激光与光电子学进展 先进成像

基于Transformer的跨年龄人脸识别方法

刘成[1]，曹良才[2]，靳业[3]，王浩威[3]，殷松峰[1*]

[1]清华大学合肥公共安全研究院，安徽 合肥 230601；

[2]清华大学精密测试技术及仪器国家重点实验室，北京 100084；

[3]合肥市公安局刑事警察支队，安徽 合肥 230601

摘要 人脸特征随着年龄变化而变化，会严重影响人脸识别的性能。提出一种基于Transformer的跨年龄人脸识别方法，首先通过改善的T2T-ViT模型提取人脸年龄和身份混合特征，然后通过残差因子分解获取人脸年龄特征和身份特征，再使用线性特征分解的去相关对抗式学习算法对人脸的年龄特征和身份特征去除相关性，从而实现年龄抗干扰性的人脸识别。相比基于卷积神经网络的DAL和MTLFace方法，所提方法在参数量、multiply-add operations(MACs)和计算耗时上均有明显降低，同时在基准数据集AgeDB-30、CACD_VS、CALFW、LFW上取得了相媲美的准确率，证明了所提方法的有效性。

关键词 人脸识别；年龄不变性；Transformer；相关性分析

中图分类号 TP751 **文献标志码** A **DOI**：10.3788/LOP220785

Transformer for Age-Invariant Face Recognition

Liu Cheng[1], Cao Liangcai[2], Jin Ye[3], Wang Haowei[3], Yin Songfeng[1*]

[1]*Hefei Institute for Public Safety Research, Tsinghua University, Hefei* 230601, *Anhui, China*;

[2]*State Key Laboratory of Precision Measurement Technology and Instruments, Tsinghua University, Beijing* 100084, *China*;

[3]*Criminal Police Detachment of Hefei Public Security Bureau, Hefei* 230601, *Anhui, China*

Abstract The change in the facial features with age is a crucial factor affecting the performance of face recognition systems. Therefore, this paper proposes a cross-age face recognition method based on a Transformer. First, the improved T2T-ViT model was used to extract mixed features considering the age and identity. The extracted age and identity features were obtained through residual factor decomposition. Subsequently, the correlation between the age and identity features was removed using a decorrelated adversarial learning algorithm with linear feature decomposition to achieve age-invariant face recognition. Compared with the convolutional neural network-based DAL and MTLFace methods, the improved model significantly reduces the number of model parameters, multiply-add operations (MACs), and calculation time. Finally, the effectiveness of the proposed method is verified using the recognition results on benchmark datasets, AgeDB-30, CACD_VS, CALFW, and LFW, and the accuracy of the proposed method is comparable to that of the DAL and MTLFace methods for age-invariant face recognition.

Key words face recognition; cross-age; Transformer; correlation analysis

1 引 言

长期以来，人脸识别一直是计算机视觉领域的研究热点。近几年，随着人工智能的快速发展，基于深度学习的人脸识别算法取得了优异的效果[1-2]。然而，在被拐儿童搜寻、嫌疑人比对等实际应用中，由于人脸特征随年龄增长发生剧烈变化，人脸识别系统的识别准确率大大降低。如何快速获取与人脸年龄无关的身份特征，实现精确的跨年龄人脸识别是亟需解决的重要问题。

近年来，研究人员在基于卷积神经网络的跨年龄人脸识别算法方面取得了诸多进展。文献[3]在跨年

收稿日期：2022-02-22；**修回日期**：2022-03-23；**录用日期**：2022-04-06；**网络首发日期**：2022-04-16

基金项目：安徽省重点研究与开发计划项目(202004d07020006)

