光学 精密工程
OPTICS AND PRECISION ENGINEERING
核心期刊
主管：中国科学院
主办：中国科学院长春光学精密机械与物理研究所
中国仪器仪表学会
主编：曹健林

对齐特征表示的跨模态人脸识别

明悦，王绍颖，范春晓，周江婉

引用本文:
明悦, 王绍颖, 范春晓, 等. 对齐特征表示的跨模态人脸识别[J]. 光学 精密工程, 2020, 28(10): 2311-2322.
MING Yue, WANG Shao-Ying, FAN Chun-Xiao, et al. Exploring aligned latent representations for cross-domain face recognition[J]. *Optics and Precision Engineering*, 2020, 28(10): 2311-2322.

在线阅读 View online: https://doi.org/10.37188/OPE.20202810.2311

您可能感兴趣的其他文章

Articles you may be interested in

多模深度卷积神经网络应用于视频表情识别

Video-based facial expression recognition using multimodal deep convolutional neural networks

光学 精密工程. 2019, 27(4): 963-970 https://doi.org/10.3788/OPE.20192704.0963

结合双模多尺度CNN特征及自适应深度KELM的浮选工况识别

Flotation performance recognition based on dual-modality multiscale CNN features and adaptive deep learning KELM

光学 精密工程. 2020, 28(8): 1785-1798 https://doi.org/10.3788/OPE.20202808.1785

多模态特征融合与多任务学习的特种视频分类

Special video classification based on multitask learning and multimodal feature fusion

光学 精密工程. 2020, 28(5): 1177-1186 https://doi.org/10.3788/OPE.20202805.1177

壳段厚度激光检测信号的变分模态分解去噪

Adaptive denoising for laser detection signal of shell thickness based on variational mode decomposition

光学 精密工程. 2017, 25(8): 2173-2181 https://doi.org/10.3788/OPE.20172508.2173

自适应编辑传播的人脸图像光照迁移

Face relighting using adaptive edit propagation

光学 精密工程. 2015, 23(5): 1450-1457 https://doi.org/10.3788/OPE.20152305.1450

文章编号 1004-924X(2020)10-2311-12

对齐特征表示的跨模态人脸识别

明 悦*,王绍颖,范春晓,周江婉
(北京邮电大学 电子工程学院,北京 100876)

摘要:跨模态人脸识别一直是人脸识别领域的研究热点,在安防、刑侦等现实场景中具有极高的应用价值和发展潜力。现有的跨模态人脸识别算法通常在图像空间或潜在空间建立不同模态人脸的联系,却忽略了二者的内在关联性,容易导致跨模态信息的丢失。为解决这一问题,本文提出基于对齐特征表示的跨模态人脸识别算法(Cross-Domain Representation Alignment,CDRA)。CDRA 算法在人脸图像空间和潜在空间、模态内和模态间探索不同模态人脸数据间的关联性:首先,为减少信息损失,CDRA 算法通过对单一模态内人脸的重建,学习到包含判别信息的模态内潜在特征表示;然后,在图像空间,CDRA 算法通过从不同模态的潜在特征表示中,跨模态地重建图像,以间接对齐不同模态的潜在特征表示,在潜在空间,CDRA 算法通过对齐不同模态数据的潜在高斯分布直接对齐不同模态的潜在特征表示,促使特征表示学习到不同模态人脸在不同空间维度多个层次的跨模态信息。实验结果表明 CDRA 算法在 Multi-Pie 数据集上的人脸识别准确率的平均值为 97.2%,在 CASIA NIR-VIS 2.0 数据集上的人脸识别准确率为 99.4%±0.2%,同时实现了跨模态人脸数据的高效互生成。CDRA 算法能够在图像空间和潜在子空间学习到更具判别能力的跨模态关联信息,有效地提高了跨模态人脸识别准确率。

关 键 词:跨模态人脸识别;变分自动编码器;人脸合成;潜在子空间
中图分类号:TP394.1;TH691.9 **文献标识码**:A **doi**:10.37188/OPE.20202810.2311

Exploring aligned latent representations for cross-domain face recognition

MING Yue*, WANG Shao-Ying, FAN Chun-Xiao, ZHOU Jiang-Wan

(*School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing* 100876, *China*)
* *Corresponding author, E-mail: myname*35875235@126.*com*

Abstract: Cross-domain face recognition (FR) has always been a research hotspot in the field of face recognition. It has high application value and development potential in real applications such as security and criminal investigation. The existing cross-domain face recognition methods usually establish the correlation between different domain faces in the image space or latent subspace, but ignore the intrinsic relation between the two, which easily leads to the loss of inter-modal correlation information. In order to solve this problem, in this paper, we propose a novel method, called Cross-Domain Representation Alignment (CDRA). CDRA algorithm explores the correlation between different domain face data in the face image space and latent space. First, in order to reduce information loss, the CDRA al-

收稿日期:2020-07-09;**修订日期**:2020-07-30.
基金项目:国家自然科学基金资助项目(No. 62076030);北京市自然科学基金资助项目(No. L182033);中央高校基本科研业务费资助(No. 2019PTB-001)

gorithm can learn the latent feature representation containing discriminant information by reconstructing the face in a single domain. Then, in image space, CDRA algorithm is used to cross domain from different domain latent features. In the latent space, CDRA directly aligns the latent feature representations of different domain by aligning the latent Gaussian distribution of different domain data, which promotes the feature representation to learn the cross domain information of different domain faces in different spatial dimensions and levels. Experimental results indicate the average face recognition accuracy rate of CDRA is 97.2% on Multi-Pie dataset, and 99.4% ± 0.2% on CASIA NIR-VIS 2.0 dataset. Simultaneously, the efficient cross-domain face synthesis is realized. The learned latent features of our CDRA method can obtain the essential cross-domain information in both image space and latent subspace for cross-domain FR task, which can effectively improve the cross-domain face recognition.

Key words: cross-domain face recognition; variational auto-encoders; face synthesis; latent subspace

1 引　言

跨模态人脸识别的目的是识别数据分布或外观差异较大的不同模态人脸图像[1]。近红外光与可见光人脸、侧脸与正脸、素描画像与照片等都是人脸的不同模态。在安防、刑侦、娱乐等场景中，跨模态人脸识别发挥着重要作用[2]。例如，在安防场景中，不可避免要识别近红外光下拍摄的人脸图像。大多数人脸识别算法，在面对跨模态人脸识别时准确率会大幅下降。因此，研究者开始深入研究不同模态人脸之间的差异，并提出多种跨模态人脸识别算法[3-8]，降低不同模态人脸之间差异。

