applied sciences

MDPI

Article

Face Gender and Age Classification Based on Multi-Task, Multi-Instance and Multi-Scale Learning

Haibin Liao [1,2], Li Yuan [1], Mou Wu [3,4,*], Liangji Zhong [3], Guonian Jin [3] and Neal Xiong [5]

[1] School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan 430200, China
[2] Jiangxi Smart City Industrial Technology Research Institute, Jiangxi Minxuan Intelligent Technology Co., Ltd., Nanchang 330096, China
[3] School of Computer Science and Technology, Hubei University of Science and Technology, Xianning 437100, China
[4] Laboratory of Optoelectronic Information and Intelligent Control, Hubei University of Science and Technology, Xianning 437100, China
[5] Department of Computer Science and Mathematics, Sul Ross State University, Alpine, TX 79830, USA
* Correspondence: mou.wu@163.com

Featured Application: Facial recognition.

Abstract: Automated facial gender and age classification has remained a challenge because of the high inter-subject and intra-subject variations. We addressed this challenging problem by studying multi-instance- and multi-scale-enhanced multi-task random forest architecture. Different from the conventional single facial attribute recognition method, we designed effective multi-task architecture to learn gender and age simultaneously and used the dependency between gender and age to improve its recognition accuracy. In the study, we found that face gender has a great influence on face age grouping; thus, we proposed a random forest face age grouping method based on face gender conditions. Specifically, we first extracted robust multi-instance and multi-scale features to reduce the influence of various intra-subject distortion types, such as low image resolution, illumination and occlusion, etc. Furthermore, we used a random forest classifier to recognize facial gender. Finally, a gender conditional random forest was proposed for age grouping to address inter-subject variations. Experiments were conducted by using two popular MORPH-II and Adience datasets. The experimental results showed that the gender and age recognition rates in our method can reach 99.6% and 96.14% in the MORPH-II database and 93.48% and 63.72% in the Adience database, reaching the state-of-the-art level.

Keywords: facial attribute recognition; feature extraction; deep learning; random forest

check for updates

Citation: Liao, H.; Yuan, L.; Wu, M.; Zhong, L.; Jin, G.; Xiong, N. Face Gender and Age Classification Based on Multi-Task, Multi-Instance and Multi-Scale Learning. *Appl. Sci.* **2022**, *12*, 12432. https://doi.org/10.3390/app122312432

Academic Editor: Douglas O'Shaughnessy

Received: 28 October 2022
Accepted: 29 November 2022
Published: 5 December 2022

1. Introduction

In daily life, gender and age classification are very important, which can help us to distinguish whether the person we contact is a "sir" or "madam", as well as "young" or "old". These behaviors rely heavily on human forecasting and the recognition of facial attributes: gender and age [1]. In addition, the gender and age attributes of faces have other real-world applications. For example, vending machines can deny selling cigarettes to minors and an electronic billboard can display advertisements based on gender and age. However, the performance index of face attribute recognition by machines is far from meeting the needs of commercial applications [2,3].

In general, face gender and age classification is a branch of face recognition; thus, generic face recognition technology can naturally be applied to this problem [4–8]. Therefore, most existing methods are based on manually designed features in this field, such as Local Binary Patterns (LBPs) [9], Gabor [10], Biologically Inspired Feature (BIF) [11] and Spatially Flexible Patches (SFP) [12]. After these manually designed feature extractions, we can resort to a classification or regression algorithm to estimate facial gender and age. Among them, Support Vector Machine (SVM)-based approaches [1,11] are used for gender classification and age grouping; Support Vector Regression (SVR) [13], linear regression [14], Canonical Correlation Analysis (CCA) [15] and Partial Least Squares (PLS) [16] regression methods are available for accurate age estimation. However, these methods can only be applied to the constrained benchmarks, rather than achieving satisfactory results in benchmarks in the wild [1,17].

Inspired by the success of ImageNet classification and face recognition [18], deep learning has been applied to gender and age classification [2,3,19,20]. Wang et al. [19] extracted discriminant features using CNN and these were combined with classification and regression approaches to estimate age based on FG-NET and MORPH. Additionally, Levi et al. [2] employed deep learning for age and gender classification together based on an uncontrolled Adience database. Niu et al. [21] took full advantage of CNN and SVM to propose a hybrid neural network with better results compared with a plain CNN. Liu et al. [22] found that the hybrid model integrating CNN and Conditional Random Field is better than other methods through extensive image segmentation experiments. What is more, Xie et al. [23] proposed a hybrid neural network by integrating CNN and SVM for scene recognition and domain adaption. Liu et al. [24] proposed a hybrid model by integrating CNN and Random Forest (RF) for facial expression recognition, in which a conditional CNN enhanced the RF for pose-aligned facial expression recognition. Recently, Guehairia [8] proposed an architecture for age estimation based on a cascade of classification tree ensembles, which have been known recently as a Deep Random Forest (DRF). The model consists of two types of DRF. The first type extends the input facial features; the second fuses all enhanced representations to consider the fuzziness of the face age. Experimental results demonstrated that it can achieve high accuracy and fast convergence with a limited amount of image data, rather than a large amount of data required by a plain CNN.

The accurate classification of face gender and age includes two important steps: feature extraction and classifier design, while the former is the key to the whole process. It not only requires the extracted features to have great differences among different classes, but also requires it to maintain invariance within the same class. Most traditional methods use manually designed features and statistical models for the recognition of gender and age [10,11,15,16], which have achieved favorable results based on the benchmarks of controlled databases, such as FG-NET [25] and MORPH [26]. However, they exhibit unsatisfactory performance based on recent benchmarks of uncontrolled databases, namely "in-the-wild" benchmarks, including Adience [1], and the apparent age dataset LAP [27], which have a variety of variations in appearance, illumination, pose and occlusion. In recent years, deep learning has been widely applied to various scenarios, such as disaster scenes [28], industrial IoT [29,30], large-scale data [31], wireless sensor networks [32–34] and healthcare monitoring [3,35]. Specifically, CNN has been extremely striking in the field of pattern recognition and computer vision due to its strong nonlinear feature extraction capability [36]. Therefore, we can enjoy great improvements brought about by CNNs [2,37] in gender and age prediction in the wild. At present, for face images in natural scenes, the recognition rates of gender and age based on depth learning can exceed 95% and 55%, respectively.

