高技术通讯 2024 年 第 34 卷 第 6 期:602-615

doi:10.3772/j.issn.1002-0470.2024.06.005

基于稠密连接的通道混合式 PCANet 的低分辨率有遮挡人脸识别①

秦 娥②* 何佳瑶* 刘银伟* 朱娅妮③** 李小薪*

(*浙江工业大学计算机科学与技术学院 杭州 310023)

(**杭州电子科技大学计算机学院 杭州 310018)

摘 要 针对低分辨率有遮挡人脸识别问题提出了基于稠密连接的通道混合式主成分分析网络(DCH-PCANet)。现有的 PCANet 模型的卷积层只使用了通道无关式卷积(CIC)。通道无关式卷积由于未考虑特征图在通道方向上的相关性,可以更好地凸显单个特征图的局部纹理特征,对于补偿因低分辨率、遮挡等因素导致的特征损失具有重要意义,但也会强化遮挡区域的特征,从而放大坏特征的影响范围;而通道相关式卷积(CDC)由于充分考虑了各特征图在通道方向上的相关性,可以较好地抑制坏特征的作用,形成较为稀疏的特征图。在 PCANet 中添加了基于通道相关式卷积的特征图提取分支,形成了通道混合式 PCANet;并且引入了稠密连接,以充分利用低阶特征提升有遮挡图像识别的鲁棒性。针对如下 4 种数据集进行了实验:受控环境、真实遮挡和模拟低分辨率的人脸数据集(AR 人脸数据集),非受控环境、真实遮挡和模拟低分辨率的人脸数据集(MFR2 和 PKU-Masked-Face),非受控环境、真实遮挡和真实低分辨率的人脸数据集(自建数据集)。实验结果表明,与现有方法相比,所提出的基于稠密连接的通道混合式 PCANet 具更好的遮挡鲁棒性和低分辨率鲁棒性,可以作为前沿方法的有效补充,提升其识别性能。

关键词 有遮挡人脸识别;主成分分析网络(PCANet);通道相关式卷积(CDC);稠密连接

近年来,以卷积神经网络(convolutional neural network,CNN)为代表的深度神经网络(deep neural network,DNN)在计算机视觉领域中获得了巨大的成功,优秀的深度学习模型在一些公开数据集上的分类能力甚至超过了人类,如在 LFW(labeled faces in the wild)人脸数据库上的认证准确率[1-2]及在 ImageNet 上的图像分类准确率[3-4]。然而,在实际情景中,待识别图像在“结构”或者“分布”上经常与训练图像存在较大的差异,这种差异会降低 DNN 的识别率。在计算机视觉领域中,这种待识别图像与训练图像或库存图像有较大差异的现象也被称作为“协变量偏移(covariate shift)”问题[5-7]。当待识别图像与训练集图像存在遮挡、光照、分辨率等幅度较大的差异时,现有的神经网络模型的识别准确率往往会大幅度下降。本文重点关注由遮挡和低分辨率造成的“偏移”问题。

为了减轻遮挡造成的性能退化,前沿的研究工作给出了 2 种不同的解决方案,即基于图像重建的方法[8-11]和基于遮挡去除的方法[12-14]。基于图像重建的方法首先修复被遮挡的面部部位,然后对修复后的人脸图像进行识别。在深度学习领域,Zhao 等人[8]提出基于长短期记忆网络(long short time memory, LSTM)自动编码器进行人脸遮挡区域的重建,并对重建后的面部图像进行识别。这种方法的主要挑战是如何在较好重建遮挡区域的同时保留人脸图像的判别信息。另一类方法旨在去除遮挡的影响,

① 浙江省自然科学基金(LGF22F020027)和国家自然科学基金(62271448)资助项目。
② 女,1978 年生,硕士,讲师;研究方向:图像处理与模式识别;E-mail: qe@zjut.edu.cn。
③ 通信作者,E-mail: 10764341@qq.com。
(收稿日期:2023-10-12)

先检测并删除被遮挡破坏的特征,然后利用剩余的干净特征进行识别。此类方法的主要挑战是如何检测到人脸图像中的遮挡位置,并找到遮挡位置与深度特征图之间的对应关系。Wan 和 Chen[12] 通过向 CNN 模型的中间层添加 MaskNet 分支来解决这个问题,该分支将较低的权重分配给被遮挡损坏的隐藏单元。Song 等人[13] 提出用二值掩码从网络的最后一个隐含层特征去除被遮挡损坏的特征,为此,需要将人脸图像划分为 $K \times K$ 块,在训练阶段学习 K^2 个遮挡字典,将输入图像中的每个遮挡块映射到遮挡字典中相应的字典原子;在测试阶段,首先检测被遮挡的块,然后检索遮挡字典中对应的字典原子,并将其用于去除测试特征中的坏特征。这种两阶段的方法必须依赖于一个用于遮挡检测的外部网络。另外,Song 等人[13] 提出的方法需要在训练阶段独立训练 K^2 个深度网络模型来学习遮挡字典,这使得训练效率低下且耗时。Qiu 等人[14] 提出一种端到端的有遮挡人脸识别方法,在训练阶段,利用遮挡的空间连续特性作为特征掩码学习的监督信号,学习掩码解码器;在测试阶段,掩码解码器会根据输入图像动态生成一个新的特征掩模,并使用该特征掩模清除特征图中被遮挡损坏的特征,从而达到鲁棒识别的目的。

