 7 2700 processor, Nvidia RTX 3050 8gb graphics card, CPU 3 GHz, 16.0 GB RAM, Windows 10 Pro 22H2 19045.3086 operating system, Google Chrome 114.0 browser .5735.135.

The FG-NET dataset consisting of 1,002 images of 82 people aged 0 to 69 was used for comparison with AE methods known from the literature [28]. The MAE was calculated between annotated age values and age values determined from facial images (Table 2). 603 unevenly illuminated images from FG-NET were selected with the help of the software posted on the link [27]. The MAE of AE by the Xception network was 4.76 years on unevenly illuminated FG-NET images (Fig. 5). The AE error increased if facial images were insufficient illuminated or blurred.

Table 2. **Mean absolute error of facial age estimation from FG-NET images by networks known from the literature**

Neural network	MAE without fine-tuning, years	MAE with fine-tuning, years
Xception	12.03	3.67
ResNet-50	11.64	3.77
Inception v3	13.26	4.08
VGG19	14.35	4.98
VGG16	10.82	6.02

Source: **compiled by the [7]**

For comparison in the Table 2 the results from [7] is shown. But it should be kept in mind that the networks were trained on all FG-NET images without taking into account uneven illumination. That is, another training set is used to obtain the results in the Table 2.

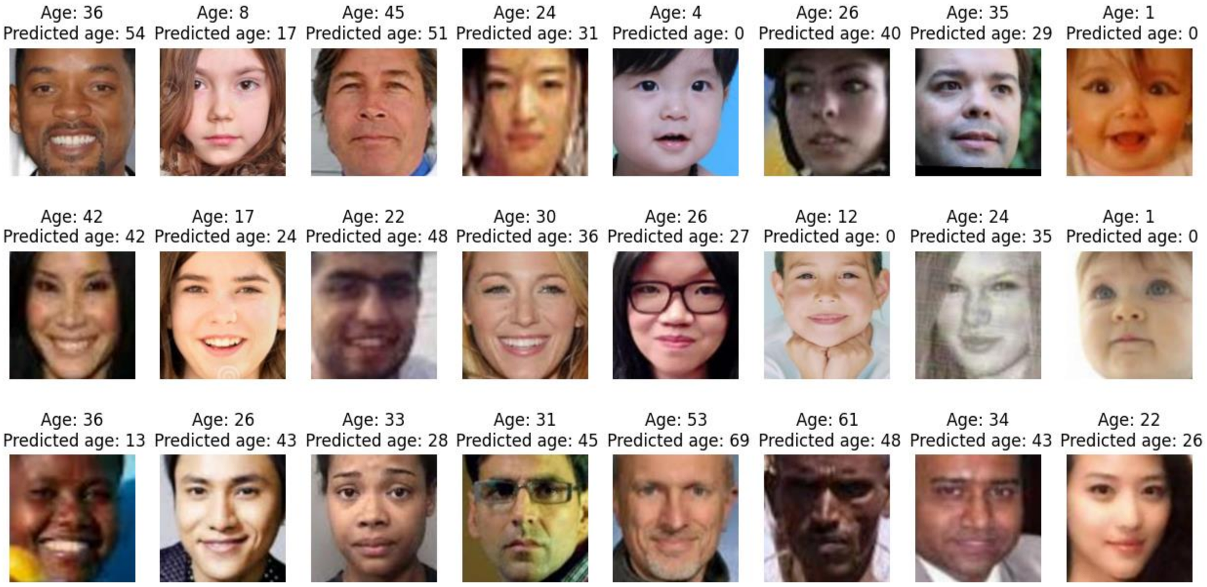

Fig. 4. **Results of facial age estimation from the uneven illuminated images of UTKF ace dataset**