Through more discriminative features and powerful classifiers, higher recognition rates can be obtained. In the CNN-based classification method, the full connection layer is the same as a common single hidden layer in the feedforward neural network (SLFN) and trained through a back-propagation (BP) algorithm. It easily causes a local minimum and over-fitting problem [38]. Therefore, in CNN-based deep learning, the generalization ability of the full connection layer is not optimal, where discriminative features can be well exploited. In order to solve these problems, a novel classifier needs to be developed by making full use of the features extracted by the convolutional layer while possessing the full connection layer or softmax classifier with similar ability. In the field of pattern recognition, three classification algorithms, including Naive Bayes [39], SVM and RF [14], have been applied extensively. To date, RF has been proven to have high generalization and big data processing ability, in addition to being easy to implement and having high speed [40]. Additionally, RF and improved RF, including its mixing approaches, have been widely used in pattern recognition tasks and have achieved excellent results [24].

Therefore, we made full use of CNN and RF to propose a hybrid deep learning architecture for facial gender and age classification. In addition, we found, in practice, that males and females have different aging models. In other words, gender has a certain impact on facial age grouping. However, this relationship between gender and age is rarely exploited by current methods. To deal with such relationships, we put facial gender and age recognition in a unified RF classification framework and proposed a gender-conditional RF to recognize facial age. Our goal is to improve both the accuracy and efficiency of facial gender and age classification in the wild. An overview of the proposed multi-instance- and multi-scale-enhanced multi-task random forests is shown in Figure 1. The robust features are extracted from face instances to overcome the variance in image resolution, illumination and occlusion. Facial gender is estimated first using RF and then, age is estimated under the conditional probability of facial gender alignment. Our contributions can be described as follows:

1. A multi-instance- and multi-scale-enhanced multi-task random forest is proposed to process gender and age classifications together, which exploits the advantages of CNN and RF.
2. We propose a multi-instance- and multi-scale-enhanced facial multi-task feature extraction model, which can alleviate the intra-subject variations in faces, such as illumination, expression, pose and occlusion.
3. We propose a gender-aligned conditional probabilistic learning model for facial age grouping to suppress inter-subject variations.

Throughout this paper, we use the following abbreviations:

- CNNs: Convolutional Neural Networks
- LBPs: Local Binary Patterns
- BIF: Biologically Inspired Feature
- SFPs: Spatially Flexible Patches
- SVM: Support Vector Machine
- SVR: Support Vector Regression
- CCA: Canonical Correlation Analysis
- PLS: Partial Least Squares
- RF: Random Forest
- DRF: Deep Random Forest
- SLFN: Feedforward Neural Network
- BP: Back Propagation
- MML: Multi-instance and Multi-scale Learning
- MMFL: Multi-scale Fusion Learning Network
- MIF: Multi-Instance Fusion
- GAP: Global Average Pooling
- FC: Fully Connected
- IRBs: Inverted Residual Blocks
- CPR: Compact Pyramid Refinement
- NCSF: Neurally Connected Split Function

The rest of the paper is organized as follows: Section 2 presents our method. Experimental results are presented in Section 3. Finally, the conclusion is provided in Section 4.

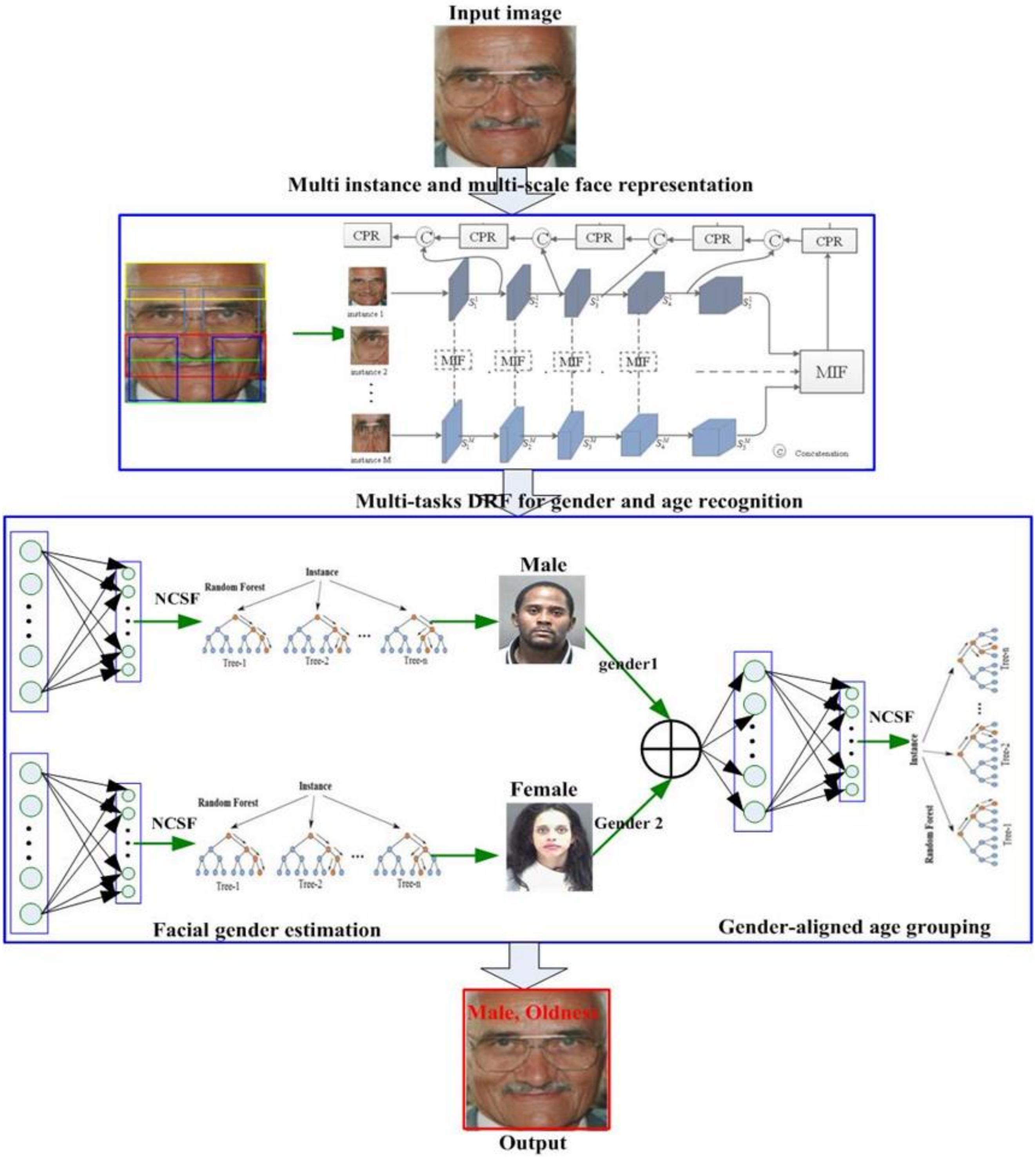

Figure 1. An overview of the proposed method for facial gender and age classification.