在跨模态人脸识别算法中，生成模型被广泛用于跨模态人脸合成和学习模态不变的特征表示[9]。生成对抗网络[10](Generative Adversarial Network, GAN)和变分自动编码器[11](Variational Auto-Encoders, VAE)是两种常用于人脸合成的基本模型。GAN 中包含生成器和判别器，二者交替训练和对抗，最终生成器生成能够欺骗过判别器的图像。然而其交替训练过程会导致训练不稳定。为克服这一缺陷，一些算法[12-13]采用 VAE 进行人脸合成。与 GAN 相比，VAE 具有更加稳定的训练过程，通过最小化重构损失函数可构建输入数据的潜在高斯分布空间，并合成逼真的人脸，从而获得具有鲁棒性和判别能力的紧凑分布，适用于跨模态人脸识别任务。因此，本文将使用 VAE 作为基本模型，学习判别性的潜在高斯分布空间。

VAE 模型能够很容易地构建出重建图像空间和潜在高斯分布空间。因此，相比于直接跨模态合成人脸或对齐潜在向量等在单一空间学习跨模态信息的算法[4-5]，本文基于 VAE 模型提出一种基于对齐特征表示的跨模态人脸识别算法(Cross-Domain Representation Alignment, CDRA)，提取不同模态人脸图像中能标识身份的特征信息和跨模态关联信息。该方法采用潜在高斯分布空间直接进行特征对齐，并在图像空间间接建立不同模态人脸间的联系方式，实现不同模态人脸特征在多空间维度的对齐，在图像空间和潜在子空间同时学习更加具有判别性的身份信息和更加丰富的多层次跨模态信息。如图 1 所示，是 CDRA 算法的框图。图中模态 A 和 B 为同一个人的可见光图像和近红外光图像。编码器通过模态内重建损失函数(L_{DSR})学习高斯潜在分布，交叉模态重建对齐损失函数(L_{CMA})和高斯分布对齐损失函数(L_{GDA})协同作用对齐潜在特征表示，将不同模态的潜在特征投影到共同的潜在子空间。该方法主要分为两部分：

(1)模态内信息提取。为减少特征学习和特征重建过程中的信息损失，CDRA 算法首先利用模态内重建(Domain-Specific-Reconstruction, DSR)损失函数来提取同一模态人脸数据的内在身份信息，主要包括同一模态人脸数据中的身份判别信息和纹理、结构等细节信息。

(2)模态间信息提取。在学习到身份信息的基础上，为减少不同模态人脸特征的差异，本文提出使用交叉模态重建对齐(Cross-Modal-Align-

ment，CMA）损失函数将潜在特征空间中某一模态的特征重构至另一模态的图像空间，学习不同模态特征间相关联的潜在信息，并利用高斯分布对齐（Gaussian-Distribution-Alignment，GDA）损失函数对齐高斯潜在分布，进一步减少不同模态人脸之间的差异。因此，CDRA 算法基于 CMA 和 GDA 损失函数，将不同模态的潜在特征表示对齐到同一潜在子空间。

本文提出的 CDRA 算法主要的贡献如下：

（1）本文提出一种端到端的跨模态特征匹配算法。该算法基于 VAE 模型构建不同模态人脸的重建图像空间和压缩的高斯分布空间来对齐跨模态潜在特征表示，并且能够很容易地扩展为同时对齐两个以上的模态。

（2）本文提出使用 DSR 损失函数，学习模态内人脸具有判别能力的身份信息，能够有效减少特征对齐过程中的信息损失。

（3）本文通过 CMA 和 GDA 损失函数在人脸图像空间和潜在高斯分布空间协同学习公共的潜在特征表示空间，从而在不同空间提取到不同模态人脸间更加丰富的关联信息。

（4）来自于公共潜在特征表示空间的特征作为输出特征，直接用于跨模态人脸识别。本文在跨模态人脸数据集 Multi-Pie 和 CASIA NIR-VIS 2.0 上进行人脸识别实验。实验结果表明，CDRA 算法获得了比现有方法更高的识别准确率，并具有良好的泛化能力。

2　跨模态人脸识别算法

跨模态人脸识别算法主要分为三类：潜在子空间方法、人脸合成方法和模态不变的特征方法。本节将从这三类方法分别综述近年来跨模态人脸识别领域的相关工作。

潜在子空间方法的目标是将不同模态的数据投影到一个公共的潜在子空间中。Wang 等[14]将 CCA（Canonical Correlation Analysis）引入到自动编码器中，学习针对不同模态特征的非线性子空间。MvDA（Multi-view Discriminant Analysis）[15]方法通过联合学习人脸多个视点的线性变换，寻找多个视点具有判别性的公共子空间。Wu 等[3]通过在跨模态变量上施加松弛约束来为不同模态的人脸特征学习公共的解构潜在空间。潜在子空间方法可以很容易地减少不同模态人脸之间的模态差异，但是在投影的过程中会存在一定程度的信息丢失。

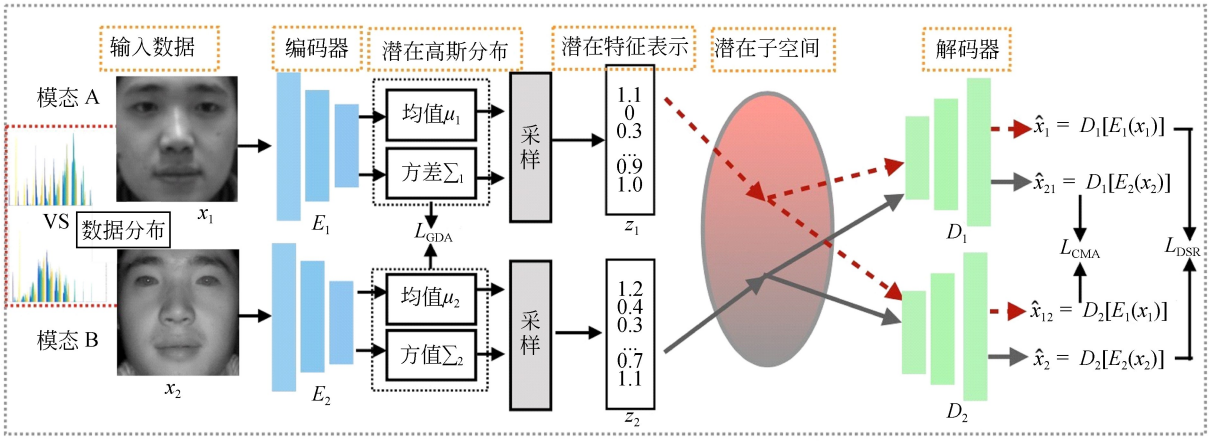

图 1　基于对齐特征表示的跨模态人脸识别算法（CDRA）的流程框图

Fig. 1　Framework of Cross-Domain Representation Alignment (CDRA) algorithm

人脸合成方法利用生成模型将人脸从一个模态合成到另外一个模态，以减少不同模态人脸数据的差异。马尔科夫网络方法[16]基于人脸的局部块合成，实现跨模态的人脸生成。Zhang 等[8]利用 Siamese 网络解构不同模态的人脸数据，通过编码-解构-解码的形式，减少不同模态人脸之间的差异。基于 GAN 的方法[7,17-18]通常通过感知图像全局或局部细节，实现不同模态间人脸的相互合成。一般人脸合成的方法对于不同的人脸生成任务需要不同的学习机制，因此人脸合成方