通信作者：*yinsongfeng@tsinghua-hf.edu.cn

龄人脸识别任务中将提取的人脸年龄特征和人脸身份特征分为两个正交部分，搭建人脸身份和年龄特征多任务网络模型。文献[4]提出了一种基于线性特征分解的去相关对抗式学习(DAL)算法，通过对抗性训练，最小化了人的身份信息和年龄信息之间的相关性，并且可以显著地减少身份信息中的年龄信息。文献[5]在残差网络中引入卷积块注意力，再利用线性回归指导年龄估计任务，以提取出年龄干扰因子，再从面部特征中将年龄干扰分离，得到与年龄无关的面部特征。文献[6]通过使用卷积注意力模块以获取年龄特征，使用批核典型相关性分析模块对分解后的身份特征和年龄特征进行相关性分析，在对抗学习的训练下，使相关性最小，实现跨年龄人脸识别。文献[7]提出了统一的多任务框架来共同处理人脸识别和人脸合成两个任务，该框架可以学习年龄不变的身份相关表征，同时实现逼真的人脸合成。

以上方法提取人脸的特征时均使用深度卷积神经网络，模型的参数量和 multiply-add operations (MACs)都比较大，增加了模型的复杂度。随着深度学习研究的深入，Transformer在分类、检测、分割等任务中取得了最优效果[8-10]，其对大数据的适配能力和全局信息提取能力更强、更加稳健，并且模型复杂度较卷积神经网络低。为解决深度卷积神经网络模型复杂度高和识别过程中计算耗时多等问题，本文在T2T-ViT模型的基础上添加特征重组模块，将T2T-ViT模型输出的序列形式转换为图像形式，设计能提取人脸特征的模型以代替传统的卷积神经网络；再使用DAL算法去除人脸中的年龄信息，实现跨年龄人脸识别，进一步提升了人脸跨年龄识别的准确率。

2　基于Transformer的人脸跨年龄识别方法

2.1　识别方法模型

解决年龄不变人脸识别性能问题的主要任务是能更快更好地提取人脸的年龄和身份特征，并通过监督式的训练去除人脸中的年龄信息。所提人脸跨年龄识别方法的模型结构如图1所示，主要分为4个模块，分别为人脸特征提取、人脸特征分解、去相关性对抗式学习、多任务训练。首先人脸图片经过人脸特征提取，获得面部的混合特征；再使用残差因子模块进行特征分解，获取人脸的年龄和身份特征；通过去相关性对抗式学习的训练，获取年龄信息和身份信息的相关性；最后通过年龄特征、身份特征、相关性正则项的多任务训练以监督优化模型。