然而,上述方案都需要对遮挡信息进行明确的处理,要么对遮挡区域进行重建,要么去除遮挡的影响。这就需要花费较大的代价对“遮挡”本身进行建模。那么,是否存在一种方法既不需要对遮挡进行显式处理,也能很好地解决遮挡问题呢?已知的方法主要有 2 种:基于图像梯度方向(image gradient orientation,IGO)域的鲁棒误差度量方法[15-16] 和基于 FPH(feature extraction-pattern map encoding-histogram)框架[17-25] 的特征提取方法。基于 IGO 的鲁棒误差度量方法由 Tzimiropoulos 等人[15] 提出,他们指出,两幅完全不同的图像在 IGO 域中的差值以非常高的置信度服从[0,2π)上的均匀分布,且总体差异之和几乎为零。利用这一现象,可以有效消除遮挡的影响。然而,正如 Li 等人[16] 所指出的,由于在实际中很难存在理想的与人脸图像“完全不同”的遮挡,在 IGO 域中的鲁棒误差度量并不能完全消除遮挡影响,还需要辅助遮挡检测或遮挡区域的重建。另一方面,与鲁棒误差度量方法不同,以主成分分析网络(principal component analysis network, PCANet)[26] 为代表的 FPH 方法能够充分利用人脸图像的局部统计特征,直接从输入图像中提取遮挡鲁棒的特征,而无须对遮挡信息进行显式建模,是一类可以很好地解决遮挡问题的方法。

然而,PCANet 的主要不足在于不能很好地识别低分辨率人脸图像。相对于高分辨率图像,低分辨率图像的一个主要问题是能够从其中提取的判别特征远远少于高分辨率图像。为了有效识别低分辨率图像,需要对其进行特征补偿。现有的 PCANet 方法[26] 利用 PCA 滤波器对输入图像进行深度滤波,从不同主方向和不同层级进行特征分解和特征提取,从而丰富了输入图像的判别特征。然而,现有的 PCANet 方法所采用的是一种通道无关式滤波方法,即在提取特征时,并未考虑特征图在通道方向上的相关性。这种滤波方式由于未考虑特征图在通道方向上的相关性,可以更好地凸显单个特征图的局部纹理特征,对于补偿因低分辨率、遮挡等因素导致的特征损失具有重要意义,但也会强化遮挡区域的特征,从而放大坏特征的影响范围;而通道相关式卷积由于充分考虑了各特征图在通道方向上的相关性,可以较好地抑制坏特征的作用,形成较为稀疏的特征图。针对这一问题,本文在原始的 PCANet 中加入了基于通道相关式卷积的特征图提取分支,形成了通道混合式 PCANet(channel-hybrid PCANet,CH-PCANet)。该方法能够有效利用特征图在通道方向上的相关性,对低分辨图像的输出特征进行补偿,从而有效提升其判别性。

另一方面,PCANet 并未充分利用所学习到的特征以获取更好的遮挡鲁棒性。PCANet 只利用了最后一个卷积层输出的特征图来生成后续的模式图和柱状图特征。正如 Wright 等人[9] 所指出的那样,特征提取会将遮挡的影响扩散到其邻域范围,因此,遮挡的影响会随着网络深度的增加而增大,这就意味着,对于有遮挡人脸识别而言,网络的底层特征往往也至关重要,需要充分利用。为了解决这一问题,受 DenseNet[27] 的启发,本文将稠密连接也引入 PCANet

的网络模型,尽可能地利用网络的底层特征,并丰富 PCANet 所提取的特征,从而提升 PCANet 的鲁棒性。

将通道混合式滤波和稠密连接相结合,本文提出了基于稠密连接的通道混合式 PCANet(dense channel-hybrid PCANet,DCH-PCANet)。实验结果表明,DCH-PCANet 更加有效地处理待识别图像中的遮挡、光照等变化,也能够对低分辨率图像的输出特征进行有效补偿,从而大幅提升了低分辨率有遮挡图像的识别性能。

1　问题定义

本文考虑低分辨率有遮挡图像的识别问题。图 1展示了来自 AR(augmented reality)人脸数据库的高分辨率的训练集 X 和低分辨率的测试集 Y 中的样本样例。已知训练集 X 中的任意一张图像 $\boldsymbol{x}_i$ 的类别标签为 $\mathrm{Id}(\boldsymbol{x}_i) = l_i$,也就是 $L = \{l_i\}_{i=1}^{N}$ 构成了训练集 X 的标签集;测试图像的分辨率小于训练图像的分辨率,即:$h' < h$ 且 $w' < w$,且测试集 Y 中的任意一张图像 y_i 的类别标签 $\mathrm{Id}(\boldsymbol{y}_i) \in L$。如何通过 $\boldsymbol{y}_i$ 与 X 中图像进行特征匹配来预测 $\mathrm{Id}(\boldsymbol{y}_i)$?