法存在泛化能力较弱的缺点。

模态不变的特征方法旨在从同一个体不同模态的人脸中学习模态不变的特征。CDL (Coupled Deep Learning)[5]提出一种跨模态的排序机制,能够最大化类间的差异和类内不同模态之间的差异。He 等[4]将 Wasserstein 距离引入到共享网络层中,度量不同模态人脸特征分布之间的差异。DFN (Deformable Face Net)[19]为可形变卷积层学习姿态感知的位移场,从而提取到姿态不变的人脸图像。但是当不同模态的人脸数据存在较大差异时,直接提取模态不变的人脸特征比较困难[20]。

不同于上述方法,本文提出的 CDRA 算法基于 VAE 模型学习潜在特征表示空间,并通过对齐潜在特征表示来实现跨模态人脸识别。首先,为减少信息丢失,CDRA 算法利用 DSR 损失函数尽可能地学习具有判别能力的人脸特征表示。在此基础上,CMA 和 GDA 损失函数分别在图像空间和潜在高斯分布空间对不同模态的人脸特征表示进行对齐。相比于单一空间对齐的方法,在图像空间和分布空间同时进行对齐的 CDRA 算法,能够获得不同模态人脸间多个空间维度不同层次的关联关系,有利于提取到更具判别能力的跨模态关联信息。并且,CDRA 算法本质是对不同模态人脸数据的特征表示进行对齐。因此,适用于不同的跨模态人脸识别任务,而不需要改变学习机制。

3 基于对齐特征表示的跨模态人脸识别算法

基于对齐特征表示的跨模态人脸识别(CDRA)算法是将两个模型学习得到的特征表示进行对齐,构建不同模态人脸特征之间相关联的公共潜在特征空间。为减少信息损失和实现更有效的特征对齐,CDRA 算法首先通过模态内重建(DSR)损失函数,学习单一模态人脸具有判别能力的信息。基于交叉重建和分布对齐原则,为实现特征在图像空间的精准映射和在特征空间不同模态特征的精准匹配,通过交叉模态重建对齐(CMA)损失函数和高斯分布对齐(GDA)损失函数实现特征对齐表示。不同于之前在单一图像或分布空间对不同模态的特征表示进行对齐。CDRA 算法利用 CMA 损失函数和 GDA 损失函数在图像空间和分布空间协同建立不同模态人脸间的联系,从而促进不同模态的潜在特征表示在多空间维度实现更加精确的对齐。接下来,本节将描述 CDRA 算法的损失函数及其数学表达式。

3.1 VAE 和模态内重建(DSR)损失函数

VAE 是 CDRA 模型的基本组成单元。在编码阶段,VAE 将输入图像 $\boldsymbol{x}$ 编码到潜在高斯分布空间得到潜在向量 $\boldsymbol{z}$,即 $\boldsymbol{z}=E(\boldsymbol{x})\sim Q(\boldsymbol{z}|\boldsymbol{x})$。然后,解码器将潜在向量 $\boldsymbol{z}$ 解码回图像 $\hat{\boldsymbol{x}}$,即 $\hat{\boldsymbol{x}}=D(\boldsymbol{z})\sim P(\boldsymbol{z}|\boldsymbol{x})$。也就是说,VAE 需要最大化 $\boldsymbol{x}$ 中每个像素的边界对数似然。因此,VAE 的重建误差损失是 $\boldsymbol{x}$ 的负期望对数似然,公式如下:

$$L_{\mathrm{re}}=-E_{z\sim Q(z|x)}\log P(\boldsymbol{z}\mid\boldsymbol{x}),\tag{1}$$

其中:$\boldsymbol{z}$ 是独立的高斯随机变量,即 $\boldsymbol{z}\in N(0,1)$。VAE 通过梯度下降算法最小化 $Q(\boldsymbol{z}|\boldsymbol{x})$ 的分布与高斯分布 $P(\boldsymbol{z})$ 的差异,即最小化二者的 KL 散度,对潜在向量 $\boldsymbol{z}$ 的分布进行控制:

$$L_{\mathrm{KL}}=D[Q(\boldsymbol{z}\mid\boldsymbol{x})\parallel P(\boldsymbol{z})].\tag{2}$$

因此,VAE 是损失函数 L_{re} 和 L_{KL} 共同组成:

$$L_{\mathrm{VAE}}=L_{\mathrm{re}}+L_{\mathrm{KL}}.\tag{3}$$

CDRA 算法的目标是学习 n 种模态的数据在公共潜在空间的特征表示。因此,CDRA 算法模型中包含 n 个 VAE 模型。为了减少信息损失和提取具有判别能力的信息,每一个 VAE 中的编码器将一种模态的数据编码到潜在高斯分布空间,解码器从潜在特征表示中重建出原始输入数据。CDRA 算法的模态内损失是 n 个 VAE 损失的总和,称为模态内重建(DSR)损失函数:

$$L_{\mathrm{DSR}}=\sum_{i=1}^{n}E_{z^{(i)}\sim Q(z^{(i)}|x^{(i)})}[\log P(z^{(i)}\mid x^{(i)})]-\beta D[Q(z^{(i)}\mid x^{(i)})\,||\,P(z^{(i)})],\tag{4}$$

其中:β 系数决定 KL 散度项的权重。通过最小化 DSR 损失,CDRA 算法中每个 VAE 模型的潜在特征表示空间能够学习到具有判别能力的模态内特征表示。

3.2 交叉模态重建对齐(CMA)损失函数

交叉模态重建对齐是通过解码来自同一个体另一模态的潜在特征表示来实现的。也就是说,模态 A 的潜在特征表示输入到模态 B 的解码器中来重构模态 B 的人脸图像,而模态 B 的潜在特

征表示输入模态 A 的解码器中来重构模态 A 的人脸图像。因此，每一个模态的解码器除了用于训练对应模态的潜在特征表示，也将用于训练另一模态的潜在特征表示。CMA 损失函数定义如下：

$$L_{\mathrm{CMA}}=\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\|x^{(j)}-D^{(j)}(E^{(i)}(x^{(i)}))\right\|_2^2, \tag{5}$$

其中：$\boldsymbol{E}^{(i)}$ 表示第 i 个模态的样本通过编码器得到特征表示，$\boldsymbol{D}^{(j)}$ 表示特征通过解码器得到的第 j 个模态的重建样本。通过 CMA 损失函数对模型进行优化，能够在图像空间中学习到不同模态之间的关联信息，并映射到潜在特征空间，从而实现将不同模态人脸图像的潜在特征表示映射到同一潜在子空间。

3.3　高斯分布对齐(GDA)损失函数

高斯分布对齐通过最小化同一个体不同模态的潜在高斯分布 Wasserstein 距离[4]实现。两个不同模态人脸数据高斯分布之间的 2-Wasserstein 距离，可构成封闭解：

$$d_{ij}=\left[\left\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right\|_2^2+Tr(\Sigma_i)+Tr(\Sigma_j)-2(\Sigma_i^{1/2}\Sigma_i\Sigma_j^{1/2})\right]^{1/2}, \tag{6}$$