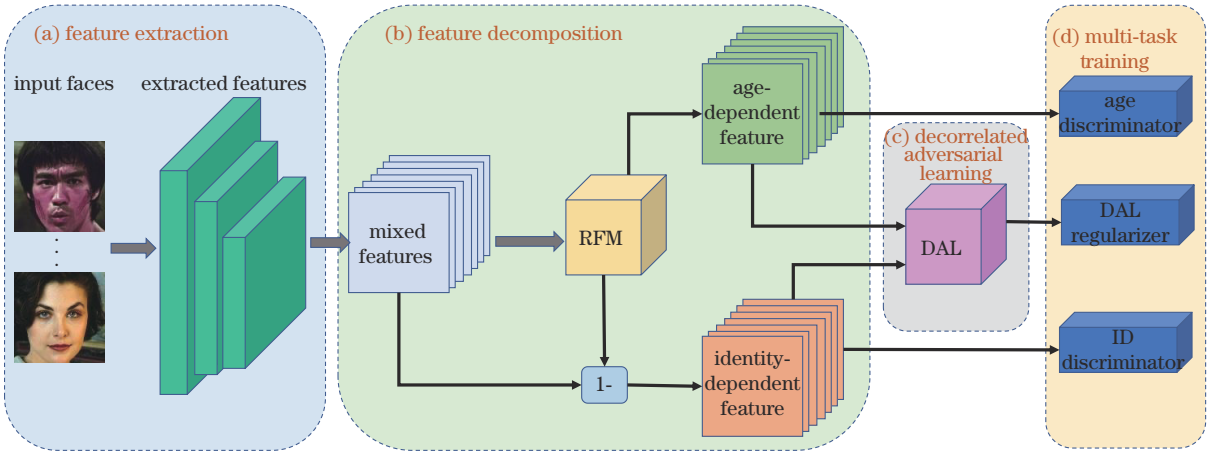

图1　T2T-DAL结构示意图

Fig. 1　Architecture of the T2T-DAL

使用的人脸特征提取模型为基于Transformer的T2T-ViT模型。Transformer基于自注意力体系结构，广泛用于自然语言处理方面，相比卷积神经网络，对大型数据表达能力强，善于捕捉全局信息，模型训练和前向速度更快，因此引入到计算机视觉领域[8]。T2T-ViT[9]改善了Transformer全局信息提取能力强而局部信息提取能力弱的问题，具有更好的性能，故本文引入T2T-ViT作为人脸特征提取基础模型。在人脸特征提取时，通过对基于Transformer的T2T-ViT模型进行改进，并有效添加特征重组模块，可以将空间维度上序列化的人脸特征重组处理为图像化形式，该方法保留了图像块的特征信息，因此可以更好地进行人脸特征分解。

2.2　人脸特征提取

主要改进T2T-ViT模型，添加特征重组模块，设计能提取人脸特征的模型，网络结构示意图如图2所示，主要分为两个部分，包括T2T-ViT模块和特征重组模块。

T2T-ViT模块主要包含软拆分(soft split)模块、重组(re-structurization)模块、类别和位置编码信息模块、TransBlocks模块。软拆分模块将图像拆分成一个个重叠块，以建模局部结构信息。经过软拆分，特征 $\boldsymbol{z}\in\mathbb{R}^{H\times W\times C}$ 变为 $\boldsymbol{T}\in\mathbb{R}^{L\times S}$，公式为

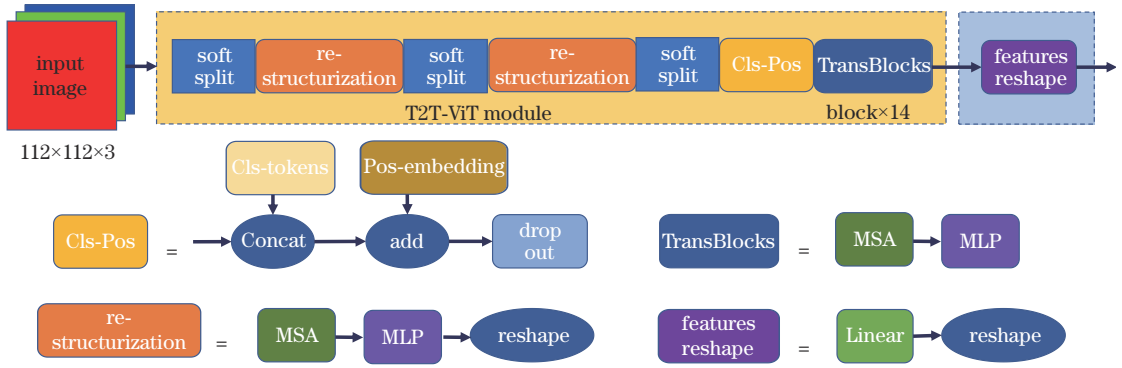

图 2　人脸特征提取模型的网络结构

Fig. 2　Network structure of face feature extraction model

$$l=\left\lfloor\frac{h+2p-k}{k-r}+1\right\rfloor\times\left\lfloor\frac{w+2p-k}{k-r}+1\right\rfloor,\tag{1}$$

$$s_{\mathrm{n}}=ck^{2},\tag{2}$$

式中：h表示特征的高度；w表示特征的宽度；c表示特征的个数；l表示序列$\boldsymbol{T}$的长度；s_{n}表示序列$\boldsymbol{T}$的个数；k为重叠块的大小；r为重叠尺寸；p为填充区(padding)大小。

