

基于图像深度先验和鲁棒马尔可夫随机场的有遮挡人脸识别

李小薪, 丁伟杰, 方怡, 张远成, 王琦晖

引用本文

李小薪, 丁伟杰, 方怡, 张远成, 王琦晖. [基于图像深度先验和鲁棒马尔可夫随机场的有遮挡人脸识别](#)[J]. 计算机科学, 2024, 51(7): 244-256.

LI Xiaoxin, DING Weijie, FANG Yi, ZHANG Yuancheng, WANG Qihui. [Occluded Face Recognition Based on Deep Image Prior and Robust Markov Random Field](#) [J]. Computer Science, 2024, 51(7): 244-256.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[局部球面规范化嵌入:PCANet的一种改进方案](#)

Local Sphere Normalization Embedding:An Improved Scheme for PCANet

计算机科学, 2018, 45(5): 238-242. <https://doi.org/10.11896/j.issn.1002-137X.2018.05.041>

[基于奇异值分解的Gabor遮挡字典学习](#)

Gabor Occlusion Dictionary Learning via Singular Value Decomposition

计算机科学, 2018, 45(6): 275-283. <https://doi.org/10.11896/j.issn.1002-137X.2018.06.049>

[基于噪声空间结构嵌入和高维梯度方向嵌入的鲁棒人脸识别方法](#)

Robust Face Recognition via Noise Spatial Structure Embedding and High Dimensional Gradient Orientation Embedding

计算机科学, 2018, 45(4): 285-290. <https://doi.org/10.11896/j.issn.1002-137X.2018.04.048>

基于图像深度先验和鲁棒马尔可夫随机场的有遮挡人脸识别

李小薪¹ 丁伟杰^{1,2} 方怡¹ 张远成¹ 王琦晖³

1 浙江工业大学计算机科学与技术学院 杭州 310023

2 浙江警察学院计算机与信息安全系 杭州 310053

3 杭州师范大学钱江学院 杭州 311121

(mordecai@163.com)

摘要 由遮挡所引发的测试数据和训练数据之间的差异,是人脸识别技术面临的重要挑战。现有的基于深度神经网络的有遮挡人脸识别方法大多需要使用大规模的有遮挡的人脸图像来训练网络模型。然而,现实世界中的任何外界物体都有可能成为遮挡,有限的训练集数据很难穷尽所有的可能性,并且使用大规模的有遮挡人脸图像训练网络模型的做法与人类视觉机制是相违背的,人眼对于遮挡区域的感知在本质上与遮挡本身并没有关系,仅依赖于无遮挡的人脸图像。为了模拟人类视觉的遮挡检测机制,将图像深度先验和鲁棒马尔可夫随机场模型结合起来,构建基于小样本数据的遮挡检测模型 DIP-rMRF,并提出了一致性零填充方法以有效利用 DIP-rMRF 的遮挡检测结果进行后续的人脸识别。在 Extended Yale B, AR 和 LFW 这 3 个人脸数据库上,针对 VGGFace, LCNN, PCANet, SphereFace, InterpretFR, FROM 这 6 种 CNN 模型的实验结果表明, DIP-rMRF 能够有效地处理遮挡以及由极端光照所引发的“类遮挡”,从而极大提升现有的深度神经网络模型对有遮挡人脸识别的性能。

关键词: 有遮挡人脸识别; 图像深度先验; 鲁棒马尔可夫随机场; 一致性零填充; 结构误差度量

中图分类号 TP391

Occluded Face Recognition Based on Deep Image Prior and Robust Markov Random Field

LI Xiaoxin¹, DING Weijie^{1,2}, FANG Yi¹, ZHANG Yuancheng¹ and WANG Qihui³

1 College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

2 Department of Computer and Information Security, Zhejiang Police College, Hangzhou 310053, China

3 Qianjiang College of Hangzhou Normal University, Hangzhou 311121, China

Abstract The occlusion-caused difference between test and training images is one of the most challenging issues for real-world face recognition system. Most of the existing occluded face recognition methods based on deep neural networks(DNNs) need to use large-scale occluded face images to train network models. However, any external object in the real world might become occlusions, and limited training data cannot exhaust all possible objects. Also, using large-scale occluded face images to train networks violates the human visual mechanism, the human eyes detect occlusions by only using small-scale unoccluded face images without seeing any occlusions. In order to simulate the occlusion detection mechanism of human vision, we combine the deep image prior with the robust Markov random field model to construct a novel occlusion detection model, namely DIP-rMRF, based on small-scale data, and propose a uniform zero filling method to effectively utilize the occlusion detection results of DIP-rMRF. Experimental results of six advanced DNN-based face recognitions methods, including VGGFace, LCNN, PCANet, SphereFace, InterpretFR and FROM, on three face datasets, including Extended Yale B, AR and LFW, show that DIP-rMRF can effectively preprocess face images with occlusions and quasi-occlusions caused by extreme illuminations, and greatly improve the performance of the existing DNN models for face recognition with occlusion.

Keywords Face recognition with occlusion, Deep image prior, Robust Markov random field, Uniform zero-filling, Structural error metric

到稿日期:2023-04-18 返修日期:2024-01-14

基金项目:浙江省自然科学基金(LGF22F020027);公安部科技强警基础工作专项项目(2020GABJC35);国家自然科学基金(62271448)

This work was supported by the Natural Science Foundation of Zhejiang Province, China(LGF22F020027), Special Project of Ministry of Public Security(2020GABJC35) and National Natural Science Foundation of China(62271448).

通信作者:丁伟杰(dingweijie@zjcx.cn)

1 引言

随着深度学习技术的日益发展,神经网络(DNN)在人脸识别领域中的应用也取得了巨大的成功^[1],甚至超过了人类视觉^[2]。然而,人脸识别是一个复杂的系统工程,除了所采用的具体技术外,人脸图像的采集环境和具体场景等对人脸识别系统也会有较大影响。当输入图像中存在较大的姿势变化^[3]、光照变化^[4-5]、表情变化^[6-7]、遮挡^[8-11]等问题时,前沿的人脸识别系统仍然面临着巨大的挑战。本文重点关注有遮挡人脸识别问题^[8-11]。

现有的基于神经网络的有遮挡人脸识别方法^[12-18]大多需要使用大规模的有“遮挡”的人脸图像来训练网络模型,以净化或过滤遮挡的影响。然而,现实世界中的任何外界物体都有可能成为“遮挡”,有限的训练集数据很难穷尽所有的可能性,当网络模型遇到从未见到过的“遮挡”时,其识别性能往往会受到很大影响。为了构建大规模的有遮挡人脸数据集,研究者们提出了一系列用于生成面部遮挡的方法^[19-22]。然而,将大规模的有遮挡人脸图像作为训练集来训练网络模型的做法与人类视觉机制是相违背的。人眼主要是通过人脸图像中的无遮挡区域进行图像识别^[8,14,23]。人眼之所以能够区分人脸图像中的遮挡和非遮挡区域并非是因为之前见过各式各样的“遮挡物”,而是因为之前见过一定数量的(可能只是极少量的)未被遮挡的人脸,并以此为基础构建了一个一般的人脸结构模型。这种结构一旦被“遮挡物”破坏,就很容易被眼感知到,从而捕捉到那些被破坏的区域。因此,人眼对于遮挡区域的感知在本质上与遮挡本身并没有关系,仅依赖于无遮挡的人脸图像。

那么,神经网络是否也可以仿照人类视觉机制,仅利用小规模人脸图像数据就可以构建鲁棒的人脸模型并进行有效的遮挡检测呢?2018年,Ulyanov等^[24]提出了图像深度先验(Deep Image Prior,DIP)的理念,指出神经网络的优越性能更多地依赖于其深层次结构及迭代优化机制,而非大规模的训练数据集。基于DIP,可以将深度学习模型的构建方式从基于大规模数据集的离线预训练转变为基于“零样本”^[25]或小样本的在线训练。相对于离线预训练,在线训练虽然需要消耗一定的计算资源和时间去拟合具体的输入图像,但具有更大的灵活性,可以更好地处理输入图像中可能存在的变化,甚至是攻击信息。例如,DIP近来被Dai等^[26]用于净化对抗样本^[27-28],并取得了比离线模型^[29]更好的效果。显然,遮挡与对抗信息一样,其存在形式也是多种多样且先验未知的,因此也更适合基于DIP来进行处理。

然而,值得注意的是,DIP主要用于图像重建^[26,30-32],其不能直接用于图像识别、目标检测等其他计算机视觉任务。这是由DIP的性质决定的。DIP基于神经网络在迭代训练过程中普遍遵循的一个规律是:首先学习目标图像的主要结构,再学习其中的次要结构和噪声信息。因此,DIP的使用

通常需要基于恰当的迭代停止条件^[33]。

对于有遮挡图像的识别,排除遮挡的影响是非常重要的任务^[14-16]。本文基于DIP进行遮挡检测。首先,基于DIP净化含遮挡的人脸图像(目标图像),再辅之以鲁棒的遮挡检测模型。基于DIP的图像重构需要重点考虑网格模型的构建和网络的输入图像,而遮挡检测模型通常需要度量重建图像与目标图像的误差图像(Error Image),并使用较为复杂的概率模型^[23,34-37]来描述误差图像和遮挡支撑¹⁾(Occlusion Support)之间的映射关系。对误差图像的度量非常重要,现有方法^[23,34,37]通常使用两幅图像的绝对差值来构建误差图像,其主要缺陷在于只考虑了两幅图像的各像素点之间的差异,不能充分反映结构差异,而这就给重建图像的精度提出了较高的要求。为了凸显由遮挡所引发的结构差异,本文提出了一类新的鲁棒结构误差度量(robust Structural Error Metric,rSEM)算子的构建方法:基于现有的鲁棒距离度量算子^[28,38-40]构建鲁棒的结构误差度量算子。进行遮挡检测,还需要构建误差图像和遮挡支撑之间的映射关系。只是使用简单的聚类方法很容易导致误检(False Positive)和漏检(False Negative),通常需要借助遮挡的先验结构,即空间连续结构,来区分遮挡和非遮挡区域。现有工作^[23,34-36]通常使用马尔可夫随机场(Markov Random Field,MRF)来对遮挡这一先验结构进行建模。我们将基于鲁棒SEM和MRF的遮挡检测模型称为鲁棒的MRF(robust MRF,rMRF),将基于DIP的重构模型和基于rMRF的遮挡检测模型联合起来进行遮挡检测,并称之为DIP-rMRF。