其中，对角协方差矩阵由编码器预测，具有可交换性。因此公式(6)可以简化为：

$$d_{ij}=\left[\left\|\mu_i-\mu_j\right\|_2^2+\left\|\Sigma_i^{1/2}-\Sigma_j^{1/2}\right\|_F^2\right]^{1/2}, \tag{7}$$

其中，F 表示 Frobenius 范数。因此，在 CDRA 算法中，GDA 损失函数写作：

$$L_{\mathrm{GDA}}=\sum_{i=1}^{n}\sum_{j\neq i}^{n}d_{ij}. \tag{8}$$

通过 GDA 损失函数能够进一步对齐不同模态的特征表示，提高 CDRA 算法模型的跨模态表达能力。

3.4　CDRA 算法损失函数

CDRA 算法的总体目标损失函数包括 DSR 损失函数、CMA 损失函数和 GDA 损失函数。DSR 损失函数能够减少信息损失，学习模态内具有判别能力的身份信息。CMA 损失函数和 GDA 损失函数能够有效地关联不同模态人脸的图像空间和潜在分布空间，学习跨模态信息。为同时学习具有判别能力的身份信息和跨模态信息，CDRA 算法将三种损失函数有机结合，学习不同模态人脸的公共潜在空间和特征表示。

$$L=L_{\mathrm{DSR}}+\gamma L_{\mathrm{CMA}}+\delta L_{\mathrm{GDA}}, \tag{9}$$

其中 γ 和 δ 系数表示 CMA 损失函数和 GDA 损失函数的权重。γ 和 δ 系数在训练的不同阶段将被设置不同的权重值，有利于逐步实现特征表示对齐。在特征对齐表示的基础上，不仅可以直接从潜在特征表示空间提取到模态不变的特征，而且可以由解码器解码潜在特征得到相应模态的生成人脸。具体细节将在下节中介绍。

4　训练算法

基于对齐特征表示的跨模态人脸识别(CDRA)算法采用基于卷积神经网络的 VAE 模型[21]学习含有高层语义信息的特征。如图 2 是基于卷积神经网络的 VAE 模型的结构框图：

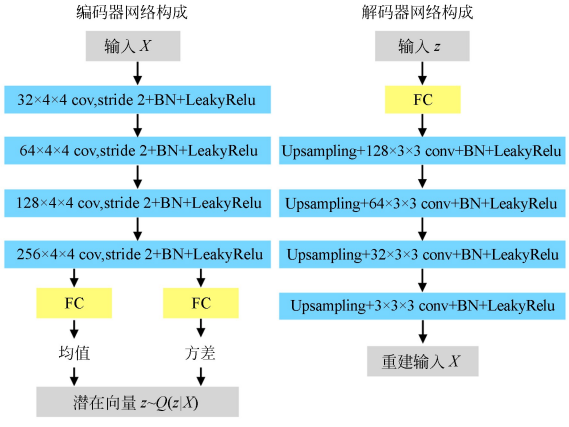

图 2　基于卷积神经网络的 VAE 模型的结构框图

Fig. 2　Framework of VAE model based on convolutional neural network

(1)编码器由 4 个卷积层组成，卷积核为 4×4，通过将步长设置为 2 实现下采样。在每个卷积层后都添加批量归一化(Batch Normalization，BN)来优化网络结构，并使用带泄露修正线性单元(Leaky ReLU)函数作为激活函数。

(2)在编码器中加入两个全连接的输出层，分别用于计算均值和方差，均值和方差将用于计算潜在特征表示和 KL 散度。

(3)解码器的卷积核设置为 3×3，步长设置为 1，通过最近邻法实现上采样。

在基于卷积神经网络的 VAE 中，编码器和解码器的结构大致对称：编码器实现学习到能够表示输入样本的潜在特征表示；解码器由潜在特征表示逐步上采样，实现从低分辨率重构样本中

重建出高分辨率的重构样本。

在模型的训练阶段，CDRA 算法模型首先通过 DSR 损失函数训练 VAE 学习模态内具有判别能力的信息。在变分自动编码器学会对特定模态进行编码之后，通过 CMA 损失函数和 GDA 损失函数约束模型将不同模态的特征映射到公共的潜在空间，实现精确的特征对齐。

CDRA 算法模型采用 warm up 策略预热损失函数的权重并使用贝叶斯优化(Bayesian Optimization)确定权重值，初始值设置均为 0，然后以不同的步长增长，如图 3 所示：δ 从第 6 个 epoch 开始到第 44 个 epoch 为止，以 0.27 为步长递增；γ 从第 21 个 epoch 开始到第 150 个 epoch 为止，以 0.022 为步长递增；对于 KL 散度损失的 β 系数，从第 0 个 epoch 开始到第 180 个 epoch 为止，以 0.001 3 为步长递增。为进一步增强潜在特征表示的判别能力，学习得到的潜在特征表示还将输入到 softmax 层。softmax 损失函数从第 50 个 epoch 开始起作用。

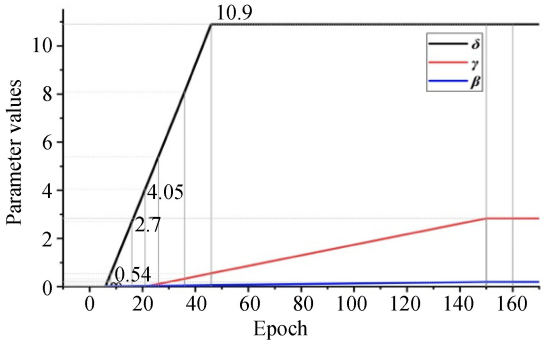

图 3　CDRA 算法损失函数中基于 warm up 更新的参数权重值

Fig. 3　Weight parameters updated by warm up strategy in CDRA method′s loss functions

在测试阶段，CDRA 算法模型通过可视化人脸生成的效果和人脸识别的准确率对学习得到的对齐潜在特征表示的效果进行验证：

(1)在人脸生成的实验中，A 模态的人脸输入模态 A 的编码器得到模态 A 到人脸特征表示，将该特征输入到模态 A 的解码器中，则能够重建出模态 A 的人脸，而输入到模态 B 的解码器，将重建出模态 B 的人脸。

(2)在人脸识别的实验中，模态 A 的人脸图像和模态 B 的人脸图像分别输入到模态 A 的编码器和模态 B 的编码器中，模态 A 的编码器和模态 B 的编码器将两种模态映射到公共的潜在特征表示空间，二者对齐的潜在特征表示将作为最终输出的人脸特征，直接用于人脸识别中。

5　实　验

本文提出 CDRA 算法在经典的姿态人脸数据集 Multi-Pie [22] 近红外光和可见光人脸数据库 CASIA NIR-VIS 2.0[23] 上进行实验，并对实验结果进行分析和总结。在 Multi-Pie [22] 和 CASIA NIR-VIS 2.0[23] 数据集中均包含两种人脸模态，因此，$n=2$。