重组模块的目的是将特征$\boldsymbol{T}\in\mathbb{R}^{L\times S}$变为$\boldsymbol{z}\in\mathbb{R}^{H\times W\times C}$，使用自注意力模块和变换处理。对于序列$\boldsymbol{T}\in\mathbb{R}^{L\times S}$，经过注意力模块处理后的序列为$\boldsymbol{T}'\in\mathbb{R}^{L\times S}$，再经过变换处理后，在空间维度上将序列形式处理为图像形式：

$$\boldsymbol{I}=R(\boldsymbol{T}')=R\left[\mathrm{MLP}(\boldsymbol{T})\right],\tag{3}$$

式中：$\boldsymbol{I}\in\mathbb{R}^{H\times W\times C}$；$C=S$，$L=H\times W$；$R$为变换处理。

在类别和位置编码信息过程中，Transformer采用文献[8]的做法，添加类别信息经过编码后对应的结果作为整个句子的表示；类似地，文献[9]图像中添加类别信息经过编码后对应的结果作为整个图的表示。Transformer中自注意力机制提取出来的特征没有位置信息，而图像中位置信息具有重要性，文献[8-9]中加入了位置编码信息以优化位置信息特征。本文使用的类别和位置编码信息与文献[9]中相同。

在TransBlocks模块中，本文采用了Transformer的多头自注意力机制(MSA)和multi-layer perceptron(MLP)模块，使用T2T-ViT的T2T-ViT-14模型，这里使用的Block的数目为14。

在特征重组模块中，序列经过TransBlocks模块后，输出的特征为序列形式，表征为二维特征，而人脸特征分解模块的输入需要三维特征。为保留图像块的特征信息，用于之后的人脸特征分解，需在空间维度上将表征为二维的序列形式处理为表征为三维的图像形式。对序列$\boldsymbol{T}\in\mathbb{R}^{L\times S}$在TransBlocks模块处理前添加了类别信息，使其变为$\boldsymbol{T}\in\mathbb{R}^{(L+1)\times S}$；经过TransBlocks模块后，输出特征$\boldsymbol{T}'\in\mathbb{R}^{(L+1)\times S}$；对$\boldsymbol{T}'$进行线性处理后，输出$\boldsymbol{T}''\in\mathbb{R}^{L\times S}$，然后进行转换处理，输出特征$\boldsymbol{F}\in\mathbb{R}^{H\times W\times C}$；$\boldsymbol{F}$通过卷积处理最终输出特征$\boldsymbol{X}$。各过程涉及的公式分别为

$$\boldsymbol{T}'=\mathrm{TransBlocks}(\boldsymbol{T}),\tag{4}$$

$$\boldsymbol{T}''=\mathrm{Linear}(\boldsymbol{T}'),\tag{5}$$

$$\boldsymbol{F}=R(\boldsymbol{T}''),\tag{6}$$

$$\boldsymbol{X}=\Re(\boldsymbol{F}),\tag{7}$$

式中：$\boldsymbol{X}\in\mathbb{R}^{k}$，$k$一般取512；$\Re(\cdot)$表示卷积处理。

2.3　人脸特征分解

通过人脸特征提取模型进行特征提取后，获得人脸的年龄信息$\boldsymbol{X}_{\mathrm{age}}$和身份信息$\boldsymbol{X}_{\mathrm{id}}$的混合特征$\boldsymbol{X}$：

$$\boldsymbol{X}=\boldsymbol{X}_{\mathrm{id}}+\boldsymbol{X}_{\mathrm{age}}。\tag{8}$$

对混合特征$\boldsymbol{X}$，利用残差因子分解模块(RFM)[4]获取人脸的年龄特征$\boldsymbol{X}_{\mathrm{age}}$：

$$\boldsymbol{X}_{\mathrm{age}}=\Re\left[\mathrm{RFM}(\boldsymbol{X})\right]。\tag{9}$$