Source: **compiled by the authors**

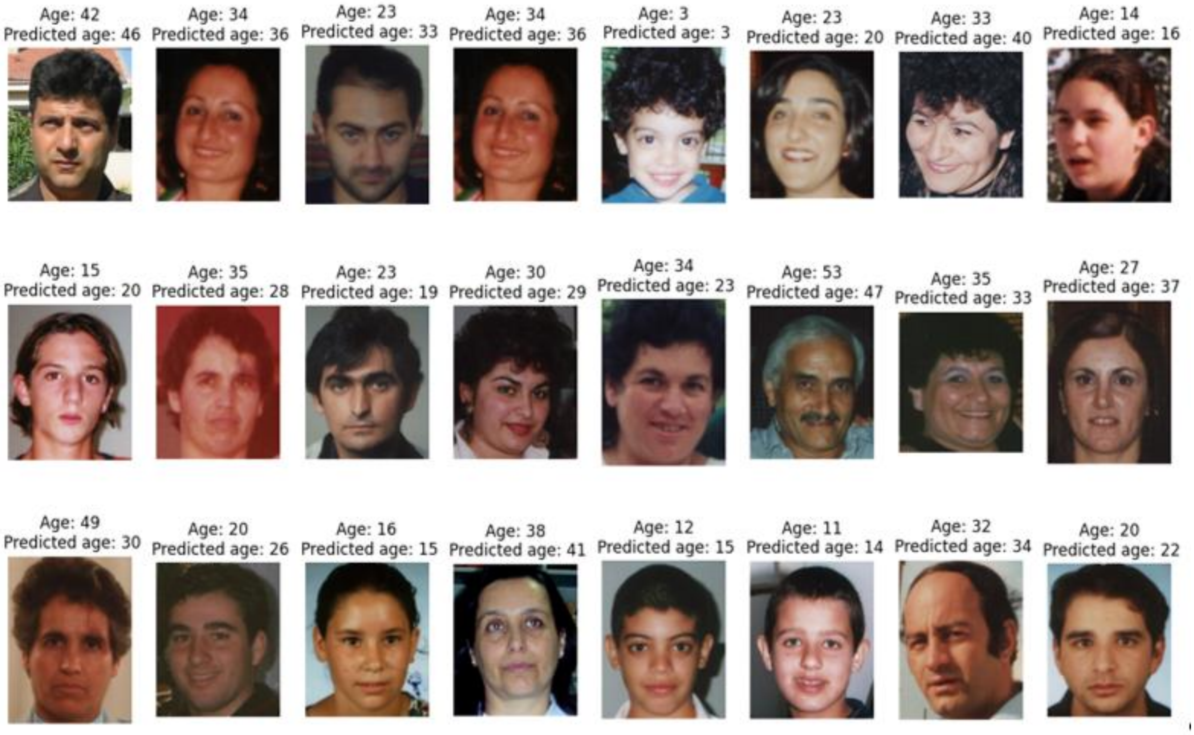

Fig. 5. **Results of facial age estimation from the uneven illuminated images of FG-NET dataset**

Source: **compiled by the authors**

The comparison is revealed that the MAE on unevenly lit images exceeds the error on all images for the Xception, ResNet-50, and Inception v3 by 29.7 %, 26.3 %, 16.7 %, respectively. Relative to VGG19, VGG16 the error decreases by 4.4 % and 26.5 %, respectively. The number of Xception network parameters is 6 times less than that of VGG-16, VGG-19, and 12 % and 5 % less than that of ResNet-50 and Inception v3, respectively.

CONCLUSIONS

For automatic facial AE in devices with limited computing resources, the transfer learning method of the Xception network received further development. This development of the transfer learning method includes the addition of an early stopping and the recovery of the network weights from the epoch with the best value of the validation loss. As a result, the MAE of facial AE was 4.76 years from unevenly illuminated FG-NET images using the Xception network. The Xception network is applied because the number of parameters of this network is 6 times less than that of VGG-16, VGG-19, and 12 % and 5% less than that of ResNet-50 and Inception v3, respectively. Then applying of the Xception network reduces the resource consumption of devices with limited computing power.

The practical significance of the obtained experimental results is that the elaborated technique of transfer learning can be recommended for the training of other networks to solve the AE problem from unevenly lit facial images.

Prospects for further research are reducing the unevenness of facial image lighting to decrease the error of age estimation. Also, to reduce the resource consumption of devices with limited computing power, it is promising to use fast transforms in the convolutional layers of the Xception network [29, 30].

REFERENCES

1. Elkarazle, K., Raman, V. & Then, P. “Facial age estimation using machine learning techniques: an overview”. *Big Data and Cognitive Computing*. 2022; 6 (4): 128. DOI: https://doi.org/10.3390/bdcc6040128.

2. Slipchenko, V. H., Poliahushko, L. H., Shatylo, V. V. & Rudyk, V. I. “Machine learning for human biological age estimation based on clinical blood analysis”. *Applied Aspects of Information Technology*. 2023; 6 (4): 431–442. DOI: https://doi.org/10.15276/aait.06.2023.29.