如何使用遮挡检测结果排除遮挡的影响进而进行图像识别也非常重要。在现有的基于DNN的有遮挡人脸识别的研究工作^[13-16]中,尚没有直接去除输入图像中的遮挡区域(假定遮挡区域已经被精确定位)来进行人脸识别的。这主要是因为大部分用于图像识别的DNN模型都不是全卷积网络(Fully Convolutional Network,FCN)^[41],在网络的最后几层一般都使用全连接,而全连接层的存在使得大部分DNN模型对输入图像的大小严格的限制。因此,现有方法^[13-16]一般是在网络输出的特征图中而不是直接在输入图像中排除遮挡的影响,此类方法需要针对具体的网络训练具体的遮挡检测模型,无法实现遮挡检测与识别模型的分离。

为了充分利用遮挡检测结果进行有遮挡人脸识别,本文提出了一种新的一致性全零填充(uniform Zero Filling,uZF)方法,将需要比对的库存人脸图像和待识别人脸图像在遮挡支撑所指向的遮挡位置上的像素值全部置为零,从而使目标图像中的遮挡信息失去判别作用。

综上,本文提出了一种新的遮挡检测模型DIP-rMRF,并提出了一致性全零填充方法uZF以充分利用遮挡检测的结果进行鲁棒人脸识别。本文的主要贡献如下:

(1)将DIP^[24]引入有遮挡人脸图像的重构,使用在线小样本训练方法对有遮挡人脸图像进行重建,可以更好地处理

¹⁾ 遮挡支撑是用于表示遮挡和非遮挡位置的二值矩阵。

现实世界中可能存在的各种遮挡。

(2) 提出了一类新的鲁棒结构误差度量 (rSEM) 方法, 并将其与马尔可夫随机场 (MRF) 模型相结合, 构建鲁棒的 MRF 模型 (rMRF), 进行遮挡检测。

(3) 为了充分利用遮挡检测结果进行鲁棒人脸识别, 提出了一致性全零填充方法。该方法可以极大程度提升现有的 DNN 模型识别有遮挡人脸图像的性能, 也可以用于对遮挡检测结果的准确性进行检验。

2 遮挡检测模型 DIP-rMRF

本文提出的 DIP-rMRF 模型结构如图 1 所示, 主要包括基于 DIP 的有遮挡人脸图像的重构模型 (DIP Reconstructor) 和基于 rMRF 的遮挡检测模型。对于输入的有遮挡人脸图像 $\mathbf{y} \in \mathbb{R}^n$, 由 DIP 重构模型净化 \mathbf{y} 中的遮挡信息, 得到重构图像 $\hat{\mathbf{y}} \in \mathbb{R}^n$, 由鲁棒结构误差度量算子 $\tilde{\mathcal{E}}(\cdot, \cdot)$ 计算 \mathbf{y} 和 $\hat{\mathbf{y}}$ 的结构误差 $\tilde{\mathbf{e}}$, 由两类均值聚类 $\mathcal{K}(\cdot)$ 和 MRF 优化遮挡支撑 $\mathbf{s} \in \{-1, 1\}^n$ ($s_i = -1$ 表示 y_i 为非遮挡点, $s_i = 1$ 表示 y_i 为遮挡点)。DIP-rMRF 模型通过求解结构误差 $\tilde{\mathbf{e}}$ 和遮挡支撑 \mathbf{s} 的联合优化问题来完成遮挡检测任务。本章将分别探讨 DIP-rMRF 模型的各个组成部分 (优化模型、DIP 重构模型、结构误差度量算子 $\tilde{\mathcal{E}}(\cdot, \cdot)$ 和 rMRF 模型) 如何构建。

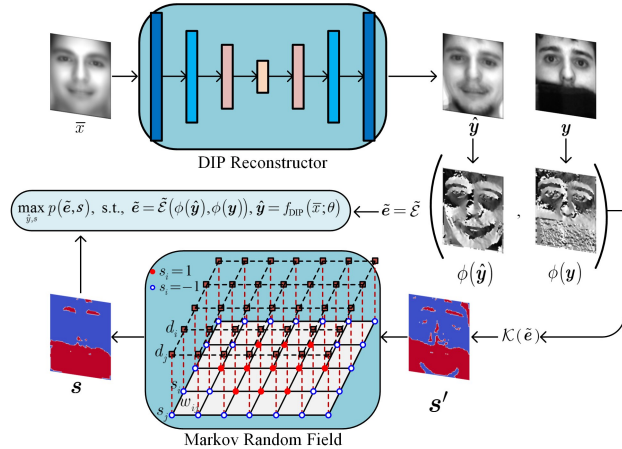


图 1 本文提出的基于 DIP-rMRF 的遮挡检测模型

Fig. 1 The proposed DIP-rMRF occlusion detection model

2.1 优化模型

遮挡检测问题在一定程度上依赖于对有遮挡图像的重建。如果遮挡区域已知, 则有遮挡人脸图像的重建问题可以视作图像修复 (Image Inpainting) 问题。由 Ulyanov 等^[24]的工作可知, 基于 DIP 的图像修复问题可用如下优化式求解:

$$\begin{aligned} \hat{\mathbf{y}}^* &= \arg \min_{\hat{\mathbf{y}}} \frac{1}{2} \|\mathbf{e} \odot (1 - \mathbf{s})\|_2^2 \\ \text{s. t. } \mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}}, \hat{\mathbf{y}} = f_{\text{DIP}}(\mathbf{z}; \theta) \end{aligned} \quad (1)$$

其中, \odot 表示 Hadamard 乘积, $f_{\text{DIP}}(\cdot; \theta)$ 表示基于 DIP 的图像重构模型, \mathbf{z} 为 $f_{\text{DIP}}(\cdot; \theta)$ 的输入图像。在经典的 DIP 模型中, \mathbf{z} 通常是由随机数组成的图像。

由式 (1) 可知, DIP 的图像修复模型旨在通过对未损坏区域的精确重建来推断损坏区域所丢失的信息。然而, 对于遮挡检测任务而言, 对遮挡区域的精确重建不是主要目标, 需要精确重建的是遮挡支撑 \mathbf{s} 。遮挡支撑 \mathbf{s} 的精确重建依赖于对重构误差 \mathbf{e} 的准确度量。显然, 误差越大的位置被遮挡的可能性也越大。在式 (1) 中, $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ 计算的是 \mathbf{y} 与其重构图像 $\hat{\mathbf{y}}$ 在各像素点的差异, 而遮挡支撑 \mathbf{s} 所表达的是遮挡的结构信息。因此, 需要度量 \mathbf{y} 与 $\hat{\mathbf{y}}$ 的结构差异 $\tilde{\mathbf{e}}$ 。另外, 由于基于 DIP 的遮挡检测需要同时考虑重构误差 $\tilde{\mathbf{e}}$ /重构图像 $\hat{\mathbf{y}}$ 与遮挡支撑 \mathbf{s} 的联合优化问题, 因此我们将这一问题转化为最大化 $\tilde{\mathbf{e}}$ 和 \mathbf{s} 的联合概率 $p(\tilde{\mathbf{e}}, \mathbf{s})$ 问题。故基于 DIP 的遮挡检测问题可用如下模型表示:

$$\begin{aligned} (\hat{\mathbf{y}}^*, \mathbf{s}^*) &= \arg \max_{\hat{\mathbf{y}}, \mathbf{s}} p(\tilde{\mathbf{e}}, \mathbf{s}) \\ \text{s. t. } \tilde{\mathbf{e}} &= \tilde{\mathcal{E}}(\phi(\hat{\mathbf{y}}), \phi(\mathbf{y})), \hat{\mathbf{y}} = f_{\text{DIP}}(\mathbf{z}; \theta) \end{aligned} \quad (2)$$

其中, $\tilde{\mathcal{E}}(\cdot, \cdot)$ 表示结构误差度量算子, $\phi(\cdot)$ 表示图像变换域。为了减小光照条件、局部姿势变化等因素的影响, 需要在图像变换域 $\phi(\cdot)$ 中度量结构误差。

2.2 DIP 重构模型

DIP 重构模型需要重点考虑网络模型的构建以及输入图像的构建。本文采用 Ulyanov 等^[24] 给出的“编码器-解码器”网络构架¹⁾来构建 DIP 重构模型, 这一网络结构在图像重建领域中使用广泛, 与之类似的有 U-Net^[42], RED-Net^[43] 等。不同的图像重建任务的编解码器配置和网络的输入图像存在一定的差异。由于有遮挡人脸图像的重建与具有较大空洞的破损图像的修复任务 (Large Hole Inpainting) 相似, 因此本文采用 Ulyanov 等^[24] 提供的用于处理这一类图像修复任务的网络配置。值得注意的是, 这一网络配置中没有使用编码器和解码器之间的跨越连接。Ulyanov 等^[24] 建议基于 DIP 的图像重建网络都要尽可能少用“短连接”, 因为短连接会将低阶语义信息传递到网络高层, 不利于构建目标图像中的结构信息。

事实上, 是否需要使用短连接与网络的输入图像 \mathbf{z} 也有着密切的关系。在 Ulyanov 等^[24] 给出的各种图像重建的范例如中, 输入图像一般都是由随机数构成的 (见图 2(a)), 因此, 低层网络输出的特征图中包含了较多的非结构信息, 不利于对目标图像的重建。也是基于这一考虑, Ulyanov 等^[24] 在对具有较大空洞的破损图像进行修复时使用了网格采样点矩阵 (Meshgrid) 作为输入图像 (见图 2(b)), 以引入平滑先验。

由于输入图像对于 DIP 重构模型至关重要, 因此本文考虑在输入图像中引入更多的先验信息, 以减少迭代次数, 获取更好的重构效果。为此, 我们使用均值脸^[44] 作为输入图像: 从训练集 X 中选取 κ 个不同的人脸图像, 构成子集 \bar{X} , 计算其均值 \bar{x} 。这里, 训练集 X 不是直接用于训练 DIP 重构模型的数据集, 而是人脸识别系统所服务的机构或团体的所有成员的人脸图像数据集, 并且待重建的有遮挡图像 \mathbf{y} 也来自于该机构或团体的某一成员。由于待重建图像也是待识别

¹⁾ <https://github.com/DmitryUlyanov/deep-image-prior>

图像,因此,通过上述方式构建均值脸作为输入图像即可将两个任务连接起来,而不是如图像修复任务般仅通过未破损区域来推断整幅图像中缺失的信息。另一方面,均值脸图像(如图2(c)和图2(d)所示)所包含的是人脸的一种模糊的轮廓结构,可用于模拟人类大脑所构建的人脸模型:不是具体的某个人,但具有人脸的五官结构。