2. Facial Gender and Age Classification Based on MML and DRF

The flowchart of the proposed method is shown in Figure 2. The robust features are extracted from multi-instance and multi-scale learning (MML) using the transferring CNN model to suppress the influence of low resolution, illumination and occlusion. DRF is used to estimate facial gender and then, age is recognized under the conditional probability of gender alignment.

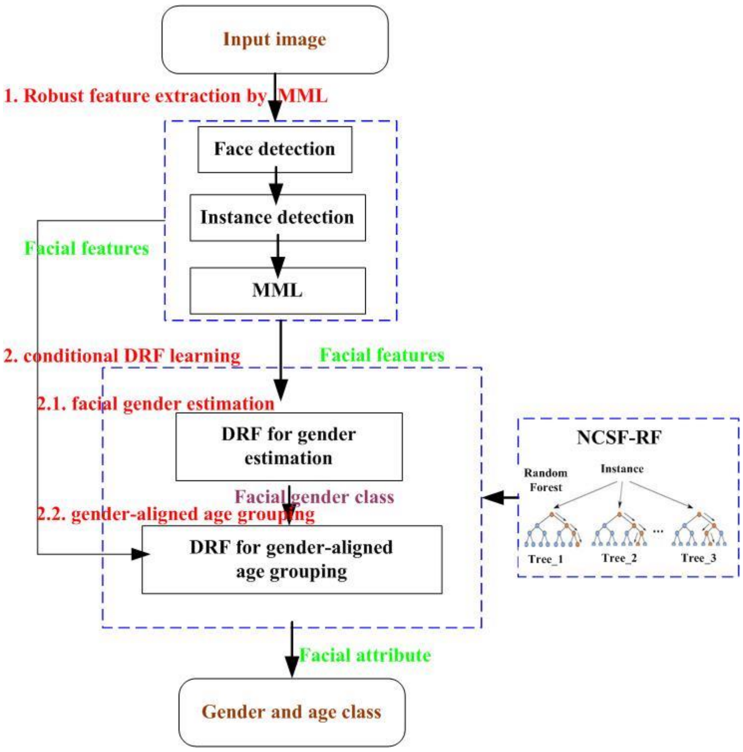
Input image
1. Robust feature extraction by MML
Face detection
Instance detection
MML
Facial features
2. conditional DRF learning
Facial features
2.1. facial gender estimation
DRF for gender estimation
Facial gender class
2.2. gender-aligned age grouping
DRF for gender-aligned age grouping
NCSF-RF
Random Forest
Instance
Tree_1
Tree_2
Tree_3
Facial attribute
Gender and age class

Figure 2. Flowchart of the proposed approach for facial gender and age classification.

2.1. Deep Feature Representation by MML

2.1.1. Facial Instance Selection

We extracted robust features from facial instances with MML. Different from randomly or densely sampled patches and the salience detection algorithm [24], we took advantage of the facial gender and age characteristics to select nine facial patches from the detected face image as facial instances, as shown in Figure 3, and instance 1 is the overall face image. The selection strategy of facial instances is based on the influence of different facial patches on face gender and age recognition. The specific facial instance selection steps can be seen as follows:

Firstly, the face detection algorithm [41] is used to cut a pure face image as in instance 1.

Secondly, according to the face detection results, the nose tip of the face is found by using face landmark localization technology.

Finally, according to the position of the nose tip and the "three eyes and five chambers" characteristics of the facial structure, eight other facial patches are selected as face instances.

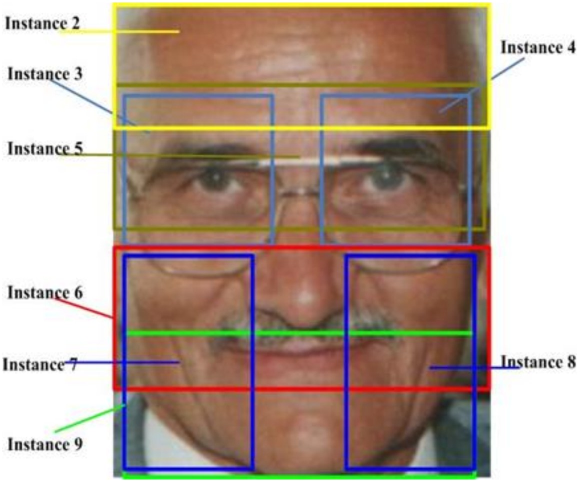

Figure 3. Facial gender and age instances in a face image.

2.1.2. Multi-Instance Learning

After selecting the facial gender and age instances, we propose a multi-instance multi-scale fusion learning network (MMFL) for robust facial feature extraction. Figure 4 depicts the MMFL architecture; we employed MobileNetV3 [42] as the backbone for multiple instance representation. A multi-instance fusion module (MIF) is applied to each scale, and the features of the top-level layer are aggregated to this level layer. For convenience, five stages in the output feature maps are denoted as S_1, S_2, S_3, S_4, S_5, with strides of $2, 2^2, 2^3, 2^4, 2^5$, respectively. We fused the extracted instance map S^1, S^2, S^3, S^4, S^5 to generate the multi-instance fuse feature. We designed a lightweight MIF for multi-instance fusion, as shown in Figure 5.