3. Morrás, M. “What is Age Verification, and why is it essential to keep your business safe and compliant”. – Available from: https://veridas.com/en/what-is-age-verification. – [Accessed: 29 September 2022].

4. “Veriff Age Estimation prevents users from accessing age restricted products or services”. – Available from: https://www.helpnetsecurity.com/2023/05/23/veriff-age-estimation. – [Accessed: 11 February 2022].

5. Nikitin, J., Burgermeister, L. C. & Freund, A. M. “The role of age and social motivation in developmental transitions in young and old adulthood”. *Frontiers in Psychology*. 2012; 3: 366. DOI: https://doi.org/10.3389/fpsyg.2012.00366.

6. Wang, H., Sanchez, V., Ouyang, W. & Li C.-T. “Using age information as a soft biometric trait for face image analysis”. In: Jiang, R., Li, CT., Crookes, D., Meng, W., Rosenberger, C. (eds) Deep Biometrics. Unsupervised and Semi-Supervised Learning. *Springer, Cham*. 2020. p. 1–20. DOI: https://doi.org/10.1007/978-3-030-32583-1_1.

7. Abdelkawy, H., Hadid, A., Othmani, A. & Taleb, A. R. “Age estimation from faces using deep learning: A comparative analysis”. *Computer Vision and Image Understanding*. 2020; 196 (8): 1029–1061. DOI: https://doi.org/10.1016/j.cviu.2020.102961.

8. Unnikrishnan, A., Ajesh, F. & Kizhakkethottam, J. “Texture-based estimation of age and gender from wild conditions”. *Procedia Technology*. 2016; 24: 1349–1357. DOI: https://doi.org/10.1016/j.protcy.2016.05.145.

9. Khajavi, M. & Ahmadyfard, A. “Human face aging based on active appearance model using proper feature set”. *Signal Image and Video Processing*. 2022; 17 (1): 1–9. DOI: https://doi.org/10.1007/s11760-022-02355-4.

10. Bartz, E., Bartz-Beielstein, T., Zaefferer, M., Mersmann, O. (eds.). “Hyperparameter tuning for machine and deep learning with R: A practical guide”. Ravensburg: Germany. *Publ.Spinger*. 2023. DOI: https://doi.org/10.1007/978-981-19-5170-1.

11. Zhang, Y. & Davison, B. “Impact of ImageNet model selection on domain adaptation”. *IEEE Winter Applications of Computer Vision Workshops (WACVW)*. 2020. p. 173–182. DOI: https://doi.org/10.1109/ WACVW50321.2020.9096945.

12. Wani, M. A., Bhat, F. A., Afzal S. & Khan, A. I. “Basics of supervised deep learning”. *Advances in Deep Learning*. 2020. p. 13–29. DOI: https://doi.org/10.1007/978-981-13-6794-6_2.

13. Dang, A. T. “Accuracy and loss: things to know about the top 1 and top 5 accuracy. Measure the performance of our model”. – Available from: https://towardsdatascience.com/accuracy-and-loss-things-to-know-about-the-top-1-and-top-5-accuracy1d6beb8f6df3#:~:text=Top%2D1%20.accuracy%20is%20the, %3D%202%2F5%20%3D%200.4. – [Accessed: 11 February, 2022].

14. Matychenko, A. D. & Polyakova, M. V. “The structural tuning of the convolutional neural network for speaker identification in mel frequency cepstrum coefficients space”. *Herald of Advanced Information Technology*. 2023; 6 (2): 115–127. DOI: https://doi.org/ten.15276/hait.06. 2023.7.

15. Leoshchenko, S. D., Oliynyk, A. O. , Subbotin, S. O., Hoffman, E. O. & Kornienko, O. V. “Method of structural adjustment of neural network models to ensure interpretability”. *Radioelectronics, Computer Science, Control*. 2021; 3: 86–96. DOI: https://doi.org/10.15588/1607-3274-2021-3-8.23.

16. Mamyrbayev, O., Alimhan, K., Oralbekova, D., Bekarystankyzy, A. & Zhumazhanov, B. “Identifying the influence of transfer learning method in developing an end-to-end automatic speech recognition system with a low data level”. *Eastern-European Journal of Enterprise Technologies.* 2022; 1 (9): 84–92, https://www.scopus.com/authid/detail.uri?authorId=55967630400.
DOI: https://doi.org/10.15587/1729-4061.2022.252801.