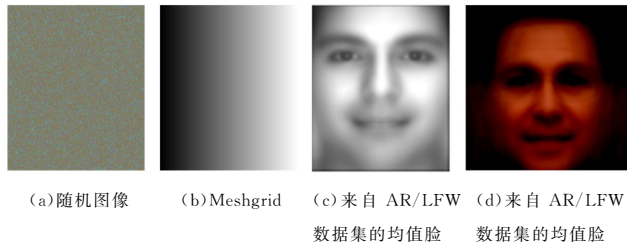


图2 DIP重构模型的4种输入图像

Fig. 2 Four typical inputs of DIP reconstruction model

在DIP模型的迭代优化过程中,子集 \tilde{X} 的构建非常重要,其所包含的图像类别决定了所生成的均值脸图像中包含的结构信息。令 κ 和 K 分别表示 \tilde{X} 和 X 中所包含的图像类别的个数。在迭代的初始阶段,为了避免过拟合,令 $\kappa=K$;在后续迭代过程中,随着重构精度和遮挡检测精度的提升,可以逐步减小 κ 的值。为此,我们使用预训练的人脸识别模型从 X 中选取与目标图像 \mathbf{y} 近邻的人脸图像来构建子集 \tilde{X} 。这里, \mathbf{y} 的近邻图像可以根据人脸识别模型所预测的 \mathbf{y} 在 X 的各个类别上的置信度来选取,选取其中置信度最高的 κ 个人的训练图像来构建 \tilde{X} 。

使用均值脸图像作为DIP重构模型的输入,即可在所使用的网络模型中加入一定数量的短连接。本文在“编码器-解码器”架构中加入从输入端到输出端的残差连接^[45],用于将均值脸所构建的结构信息直接传递给输出端。

2.3 结构误差度量

遮挡检测重点关注目标图像与其重构图像在结构上的差异,而非各像素点之间的差异。为了度量两幅图像在结构上的差异,本文针对遮挡检测问题,提出了一种新的结构误差度量方法:由鲁棒的距离度量算子构建鲁棒的结构误差度量算子。

鲁棒的距离度量在本质上是一种加权误差度量,并且权重通常是由权重函数根据两个向量的各元素之差隐式计算的。具体地,任意两个向量 $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ 之间的鲁棒距离定义如下:

$$\mathcal{D}(\mathbf{a}, \mathbf{b}) \triangleq 1 - \frac{1}{n} \sum_{i=1}^n \rho(e_i) \quad (3)$$

其中, $e_i = a_i - b_i$, $\rho(\cdot)$ 表示权重函数。常用的权重函数有M估计中的Welsch权重函数、Cauchy权重函数等;也可以使用核函数,如高斯核函数^[38]、余弦核函数^[39]等,在高维空间中度量局部相似性。

由于鲁棒距离度量算子 $\mathcal{D}(\cdot, \cdot)$ 考虑的是任意向量之间的距离,而遮挡检测需要考虑的是两幅图像(\mathbf{y} 与其重构图像 $\hat{\mathbf{y}}$)之间的结构差异。因此,可以改造 $\mathcal{D}(\cdot, \cdot)$,使其作用范围收缩到单个像素点的邻域上,使用 $\mathcal{D}(\cdot, \cdot)$ 在该点邻域上的计算结果作为该点的重构误差。具体地, \mathbf{y} 及其重构

图像 $\hat{\mathbf{y}}$ 中的任一像素点 i 的结构差异 \tilde{e} ,可通过如下结构误差度量算子进行计算:

$$\mathcal{E}(\mathbf{y}_i, \hat{\mathbf{y}}_i) \triangleq 1 - \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \rho(\mathbf{y}_j - \hat{\mathbf{y}}_j) \quad (4)$$

其中, $\mathcal{N}(i)$ 表示像素点 i 的邻域像素点的索引集。

直接由 $\mathcal{E}(\cdot, \cdot)$ 计算 \tilde{e} 需要两重循环,计算代价较高。然而,容易看出,式(4)实际上相当于对 $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ 的变换域 $\hat{\mathbf{e}} = 1 - \rho(\mathbf{e})$ 施加均值滤波。这里, $\rho(\mathbf{e})$ 表示将 $\rho(\cdot)$ 作用于 \mathbf{e} 的每一个元素得到的与 \mathbf{e} 维度相同的矩阵。因此,结构差异 \tilde{e} 可以通过如下误差度量算子快速计算:

$$\mathcal{E}_{\rho}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq (1 - \rho(\mathbf{y} - \hat{\mathbf{y}})) * \mathbf{1}_{|\mathcal{N}| \times |\mathcal{N}|} \quad (5)$$

其中,*表示卷积运算, $\mathbf{1}_{|\mathcal{N}| \times |\mathcal{N}|}$ 表示 $|\mathcal{N}| \times |\mathcal{N}|$ 的全1矩阵。

为了获取具有特殊结构的结构误差,可以引入新的权重计算函数 $\rho'(\cdot)$,并定义新的邻域范围 \mathcal{N}' ,构建新的具有递归结构的结构误差度量算子。

$$\mathcal{E}_{\rho \rightarrow \rho'}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq (1 - \rho'(\mathcal{E}_{\rho}(\mathbf{y}, \hat{\mathbf{y}}))) * \mathbf{1}_{|\mathcal{N}'|} \quad (6)$$

也可以在变换域 $\phi(\cdot)$ 中计算两幅输入图像的结构差异。

$$\mathcal{E}_{\phi \rightarrow \rho}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq (1 - \rho(\phi(\mathbf{y}) - \phi(\hat{\mathbf{y}}))) * \mathbf{1}_{|\mathcal{N}|} \quad (7)$$

或者,进一步联立公式 $\mathcal{E}_{\rho'}(\cdot)$ 和 $\mathcal{E}_{\phi}(\cdot)$ 构建变换域中的具有递归结构的结构误差度量算子。

$$\mathcal{E}_{\phi \rightarrow \rho'}(\mathbf{y}, \hat{\mathbf{y}}) \triangleq (1 - \rho'(\mathcal{E}_{\phi \rightarrow \rho}(\mathbf{y}, \hat{\mathbf{y}}))) * \mathbf{1}_{|\mathcal{N}'|} \quad (8)$$

为了凸显遮挡结构,需要对 $\phi(\cdot)$, $\rho(\cdot)$ 和 $\rho'(\cdot)$ 进行特别的设置。由Tzimiropoulos等的工作^[39]可知,在图像梯度方向域中,使用余弦核函数进行距离度量,可以有效消除局部遮挡的影响。利用这一特性,本文将 $\phi(\cdot)$ 定义为图像梯度方向变换,将 $\rho(\cdot)$ 定义为余弦核函数 $\cos(\cdot)$ 。进一步,由He等的工作^[38]可知,使用高斯核函数有利于在高维空间中进行局部相似性度量,因此,将 $\rho'(\cdot)$ 定义为高斯核函数 $g(\cdot)$ 。根据式(8),定义具体的结构化误差度量算子为 $\mathcal{E}_{\phi \rightarrow \cos \rightarrow g}$,并记度量结果为 $\tilde{e}_{\phi \rightarrow \cos \rightarrow g}$ 。图3给出了 $\tilde{e}_{\phi \rightarrow \cos \rightarrow g}$ 的生成过程,从中可以看出算子 $\phi(\cdot)$, $\cos(\cdot)$ 和 $g(\cdot)$ 在生成结构误差 $\tilde{e}_{\phi \rightarrow \cos \rightarrow g}$ 的过程中所起的作用。

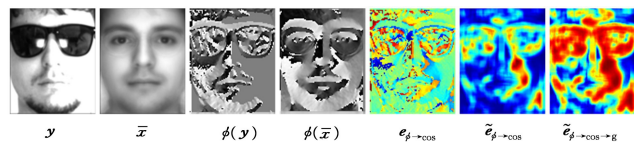


图3 结构误差 $\tilde{e}_{\phi \rightarrow \cos \rightarrow g}$ 的生成过程示例

Fig. 3 Illustration of the generation procedure for structured error

$$\tilde{e}_{\phi \rightarrow \cos \rightarrow g}$$

2.4 误差图像与遮挡支撑的联合概率生成模型

进一步考虑结构误差 \tilde{e} 和遮挡支撑 s 的联合概率模型 $p(\tilde{e}, s)$ 。由于 $p(\tilde{e}, s) = p(\tilde{e} | s)p(s)$,因此下面分别讨论遮挡支撑 s 的先验分布 $p(s)$ 和结构误差 \tilde{e} 的条件概率分布 $p(\tilde{e} | s)$ 。

遮挡支撑 s 最重要的先验是具有空间连续性,即:如果当前像素点被遮挡,则其邻域像素点被遮挡的可能性也很大。现有方法^[23,46-47]主要是通过马尔可夫随机场(MRF)对 $p(s)$ 建模。

$$p_{\text{Ising}}(\mathbf{s}; \Theta) \propto \exp(\mathbf{d}^T \mathbf{s} + \zeta(\mathbf{s}; \mathbf{w})) \quad (9)$$

其中, $\zeta(\mathbf{s}; \mathbf{w}) = \sum_{i \in \mathcal{P}, j \in \mathcal{N}(i)} w_{i,j} s_i s_j$, $\Theta = \{\mathbf{d}, \mathbf{w}\}$, \mathbf{d} 和 \mathbf{w} 分别为数据权重和对遮挡的空间连续性进行建模的平滑权重。

对 $p(\tilde{\mathbf{e}} | \mathbf{s})$ 的建模需要分别考虑遮挡区域 ($s_i = 1$) 和非遮挡区域 ($s_i = -1$) 的 $\tilde{\mathbf{e}}$ 的分布。令 $\tilde{\mathcal{E}} = \mathcal{E}_{\tilde{\mathbf{e}} \rightarrow \cos \rightarrow g}$, $\tilde{\mathbf{e}} = \tilde{\mathcal{E}}_{\tilde{\mathbf{e}} \rightarrow \cos \rightarrow g}$, 从图 3 所展示的结构误差图像 $\tilde{\mathbf{e}}$ 来看, 结构误差度量算子 $\tilde{\mathcal{E}}$ 可以较好地地区分遮挡和非遮挡区域。在遮挡区域(太阳镜遮挡区域和强光光照区域), $\tilde{\mathbf{e}}$ 的数值分布较为均匀, 可以认为 $\tilde{\mathbf{e}}$ 在这一区域服从 $(\tau, 1]$ 上的均匀分布, 即 $\tilde{e}_i \sim \frac{1}{u}$ 。这里, $u = 1 - \tau$, $\tau > 0$ 为预设的某一阈值。在非遮挡区域, $\tilde{\mathbf{e}}$ 的数值大部分趋于 0。由于 $\tilde{\mathbf{e}}$ 是非负的, 可以认为 $\tilde{\mathbf{e}}$ 在这一区域服从指数分布, 即 $\tilde{e}_i \sim \gamma \exp(-\gamma \tilde{e}_i)$, 其中 $\gamma > 0$ 为指数分布的超参数。因此, 可以得到结构误差 $\tilde{\mathbf{e}}$ 的条件概率密度:

$$\begin{aligned} p(\tilde{\mathbf{e}} | \mathbf{s}) &= \prod_{i \in \mathcal{P}} p(\tilde{e}_i | s_i = 1) p(\tilde{e}_i | s_i = -1) \\ &= \prod_{i \in \mathcal{P}} u^{-\frac{1+s_i}{2}} (\gamma \exp(-\gamma \tilde{e}_i))^{-\frac{1-s_i}{2}} \\ &\propto \exp(-\gamma \mathbf{e}^T (1 - \mathbf{s}) - \nu \mathbf{1}^T \mathbf{s}) \end{aligned} \quad (10)$$