5.1　实验数据集

Multi-Pie：Multi-Pie[22] 数据集用于姿态人脸对正脸的识别。数据集中前 200 人的图像(共计 161 460 张)作为训练集，剩余 137 人的图像作为测试集，包括 probe 集 (共 72 000 张)和 gallery 集(共 137 张)。其中正脸作为一种模态，包含姿态变化的人脸作为另一种模态。

CASIA NIR-VIS 2.0：CASIA NIR-VIS 2.0[23] 数据集是目前最大和最具挑战性的可见光(VIS)和近红外光(NIR)异构人脸识别数据库。它包括 725 人，每个人有 1～22 张可见光和 5～50 张近红外光图像，分为 10 个子集。训练集含有来自 360 人的大约 2 500 张可见光和 6 100 张近红外图像。在测试集中，gallery 集中包含 358 人的可见光图像，每个人只有一张图像，probe 集包含着 358 人的 6 000 多张近红外图像。

5.2　潜在特征表示维度的影响

潜在特征表示维度是 CDRA 算法模型中，唯一需要进行手动选择的参数。因此，本节通过实验分析模型中潜在特征表示的维度对模型性能的影响，从而确定模型中公共潜在空间的最佳特征维度。

实验结果如图 4 所示，随着特征维度的增加，人脸识别的准确率总体呈现先上升后下降的趋势。在维度为 128 时，人脸识别准确率在两个数据库上均到达顶峰。原因主要有以下两点：(1)潜在特征表示的维度越大，模型的复杂程度和灵活度也越高，就能够学习到性能更好的特征表示；(2)潜在特征表示是对输入人脸数据的压缩表示，

能够学习到人脸数据中最重要的特征表示。但是，如果维度太大，潜在特征空间会学习到人脸数据中不太重要的信息，反而会降低模型的特征表示能力。

潜在特征表示维度的选取需要兼顾模型的复杂度和性能，因此，根据实验结果和分析，在后续实验中，选取的特征维度为 128。

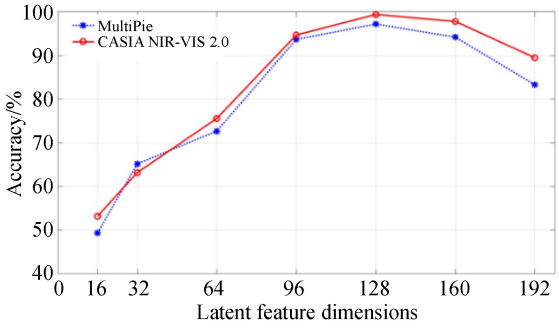

图 4　潜在特征表示维度对人脸识别准确率的影响

Fig. 4　Face recognition accuracy rates with different latent feature dimensions

5.3　DSR，CMA，GDA 损失函数的影响

为了确定 DSR，CMA 和 GDA 损失函数的影响，在不改变网络结构的前提下，本实验将采用不同损失函数的组合对网络模型进行训练和测试。不同损失函数的组合包括 L_{DSR}，$L_{\mathrm{DSR}}+\gamma L_{\mathrm{CMA}}$，$L_{\mathrm{DSR}}+\delta L_{\mathrm{GDA}}$ 和 $L_{\mathrm{DSR}}+\gamma L_{\mathrm{CMA}}+\delta L_{\mathrm{GDA}}$。

如图 5 所示，由于基于 L_{DSR} 训练的模型仅在模态内学习表示单一模态的信息，而不能获取不同模态之间的相关性，因而学习得到的模型的跨模态人脸识别准确率最低。基于 $L_{\mathrm{DSR}}+\gamma L_{\mathrm{CMA}}$ 和 $L_{\mathrm{DSR}}+\delta L_{\mathrm{GDA}}$ 训练的模型通过在图像空间或潜在分布空间对齐潜在特征表示，提高了跨模态人脸识别准确率。而基于 $L_{\mathrm{DSR}}+\gamma L_{\mathrm{CMA}}+\delta L_{\mathrm{GDA}}$ 训练的模型相比于基于 $L_{\mathrm{DSR}}+\gamma L_{\mathrm{CMA}}$ 和 $L_{\mathrm{DSR}}+\delta L_{\mathrm{GDA}}$ 训练的模型在 Multi-Pie 和 CASIA NIR-VIS 2.0 数据集上准确率均有较大幅度提升。这证明在图像空间和潜在分布空间同时对齐分布，能够建立图像空间和潜在分布空间的内在联系，有利于潜在特征表示学习到更具有判别能力的跨模态信息。

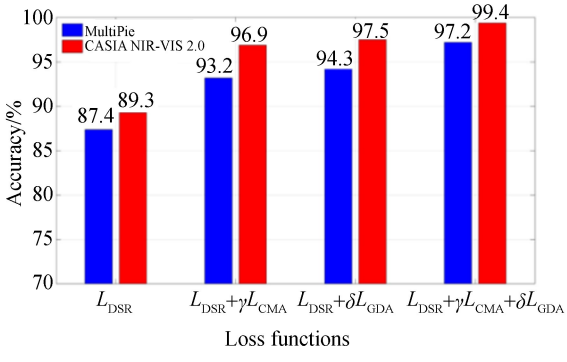

图 5　损失函数对人脸识别准确率的影响

Fig. 5　Face recognition accuracy rates with different loss functions

5.4　Multi-Pie 数据集上的实验结果

CDRA 算法在姿态人脸数据集 Multi-Pie 上不同角度变化的人脸对正脸的识别。在该数据集上，本文使用人脸识别的准确率（识别正确的样本/样本总数）作为评价指标，实验结果如表 1 所示。

表 1　Multi-Pie 数据库上的人脸识别准确率

Tab. 1　Face recognition accuracy rates on Multi-Pie dataset　(%)

算法	人脸角度变化				
	±60°	±45°	±30°	±15°	Avg
FIP+LDA [24]	45.9	64.1	80.7	90.7	70.4
MVP+LDA [25]	60.1	72.9	83.7	92.8	77.4
CPF [26]	61.9	79.9	88.5	95.0	81.3
DR-GAN [17]	83.2	86.2	90.1	94.0	88.4
CAPG-GAN [27]	90.6	97.3	99.5	99.8	96.8
CDRA	91.8	97.5	99.6	99.8	97.2

实验结果表明，随着人脸变化角度的增加，人脸纹理信息丢失的越来越多，因而所有方法的人脸识别准确性都随着角度的增加而下降。FIP+LDA[24] 和 MVP+LDA[25] 算法在提取到对姿态鲁棒的特征后，利用 LDA 进一步提高特征的判别能力。而 CPF[26]，DR-GAN[17] 和 CAPG-GAN[27] 算法通过对姿态进行编码，指导网络合成正脸。不同于上述方法，CDRA 算法通过 DSR 损失函数学习具有判别能力的信息，然后通过 CMA 和 GDA 损失函数在图像空间和潜在分布空间学习跨模态信息，从而减少含有角度变化的人脸与正脸之间的潜在特征表示差异。因此，CDRA 算法不仅能够在潜在特征分布空间学习到对姿态鲁棒的人脸特征，而且能够从公共的潜在特征空间中解码重建出正脸图像。如图 6 所示，是 CDRA 算法人脸合成的效果图。其中，a，c，e 行是含有姿态变化的原始人脸，b，d，f 行