则根据式(8)得到人脸的身份特征为

$$\boldsymbol{X}_{\mathrm{id}}=\boldsymbol{X}-\boldsymbol{X}_{\mathrm{age}}=\boldsymbol{X}-\Re\left[\mathrm{RFM}(\boldsymbol{X})\right]。\tag{10}$$

RFM结构如图3所示。

图 3　RFM结构

Fig. 3　RFM architecture

2.4　去相关性对抗式学习

通过人脸特征分解获取了人脸的身份信息$\boldsymbol{X}_{\mathrm{id}}$和年龄信息$\boldsymbol{X}_{\mathrm{age}}$，这两个特征信息之间可能存在着潜在的线性相关性的关系。为了更好地消除人脸的身份信息$\boldsymbol{X}_{\mathrm{id}}$和年龄信息$\boldsymbol{X}_{\mathrm{age}}$之间的潜在关系，对$\boldsymbol{X}_{\mathrm{id}}$和$\boldsymbol{X}_{\mathrm{age}}$进行去相关对抗学习[4]，通过计算人脸的身份信息和年龄信息之间的相关性，再通过监督训练以减少身份信息中的年龄信息。

首先使用线性规范化映射模块(CMM)，对$\boldsymbol{X}_{\mathrm{id}}$和$\boldsymbol{X}_{\mathrm{age}}$进行规范化并映射到$\boldsymbol{V}_{\mathrm{id}}$和$\boldsymbol{V}_{\mathrm{age}}$：

$$\boldsymbol{V}_{\mathrm{id}}=\mathrm{CMM}\left(\boldsymbol{X}_{\mathrm{id}}\right)=\boldsymbol{w}_{\mathrm{id}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{id}},\tag{11}$$

$$\boldsymbol{V}_{\mathrm{age}}=\mathrm{CMM}\left(\boldsymbol{X}_{\mathrm{age}}\right)=\boldsymbol{w}_{\mathrm{age}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{age}},\tag{12}$$

式中：$\boldsymbol{w}_{\mathrm{id}}^{\mathrm{T}}$和$\boldsymbol{w}_{\mathrm{age}}^{\mathrm{T}}$为神经网络模型参数。然后，计算$\boldsymbol{V}_{\mathrm{id}}$和$\boldsymbol{V}_{\mathrm{age}}$之间的相关性$\rho$：

$$\rho=\frac{\mathrm{Cov}\left(\boldsymbol{V}_{\mathrm{id}},\boldsymbol{V}_{\mathrm{age}}\right)}{\sqrt{\mathrm{Var}\left(\boldsymbol{V}_{\mathrm{id}}\right)\mathrm{Var}\left(\boldsymbol{V}_{\mathrm{age}}\right)}}=\frac{\left(\left|\boldsymbol{V}_{\mathrm{id}}\right|-\mu_{\mathrm{id}}\right)\left(\left|\boldsymbol{V}_{\mathrm{age}}\right|-\mu_{\mathrm{age}}\right)}{\sqrt{\sigma_{\mathrm{id}}^2+\varepsilon}\sqrt{\sigma_{\mathrm{age}}^2+\varepsilon}},\tag{13}$$

式中：μ_{id}和σ_{id}^2分别为$\left|\boldsymbol{V}_{\mathrm{id}}\right|$的均值和方差，$\mu_{\mathrm{age}}$和$\sigma_{\mathrm{age}}^2$分别为$\left|\boldsymbol{V}_{\mathrm{age}}\right|$的均值和方差；$\varepsilon$是一个数值稳定的常量，一般取0.000001。

2.5　多任务训练

使用多任务训练策略以监督并优化人脸特征的学习，主要包括三个监督信息：人脸年龄特征判别器、人脸身份特征判别器、人脸年龄和身份特征去相关性对抗学习的正则项。

在人脸年龄特征判别器中，对于年龄信息的优化学习，将年龄信息划分为8组：

$$g_{\mathrm{age}}=\begin{cases}0, & a\leqslant 12\\ 1, & 12<a\leqslant 18\\ 2, & 18<a\leqslant 25\\ 3, & 25<a\leqslant 35\\ 4, & 35<a\leqslant 45\\ 5, & 45<a\leqslant 55\\ 6, & 55<a\leqslant 65\\ 7, & a>65\end{cases},\tag{14}$$