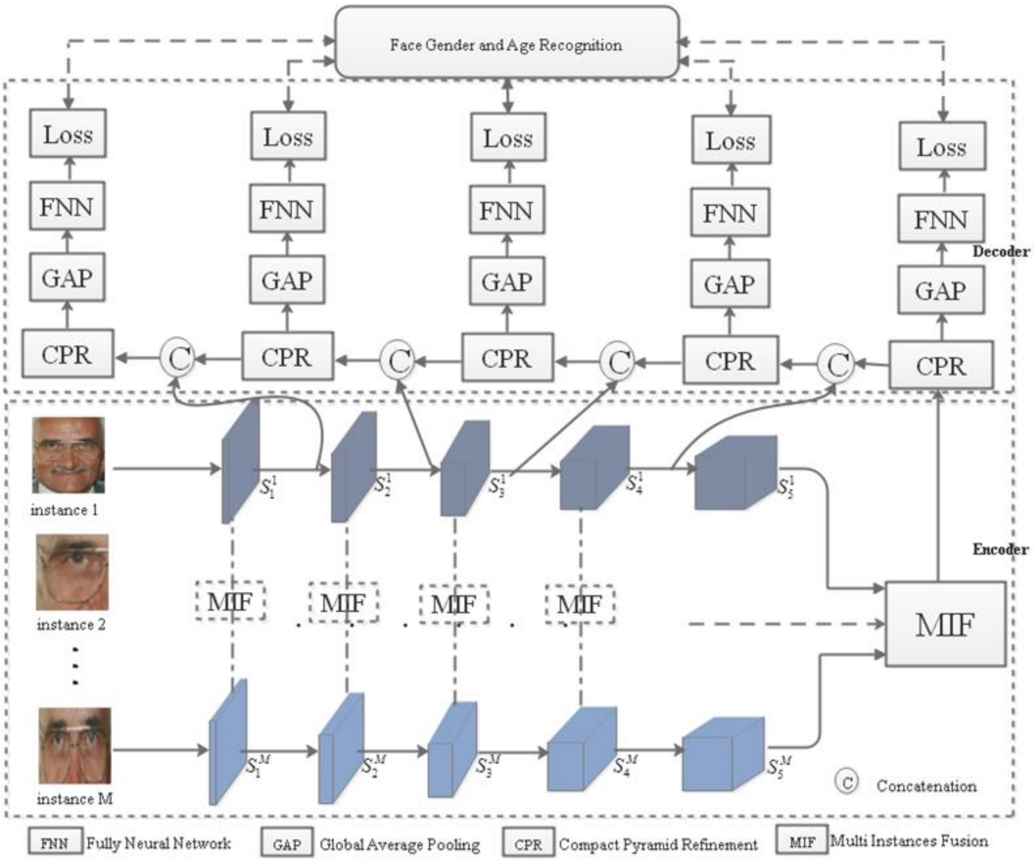

Figure 4. Multi-instance multi-scale fusion learning network.

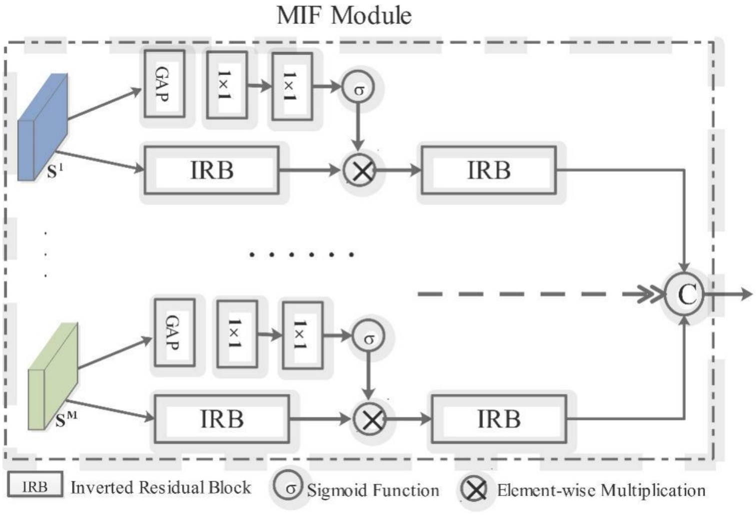

Figure 5. Multi-instance fusion module.

Specifically, we first obtained a vector by using a global average pooling (GAP) layer to S^i, followed by an attention vector v^i computed based on two fully connected (FC) layers:

$$v^i = \sigma\Big(\mathrm{FC}_2(\mathrm{ReLU}(\mathrm{FC}_1(GAP(S^i))))\Big) \tag{1}$$

where ReLU and σ indicate ReLU layers and the standard sigmoid function, respectively. At the same time, the S^i is sent to Inverted Residual Blocks (IRBs) [6] to derive the feature maps $\mathbb{N}^i = \mathrm{IRB}\big(S^i\big)$. With $\mathbb{N}^i$ and v^i, the multiplication of $\mathbb{N}^i$ and v^i is fed into an IRB, such as

$$\mathbb{C}^i = \mathrm{IRB}(v^i \otimes \mathbb{N}^i) \tag{2}$$

Note that attention v^i is replicated to the same shape as $\mathbb{N}^i$ before multiplication and the v^i is used to recalibrate the instance features. We than combined each $\mathbb{C}^i$ through concatenation to derive the instance fusion feature $\mathbb{C} = [\mathbb{C}^1, \mathbb{C}^2, \cdots, \mathbb{C}^M]$.

2.1.3. Multi-Scale Integration Learning

It is generally believed that in the backbone network, high-level features contain more semantic abstract information, while low-level features contain more detailed information. For facial gender and age recognition, we designed a lightweight decoder using the Compact Pyramid Refinement (CPR) [43] module as the basic unit. Because different levels of features correspond to different scales, multi-scale learning integrates features of different scales, so that the final extracted features not only have semantic abstract information, but also have details. Hence, we designed a multi-scale integration learning strategy to enhance facial feature extraction.

Suppose that the input of a CPR module is $\mathbb{C}$. First, the channels of $\mathbb{C}$ are expanded M times by using a 1×1 convolution. Second, we apply three depth-wise separable convolutions with dilation rates of 1, 2 and 3 to obtain three different scale features. Finally, these multi-scale features are connected with a multi-scale fusion strategy, which can be denoted as:

$$\begin{aligned}
&\mathbb{C}_1 = \mathrm{Conv}_{1\times1}(\mathbb{C}),\\
&\mathbb{C}_2^{d_1} = \mathrm{Conv}_{3\times3}^{d_1}(\mathbb{C}_1),\\
&\mathbb{C}_2^{d_2} = \mathrm{Conv}_{3\times3}^{d_2}(\mathbb{C}_1),\\
&\mathbb{C}_2^{d_3} = \mathrm{Conv}_{3\times3}^{d_3}(\mathbb{C}_1),\\
&\mathbb{C}_2 = \mathrm{ReLU}(\mathrm{BN}(\mathbb{C}_2^{d_1} + \mathbb{C}_2^{d_2} + \mathbb{C}_2^{d_3})),
\end{aligned} \tag{3}$$

where d_1, d_2 and d_3 are dilation rates. BN indicates batch normalization. Next, we used a 1×1 convolution to compress channels of $\mathbb{C}_2$ to the same number as the input:

$$\mathbb{C}_3 = \mathrm{Conv}_{1\times1}(\mathbb{C}_2) + \mathbb{C} \tag{4}$$

Then, an attention vector v' is computed by applying the attention mechanism in Equation (1), so that we have:

$$X = v' \otimes \mathrm{Conv}_{1\times1}(\mathbb{C}_3) \tag{5}$$

Equation (5) uses global contextual information to recalibrate the multi-scale fusion features. As shown in Figure 4, at each decoding phase, the feature maps of the top decoder and the corresponding encoder are concatenated and then, the CPR module is used for fusion. In this way, the decoder can aggregate multi-level features from top to bottom.