是合成的正脸。经观察可知，CDRA 算法对人脸的一些外观细节实现了较为真实的合成，这表明不同模态的潜在特征表示不仅实现了精准对齐，而且包含具有判别能力的身份信息和结构信息。

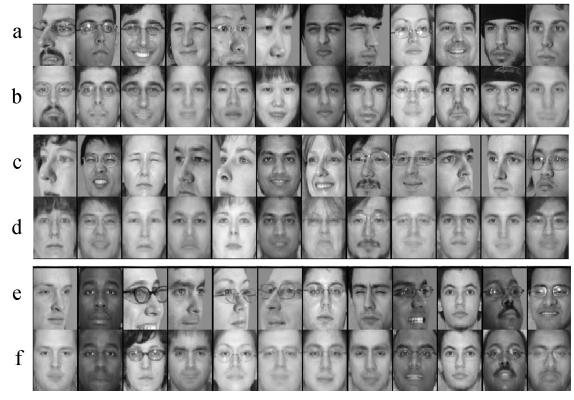

图 6 CDRA 算法的人脸合成效果

Fig. 6 Visualization of face synthesis of CDRA method

5.5 CASIA NIR-VIS 2.0 数据集上的实验结果

CDRA 算法在可见光(VIS)和近红外光(NIR)人脸图像数据集 CASIA NIR-VIS 2.0 上进行 VIS-NIR 人脸识别实验，并与现有的最好的算法进行比较。在该数据集上，本文使用人脸识别的准确率和当假正类率(FAR)=0.1%时的真正类率(TAR)值作为评价指标，实验结果如表 2 所示。实验结果表明，CDRA 算法能够将可见光与近红外光人脸图像映射到公共的潜在特征表示空间，有效地减少可见光与近红外光人脸图像之间的差异，提高了 VIS-NIR 人脸识别的准确率。

基于传统手工设计特征的方法 KDSR[28] 难以克服不同模态人脸间数据分布的差异，学习到具有模态不变性的特征。基于深度学习的方法 Gabor + RBM[29]，IDNet[30]，CDL[5]，ADFL[31]，DVR[3] 和 Peng 等.[33] 得益于深度特征具有更强的表达能力[32]，在 VIS-NIR 人脸识别中表现出较为出色的性能。本文提出的 CDRA 算法不仅在图像空间对可见光和近红外光人脸进行对齐，而且在潜在高斯分布空间对可见光和近红外光人脸的潜在特征分布进行对齐，从而在不同的空间学习到不同模态之间更强的关联信息。

表 2 CASIA NIR-VIS 2.0 数据库上的实验结果

Tab. 2 Experimental results of face recognition on CASIA NIR-VIS 2.0 dataset (%)

算法	Rank-1	TAR(FAR=0.1%)
KDSR[28]	37.5	9.3
Gabor+RBM[29]	86.2 ± 1.0	81.3 ± 1.8
IDNet[30]	87.1 ± 0.9	74.5
CDL[5]	98.6 ± 0.2	98. 3± 0.1
ADFL[31]	98.2 ±0.3	97.2 ±0.3
DVR[3]	99.1 ± 0.2	98.6 ± 0.2
Peng et al.[32]	98.7	96.5
CDRA	99.4 ± 0.2	98.9 ± 0.1

5.6 人脸生成实验结果

CDRA 方法在 CUHK-CUFS[34] 数据集上进行人脸生成的实验。该数据集包含素描人脸和照片人脸，CDRA 模型中的编码器将素描人脸和照片人脸映射到同一潜在特征空间，而解码器将潜在特征解码为照片人脸和素描人脸。因此，CDRA 模型可同时实现由人脸照片和人脸素描画像的互相转换，即由照片生成素描人脸和由素描人脸生成照片。

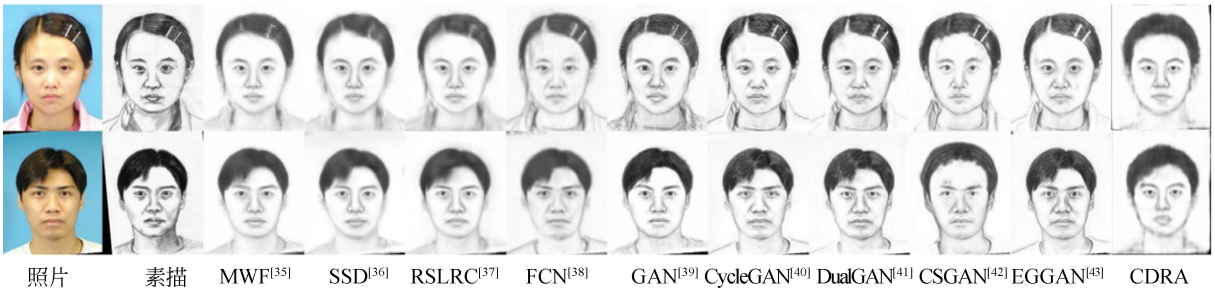

图 7 CUHK-CUFS 数据集中由照片生成素描人脸的效果图

Fig. 7 Visualization of the sketch face synthesis from photos in CUHK-CUFS dataset

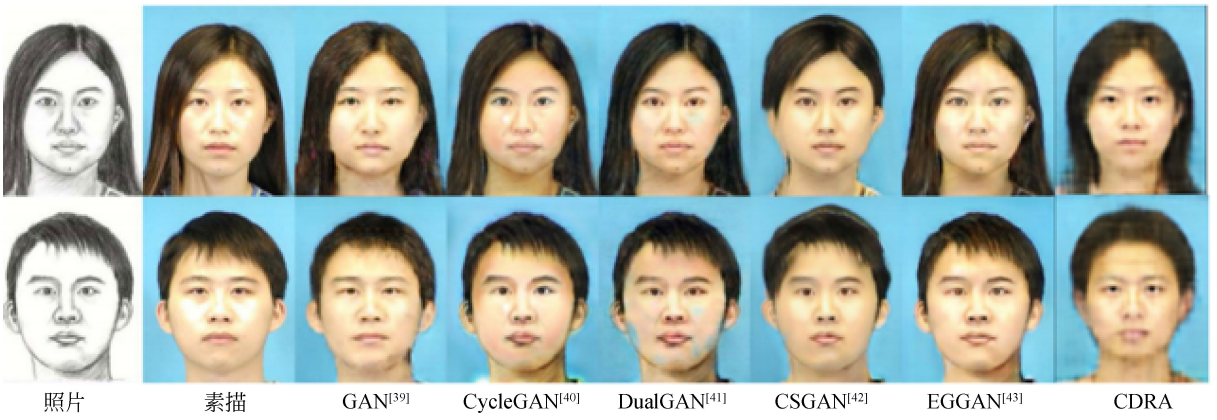

图 8　CUHK-CUFS 数据集中由素描生成照片人脸的效果图

Fig. 8　Visualization of the photo face synthesis from sketches in CUHK-CUFS dataset