17. Hamza, A. H., Hussein, S. A., Ismaeel, G. A., Abbas, S. Q., Zahra, M. M. A. & Sabry, A. H. “Developing three dimensional localization system using deep learning and pre-trained architectures for IEEE 802.11 Wi-Fi”. *Eastern-European Journal of Enterprise Technologies.* 2022; 4 (9): 41–47, https://www.scopus.com/authid/detail.uri?authorId=57214125418. DOI: https://doi.org/10.15587/1729-4061.2022.263185.

18. Lin, Y.-P. & Jung, T.-P. “Improving EEG-based emotion classification using conditional transfer learning”. *Front. Hum. Neurosci*. 2017; 11: 334. DOI: https://doi.org/10.3389/fnhum.2017.00334.

19. Brownlee, J. “Use early stopping to halt the training of neural networks at the right time”. – Available from: https://machinelearningmastery.com/how-to-stop-training-deep-neural-networks-at-the-right-time-using-early-stopping. – [Accessed: 30 September 2022].

20. Polyakova, M. V. “RCF-ST: Richer convolutional features network with structural tuning for the edge detection on natural images”. *Radio Electronics, Computer Science, Control.* 2023; (4): 122–134. DOI: https://doi.org/10.15588/1607-3274-2023-4-12

21. Chollet, F. “Xception: Deep learning with depthwise separable convolutions”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. p. 1800–1807. DOI: https://doi.org/ 10.1109/CVPR.2017.195.

22. Malve, P. & Gulhane, V. S. “Breast cancer data classification using Xception-based neural network”. *Computer Science*. 2023, 4 (6): 734. DOI: https://doi.org/10.1007/s42979-023-02205-1.

23. Download ImageNet Data. – Available from: https://www.image-net.org. – [Accessed: 30 September 2022].

24. He, K., Girshick, R. & Dollár, P. “Rethinking ImageNet pre-training”. *IEEE/CVF International Conference on Computer Vision (ICCV).* 2019. p. 4918–4927. DOI: https://doi.org/10.1109/ ICCV.2019.00502

25. UTKFace Large Scale Face Dataset. – Available from: https://susanqq.github.io/UTKFace. – [Accessed: 30 September 2022].

26. Dozat, T. “Incorporating Nesterov momentum into Adam”. *4th International Conference on Learning Representations Workshops (ICLR)*. 2016. p. 1–4.

27. “Convolutional neural networks to predict the aesthetic and technical quality of images”. – Available from: https://github.com/idealo/image-quality-assessment. – [Accessed: 11 February 2022].

28. “FG-NET dataset by Yanwei Fu”. – Available from: http://yanweifu.github.io/FG_NET_data/FGNET.zip. – [Accessed: 30 September 2022].

29. Cariow, A., Papliński, J. & Makowska, M. “VLSI-friendly filtering algorithms for deep neural networks”. *Appl. Sci.* 2023; 13 (15): 9004. DOI: https://doi.org/10.3390/app13159004.

30. Cariow, A. & Cariowa, G. “Minimal filtering algorithms for convolutional neural networks”. In: van Gulijk, C., Zaitseva, E. (eds) Reliability Engineering and Computational Intelligence. Studies in Computational Intelligence. *Springer, Cham.* 2021; 976: 73–88. DOI: https://doi.org/10.1007/978-3-030-74556-1_5.

Conflicts of Interest:the authors declare no conflict of interest

Received 06.12.2023
Received after revision 27.02.2024
Accepted 12.03.2024

DOI: https://doi.org/10.15276/aait.07.2024.6
УДК 004.93

Передавальне навчання мережі Xception з ранньою зупинкою для визначення віку людини

Полякова Марина Вячеславівна[1)]
ORCID: https://orcid.org/0000-0001-7229-7657; marinapolyakova943@gmail.com. Scopus Author ID: 57017879200
Рогачко Владислав Володимирович[1)]
ORCID: https://orcid.org/0009-0005-9691-445X; rohachko.8089583@stud.op.edu.ua.
Нестерюк Олександр Геннадійович[1)]
ORCID: https://orcid.org/0000-0002-0806-8259; nesteryuk@op.edu.ua. Scopus Author ID: 57207314663
Гуляєва Наталя Анатоліївна[1)]
ORCID: https://orcid.org/0000-0001-6485-3094; gulyaeva.nata72@gmail.com. Scopus Author ID: 57202228675
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