2.2. DRF Model

In general, gender recognition is easier than age grouping. In the facial age grouping field, due to the existence of facial gender factors, the facial age grouping in the feature space is different, which makes it difficult to construct a facial age classifier with high accuracy. Therefore, by putting a gender and age recognition study together and using facial gender as an implicit condition to divide the face data space, we propose a face age grouping method based on conditional random forest. The implementation steps for the DRF model are as follows:

Step 1. A face gender classifier based on random forest is developed by using all face data.

MMFL was used to extract the robust features y, and T^G was used to estimate face gender g, where T^G training uses uncertainty measures:

$$H(y) = -\sum_g p(g|y)\log_2(p(g|y)) \tag{6}$$

The uncertainty measure guides each node to choose the best binary test from the candidate library of binary tests to ensure that the current node can be divided into two subnodes with reduced uncertainty. Face gender is stored on each leaf node l of T^G with a Gaussian model:

$$p(g|l(y)) = N(g;\bar{g_l};\sigma_l) \tag{7}$$

where $\bar{g_l}$ and σ_l are the mean and covariance, respectively. While Equation (7) models the probability for a sample feature y ending in a leaf l, the gender probability of the forest is obtained by averaging over all trees:

$$p(g|y) = \frac{1}{M}\sum_m p(g|l_m(y)) \tag{8}$$

where l_m is the corresponding leaf for a tree and M is the number of trees.

Step 2. We classified the face dataset according to face gender and trained a series of gender-conditional random forest decision trees.

Each decision tree in a conditional random forest $\{T^S(\Omega_n)\}_{n=1}^2$ is independently trained by using the same method. In order to build each decision tree $T_t^S(\Omega_n)$, first, randomly select images from the corresponding data subset S_{Ω_n} to form a training dataset; then, randomly extract a series of sub-features $\{y_i = (a_i, I_i)\}$ from each training image feature y, where a_i is the face age class and $I_i = \{I_i^1, I_i^2, \cdots, I_i^F\}$ is a set of sub-features selected from y; finally, the selected sub-features are used to split the decision tree nodes to generate the final decision tree.

We used a Neurally Connected Split Function (NCSF) splitting model to reinforce the learning capability of a splitting node by combining the Information Gain of the decision tree and the loss function of the deep network model [24]. The connection function f_n of a hidden layer in MMFL is used to enhance the conditional feature presentation y of a face sample; meanwhile, the enhanced feature presentation is used as the node feature selection of the network-enhanced forest:

$$d_n(y, K|\Omega_g) = \sigma(f_n(y, K|\Omega_g)) \tag{9}$$

where $\sigma(x) = (1+e^{-x})^{-1}$ is the sigmoid function, K is the parametrization of the network, the Adaptive Moment Estimation approach is used to minimize the risk with respect to K, Ω_g is the age sub-forest with different genders and n is a decision node. We employed an Information Gain approach to split a node into its left and right child nodes in the tree construction:

$$\widetilde{\varphi} = \underset{\varphi}{\mathrm{argmax}}(H(d_n) - \textstyle\sum_{S\in\{N_r,N_l\}} \frac{|d_n^S|}{|d_n|} H(d_n)) \tag{10}$$

where $\frac{d_n^S}{d_n}, S \in \{R, L\}$ is the probability between the number of feature samples in d_n^L (the left child node) and d_n^R (the right child node) and $H(d_n)$ is the entropy of d_n.

Step 3. The conditional random-forest-based face age classifier is dynamically constructed according to face gender.

Under the condition of face gender $g \in \Omega_n$, we can model the conditional probability $p(a|\Omega_n, y)$ of facial age by voting on all trees in the random forest T^A:

$$p(a|\Omega_n, y) = \frac{1}{M}\sum_m p(a|\Omega_n, l_m(y)) \tag{11}$$

In the case of unknown face gender g, we can model the probability $p(a|y)$ of facial age:

$$\begin{aligned} p(a|y) &= \sum_n p(a|\Omega_n, y)\int_{g\in\Omega_n} p(g|y)dg \\ &= \sum_n \Big(\frac{1}{M}\sum_m p(a|\Omega_n, l_m(y))\Big)\int_{g\in\Omega_n} p(g|y)dg \\ &\approx \frac{1}{M}\sum_n \sum_{m=1}^{k_n} p(a|\Omega_n, l_{m,\Omega_n}(y)) \end{aligned} \tag{12}$$

where $k_n \approx M\int_{g\in\Omega_n} p(g|y)dg$.

It can be seen from the above equation, in the facial age classification, k_n decision trees are randomly selected from the conditional random forest $T^S(\Omega_n)$ to construct the random forest T^A dynamically according to the results of face gender estimation, and then, the age probability $p(a|y)$ of the test image feature y is obtained by voting on each decision tree in T^A.

3. Experimental Results

3.1. Datasets and Settings

In order to evaluate the performance of our model, we used two publicly available benchmarking databases, namely MORPH-II [26] and Adience [1].

The MORPH-II database is the largest public dataset of non-celebrities marked by gender and age, including 46,645 male images and 8487 female images, ranging in age from 16 to 77. We split the selected image sets from the MORPH II datasets into three age groups: 16–30, 31–45 and 46–60+.

The Adience database consists of images that are automatically uploaded to Flickr from smart-phone devices, which are collected for age and gender classification. Because these images are not manually filtered before they are uploaded, as in the case of media websites or social networking sites, these images are collected in an uncontrollable environment, reflecting many challenges in the real world of faces appearing in Internet images. Therefore, Adience images have extreme environmental variations, such as illumination conditions, pose and resolution changes. The Adience database includes roughly 26 K images of 2284 subjects. Table 1 shows the dataset by age category. Tests classified by age or gender are performed by using a standard five-fold, subject-exclusive cross-validation protocol, defined in [1].

Table 1. Adience faces benchmark.