本文将 CDRA 方法与非 GAN 类方法和 GAN 类方法的生成人脸进行可视化对比,实现结果如图 7 和图 8 所示。非 GAN 类方法(MWF[35],SSD[36],RSLCR[37],FCN[38])生成的图像通常呈现较为模糊的效果,而 GAN 类方法(GAN[39],CycleGAN[40],DualGAN[41],CSGAN[42],EGGAN[43])生成的图像包含较为丰富的纹理和细节信息。但是非 GAN 类方法生成的图像与原始图像的相似性更高,而 GAN 类方法生成的图像在相似性保持方面表现不足。本文提出的 CDRA 方法更倾向于保持与原始图像的相似性,对于眼睛、鼻子等部分的细节信息生成效果较好,但是头发和衣服部分的生成图像较为粗糙。这是因为变分自动编码器的潜在特征空间是学习人脸的压缩表示,会提取到人脸中结构和五官等重要的信息,忽略不太重要的头发、配饰等信息。

本文使用 SSIM 作为生成图像的质量评测标准。SSIM 用于测量原始人脸与生成人脸之间的结构相似性。

表 3 和表 4 是 CDRA 与现有方法在 CUHK-CUFS 和 CUHK-CUFSF 数据集上的 SSIM 值。本文不仅测试了由人脸照片生成的人脸素描图像的 SSIM 值,而且测试了由人脸素描图像生成的人脸照片的 SSIM 值。SSIM 的取值范围是 0～1,SSIM 值越大表示两张图片越相似。实验结果表明,由人脸照片生成的人脸素描图像的 SSIM 值要整体低于由人脸素描图像生成的人脸照片的 SSIM 值。这是因为在绘制人脸的素描图像时,绘制者的手法不同,但是模型是所有绘制手法的统一表示。因此,由人脸照片生成素描图像比由素描图像生成人脸照片更加困难。

非 GAN 类方法在 SSIM 上的表示要优于 GAN 类方法,原因是 GAN 类方法倾向于合成具有清晰纹理的图像,却容易忽略保持人脸的结构相似性。本文提出 CDRA 方法在非 GAN 类方法和 GAN 类方法中均获得较高的 SSIM 值。CDRA 在图像空间的对齐使得图像获得纹理信息,在特征空间的对齐保证同一个体的人脸保持相似性信息,从而使得生成的人脸图像获得了较好的结构相似性。

表 3　由照片人脸生成素描人脸的 SSIM 值

Tab. 3　SSIM value of sketch face synthesis from photos

算法	SSIM	
	CUHK-CUFS	CUHK-CUFSF
MWF[35]	0.539 3	0.429 9
SSD[36]	0.542 0	0.440 9
RSLCR[37]	0.554 7	0.449 6
FCN[38]	0.521 4	0.362 2
GAN[39]	0.493 8	0.366 5
CycleGAN[40]	0.506 4	0.344 8
DualGAN[41]	0.516 1	0.374 0
CSGAN[42]	0.452 8	0.342 2
EGGAN[43]	0.518 2	0.374 8
CDRA	0.610 6	0.444 3

表 4　由素描人脸生成照片人脸的 SSIM 值

Tab. 4　SSIM value of photo face synthesis from sketches

算法	SSIM	
	CUHK-CUFS	CUHK-CUFSF
GAN[39]	0.616 8	0.603 9
CycleGAN[40]	0.611 1	0.583 5
DualGAN[41]	0.619 2	0.573 2
CSGAN[42]	0.583 1	0.603 8
EGGAN[43]	0.632 5	0.584 2
CDRA	0.592 0	0.577 1

6　结　论

本文提出了一种基于对齐特征表示的跨模态人脸识别算法(CDRA)。该算法基于 VAE 模型，利用 DSR 损失函数，促使 CDRA 算法模型从每一种人脸模态中学习到具有判别能力的身份信息。在此基础上，CMA 和 GDA 的损失函数协同作用，在图像空间和潜在分布空间对不同人脸模态的潜在特征表示进行了有效的对齐，从而在不同的空间维度进一步增强了不同模态间的关联性。CDRA 算法在不同的跨模态人脸识别任务中均表现出色，在 Multi-Pie 数据集上的人脸识别的准确率的平均值为 97.2%，在 CASIA NIR-VIS 2.0 数据集上的人脸识别准确率为 99.4%±0.2%，同时在 CUHK-CUFS 数据集的人脸跨模态生成实验中表现出较好的人脸结构相似性和局部细节描述能力。综上所述，CDRA 算法具有良好的判别能力和泛化能力。

参考文献：

[1] CAO B, WANG N N, GAO X B, *et al*.. Multi-margin based decorrelation learning for heterogeneous face recognition[C]. *International Joint Conference on Artificial Intelligence*, 2019: 680-686.

[2] 张军，何昕，魏仲慧，等. 基于多特征匹配的快速星图识别[J]. 光学 精密工程, 2019, 27(8): 1870-1879.
ZHANG J, HE X, WEI ZH H, *et al*.. Fast star identification algorithm based on multi-feature matching[J]. *Opt. Precision Eng.*, 2019, 27(4): 963-970. (in Chinese)

[3] WU X, HUANG H B, PATEL V M, *et al*.. Disentangled variational representation for heterogeneous face recognition[C]. *Thirty-third AAAI Conference on Artificial Intelligence*, 2019: 9005-9012.

[4] HE R, WU X, SUN Z N, *et al*.. Wasserstein CNN: Learning invariant features for nir-vis face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 41(7): 1761-1773.

[5] WU X, SONG L X, HE R, *et al*.. Coupled deep learning for heterogeneous face recognition[C]. *Thirty-second AAAI Conference on Artificial Intelligence*, 2018.

[6] OLIVEIRA J S, SOUZA G B, ROCHA A R, *et al*.. Cross-domain deep face matching for real banking security systems[C]. *arXiv preprint*: 1806.07644, 2018.

[7] HE R, CAO J, SONG L X, *et al*.. Cross-spectral face completion for nir-vis heterogeneous face recognition[C]. *arXiv preprint*: 1902.03565, 2019.

[8] ZHANG T, WANG H, DONG Q L. Deep disentangling siamese network for frontal face synthesis under neutral illumination[J]. *IEEE Signal Processing Letters*, 2018, 25(9): 1344-1348.

[9] 潘仙张，张石清，郭文平. 多模深度卷积神经网络应用于视频表情识别[J]. 光学 精密工程, 2019, 27(4): 963-970.
PAN X ZH, ZHANG SH Q, GUO W P. Video-based facial expression recognition using multimodal deep convolutional neural networks [J]. *Opt. Precision Eng.*, 2019, 27(4): 963-970. (in Chinese)

[10] GOODFELLOW I, JEAN P A, MIRZA M, *et al*.. Generative adversarial nets[C]. *Conference and Workshop on Neural Information Processing Systems*, 2014: 2672-2680.