其中, $\nu = \log(\gamma u)$ 。

结合式(9)和式(10), 可得到联合概率生成模型:

$$p(\tilde{\mathbf{e}}, \mathbf{s}; \Theta) \propto \exp(-\tilde{\gamma} \mathbf{e}^T (1 - \mathbf{s}) + \tilde{\mathbf{d}}^T \mathbf{s} + \zeta(\mathbf{s}; \mathbf{w})) \quad (11)$$

其中, $\tilde{\mathbf{d}} = \mathbf{d} - \nu$, $\Theta = \{\mathbf{d}, \mathbf{w}\}$ 。

2.5 模型优化

将式(11)代入目标式(2), 求解 $\tilde{\mathbf{e}}$ 和 \mathbf{s} 。由于式(11)还包含了 \mathbf{d} 和 \mathbf{w} 两个参数, 同时求解 4 个未知变量是非常困难的, 可以采用交替迭代的方法求得局部最优解。

$$\tilde{\mathbf{e}}^{(t)} = \arg \min_{\tilde{\mathbf{e}}} \gamma \mathbf{e}^T (1 - \mathbf{s}^{(t-1)}) \quad (12)$$

$$\text{s. t. } \tilde{\mathbf{e}} = \tilde{\mathcal{E}}(\phi(\mathbf{y}) - \phi(\hat{\mathbf{y}}))$$

$$\mathbf{s}^{(t)} = \arg \max_{\mathbf{s}} (\gamma \tilde{\mathbf{e}}^{(t)} + \tilde{\mathbf{d}}^{(t-1)})^T \mathbf{s} + \zeta(\mathbf{s}; \mathbf{w}^{(t-1)}) \quad (13)$$

$$(\mathbf{d}^{(t)}, \mathbf{w}^{(t)}) = \arg \max_{\mathbf{d}, \mathbf{w}} (\gamma \tilde{\mathbf{e}}^{(t)} + \tilde{\mathbf{d}})^T \mathbf{s}^{(t)} + \zeta(\mathbf{s}; \mathbf{w}) \quad (14)$$

其中, $\tilde{\mathcal{E}} = \mathcal{E}_{\tilde{\mathbf{e}} \rightarrow \cos \rightarrow g}$, 上标 (t) 用于标识迭代次数。

式(12)中的结构误差度量算子 $\tilde{\mathcal{E}} = \mathcal{E}_{\tilde{\mathbf{e}} \rightarrow \cos \rightarrow g}$ 用到的 3 个核心函数 $\phi(\cdot)$, $\cos(\cdot)$ 和 $g(\cdot)$ 都是可微的, 可以用梯度下降法求解 $\tilde{\mathbf{e}}^{(t)}$ 。式(13)和(14)用于求解遮挡支撑 \mathbf{s} 及相关变量。式(13)是经典的伊辛模型, 可以通过图割(Graph Cut)^[48] 方法进行求解; 式(14)是常规的线性模型, 可以用梯度下降方法求解。值得注意的是, 式(13)的数据权重 $\tilde{\mathbf{d}}$ 的计算依赖于阈值 τ 。为了避免对 τ 的取值进行手工设置, 可以通过对 $\tilde{\mathbf{e}}$ 进行聚类来计算 τ 的值。具体地, $\tau^{(t)} = \min(\{\tilde{e}_i^{(t)} | s_i^{(t)} = 1\})$, 其中, $\mathbf{s}^{(t)} = \mathcal{H}(\tilde{\mathbf{e}}^{(t)})$, $\mathcal{H}(\cdot)$ 表示两类均值聚类。由于在计算 τ 值时, 已经考虑了结构误差 $\tilde{\mathbf{e}}$ 对 \mathbf{s} 的影响, 因此, 实际计算时, 可以将式(13)和(14)简化为:

$$\mathbf{s}^{(t)} = \arg \max_{\mathbf{s}} (\tilde{\mathbf{d}}^{(t-1)})^T \mathbf{s} + \zeta(\mathbf{s}; \mathbf{w}^{(t-1)}) \quad (15)$$

$$(\mathbf{d}^{(t)}, \mathbf{w}^{(t)}) = \arg \max_{\mathbf{d}, \mathbf{w}} \tilde{\mathbf{d}}^T \mathbf{s}^{(t)} + \zeta(\mathbf{s}; \mathbf{w}) \quad (16)$$

另外, 求解式(12)、式(15)、式(16)还须对 \mathbf{s} , \mathbf{d} 和 \mathbf{w} 进行初始化。本文设置 $\mathbf{d}^{(0)} = \mathbf{1}$, $\mathbf{w}_{i,j}^{(0)} = 0.5$, $\mathbf{s}^{(0)} = \mathcal{H}(\tilde{\mathcal{E}}(\phi(\mathbf{y}) - \phi(\bar{\mathbf{y}}^{(0)})))$ 。这里, $\bar{\mathbf{y}}^{(0)}$ 为初始均值脸, 详见 2.2 节。

3 基于一致性零填充的图像识别

使用遮挡检测结果进行图像识别是遮挡检测的最终目标, 也可以用来检验遮挡检测结果的准确性。在传统的机器学习方法^[8, 23, 34]中, 可以在输入图像中直接去除遮挡区域, 只使用无遮挡区域进行图像识别。然而, 对于神经网络(DNN)而言, 这种方法一般是不可行的, 因为全连接层的存在使得大部分 DNN 模型对输入图像的大小有严格的限制。因此, 现有的方法, 或者是直接对输入图像进行重建^[12], 或者需要在特征域中排除遮挡的影响^[13-16]。前者对重建图像的精度有着极高的要求, 不仅要能够重构出遮挡区域丢失的信息, 还需要保证对遮挡区域的重建不会引入其他类别的信息^[12]; 而后者需要构建出较为复杂的输入图像中的遮挡块与网络所生成的特征图中的特征块之间的映射关系。这两种方法都需要构建大规模的有遮挡人脸数据集。

为了充分利用遮挡检测结果使用预训练的 DNN 模型 $f_{\text{DNN}}(\cdot)$ 进行有遮挡人脸识别, 本文提出了一种新的一致性全零填充(uniform Zero Filling, uZF)方法。根据遮挡检测的结果 \mathbf{s} , 将待比对的库存人脸图像数据集 X 和待识别的有遮挡人脸图像 \mathbf{y} 中的遮挡支撑 \mathbf{s} 所指向的 $s_i = 1$ 的像素值全部置为 0, 再分别使用 $f_{\text{DNN}}(\cdot)$ 提取特征。具体地, 令 $f_{\text{DNN}}(\cdot)$ 的输出为其最后一个隐藏层的特征, 对所有图像 $\mathbf{z} \in X \cup \{\mathbf{y}\}$ 进行如下操作:

$$\mathbf{z}' = f_{\text{DNN}}\left(\mathbf{z} \odot \frac{1 - \mathbf{s}}{2}\right) \quad (17)$$

获得 X 和 \mathbf{y} 的特征 X' 和 \mathbf{y}' , 再进行特征相似性匹配, 从而获得 \mathbf{y} 的类别信息。

使用零填充是为了保证在网络前向传播过程中, 由遮挡所造成的影响尽可能小地向邻域范围和下一层网络扩散。值得注意的是, 只对待识别图像的遮挡区域进行零填充, 即目标零填充(target Zero Filling, tZF), 相当于在目标图像中引入了新的全零遮挡, 无益于识别性能的提升。图 4 和表 1 给出了两种全零填充方法的对比结果, 可以看出, 相比 tZF, uZF 可以充分利用遮挡检测的结果, 大幅度提升识别性能。这也意味着 uZF 也可以用来检验遮挡检测的精度。



图 4 目标零填充和一致性零填充示例

Fig. 4 Examples of target zero filling and uniform zero filling

表1 FaceNet^[49]使用两种零填充方法在LFW上的识别性能对比Table 1 Recognition performance comparison of FaceNet^[49]

using two zero filling methods on LFW		
遮挡比例	目标零填充/%	一致性零填充/%
0%	99.13	99.13
20%	92.26	95.23
50%(下)	84.43	95.00
50%(上)	45.60	94.13

4 仿真与实验

4.1 实验设置

本节通过仿真和实验验证本文提出的DIP-rMRF模型的有效性。首先,验证DIP-rMRF模型对有遮挡人脸图像的重建和遮挡检测效果(详见4.2节);其次,将DIP-rMRF模型的输出用于已有人脸识别模型,验证其对已有人脸识别模型的识别性能的影响(详见4.3节—4.5节)。DIP-rMRF中的超参数只有一个用于定义结构误差 \tilde{e} 的条件概率密度的 γ ,本节实验将其设置为1。DIP重构模型的优化使用Adam算法,初始学习率设置为0.01,最大迭代次数设置为1000,DIP重构模型使用5个编码层和4个解码层,其他参数的详细配置详见DIP官网所公布的代码。使用NVIDIA RTX 2080Ti GPU(显存大小为11GB)进行加速。

为了验证本文方法的有效性,选取VGGFace^[48],LCNN^[50]和SphereFace^[51-52]这3种未对遮挡图像进行显式处理的人脸识别模型,以及PCANet^[53],InterpretFR^[15]和FROM^[16]这3种遮挡鲁棒的人脸识别模型,比较它们在使用本文提出的DIP-rMRF模型进行遮挡检测并排除遮挡影响前后的性能,并用带有下划线的VGGFace,LCNN,SphereFace,PCANet,InterpretFR,FROM表示使用了DIP-rMRF的DNN模型。VGGFace^[49]基于经典的VGG网络架构,是一种广泛使用的人脸识别模型;LCNN是由Wu等^[50]提供的一种轻量级的网络模型,其网络模型的参数规模大约只有VGGFace的1/18,但其在LFW^[54]和YTF^[55]等数据集上取得了与VGGFace^[49],DeepFace^[56]等相当或者更好的识别性能;SphereFace^[51]采用了更具有判别性的角度损失函数A-Softmax来训练网络模型,近年来流行的人脸识别模型如CosFace^[57],ArcFace^[58],FROM^[16]等基本上都是以A-Softmax为基础构建损失函数的。SphereFace的作者近来对A-Softmax损失进行了进一步改进^[52],使其更容易优化。PCANet^[53],InterpretFR^[15]和FROM^[16]是3种对遮挡鲁棒的人脸识别模型。其中,PCANet^[53]是将卷积神经网络模型与传统的FPH框架^[8]相结合的网络模型,其并未对遮挡进行显式