	0–2	**4–6**	**8–13**	**15–20**	**25–32**	**38–43**	**48–53**	**60−**	**Total**
Male	745	928	934	734	2308	1294	392	442	8192
Female	682	1234	1360	919	2589	1056	433	427	9411
Both	1427	2162	2294	1653	4897	2350	825	869	17,603

Figure 6 shows the facial gender and age classification examples of MORPH-II and Adience. We used the Pytorch framework for implementing MMFL. In the training process, random translation and mirror data augmentation methods are introduced. The key training parameters in the experiments include the learning rate (0.001), epochs (6000), splitting interactive times (1500) and tree depths (20).

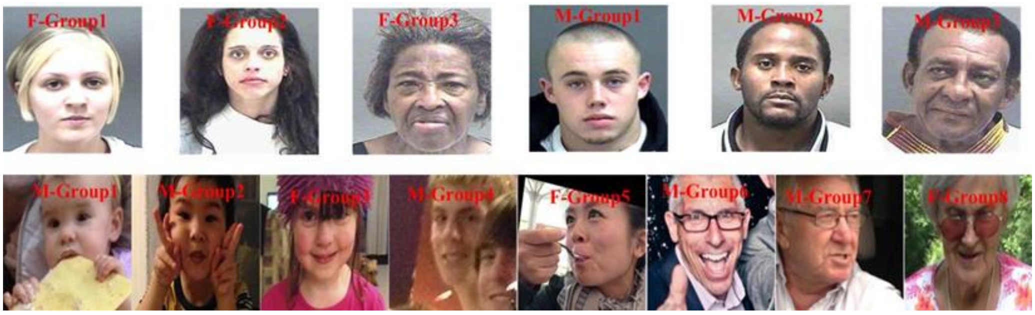

Figure 6. Examples of recognition results from MORPH-II and Adience datasets. Top row: results of MORPH-II. Bottom row: results of Adience.

3.2. Face Feature Extraction Experiments

In order to evaluate the influence of feature representation, the common feature extraction methods used in facial gender and age recognition were selected for comparative analysis, including deep learning features, Gabor, LBP and BIF. The comparative results with six features based on the Adience datasets are shown in Table 2. The results show that our MMFL features achieve the best results. In the challenging dataset with SVM, the gender and age recognition rates reached 92.35% and 55.24% by using the MMFL features, which were improved by about 4% with respect to the second-best result. Additionally, compared with SVM, the DRF has better recognition performance.

Table 2. Gender and age classification accuracy (%) of SVM/DRF using different image features.

Features	SVM (Gender/Age)	DRF (Gender/Age)
MMFL	**92.35/55.24**	**93.48/63.72**
Gabor [10]	82.61/42.72	82.45/48.62
LBP [9]	84.52/41.47	85.06/47.67
BIF [11]	83.48/44.06	83.67/50.61
Plain CNN [2]	86.83/50.75	87.14/55.32
ResNet50 [44]	88.21/51.58	89.84/58.05

3.3. Facial Gender and Age Recognition

- Facial Gender Estimation:

We evaluated the method based on MORPH-II and Adience databases, in comparison with the state-of-the-art facial gender estimation and age grouping methods. Table 3 lists the comparison results of our method, plain CNN [2], RoR [20] and CNN-ELM [3] for gender estimation. For the Adience database, we directly selected the experimental results of plain CNN, RoR and CNN-ELM. For the MORPH-II database, as plain CNN, RoR and CNN-ELM did not conduct experiments based on this database, we reproduced these methods and took the best results for comparison. The plain CNN uses AlexNet architecture to obtain an average accuracy of 98.7% and 86.8% based on MORPH-II and Adience databases, respectively. For Residual Networks of Residual Networks (RoR), which use the basic block and bottleneck block to construct the training network, the average accuracy using RoR is 99.5% and 92.43% based on MORPH-II and Adience databases, respectively. The CNN-ELM combines Convolutional Neural Networks and the Extreme Learning Machine in a hierarchical fashion, which takes advantage of CNN and ELM; the average accuracy using CNN-ELM is 98.5% and 88.2% based on MORPH-II and Adience databases, respectively. Our method achieves an average accuracy of 99.6% and 93.48% based on MORPH-II and Adience databases, respectively, which is competitive with the methods mentioned above. It should be pointed out that the accuracy of the ROR method is similar to that of our method. However, its network is deeper and more complex and its training time is longer. Comparing DRF with RoR, the training time of DRF is less than one-tenth of the time of RoR and the testing time of DRF is also much less than that of RoR.

Table 3. Gender estimation accuracy (%) by using different methods based on two datasets.

Methods	**Accuracy**	
	MORPH-II	**Adience**
plain CNN	98.7	86.8
RoR	99.5	92.43
CNN-ELM	98.5	88.2
Ours	**99.6**	**93.48**

- Facial age grouping:

In comparison with the state-of-the-art facial age grouping methods, Table 4 shows the average age grouping accuracy based on MORPH-II and Adience datasets. The plain CNN achieves an average accuracy of 89.15% and 50.7% based on MORPH-II and Adience databases, respectively. RoR achieves an average accuracy of 94.86% and 62.34%, respectively. The CNN-ELM achieves an average accuracy of 92.58% and 52.3%, respectively. Our method achieves an average accuracy of 96.14% and 63.72% based on MORPH-II and Adience databases, respectively. It is shown that the accuracy of DRF is greater than that of the other methods.

Table 4. Age grouping accuracy (%) by using different methods based on two datasets.

Methods	**Accuracy**	
	MORPH-II	**Adience**
plain CNN	89.15	50.7
RoR	94.86	62.34
CNN-ELM	92.58	52.3
Ours	**96.14**	**63.72**

Age grouping confusion matrixes with MORPH-II and Adience datasets are shown in Tables 5 and 6. The accuracies are all above 93% with an average accuracy of 96.14% for the MORPH-II database. In the Adience database, the average accuracy was 63.72% and the highest accuracy was 66.9% for group1 (0–2), followed by that of group5 and group8. The lowest accuracy was 59.29% for group7.

Table 5. Face age grouping confusion matrix in MORPH-II.

	Group1: 16–30	**Group2: 31–45**	**Group3: 46–60+**
Group1: 16–30	**97.8**	1.4	0.8
Group2: 31–45	1.8	**96.6**	1.6
Group3: 46–60+	3.2	2.78	**94.02**

Table 6. Facial age grouping confusion matrix in Adience.