[11] KINGMA D P, WELLING M. Auto-encoding variational bayes[C]. *International Conference on Machine Learning*, 2013.

[12] HUANG H B, HE R, SUN Z N, *et al*.. Introvae: Introspective variational autoencoders for photographic image synthesis[C]. *Conference and Workshop on Neural Information Processing Sys-*

tems, 2018: 52-63.

[13] SUN H Z, XU W D, DENG C, *et al*.. Multi-digit image synthesis using recurrent conditional variational autoencoder[C]. *International Joint Conference on Neural Networks*, 2016: 375-380.

[14] WANG W R, ARORA R, LIVERSCU K, *et al*.. On deep multi-view representation learning[C]. *International Conference on Machine Learning*, 2015: 1083-1092.

[15] KAN M N, SHAN S G, ZHANG H H, *et al*.. Multi-view discriminant analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(1): 188-194.

[16] WANG N, GAO X, SUN L, *et al*.. Bayesian face sketch synthesis[J]. *IEEE Transactions on Image Processing*, 2017, 26 (3): 1264-1274.

[17] TRAN L, YIN X, LIU X M. Disentangled representation learning gan for pose-invariant face recognition[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1415-1424.

[18] QIAN Y, DENG W, HU J. Unsupervised face normalization with extreme pose and expression in the wild[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2019: 9851-9858.

[19] HE M, ZHANG J, SHAN S, *et al*.. Deformable face net for pose invariant face recognition[J]. *Pattern Recognition*, 2020, 100: 10711.

[20] 范丽丽,赵宏伟,赵浩宇, 等. 基于卷积神经网络的目标检测研究综述[J]. 光学 精密工程,2020,28(5): 1153-1164.
FAN L L, ZHAO H W, ZHAO H Y, *et al*.. Survey of target detection based on deep convolutional neural networks [J]. *Opt. Precision Eng.*, 2020,28(5): 1153-1164. (in Chinese)

[21] HOU X X, SHEN L L, SUN K, *et al*.. Deep feature consistent variational autoencoder[C]. *Winter Conference on Applications of Computer Vision*, 2017: 1133-1141.

[22] GROSS R, MATTHEWS I, COHN J, *et al*.. Multi-pie[J]. *Image and Vision Computing*, 2010, 28(5): 807-813.

[23] LI S, YI D, LEI Z, *et al*.. The casia nir-vis 2.0 face database[C]. *IEEE Conference on Computer Vision and Pattern Recognition workshops*, 2013: 348-353.

[24] ZHU Z Y, LUO P, WANG X G, *et al*.. Deep learning identity-preserving face space[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 113-120.

[25] ZHU Z Y, LUO P, WANG X G, *et al*.. Multi-view perceptron: a deep model for learning face identity and view representations[C]. *Conference and Workshop on Neural Information Processing Systems*, 2014: 217-225.

[26] YIM J, JUNG H, YOO B, *et al*.. Rotating your face using multi-task deep neural network[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 676-684.

[27] HU Y B, WU X, YU B, *et al*.. Pose-guided photorealistic face rotation[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 8398-8406.

[28] HUANG X S, LEI Z, FAN M Y, *et al*.. Regularized discriminative spectral regression method for heterogeneous face matching[J]. *IEEE Transactions on Image Processing*, 2012, 22(1): 353-362.

[29] YI D, LEI Z, LI S Z. Shared representation learning for heterogenous face recognition[C]. *IEEE international conference and workshops on automatic face and gesture recognition*, 2015: 1-7.

[30] REALE C, NASRABADI N M, KWON H, *et al*.. Seeing the forest from the trees: A holistic approach to near-infrared heterogeneous face recognition[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 54-62.

[31] SONG L, ZHANG M, WU X, *et al*.. Adversarial discriminative heterogeneous face recognition[C]. *International Joint Conference on Artificial Intelligence*, 2018.

[32] 任克强,胡慧. 角度空间三元组损失微调的人脸识别[J]. 液晶与显示,2019,34(1): 110-117.
REN K Q, HU H. Face recognition of triple loss fine-tuning in angular space[J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(1): 110-117. (in Chinese)

[33] PENG C, WANG N, LI J, *et al*.. Re-ranking high-dimensional deep local representation for NIR-VIS face recognition[J]. *IEEE Transactions on Image Processing*, 2019, 28(9): 4553-4565.

[34] MESSER K, KITTLER J, SADEGHI M, *et al*.. Face verification competition on the XM2VTS database[C]. *International Conference on Audio- and Video-Based Biometric Person Authentica-*

tion, Springer, 2003: 964-974.

[35] ZHANG M, WANG N, GAO X, *et al*.. Markov Random Neural Fields for Face Sketch Synthesis [C]. *International Joint Conference on Artificial Intelligence*, 2018: 1142-1148.

[36] SONG Y, BAO L, YANG Q, *et al*.. Real-time exemplar-based face sketch synthesis[C]. *European Conference on Computer Vision*, *Springer*, 2014: 800-813.

[37] WANG N, GAO X, LI J. Random sampling for fast face sketch synthesis[J]. *Pattern Recognition*, 2018, 76(1): 215-227.

[38] ZHANG L, LIN L, WU X, *et al*.. End-to-end photo-sketch generation via fully convolutional representation learning[C]. *The 5th ACM on International Conference on Multimedia Retrieval*, *ACM Press*, 2015: 627-634.

[39] GOODFELLOW I, POUGET-ADADIE J, MIRZA M, *et al*.. Generative adversarial nets[C]. *Advances in neural information processing systems*, *MIT Press*, 2014: 2672-2680.

[40] ZHU J Y, PARK T, ISOLA P, *et al*.. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, *Piscataway*: *IEEE*, 2017: 2223-2232.

[41] YI Z, ZHANG H, TAN P, *et al*.. Dualgan: Unsupervised dual learning for image-to-image translation [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, *IEEE*, 2017: 2849-2857.

[42] KANCHARAGUNTA K B, DUBEY S R. Csgan: cyclic-synthesized generative adversarial networks for image-to-image transformation[J]. *arXiv preprint arXiv*:1901.03554, 2019.

[43] ZHENG J, SONG W, WU Y, *et al*.. Feature encoder guided generative adversarial network for face photo-sketch synthesis[J]. *IEEE Access*, 2019, 7(1): 154971-154985.

作者简介：

明　悦(1984－)，女，北京，副教授，博士生导师，2006 年于北京交通大学获得学士学位，2008 年于北京交通大学获得硕士学位，2013 年于北京交通大学获得博士学位，主要从事模式识别与机器学习方面的研究。E-mail: myname35875235@126.com

王绍颖(1994－)，女，山东，硕士研究生，2017 年于中国传媒大学获得学士学位，2020 年于北京邮电大学获得硕士学位，主要从事人脸识别方面的研究。E-mail: 13693378978@163.com