建模,但取得了很好的遮挡鲁棒性^[8];InterpretFR^[15]和FROM^[16]是近年来提出的利用有遮挡数据集对遮挡进行显式建模的两种网络模型,其中,InterpretFR^[15]使用两种多样性损失提升滤波器在人脸各个区域的激活响应(其中的特征激活多样性损失本质上是一种基于阈值的遮挡检测模型),而FROM^[16]在网络最深的隐藏层构建遮挡检测模型,以去除遮挡在深层特征上所造成的影响。类似的方法还有MaskNet^[13]和PDSN^[14],而FROM^[16]是最为先进的方法。值得注意的是,InterpretFR^[15]和FROM^[16]与本文的遮挡检测方法密切相关,但不同的是,本文方法是直接针对输入图像检测遮挡,而InterpretFR^[15]和FROM^[16]是针对特征层进行遮挡检测。尽管InterpretFR和FROM能够在特征层过滤遮挡的影响,但从本文的实验结果来看,基于本文方法在输入端过滤遮挡的影响对有遮挡人脸图像识别性能的提升仍然具有显著作用。

VGGFace,LCNN,SphereFace,PCANet,InterpretFR和FROM都提供了可以公开获取的网络模型¹⁾。我们将神经网络模型的最后一个隐藏层的输出作为人脸图像的特征,使用协同表示分类器(Collaborative Representation Classifier,CRC)进行分类;根据PCANet作者^[53]的建议,使用最近邻分类器(Nearest Neighbor,NN)对PCANet特征进行分类,并且采用卡方距离度量图像间的距离。

选取Extended Yale B^[59],AR^[60]和LFW^[54]这3个基准人脸数据库进行实验。由于人脸识别是一个复杂的系统工程,人脸图像中可能同时存在姿势^[3]、光照^[4-5]、表情^[6-7]和遮挡^[8-11]等变化因素,面向现实的人脸识别系统需要综合考虑这些变化因素。Extended Yale B^[59]和AR^[60]中的人脸图像在受控环境下采集,充分考虑了光照变化,前者包含了不同程度和不同区域的弱光照变化,后者包含了不同程度和不同区域的强光照变化,并且AR数据集还充分考虑了表情变化和遮挡问题,遮挡主要包括太阳镜、围巾等现实遮挡。目前,在Extended Yale B上施加不同比例(60%~90%)的模拟遮挡以及使用AR数据集进行有光照变化和遮挡的人脸识别实验是众多研究者广泛采用的方法^[14,16,23,36,61]。LFW^[54]中的人脸图像在非受控环境下采集(来自于互联网),包含了不同程度的光照、姿势、遮挡、表情等变化。为了测试本文算法对遮挡的鲁棒性,我们借鉴现有文献的做法^[8,36,61],进一步在LFW^[54]中添加了不同比例(0%~60%)的模拟遮挡。

通常情况下,识别率是衡量分类算法的识别性能的最常用的和最重要的量化指标。在此基础上,存在遮挡人脸识别问题的特殊性,因此本文进一步提出用于衡量识别算法稳定性的3个非量化的指标,即性能收敛性、性能衰减性和性能震荡性,以及用于衡量识别算法鲁棒性的两个量化指标,即性能

1) 各预训练的网络模型可以分别从如下网址获得:

VGGFace:http://www.robots.ox.ac.uk/~vgg/software/vgg_face/

LCNN:https://github.com/AlfredXiangWu/face_verification_experiment

SphereFace:<https://github.com/wyliu/sphereface>

PCANet:https://mx.nthu.edu.tw/~tsunghan/download/PCANet_demo_pyramid.rar

InterpretFR:<http://cvlab.cse.msu.edu/project-interpret-FR>

FROM:<https://github.com/haibo-qiu/FROM>

崩溃点和性能退化点。性能收敛指分类算法的识别性能随着迭代次数的增加而逐渐趋于稳定,与之相对应的概念是性能衰减(随着迭代次数的增加,识别性能下降)或性能震荡(随着迭代次数的增加,识别性能出现上下浮动)。性能崩溃点(Breakdown Point)^[62]指随着识别难度的增加,分类算法的识别率开始低于某一阈值并出现性能急剧衰减的临界点。这一临界点是相对于可量化的识别难度来说的,例如,识别难度可以用遮挡比例来量化,遮挡比例越高识别难度越大,那么,导致识别算法的性能崩溃的某一遮挡比例就可以称为性能崩溃点。与性能崩溃点相对应的概念是性能退化点。一般来说,随着图像维度的提升,识别算法的性能也会增强,但当图像维度高到一定程度时,识别算法的性能会下降,我们将这一现象称为“性能退化”。性能退化最为突出的例子是维度灾难(Curse of Dimensionality)^[63]。

4.2 遮挡检测模型对比

首先验证本文提出的 DIP-rMRF 模型进行图像重构和遮挡检测的有效性。为此,我们构造了 DIP-MRF 模型进行对比。DIP-MRF 与 DIP-rMRF 的主要区别在于,在构建误差图像时,前者没有使用本文提出的结构误差度量,而是使用两幅输入图像的绝对误差。

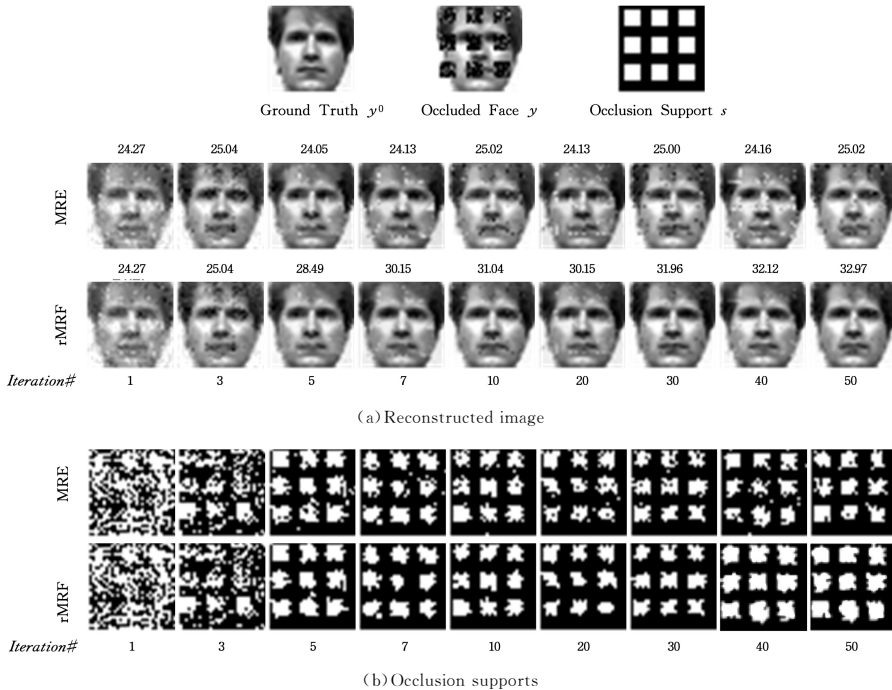


图5 DIP-MRF 和 DIP-rMRF 对重构图像 \hat{y} 和遮挡支撑 s 的估计效果对比

Fig. 5 Comparison of DIP-MRF and DIP-rMRF for estimation of reconstructed image \hat{y} and occlusion support s

4.3 基于 AR 的有强光照和两类实际遮挡的人脸识别

本节基于 AR 人脸数据库^[60]在有实际遮挡(并且混合了光照变化)的情形下测试 DIP-rMRF 对 6 种 CNN 模型的识别性能的影响。AR 数据库包含了 126 个人的 3276 张图像,平均每人 26 张,分别在两个不同的时间段采集,包含了不同的表情、光照和遮挡等变化。我们按照如下方式部署实验环境:从 AR 数据库中选取 119 个人(65 个男性和 54 个女性)的人脸图像作为实验数据;训练集选取 $119 \times 8 = 952$ 张不含遮挡和光照变化的人脸图像,如图 6(a)所示;测试集 I 选取 $119 \times 3 = 357$ 张

图 5 给出了 DIP-MRF 和 DIP-rMRF 在迭代过程中所生成的重构图像 \hat{y} 和遮挡支撑 s 的对比结果。从迭代过程中产生的重构图像来看(见图 5(a)),DIP-rMRF 由于能够很好地排除遮挡的影响,其重构质量相对于 DIP-MRF 有明显的提升,收敛速度也更快。

从各次迭代所估计的遮挡支撑(见图 5(b))来看,DIP-MRF 由于没有充分考虑 \hat{y} 和 y 的结构差异,因此所估计的遮挡块的中间出现了许多孔洞,而这些孔洞只是通过考虑空间连续性(由 MRF 建模)无法进行填充,而得到了许多伪负(False Negative)点;同时,也是因为没有考虑结构差异,DIP-MRF 的检测结果出现了很多零星的伪正(False Positive)点,也就是错误地把一些非遮挡的像素点估计为遮挡像素点。相对于 DIP-MRF,DIP-rMRF 由于使用了结构误差度量,检测结果有了很大的改善,在初始的 30 次迭代的检测结果中漏掉了一些真正(True Positive)的遮挡像素点,但随着迭代次数的增加,几乎所有的真正点都被覆盖了。从 DIP-rMRF 和 DIP-MRF 最终所检测到的遮挡块来看,前者比后者的形状稍大些,也比实际的遮挡支撑 s 中所呈现的各遮挡块更大些,这是因为 DIP-rMRF 所采用的结构误差度量充分考虑了邻域信息(详见式(5))。

有左侧强光、右侧强光和正面强光 3 种光照变化的人脸图像,如图 6(b)所示;测试集 II 选取 357 张有围巾遮挡且混合了正常光照、左侧强光和右侧强光 3 种光照条件的人脸图像,如图 6(c)所示;测试集 III 选取 357 张有太阳镜遮挡且分别包含了正常光照、左侧强光和右侧强光 3 种光照条件的人脸图像,如图 6(d)所示。所有图像首先被裁剪和对齐为 112×92 的像素矩阵。为了系统测试各识别方法的性能退化程度,对所有图像进行了 8 倍、4 倍、2 倍和 1 倍下采样,对应的维数分别为 $14 \times 11 = 154$, $28 \times 23 = 644$, $56 \times 46 = 2576$ 和 $112 \times 92 = 10304$ 。