式中：a为年龄。使用交叉熵损失来优化人脸年龄特征判别器的学习：

$$L_{\mathrm{age}}=-\log\frac{\mathrm{e}^{y_i}}{\sum_{j=1}^{n}\mathrm{e}^{y_j}}=-y_i+\log\left(\sum_{j=1}^{n}\mathrm{e}^{y_j}\right),\tag{15}$$

式中：n为类别数，这里为8；y_i为类别i对应的年龄标签。

在人脸身份特征判别器中，使用ArcFace[2]来监督并优化人脸身份特征判别器的学习：

$$L_{\mathrm{id}}=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{\mathrm{e}^{s\cos\left(\theta_{y_i}+m\right)}}{\mathrm{e}^{s\left[\cos\left(\theta_{y_i}+m\right)\right]}+\sum_{j=1,j\neq y_i}^{n}\mathrm{e}^{s\cos\theta_j}},\tag{16}$$

式中：N为样本数；y_i为样本i对应的身份标签；s为超参数，标量值，一般为64；m为超参数，一般为0.5。

在去相关对抗学习正则项中，正则项即为人脸年龄和人脸身份的相关性：

$$L_\rho=\rho。\tag{17}$$

综上，多任务训练监督并优化的总损失函数为

$$L=L_{\mathrm{id}}+\alpha L_{\mathrm{age}}+\beta L_\rho,\tag{18}$$

式中：α和β均为平衡比例系数，一般均取值0.1。

3　算法验证

3.1　实验数据

使用的实验数据如表1所示。训练数据集与文献[2]相同，称为faces emore[11]。参考文献[7]，使用公开的Azure Facial API[12]来估计数据集中的人脸年龄，最终获取带有人脸年龄信息的85742个人脸，共5774205张图片。将faces emore数据集中的年龄信息按照式(14)分为8组：12－、13～18、19～25、26～35、36～45、46～55、56～65、65＋。测试数据集包括基准数据集AgeDB-30[13]、CACD_VS[14]、CALFW[15]、LFW[16]，具体信息如表1所示。

表1　使用的数据集的信息

Table 1　Information about the used dataset

Dataset	Number of identities	Number of images
faces emore[11]	85742	5774205
AgeDB-30[13]	6000 pairs	12000
CACD_VS[14]	4000 pairs	8000
CALFW[15]	6000 pairs	12000
LFW[16]	6000 pairs	12000

3.2　训练细节

在模型训练时，设置batch为512，设置训练迭代次数为20；使用SGD优化网络模型；初始学习率为0.01，在迭代次数为12、15、18时，学习率分别降为上一次学习率的0.1，即0.01、0.001、0.0001、0.00001；设置momentum为0.9；设置ArcFace的超参数s为64，m为0.5；设置损失函数的超参数α、β分别为0.1、0.1；使用单卡NVIDIA V100显卡进行训练。DAL结构训练时需要对DAL模块和其他模型交替进行训练，每个迭代中，每训练70个batch时，训练DAL模块20个batch，其他模块训练50个batch。

测试时，使用批处理大小为16(batch为16)进行测试。测试环境：Python3、PyTorch1.8.1、显卡为NVIDIA GeForce RTX 2060、CUDA版本为11.4。