	0–2	**4–6**	**8–13**	**15–20**	**25–32**	**38–43**	**48–53**	**60−**
0–2	**66.90**	24.35	8.50	0.25	0	0	0	0
4–6	21.44	**63.03**	14.36	1.17	0	0	0	0
8–13	2.57	15.36	**62.97**	18.76	0.34	0	0	0
15–20	0	0.79	16.56	**64.20**	15.97	2.48	0	0
25–32	0	0	0.74	13.73	**65.35**	19.15	1.03	0
38–43	0	0	0	0.8	18.38	**61.78**	17.46	1.58
48–53	0	0	0	1.82	3.46	15.28	**60.29**	19.15
60−	0	0	0	0.44	4.53	10.63	19.16	**65.24**

3.4. Facial Gender Alignment Analysis

An experimental comparison of age grouping with and without gender-aligned conditional probability is shown in Figure 7. This shows that the proposed method using gender-aligned conditional probability outperformed the other without gender-aligned conditional probability based on both MORPH-II and Adience datasets. The recognition rate was improved by about 8% based on the Adience dataset. This demonstrates that facial gender and age exhibit a mutual influence and interaction and it is helpful to study them together to improve the recognition rate.

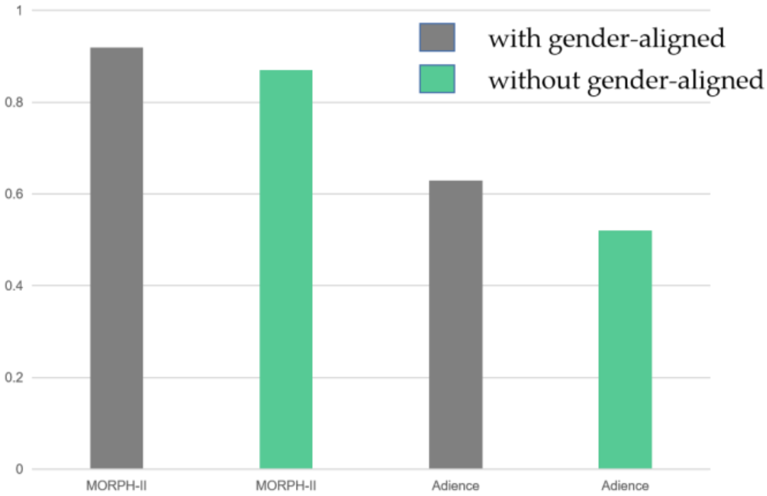

Figure 7. Age grouping with and without gender-aligned conditional probability.

4. Conclusions and Future Work

We present a novel deep-learning-enhanced multi-task random forest method for facial gender and age recognition. The facial robust features are extracted using multi-instances and multi-scale deep learning, and the facial gender and age are recognized together using a multi-task random forest. The proposed approach achieves good results owing to transfer learning, multi-instance multi-scale learning and multi-task conditional random forest learning. The multi-instance multi-scale learning features can alleviate the problem of intra-person variation, such as low image resolution, illumination and occlusion; the multi-task random forest can alleviate the inter-subject variations existing due to different personal attributes, such as gender, ethnic backgrounds and level of expressiveness.

In the future, we plan to consider other factors in our model. In reality, facial age is not only related to gender, but also to other attributes, such as ethnicity, expressions and poses, etc. If we can take all facial attributes into account and learn the relationship between the attributes and age, it will definitely help us to improve the facial age grouping accuracy. In addition, using the interdependence between face attributes, multi-task learning can identify multiple attributes, such as gender, race, age, expression and the pose of a face at one time, to achieve the goal of a double win.

Author Contributions: Conceptualization, H.L. and L.Y.; methodology, H.L. and M.W.; validation, H.L., L.Z. and G.J.; formal analysis, H.L. and M.W.; writing—original draft preparation, H.L. and M.W.; writing—review and editing, L.Z., G.J. and N.X.; supervision, M.W., G.J. and N.X.; funding acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the natural science foundation of Hubei province, grant number No. 2021CFB388, the outstanding young and middle-aged science and technology innovation team of universities in Hubei Province, grant number No. T2022032, the innovation team of Hubei University of Science and Technology, grant number No. 2022T04 and the science and technology planning project of Xianning city, grant number No. 2022GXYF056.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Eidinger, E.; Enbar, R.; Hassner, T. Age and gender estimation of unfiltered faces. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2170–2179. [CrossRef]
2. Levi, G.; Hassner, T. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 34–42.
3. Wu, C.; Luo, C.; Xiong, N.; Zhang, W.; Kim, T.H. A greedy deep learning method for medical disease analysis. *IEEE Access* **2018**, *6*, 20021–20030. [CrossRef]
4. Gupta, S.K.; Yesuf, S.H.; Nain, N. Real-Time Gender Recognition for Juvenile and Adult Faces. *Comput. Intell. Neurosci.* **2022**, *2022*, 1503188. [CrossRef]
5. Sendik, O.; Keller, Y. DeepAge: Deep learning of face-based age estimation. *Signal Process. Image Commun.* **2019**, *78*, 368–375. [CrossRef]
6. Guehairia, O.; Ouamane, A.; Dornaika, F.; Taleb-Ahmed, A. Feature fusion via Deep Random Forest for facial age estimation. *Neural Netw.* **2020**, *130*, 238–252. [CrossRef]
7. Gupta, S.K.; Nain, N. Single attribute and multi attribute facial gender and age estimation. *Multimed. Tools Appl.* **2022**, 1–23. [CrossRef]
8. Dantcheva, A.; Elia, P.; Ross, A. What else does your biometric data reveal? A survey on soft biometrics. *IEEE Trans. Inf. Forensics Secur.* **2015**, *11*, 441–467. [CrossRef]
9. Gunay, A.; Nabiyev, V.V. Automatic age classification with LBP. In Proceedings of the 2008 23rd International Symposium on Computer and Information Sciences, Istanbul, Turkey, 27–29 October 2008; pp. 1–4.
10. Gao, F.; Ai, H. Face age classification on consumer images with gabor feature and fuzzy lda method. In Proceedings of the International Conference on Biometrics, Alghero, Italy, 2–5 June 2009; Springer: Cham, Switzerland, 2009; pp. 132–141.
11. Guo, G.; Mu, G.; Fu, Y.; Huang, T.S. Human age estimation using bio-inspired features. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 112–119.
12. Yan, S.; Liu, M.; Huang, T.S. Extracting age information from local spatially flexible patches. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 737–740.
13. Huang, H.; Wei, X.; Zhou, Y. An overview on twin support vector regression. *Neurocomputing* **2022**, *490*, 80–92. [CrossRef]
14. Karthikeyan, V.; Priyadharsini, S.S. Adaptive boosted random forest-support vector machine based classification scheme for speaker identification. *Appl. Soft Comput.* **2022**, *131*, 109826.
15. Guo, G.; Mu, G. Joint estimation of age, gender and ethnicity: CCA vs. PLS. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–6.
16. Guo, G.; Mu, G. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 657–664.
17. Greco, A.; Saggese, A.; Vento, M.; Vigilante, V. Effective training of convolutional neural networks for age estimation based on knowledge distillation. *Neural Comput. Appl.* **2021**, *34*, 21449–21464. [CrossRef]
18. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, present, and future of face recognition: A review. *Electronics* **2020**, *9*, 1188. [CrossRef]
19. Wang, X.; Guo, R.; Kambhamettu, C. Deeply-learned feature for age estimation. In Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2015; pp. 534–541.
20. Zhang, K.; Gao, C.; Guo, L.; Sun, M.; Yuan, X.; Han, T.X.; Zhao, Z.; Li, B. Age group and gender estimation in the wild with deep RoR architecture. *IEEE Access* **2017**, *5*, 22492–22503. [CrossRef]
21. Niu, X.X.; Suen, C.Y. A novel hybrid CNN–SVM classifier for recognizing handwritten digits. *Pattern Recognit.* **2012**, *45*, 1318–1325. [CrossRef]
22. Liu, F.; Lin, G.; Shen, C. CRF learning with CNN features for image segmentation. *Pattern Recognit.* **2015**, *48*, 2983–2992. [CrossRef]
23. Xie, G.S.; Zhang, X.Y.; Yan, S.; Liu, C.L. Hybrid CNN and dictionary-based models for scene recognition and domain adaptation. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *27*, 1263–1274. [CrossRef]
24. Liu, Y.; Yuan, X.; Gong, X.; Xie, Z.; Fang, F.; Luo, Z. Conditional convolution neural network enhanced random forest for facial expression recognition. *Pattern Recognit.* **2018**, *84*, 251–261. [CrossRef]
25. Lanitis, A.; Draganova, C.; Christodoulou, C. Comparing different classifiers for automatic age estimation. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2004**, *34*, 621–628. [CrossRef]
26. Ricanek, K.; Tesafaye, T. Morph: A longitudinal image database of normal adult age-progression. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, UK, 10–12 April 2006; pp. 341–345.
27. Escalera, S.; Gonzalez, J.; Baró, X.; Pardo, P.; Fabian, J.; Oliu, M.; Escalante, H.J.; Huerta, I.; Guyon, I. Chalearn looking at people 2015 new competitions: Age estimation and cultural event recognition. In Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 12–17 July 2015; pp. 1–8.
28. Wu, C.; Ju, B.; Wu, Y.; Lin, X.; Xiong, N.; Xu, G.; Li, H.; Liang, X. UAV autonomous target search based on deep reinforcement learning in complex disaster scene. *IEEE Access* **2019**, *7*, 117227–117245. [CrossRef]
29. Fu, A.; Zhang, X.; Xiong, N.; Gao, Y.; Wang, H.; Zhang, J. VFL: A verifiable federated learning with privacy-preserving for big data in industrial IoT. *IEEE Trans. Ind. Inform.* **2022**, *18*, 3316–3326. [CrossRef]