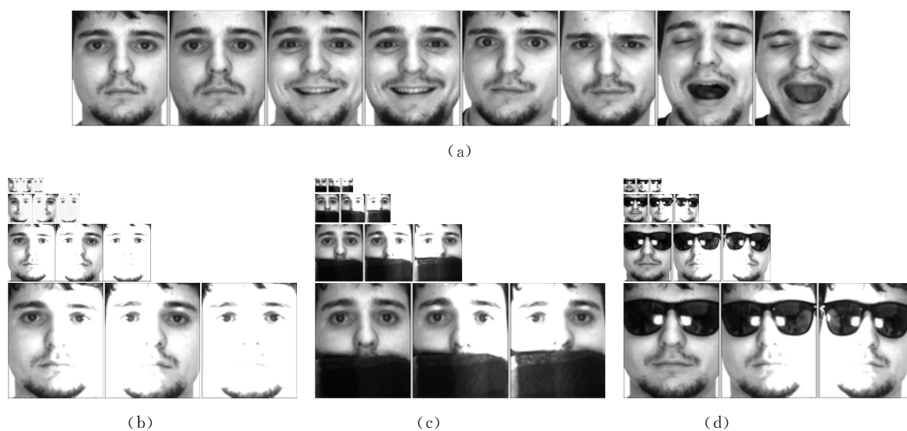


图6 来自于AR的训练和测试样本

Fig. 6 Training and test examples from AR database

表2比较了在AR库的3个测试集上将DIP-rMRF和一致性零填充施加于6种已有CNN模型前后的性能变化。可以看出,本文方法(DIP-rMRF和一致性零填充)几乎在各种情形下都提升了已有模型的识别率。值得注意的是:

(1)在输入图像的像素维度较低的情形下,本文方法带来的识别性能提升最为显著。例如,相对于PCANet,当输入图像的像素维为154维时,DIP-rMRF将PCANet的识别率在3个测试子集上依次提升了10.92%,12.44%,13.44%;而当输入图像的像素维为10304维时,DIP-rMRF只是将PCANet的识别率提升了3.35%,2.62%,-3.43%。产生这一现象的主要原因是:相对于高维的输入图像,低维的输入图像中所包含的判别性信息本身就已经显著下降,此时更需要有效排除遮挡等坏特征的影响。

(2)本文方法也可能会引发性能退化现象。从表中可以清楚地看到:尽管6种已有CNN模型的识别性能各不相同,但它们在各个测试子集上几乎都没有出现性能退化现象,除了LCNN(它在测试集I上有轻微的性能退化)。然而,使用

本文方法却可能导致“性能退化”。例如,将VGGFace与本文方法相结合在3个测试子集上的2576像素维(而非最高的10304维)下达到了最优的识别性能。这说明基于DIP-rMRF的遮挡检测模型可能需要在特定的维度下工作才能发挥更大的作用。关于DIP-rMRF和输入像素维之间的关系将留待以后进一步研究。

(3)在测试集III的最高像素维上,本文方法能带来的识别性能提升相对较小,尤其是对于PCANet,本文方法反而导致其识别性能下降。产生这一现象的原因是多方面的,对于PCANet而言,主要是超高维的特征所致。经过深度滤波、模式图编码和局部柱状图特征提取,PCANet将输入图像变换到了维度非常高的特征空间,约为输入图像维度的40倍。在如此高维的特征空间中,足以“稀释”高维输入图像中所包含的噪声特征。换言之,PCANet是用更高维的变换特征来避免高维的输入图像所带来的维度灾难问题。在此情形下,对遮挡位置进行检测就没有那么重要了,反而会由于“误检”而出现识别性能下降。

表2 在AR的3个测试集上的识别率

Table 2 Recognition rate of each CNN method on three test sets

像素维度	测试集 I				测试集 II				测试集 III			
	154	644	2576	10304	154	644	2576	10304	154	644	2576	10304
PCANet	50.14	67.79	74.23	[82.35]	26.89	54.90	72.55	92.72	13.17	36.13	70.31	[90.54]
PCANet	61.06	78.99	76.51	[85.70]	39.33	63.26	[84.15]	[95.34]	26.61	54.62	75.27	[87.11]
VGGFace	23.53	59.94	66.11	[67.23]	9.52	54.90	68.35	[68.35]	4.48	9.52	22.97	[28.01]
VGGFace	52.91	70.11	[75.35]	73.62	26.61	61.62	[73.95]	73.86	19.49	22.18	[33.55]	32.90
LCNN	61.90	73.39	[75.63]	74.51	38.94	59.94	70.03	[74.23]	18.49	40.90	54.34	[64.15]
LCNN	69.62	75.91	[80.14]	78.07	45.45	61.06	75.15	[79.18]	28.57	61.62	66.39	[71.66]
SphereFace	64.15	76.71	81.79	[82.07]	32.49	61.90	[78.43]	72.83	25.77	54.34	57.98	[70.31]
SphereFace	71.34	79.83	[89.83]	88.63	44.53	72.27	[82.20]	81.24	26.89	59.94	63.87	[82.35]
InterpretFR	72.93	74.05	80.26	[88.43]	63.26	75.05	84.32	[85.71]	36.92	75.59	[79.59]	82.53
InterpretFR	75.59	78.05	81.63	[89.55]	69.05	78.91	86.55	[87.55]	40.85	79.67	82.82	[89.72]
FROM	73.56	74.79	80.65	[89.25]	65.67	76.35	85.43	[86.28]	37.25	76.79	[81.39]	85.78
FROM	76.92	79.32	82.69	[89.85]	70.32	79.58	87.32	[88.65]	43.59	80.82	83.92	[91.57]

注:方括号表示各识别方法在输入图像的各个维度上所能达到的最优识别率。

导致本文方法对已有CNN模型性能提升有限的另一个主要原因是测试子集III的识别是最具有挑战性的。与下文实验中将要使用的模拟遮挡不同,现实遮挡所形成的局部特征往往可能与人脸图像的真实特征相混淆(模拟遮挡仅仅存在于测试集中)。图7(a)~图7(c)分别展示了训练集中人脸

图像的真实特征(如肤色、胡须和眼镜)容易与测试集I,II和III中的实际遮挡(如强光光照、围巾遮挡、太阳镜遮挡)相混淆的例子。显然,对于某一待识别的有遮挡图像y而言,其识别的难度并不取决于训练集与y之间所能够产生的混淆的最大程度,而取决于训练集中容易与y产生混淆的图像的个数。

在 AR 库中,肤色与光照发生混淆的情况要远远少于胡须与围巾发生混淆的情况,而胡须与围巾发生混淆的情况又要远远少于眼镜与太阳镜遮挡发生混淆的情况。因此,从测试 I 到测试集 III,识别难度依次增加,表 2 的实验结果也验证了这一点。



图 7 训练集中容易与测试样本发生“混淆”的数据样例

Fig. 7 Examples of data that are easily “confused” with test samples in training set

4.4 基于 Extended Yale B 的有光照变化与模拟遮挡的人脸识别

本节基于 Extended Yale B^[59]人脸数据库测试 DIP-rMRF 及所比较的 8 种分类器对不同程度的光照变化和不同

程度的遮挡的承受能力。Extended Yale B 数据库包含 38 个人的具有不同光照变化的脸部图像,按照光照变化条件可划分为 5 个子集:I(共 262 张)、II(共 455 张)、III(共 453 张)、IV(共 524 张)、V(共 712 张),从子集 I 到子集 V,光照逐渐变暗,如图 8 第一列所示。为了模拟不同程度的光照和不同程度(百分比)的遮挡相混合的情况,我们在子集 II、IV 和 V 的每张图像中分别添加 0%~90%(以 10%为间隔)的模拟遮挡(狒狒),且遮挡的位置是随机的,如图 8(b)~图 8(d)所示,最终得到的子集 III、IV 和 V 中共有 $(453+524+712) \times 10 = 16890$ 张包含了不同程度的光照变化和遮挡的图像。我们选取子集 I 和 II 的所有图像(共 717 张)作为训练集,选取扩充后的子集 III(共 4530 张)、IV(共 5240 张)和 V(共 7120 张)分别作为测试集 I、II 和 III。所有的训练集和测试集图像均被裁剪并对齐为 96×84 的像素矩阵。注意,本文实验环境与 Wei 等^[64]的实验环境类似。



图 8 来自 Extended Yale B 的训练和测试样本

Fig. 8 Examples of training and test images from Extended Yale B

表 3 列出了在 Extended Yale B 的 3 个测试集上将 DIP-rMRF 和一致性零填充施加于 6 种已有 CNN 模型前后的识别性能对比结果。为了展示大面积遮挡对识别性能的影响,表 3 只列出了当遮挡比例从 60% 上升到 90% 时的识别率。可以看出,当存在大面积遮挡时,大部分 CNN 模型的性能都降低到了 50% 以下,这是由于预训练时所采用的训练集样本都是在非受控环境下训练得到的,而 Extended Yale B 数据集在严格的受控环境下采集,训练样本的分布存在巨大的差异。

在 4.5 节,我们将看到 CNN 模型的识别性能优势在非受控环境下非常明显。Mehdipour 等^[66]的实验也给出了类似的结果(VGGFace 和 LCNN 在受控环境下的识别性能要远远弱于非受控环境下的性能)。但不同的是,Mehdipour 等的实验指出 VGGFace 对新数据的适应能力优于 LCNN,而从本节的实验结果来看,结论正好相反,这可能是由于 Mehdipour 等并没有采用最新的 LCNN 模型(目前,LCNN 模型已经更新到了 C 版)。

表 3 在 Extended Yale B 的 3 个测试集上的识别率

Table 3 Recognition rates on three test sets of Extended Yale B

occ	测试集 I				测试集 II				测试集 III			
	60	70	80	90	60	70	80	90	60	70	80	90
PCANet	100.00	100.00	98.46	25.49	99.81	97.34	84.03	16.73	85.15	59.24	27.31	3.64
PCANet	100.00	100.00	99.85	55.32	100.00	98.59	86.27	44.18	89.80	82.72	76.67	42.38
VGGFace	7.91	3.96	2.42	2.86	5.51	3.04	3.23	3.04	3.64	2.38	2.38	1.82
VGGFace	62.00	56.30	45.48	15.49	59.34	41.44	33.19	11.38	51.97	39.33	23.14	10.03
LCNN	13.85	6.59	2.86	2.42	5.89	3.04	3.61	2.85	2.38	2.10	2.38	2.24
LCNN	71.20	65.46	55.64	26.59	63.54	58.71	40.76	14.45	58.15	46.08	30.37	19.66
SphereFace	39.96	27.34	19.28	8.67	33.72	25.53	17.30	7.49	31.81	23.59	15.67	7.04
SphereFace	77.01	66.57	56.70	31.21	76.62	63.16	52.11	26.08	70.03	50.70	33.25	22.66
InterpretFR	50.15	47.87	39.56	25.60	49.24	39.46	31.08	21.27	42.52	37.67	27.86	11.53
InterpretFR	79.56	71.58	66.81	45.16	78.30	73.46	61.94	32.09	73.81	67.42	59.83	27.23
FROM	52.32	48.58	42.39	28.72	50.32	42.86	32.58	23.79	43.22	38.57	28.76	15.39
FROM	82.35	75.72	68.68	48.58	79.52	76.59	63.59	35.57	75.92	69.38	71.59	29.58