Швидкий розвиток глибокого навчання привертає більше уваги до аналізу зображень обличчя людини. Методи оцінки віку людини за зображеннями обличчя на основі глибокого навчання більш ефективні порівняно з методами на основі антропометричних моделей, моделей активного зовнішнього вигляду, текстурних моделей, підпростору шаблонів старіння. Однак мережі глибокого навчання потребують більшої обчислювальної потужності для обробки зображень. Попередньо навчені моделі працюють без потреби у великій кількості зразків, а час навчання менший. Однак значення параметрів, отриманих в результаті навчання, значно впливають на ефективність попередньо навченої нейронної мережі. Також потрібно враховувати особливості оброблюваних зображень, зокрема, умови, у яких їх отримано. Останнім часом визначення віку людини за зображенням обличчя реалізується у додатках пристроїв з обмеженим джерелом обчислень, наприклад, у смартфоні. Під час фотографування обличчя людини камерою смартфона дуже складно забезпечити рівномірне освітлення. Метою дослідження, є зниження помилки оцінки віку людини за нерівномірно освітленим зображенням обличчя шляхом застосування ранньої зупинки передавального навчання мережі Xception. Запропоновано методику попереднього навчання нейронних мереж, яка використовує ранню зупинку навчання, якщо покращення результатів не спостерігається протягом певної кількості епох. Потім відновлюються мережеві ваги з епохи з найкращим значенням функції втрат на валідаційних даних. У результаті середня абсолютна помилка оцінки віку шляхом застосування навченою за запропонованою методикою мережею Xception за нерівномірно освітленими тестовими зображеннями становила близько п'яти років. Обрання Xception обумовлене тим, що кількість параметрів цієї мережі менша, ніж у інших мереж, які застосовувалися для оцінювання віку людини. Це зменшує споживання ресурсів пристроїв з обмеженими обчислювальними можливостями. Перспективами подальших досліджень є зменьшення рівня нерівномірності освітлення зображеннь обличчя для зменьшення помилки оцінки віку людини. Також для зменшення споживання обчислювальних ресурсів перспективним є застосування швидких перетворень у згорткових шарах мережі Xception.

Ключові слова: визначення віку людини; мережа Xception; налаштування параметрів; рання зупинка; нерівномірно освітлені зображення; передавальне навчання

ABOUT THE AUTHORS

Marina V. Polyakova - Doctor of Engineering Sciences, Associated Professor, Professor of Department of Applied Mathematics and Information Technology. Odessa Polytecnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine.
ORCID: https://orcid.org/0000-0001-7229-7657; marinapolyakova943@gmail.com. Scopus Author ID: 57017879200
Research field**:** Intelligent data analysis; machine learning; digital image processing

Полякова Марина Вячеславівна - доктор технічних наук, доцент, професор кафедри Прикладної математики та інформаційних технологій. Національний університет «Одеська політехніка», пр. Шевченка,1. Одеса, 65044, Україна

Vladyslav V. Rogachko - bachelor, graduate student of Department of Applied Mathematics and Information Technology. Odessa Polytecnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine.
ORCID: https://orcid.org/0009-0005-9691-445X; rohachko.8089583@stud.op.edu.ua.
Research field**:** Intelligent data analysis; machine learning; digital image processing

Рогачко Владислав Володимирович – бакалавр, магістрант кафедри Прикладної математики та інформаційних технологій. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

Oleksandr H. Nesteriuk - PhD, Associated Professor of Department of Computer Systems. Odessa Polytecnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-0806-8259; nesteryuk@op.edu.ua. Scopus Author ID: 57207314663
Research field**:** Hybrid systems; digital image processing; pattern recognition

Нестерюк Олександр Геннадійович – кандидат технічних наук, доцент кафедри Комп'ютерних систем. Національний університет «Одеська політехніка», пр. Шевченка,1. Одеса, 65044, Україна

Natalia A. Huliaieva - Senior Lecturer, Department of Applied Mathematics and Information Technologies. Odessa Polytecnic National University, 1, Shevchenko Ave., Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-6485-3094; gulyaeva.nata72@gmail.com. Scopus Author ID: 57202228675
Research field**:** Intelligent data analysis; mathematical economics

Гуляєва Наталя Анатоліївна - старший викладач кафедри Прикладної математики та інформаційних технологій. Національний університет «Одеська політехніка», пр. Шевченка,1. Одеса, 65044, Україна