3.3　模型复杂度和耗时测试结果

所提模型与文献[4]中的DAL模型、文献[7]中的MTLFace模型在参数量、MACs和耗时上的结果如表2所示。在MACs上，所提模型为1.2 GB，相比DAL(OE-CNNs)模型和MTLFace模型，分别减少了85.28%和81.01%；在参数量上，所提模型为40.71×10^6，相比DAL(OE-CNNs)和MTLFace模型，分别减少了28.68%和13.84%。在测试单张图片平均耗时方面，测试了44000张图片，图片尺寸为112×112。在batch为1时，所提模型的单张图片平均耗时为13.75 ms，相比DAL(OE-CNNs)和MTLFace模型，分别降低了45.02%和37.47%；在batch为16时，所提模型的单张图片平均耗时为1.19 ms，相比DAL(OE-CNNs)和MTLFace模型，分别降低了55.76%和53.15%。因此，相比传统卷积神经网络模型的DAL(OE-CNNs)和MTLFace，所提模型减少了参数量和MACs，降低了模型的复杂度，同时大大地降低了人脸识别过程的耗时。

表 2 不同模型的参数的对比

Table 2 Comparison of parameters of different models

Parameter		Proposed model	DAL(OE-CNNs)[4]	MTLFace[7]
MACs /GB		1.2	8.15	6.32
Params /10^6		40.71	57.08	47.25
Inference speed /ms	Batch size is 1	13.75	25.01	21.99
	Batch size is 16	1.19	2.69	2.54

3.4 基准数据集测试结果

使用 faces emore 数据集进行模型训练，使用训练好的模型在公开数据集 AgeDB-30、CACD_VS、CALFW、LFW 上进行测试。在测试每个数据集时，参看同类论文中的测试方法，并严格遵循数据集的测试协议。本文复现了文献[4]中的 DAL(OE-CNNs)方法，记为 DAL(OE-CNN$_s$)-Re。

对于 AgeDB-30、CACD_VS、CALFW 三个人脸跨年龄数据集，将每个数据集分割为 10 个部分，并遵循测试协议进行 10 折交叉验证。在三个数据集上测试的结果如表 3 所示。对于 AgeDB-30 数据集，所提方法取得了 96.25% 的准确率，略高于 MTLFace 方法，比 DAL(OE-CNN$_s$)-Re 高 0.27 个百分点。对于 CACD_VS 数据集，所提方法的准确率为 99.35%，比 DAL(OE-CNN$_s$)-Re 略低 0.03 个百分点，比文献[4]中的 DAL(OE-CNNs)略低 0.05 个百分点，比 MTLFace 方法低 0.2 个百分点。对于 CALFW 数据集，所提方法的准确率为 95.48%，与 DAL(OE-CNN$_s$)-Re 方法相同，比 MTLFace 方法低 0.14 个百分点。综合测试结果，所提方法在识别准确率上略优于 DAL(OE-CNNs)方法，略差于 MTLFace 方法。相比于 MTLFace 方法，所提方法在识别准确率减小很小的情况下，显著降低了模型的复杂度，有效提升了人脸识别的实时性。

为了验证所提方法对一般人脸识别的泛化能力，在 LFW 数据集上进行了测试。LFW 是最受欢迎的一般人脸识别公共基准数据集，包含了 5749 个个体，共 13233 张人脸图像。使用公开处理好的 6000 对人脸进行测试，将数据集分割为 10 个部分，遵循测试协议进行 10 折交叉验证。测试结果如表 3 所示，所提方法在 LFW 数据集上取得了 99.63% 的准确率，比文献[4]中 DAL(OE-CNNs)提高 0.16 个百分点，比 MTLFace 方法提高 0.11 个百分点，比 DAL(OE-CNN$_s$)-Re 提高 0.07 个百分点。测试结果表明所提方法在一般人脸识别上具有很强的泛化能力。

表 3 不同方法在四个数据集上的准确率

Table 3 Accuracy of different methods on four datasets unit: %

Method	AgeDB-30[13]	CACD_VS[14]	CALFW[15]	LFW[16]
DAL(OE-CNNs)[4]		99.4		99.47
DAL(OE-CNN$_s$)-Re	95.98	99.38	95.48	99.56
MTLFace[7]	96.23	99.55	95.62	99.52
Proposed method	96.25	99.35	95.48	99.63