(%)

然而,PCANet在这3个测试集上对“遮挡”的容忍度达到了令人惊讶的程度。例如,在测试集I上,即使存在80%的遮挡,PCANet仍然能达到惊人的98.46%的准确率;而在测试集III上,当存在极端的光照变化和60%的遮挡时,PCANet的识别率仍然能达到85.15%。这首先是因为PCANet没有使用全连接,采用的是局部统计特征;其次,PCANet的网络模型(PCANet的前两层的卷积核)不需要大规模的训练数据(通过实验可以发现大规模数据对PCANet的性能提升影响并不大)。因此,其可以较好地适用于Extended Yale B中的人脸数据。

然而,不论是对预训练的CNN模型,还是对PCANet,在施加了DIP-rMRF和一致性零填充后,在这3个测试集的4种不同的遮挡比例下,识别性能均得到了显著提升,尤其是在更大的遮挡比例或更差的光照条件下,DIP-rMRF和一致性零填充带来的性能提升更为明显。

对比表2和表3中本文方法对已有CNN模型的性能提升效果可以发现,本文方法在处理实际遮挡问题上的优势更为明显。而正如4.3节所述,与模拟遮挡不同的是,实际遮挡很容易引发“混淆”现象,能够更好地处理实际遮挡表明本文方法具有很强的抗“混淆”的能力。

4.5 基于LFW的非受控环境下的有遮挡人脸识别

LFW数据库^[54]包含了5749个人的13233张人脸图像,所有图像均采集于互联网。该数据集被广泛应用于非受控环境下的人脸识别与认证算法的有效性验证^[48-49,56,66-71]。由于LFW的人脸图像包含了较大的姿势、表情、遮挡和光照等变化,因此,图像的预处理对于算法性能的提升特别重要。本文采用由Huang等^[72]提供的基于深度学习方法对齐的LFW人脸图像数据集LFW-deepfunneled,所有图像被裁剪和变换为62×58的灰度像素矩阵。选取样本个数大于8的217个人用于识别实验,其中,训练集选取具有较少的光照和姿势变化且不含遮挡的217×8=1736张图像,测试集选取这217个人剩余的3086张图像,并且为了模拟不同程度的遮挡,在测试集的每张图像中分别添加0%~60%(以10%为间隔)的模拟遮挡(狒狒),具体选取策略如图9所示。



(a)训练集



(b)测试集

图9 来自LFW的训练数据和测试数据样例

Fig. 9 Training and test samples from LFW database

表4列出了将DIP-rMRF和一致性零填充施加于6种已有CNN模型前后在LFW人脸库上的识别率。可以看出,在不排除遮挡干扰的情况下,各CNN模型的识别性能会迅速下降。相对而言,PCANet, InterpretFR和FROM的识别

性能下降速度较慢。这是因为PCANet没有全连接层,它所提取的是一种局部统计特征,遮挡的干扰相对较小,而InterpretFR和FROM本身就是针对有遮挡人脸识别问题设计的,所以它们在各种遮挡比例下的识别率也是最高的。即便如此,30%的遮挡就已经使得InterpretFR和FROM的识别性能从原本的99.05%和99.25%下降到了67.05%和68.23%,这表明只是在深层特征上去除遮挡的影响是远远不够的,还需要直接在输入图像中排除遮挡的影响。从表4的结果来看,在使用本文提出的DIP-rMRF进行遮挡检测并使用一致性零填充排除遮挡的影响后,各CNN模型的性能均有显著提升。

表4 各识别方法在LFW人脸库上的识别率

Table 4 Recognition rate of each recognition method on LFW face database

	Occ	0	10	20	30	40	50	60
PCANet	67.41	56.71	50.77	42.67	34.18	24.64	16.87	
PCANet	—	67.04	60.16	50.69	49.01	30.31	25.32	
VGGFace	98.07	88.48	62.72	29.05	8.24	2.37	1.31	
VGGFace	—	97.41	83.08	65.24	42.86	31.22	14.88	
LCNN	96.10	88.74	68.99	43.48	18.8	5.98	1.86	
LCNN	—	92.40	85.29	63.37	46.78	25.21	16.38	
SphereFace	99.42	93.54	75.08	50.05	31.01	21.64	15.23	
SphereFace	—	98.17	87.71	78.75	65.47	43.78	29.35	
InterpretFR	99.05	94.45	80.66	67.05	40.33	35.50	22.86	
InterpretFR	—	99.02	93.31	91.73	89.14	65.82	48.93	
FROM	99.25	95.52	81.76	68.23	43.52	38.72	25.95	
FROM	—	99.10	95.42	93.53	90.52	67.39	50.29	

假定以遮挡比例每上升10%、识别率下降20%为性能崩溃点,表5比较了施加DIP-rMRF前后各CNN模型的性能崩溃点。可以看出:在施加DIP-rMRF前,大部分方法的性能崩溃点出现在遮挡比例从20%向30%过渡时;施加DIP-rMRF后,大部分方法(除了VGGFace)的性能崩溃点都大幅度后移到了遮挡比例从40%向50%过渡时。PCANet本身没有性能崩溃点,但当遮挡比例从40%增加到50%时,PCANet的性能有18.70%的下降,接近于性能崩溃点。这说明,当遮挡比例增加到50%时,在非受控环境下,大部分方法已经很难正确执行识别任务了。

表5 各方法在LFW上施加了DIP-rMRF进行遮挡检测前后的性能崩溃点

Table 5 Performance breakdown points of each comparison method before and after applying DIP-rMRF

	Occ	0~10	10~20	20~30	30~40	40~50	50~60
PCANet	10.70	5.94	8.10	8.49	9.54	7.77	
VGGFace	9.59	25.76	33.67	20.81	5.87	1.06	
LCNN	7.36	19.75	25.51	24.68	12.82	4.12	
SphereFace	5.88	18.46	25.03	19.04	9.37	6.41	
InterpretFR	4.60	13.79	13.61	26.72	4.83	12.64	
FROM	3.73	13.76	13.53	24.71	4.8	12.77	
PCANet	0.37	6.88	9.47	1.68	18.70	4.99	
VGGFace	0.66	14.33	17.84	22.38	11.64	16.34	
LCNN	3.70	7.11	21.92	16.59	21.57	8.83	
SphereFace	1.25	10.46	8.96	13.28	21.69	14.43	
InterpretFR	0.03	5.71	1.58	2.59	23.32	16.89	
FROM	0.15	3.68	1.89	3.01	23.13	17.10	

结束语 由遮挡引发的测试数据和训练数据之间的差异,是有遮挡人脸识别技术面临的重要挑战。受训练数据规模的限制,在非受控环境下训练得到的神经网络难以适用于受控环境下的人脸识别,特别是当人脸图像中存在遮挡时。为了降低遮挡对识别性能的影响,本文将 DIP 重构模型、鲁棒结构误差度量和马尔可夫随机场模型结合起来,提出了一种新的遮挡检测模型 DIP-rMRF,以有效处理人脸识别中由遮挡引发的问题;并提出了一致性零填充方法,以有效利用 DIP-rMRF 的遮挡检测结果。在 Extended Yale B, AR 和 LFW 这 3 个人脸数据库中针对 VGGFace, LCNN, PCANet, SphereFace, InterpretFR, FROM 这 6 种 CNN 模型的实验结果表明, DIP-rMRF 能够有效地处理遮挡以及由极端光照所引发的“类遮挡”,从而可以极大地提升现有 CNN 模型对有遮挡人脸识别的性能。因此,本文提出的 DIP-rMRF 为神经网络模型识别有遮挡人脸图像提供了一种重要的预处理方法。由于 DIP 重构模型需要在线训练,会极大影响识别速度,因此,如何进一步提升在线训练速度是下一步的工作。