30. Kumar, P.; Kumar, R.; Srivastava, G.; Gupta, G.P.; Tripathi, R.; Gadekallu, T.R.; Xiong, N.N. PPSF: A privacy-preserving and secure framework using blockchain-based machine-learning for IoT-driven smart cities. *IEEE Trans. Netw. Sci. Eng.* **2021**, *8*, 2326–2341. [CrossRef]
31. Chen, Y.; Zhou, L.; Pei, S.; Yu, Z.; Chen, Y.; Liu, X.; Du, J.; Xiong, N. KNN-BLOCK DBSCAN: Fast clustering for large-scale data. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *51*, 3939–3953. [CrossRef]
32. Zhao, J.; Huang, J.; Xiong, N. An effective exponential-based trust and reputation evaluation system in wireless sensor networks. *IEEE Access* **2019**, *7*, 33859–33869. [CrossRef]
33. Xia, F.; Hao, R.; Li, J.; Xiong, N.; Yang, L.T.; Zhang, Y. Adaptive GTS allocation in IEEE 80.215. 4 for real-time wireless sensor networks. *J. Syst. Archit.* **2013**, *59*, 1231–1242.
34. Yao, Y.; Xiong, N.; Park, J.H.; Ma, L.; Liu, J. Privacy-preserving max/min query in two-tiered wireless sensor networks. *Comput. Math. Appl.* **2013**, *65*, 1318–1325. [CrossRef]
35. Gao, Y.; Xiang, X.; Xiong, N.; Huang, B.; Lee, H.J.; Alrifai, R.; Jiang, X.; Fang, Z. Human action monitoring for healthcare based on deep learning. *IEEE Access* **2018**, *6*, 52277–52285. [CrossRef]
36. Cheng, H.; Xie, Z.; Shi, Y.; Xiong, N. Multi-step data prediction in wireless sensor networks based on one-dimensional CNN and bidirectional LSTM. *IEEE Access* **2019**, *7*, 117883–117896. [CrossRef]
37. Saggu, G.S.; Gupta, K.; Mann, P.S. Efficient Classification for Age and Gender of Unconstrained Face Images. In Proceedings of the International Conference on Computational Intelligence and Emerging Power System, Ajmer, India, 9–10 March 2021; Springer: Singapore, 2022; pp. 13–24.
38. Yang, Z.; Zhang, H.; Sudjianto, A.; Zhang, A. An effective SteinGLM initialization scheme for training multi-layer feedforward sigmoidal neural networks. *Neural Netw.* **2021**, *139*, 149–157. [CrossRef]
39. Dikananda, A.R.; Ali, I.; Fathurrohman; Rinaldi, R.A.; Iin. *Genre e-sport gaming tournament classification using machine learning technique based on decision tree, Naive Bayes, and random forest algorithm*; IOP Conference Series: Materials Science and Engineering; IOP Publishing: Bristol, UK, 2021; Volume 1088, p. 012037.
40. Bai, J.; Li, Y.; Li, J.; Yang, X.; Jiang, Y.; Xia, S.T. Multinomial random forest. *Pattern Recognit.* **2022**, *122*, 108331. [CrossRef]
41. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
42. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
43. Wu, Y.H.; Liu, Y.; Xu, J.; Bian, J.W.; Gu, Y.C.; Cheng, M.M. MobileSal: Extremely efficient RGB-D salient object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 10261–10269. [CrossRef]
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.