4 结 论

在 DAL 人脸跨年龄识别的结构上，改进基于 Transformer 的 T2T-ViT 模型，添加特征重组模块，提出了一种人脸特征提取模型。通过使用去相关性对抗学习算法和多任务训练策略进行训练调优，获得了一种基于 Transformer 的人脸跨年龄识别模型。相比基于传统卷积神经网络的 DAL(OE-CNNs)方法和 MTLFace 方法，该模型减少了参数量和 MACs，降低了模型的复杂度，大大地降低了人脸识别过程的耗时。所提模型在公开数据集上进行训练，并在公开数据集 AgeDB-30、CACD_VS、CALFW、LFW 上进行测试，获得了与现阶段人脸跨年龄识别优秀方法相媲美的结果。结果表明所提针对人脸跨识别而改进 T2T-ViT 的方法具有一定的优势，在公共安全领域具有广阔的应用前景。

参 考 文 献

[1] Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 815-823.

[2] Deng J K, Guo J, Xue N N, et al. ArcFace: additive angular margin loss for deep face recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 4685-4694.

[3] Wang Y T, Gong D H, Zhou Z, et al. Orthogonal deep features decomposition for age-invariant face recognition [EB/OL]. (2018-10-17)[2022-02-03]. https://arxiv.org/abs/1810.07599.

[4] Wang H, Gong D H, Li Z F, et al. Decorrelated adversarial learning for age-invariant face recognition[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern

Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3522-3531.

[5] 何星辰, 郭勇, 李奇龙, 等. 基于深度学习的抗年龄干扰人脸识别[J]. 自动化学报, 2022, 48(3): 877-886.
He X C, Guo Y, Li Q L, et al. Age invariant face recognition based on deep learning[J]. Acta Automatica Sinica, 2022, 48(3): 877-886.

[6] 孙文斌, 王荣, 孙连烛, 等. 基于深度学习的跨年龄人脸识别[J]. 激光与光电子学进展, 2022, 59(2): 0215001.
Sun W B, Wang R, Sun L Z, et al. Deep learning for cross-age face recorgnition[J]. Laser & Optoelectronics Progress, 2022, 59(2): 0215001.

[7] Huang Z Z, Zhang J P, Shan H M. When age-invariant face recognition meets face age synthesis: a multi-task learning framework[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 7278-7287.

[8] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2020-10-22)[2022-02-03]. https://arxiv.org/abs/2010.11929.

[9] Yuan L, Chen Y P, Wang T, et al. Tokens-to-token ViT: training vision transformers from scratch on ImageNet[EB/OL]. (2021-01-28)[2022-02-05]. https://arxiv.org/abs/2101.11986.

[10] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[EB/OL]. (2021-03-25)[2022-02-05]. https://arxiv. org/abs/2103.14030.

[11] Faces_emore face database[EB/OL]. [2022-02-05]. https://github.com/TreB1eN/InsightFace_Pytorch.

[12] AzureMicrosoft. Microsoft azure cognitive services facial recognition[EB/OL]. [2022-02-05]. https://azure.microsoft.com/enus/services/cognitive-services/face/.

[13] Moschoglou S, Papaioannou A, Sagonas C, et al. AgeDB: the first manually collected, in-the-wild age database[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1997-2005.

[14] Chen B C, Chen C S, Hsu W H. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset[J]. IEEE Transactions on Multimedia, 2015, 17(6): 804-815.

[15] Zheng T Y, Deng W H, Hu J N. Cross-age LFW: a database for studying cross-age face recognition in unconstrained environments[EB/OL]. (2017-08-28)[2022-03-05]. https://arxiv.org/abs/1708.08197.

[16] Huang G B, Mattar M, Berg T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments[EB/OL]. [2022-02-05]. http://vis-www.cs.umass.edu/papers/lfw.pdf.