参 考 文 献

- [1] WANG M, DENG W. Deep Face Recognition: A Survey [J]. *Neurocomputing*, 2021, 429: 215-244.
- [2] RANJAN R, SANKARANARAYANAN S, BANSAL A, et al. Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, Than Humans [J]. *IEEE Signal Processing Magazine*, 2018, 35(1): 66-83.
- [3] DING C, TAO D. A Comprehensive Survey on Pose - Invariant Face Recognition [J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2016, 7(3): 1-42.
- [4] BENDJILLALI R I, BELADGHAM M, MERIT K, et al. Illumination-Robust Face Recognition Based on Deep Convolutional Neural Networks Architectures [J]. *Indonesian Journal of Electrical Engineering and Computer Science*, 2020, 18(2): 1015-1027.
- [5] HAN X, LIU Y, YANG H, et al. Normalization of Face Illumination with Photorealistic Texture Via Deep Image Prior Synthesis [J]. *Neurocomputing*, 2020, 386: 305-316.
- [6] WANG K, PENG X, YANG J, et al. Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition [J]. *IEEE Transactions on Image Processing*, 2020, 29: 4057-4069.
- [7] REVINA I M, EMMANUEL W S. A Survey on Human Face Expression Recognition Techniques [J]. *Journal of King Saud University-Computer and Information Sciences*, 2021, 33(6): 619-628.
- [8] LI X, LIANG R. A Review for Face Recognition with Occlusion: From Subspace Regression to Deep Learning [J]. *Chinese Journal of Computers*, 2018, 41(1): 177-207.
- [9] ZENG D, VELDHUIS R, SPREEUWERS L. A Survey of Face Recognition Techniques under Occlusion [J]. *IET Biometrics*, 2021, 10(6): 581-606.
- [10] DAMER N, GREBE J H, CHEN C, et al. The Effect of Wearing a Mask on Face Recognition Performance: An Exploratory Study [C] // *Proceedings of the International Conference of the Biometrics Special Interest Group*. New York: IEEE Press, 2020: 1-6.
- [11] DAMER N, BOUTROS F, SÜBILCH M, et al. Extended Evaluation of the Effect of Real and Simulated Masks on Face Recognition Performance [J]. *IET Biometrics*, 2021, 10(5): 548-561.
- [12] ZHAO F, FENG J, ZHAO J, et al. Robust Lstm-Autoencoders for Face De-Occlusion in the Wild [J]. *IEEE Transactions on Image Processing*, 2017, 27(2): 778-790.
- [13] WAN W, CHEN J. Occlusion Robust Face Recognition Based on Mask Learning [C] // *Proceedings of the IEEE/CVF International Conference on Image Processing*. New York: IEEE Press, 2017: 3795-3799.
- [14] SONG L, GONG D, LI Z, et al. Occlusion Robust Face Recognition Based on Mask Learning with Pairwise Differential Siamese Network [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. New York: IEEE Press, 2019: 773-782.
- [15] YIN B, TRAN L, LI H, et al. Towards Interpretable Face Recognition [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. New York: IEEE Press, 2019: 9348-9357.
- [16] QIU H, GONG D, LI Z, et al. End2end Occluded Face Recognition by Masking Corrupted Features [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(10): 6939-6952.
- [17] ZHANG Y, WANG X, SHAKEEL M S, et al. Learning Upper Patch Attention Using Dual-Branch Training Strategy for Masked Face Recognition [J]. *Pattern Recognition*, 2022, 126: 108522.
- [18] HUANG B, WANG Z, WANG G, et al. Plface: Progressive Learning for Face Recognition with Mask Bias [J]. *Pattern Recognition*, 2023, 135: 109142.
- [19] DENG H, FENG Z, QIAN G, et al. Mfcoface: A Masked-Face Recognition Algorithm Based on Large Margin Cosine Loss [J]. *Applied sciences*, 2021, 11(16): 7310.
- [20] WANG C, FANG H, ZHONG Y, et al. Mlfw: A Database for Face Recognition on Masked Faces [C] // *Proceedings of the Chinese Conference on Biometric Recognition*. New York: Springer, 2022: 180-188.
- [21] DU H, SHI H, LIU Y, et al. Towards Nir-Vis Masked Face Recognition [J]. *IEEE Signal Processing Letters*, 2021, 28: 768-772.
- [22] ZHAO W, ZHU X, SHI H, et al. Consistent Sub-Decision Network for Low-Quality Masked Face Recognition [J]. *IEEE Signal Processing Letters*, 2022, 29: 1147-1151.
- [23] LI X X, DAI D Q, ZHANG X F, et al. Structured Sparse Error Coding for Face Recognition with Occlusion [J]. *IEEE Transactions on Image Processing*, 2013, 22(5): 1889-1900.
- [24] ULYANOV D, VEDALDI A, LEMPITSKY V. Deep Image Prior [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press,

- 2018;9446-9454.
- [25] POURPANAH F, ABDAR M, LUO Y, et al. A Review of Generalized Zero-Shot Learning Methods [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(4): 4051-4070.
- [26] DAI T, FENG Y, CHEN B, et al. Deep Image Prior Based Defense against Adversarial Examples [J]. *Pattern Recognition*, 2022, 122: 108249.
- [27] YUAN X, HE P, ZHU Q, et al. Adversarial Examples: Attacks and Defenses for Deep Learning [J]. *IEEE Transactions on Neural Networks And Learning Systems*, 2019, 30(9): 2805-2824.
- [28] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and Harnessing Adversarial Examples [C]// *Proceedings of the International Conference on Learning Representations*, 2015.
- [29] JIA X, WEI X, CAO X, et al. Comdefend: An Efficient Image Compression Model to Defend Adversarial Examples [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2019: 6084-6092.
- [30] MATAEV G, MILANFAR P, ELAD M. Deepred: Deep Image Prior Powered by Red [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. IEEE Press, 2019.
- [31] LIU J, SUN Y, XU X, et al. Image Restoration Using Total Variation Regularized Deep Image Prior [C]// *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. New York: IEEE Press, 2019: 7715-7719.
- [32] OTE K, HASHIMOTO F, ONISHI Y, et al. List-Mode Pet Image Reconstruction Using Deep Image Prior [J]. *arXiv*: 2204.13404, 2023.
- [33] ZHOU Q, ZHOU C, HU H, et al. Towards the Automation of Deep Image Prior [J]. *arXiv*: 191107185, 2019.
- [34] ZIHAN Z, WAGNER A, MOBAHI H, et al. Face Recognition with Contiguous Occlusion Using Markov Random Fields [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. New York: IEEE Press, 2009: 1050-1057.
- [35] LI X X, LIANG R, FENG Y, et al. Robust Face Recognition with Occlusion by Fusing Image Gradient Orientations with Markov Random Fields [C]// *Proceedings of the International Conference on Intelligent Science and Big Data Engineering*. New York: Springer, 2015: 431-440.
- [36] LI X X, HAO P, HE L, et al. Image Gradient Orientations Embedded Structural Error Coding for Face Recognition with Occlusion [J]. *Journal of Ambient Intelligence and Humanized Computing*, 2020, 11(6): 2349-2367.
- [37] DAHUA L, XIAOOU T. Quality-Driven Face Occlusion Detection and Recovery [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2007: 1-7.
- [38] HE R, ZHENG W, HU B. Maximum Correntropy Criterion for Robust Face Recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1561-1576.
- [39] TZIMIROPOULOS G, ZAFEIRIOU S, PANTIC M. Subspace Learning from Image Gradient Orientations [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(12): 2454-2466.
- [40] HE R, ZHENG W S, TAN T, et al. Half-Quadratic-Based Iterative Minimization for Robust Sparse Representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 36(2): 261-275.
- [41] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431-3440.
- [42] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]// *Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention*. New York: Springer, 2015: 234-241.
- [43] MAO X, SHEN C, YANG Y B. Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections [C]// *Proceedings of the Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2016: 2802-2810.
- [44] MENG Y, LEI Z, JIAN Y, et al. Robust Sparse Coding for Face Recognition [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2011: 625-632.
- [45] KIM J, LEE J K, LEE K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2016: 1646-1654.
- [46] ZHOU Z, WAGNER A, MOBAHI H, et al. Face Recognition with Contiguous Occlusion Using Markov Random Fields [C]// *Proceedings of the IEEE International Conference on Computer Vision*, 2009: 1050-1057.
- [47] LIANG R, LI X X. Mixed Error Coding for Face Recognition with Mixed Occlusions [C]// *Proceedings of the International Joint Conference on Artificial Intelligence*. Washington, DC: AAAI, 2015.
- [48] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep Face Recognition [C]// *Proceedings of the British Machine Vision Conference*. Durham: BMVA Press, 2015.
- [49] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A Unified Embedding for Face Recognition and Clustering [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2015: 815-823.
- [50] WU X, HE R, SUN Z. A Lightened CNN for Deep Face Representation [J]. *arXiv*: 1511.02683, 2015.
- [51] LIU W, WEN Y, YU Z, et al. SpheroFace: Deep Hypersphere Embedding for Face Recognition [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2017: 212-220.
- [52] LIU W, WEN Y, RAJ B, et al. SpheroFace Revived: Unifying Hyperspherical Face Recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(2): 2458-2474.

- [53] CHAN T H, JIA K, GAO S, et al. PCANet: A Simple Deep Learning Baseline for Image Classification? [J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 5017-5032.
- [54] HUANG G B, MATTAR M, BERG T, et al. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments [J/OL]. <https://inria.hal.science/inria-00321923/document>.
- [55] WOLF L, HASSNER T, MAOZ I. Face Recognition in Unconstrained Videos with Matched Background Similarity [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 2011: 529-534.
- [56] TAIGMAN Y, YANG M, RANZATO M A, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2014: 1701-1708.
- [57] WANG H, WANG Y, ZHOU Z, et al. Cosface: Large Margin Cosine Loss for Deep Face Recognition [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2018: 5265-5274.
- [58] DENG J, GUO J, XUE N, et al. Arcface: Additive Angular Margin Loss for Deep Face Recognition [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2019: 4690-4699.
- [59] LEE K C, HO J, KRIEGMAN D J. Acquiring Linear Subspaces for Face Recognition under Variable Lighting [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2005(5): 684-698.
- [60] MARTINEZ A M. The Ar Face Database [C]// *CVC*. 1998.
- [61] LI X X, HE L, HAO P, et al. Adaptive Weberfaces for Occlusion-Robust Face Representation and Recognition [J]. *IET Image Processing*, 2017, 11(11): 964-975.
- [62] FIDLER S, SKOCAJ D. Combining Reconstructive and Discriminative Subspace Methods for Robust Classification and Regression by Subsampling [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006(3): 337-350.
- [63] MURPHY K P. *Machine Learning: A Probabilistic Perspective* [M]. Massachusetts: MIT Press, 2012.
- [64] WEI X, LI C T, HU Y. Robust Face Recognition under Varying Illumination and Occlusion Considering Structured Sparsity [C]// *Proceedings of the International Conference on Digital Image Computing Techniques and Applications*. New York: IEEE Press, 2012: 1-7.
- [65] MEHDIPOUR G M, KEMAL E H. A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. New York: IEEE Press, 2016: 34-41.
- [66] GUILLAUMIN M, VERBEEK J, SCHMID C. Is That You? Metric Learning Approaches for Face Identification [C]// *Proceedings of the IEEE International Conference on Computer Vision*. New York: IEEE Press, 2009: 498-505.
- [67] PINTO N, DICARLO J J, COX D D. How Far Can You Get with a Modern Face Recognition Test Set Using Only Simple Features? [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2009: 2591-2598.
- [68] SUN Y, CHEN Y, WANG X, et al. Deep Learning Face Representation by Joint Identification-Verification [C]// *Proceedings of the Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2014: 1988-1996.
- [69] SUN Y, WANG X, TANG X. Deep Learning Face Representation from Predicting 10,000 Classes [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2014: 1891-1898.
- [70] SUN Y, WANG X, TANG X. Deeply Learned Face Representations Are Sparse, Selective, and Robust [C]// *Proceedings of the IEEE/CVF Conference on Computer vision and Pattern Recognition*. New York: IEEE Press, 2015: 2892-2900.
- [71] SUN Y, WANG X, TANG X. Sparsifying Neural Network Connections for Face Recognition [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2016: 4856-4864.
- [72] HUANG G, MATTAR M, LEE H, et al. Learning to Align from Scratch [C]// *Proceedings of the Advances in Neural Information Processing Systems*. Cambridge: MIT Press, 2012: 764-772.



LI Xiaoxin, born in 1980, Ph.D, associate professor, master supervisor, is a member of CCF(No. 80065M). His main research interests include image processing and pattern recognition.



DING Weijie, born in 1981, Ph.D, professor. His main research interests include information visualization and network security.

(责任编辑:何杨)