(a) 训练样本:8张高分辨率无遮挡和光照变化的人脸图像　(b) 测试样本:9张低分辨率有遮挡和光照变化的人脸图像

图 1　低分辨有遮挡人脸识别问题示例

由于测试图像与训练图像不仅在像素分布上而且在分辨率上都存在较大的差异,直接在像素域中进行特征匹配是不可行的。本文探讨如何基于 DCH-PCANet 有效补偿低分辨率有遮挡人脸图像中丢失的判别信息,从而达到鲁棒识别的目的。基于于卡方距离[26] $\chi^2(\cdot,\cdot)$ 的最近邻分类器进行分类,对于 $\forall \boldsymbol{y} \in Y$,从 X 中寻求满足如下优化式的 $\boldsymbol{x}$:

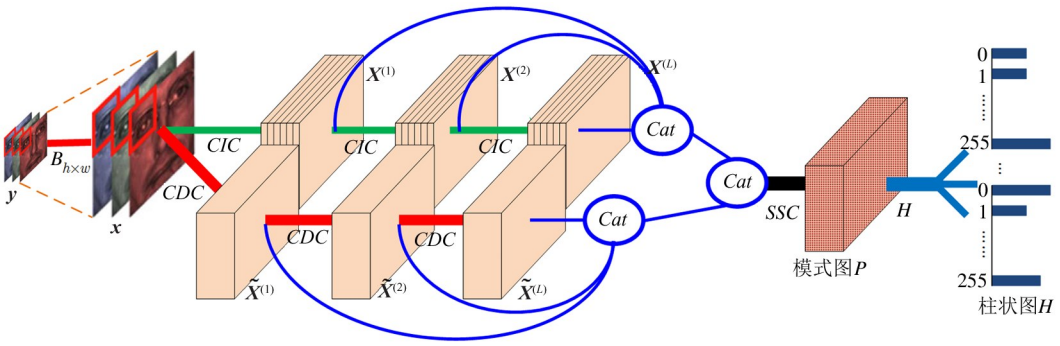
$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in X} \chi^2(f_{\text{CH-PCANet}}(\boldsymbol{x}), f_{\text{CH-PCANet}}(B_{h\times w}(\boldsymbol{y}))) \tag{1}$$

从而确定 $\boldsymbol{y}$ 的类别:$\mathrm{Id}(\boldsymbol{y}) = \mathrm{Id}(\boldsymbol{x}^*)$。式(1)中的 $B_{h\times w}(\cdot)$ 表示双三次插值函数(Bicubic),对低分辨率测试图像进行插值预处理,使得插值后的图像的分辨为 $h \times w$。

2　通道混合式 PCANet

2.1　网络设计

本文所提出的基于稠密连接的通道混合式 PCANet 的总体网络结构图如图 2 所示。正如 Lim 等人[28]所指出的那样,提升网络性能的关键在于增加网络的规模,即网络的深度和宽度。然而,由于 PCANet 的各层滤波器存在的冗余性,增加 PCANet 的深度并不能有效提升其性能,本文考虑增加其宽度。原始的 PCANet 在特征提取阶段采用"通道无关式卷积(channel-independent convolution,CIC)",即不论是其 PCA 滤波器的训练,还是其卷积运算都没有考虑特征图的各通道之间的相互影响;DCH-PCANet 在此基础上增加了"通道相关式卷积(channel-dependent convolution,CDC)"分支,以补偿通道无关式卷积所丢失的对通道相关性信息的描述。



CIC 表示通道无关式卷积运算,*CDC* 表示通道相关式卷积运算,*Cat* 表示对特征图子集的拼接运算,*H* 表示对模式图进行分块柱状图特征提取

图 2　基于稠密连接的通道混合式 PCANet(DCH-PCANet)的总体网络结构图

令 $\boldsymbol{X}^{(l)}$ 和 $\boldsymbol{X}^{\sim(l)}$ 分别表示 CH-PCANet 在第 l 层所产生的通道无关式和通道无关式特征图，$X^{(0)} = x/B_{h\times w}(y)$ 表示输入图像；CIC 表示通道无关式卷积算子，$\boldsymbol{W}^{(l)}$ 表示其卷积核参数；CDC 表示通道相关式卷积算子，$\widetilde{\boldsymbol{W}}^{(l)}$ 表示其卷积核参数；CH-PCANet 的特征提取过程可表示为

$$f_{\text{CH-PCANet}}(\boldsymbol{x}) = H(SSC(Cat(\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)},\cdots,\boldsymbol{X}^{(L)},\widetilde{\boldsymbol{X}}^{(1)},\widetilde{\boldsymbol{X}}^{(2)},\cdots,\widetilde{\boldsymbol{X}}^{(L)}))) \tag{2}$$

其中，$Cat(\cdot,\cdots)$ 表示特征图在通道方向上的拼接操作，$X^{(l)}$ 和 $\tilde{X}^{(l)}$ 递归定义如下：

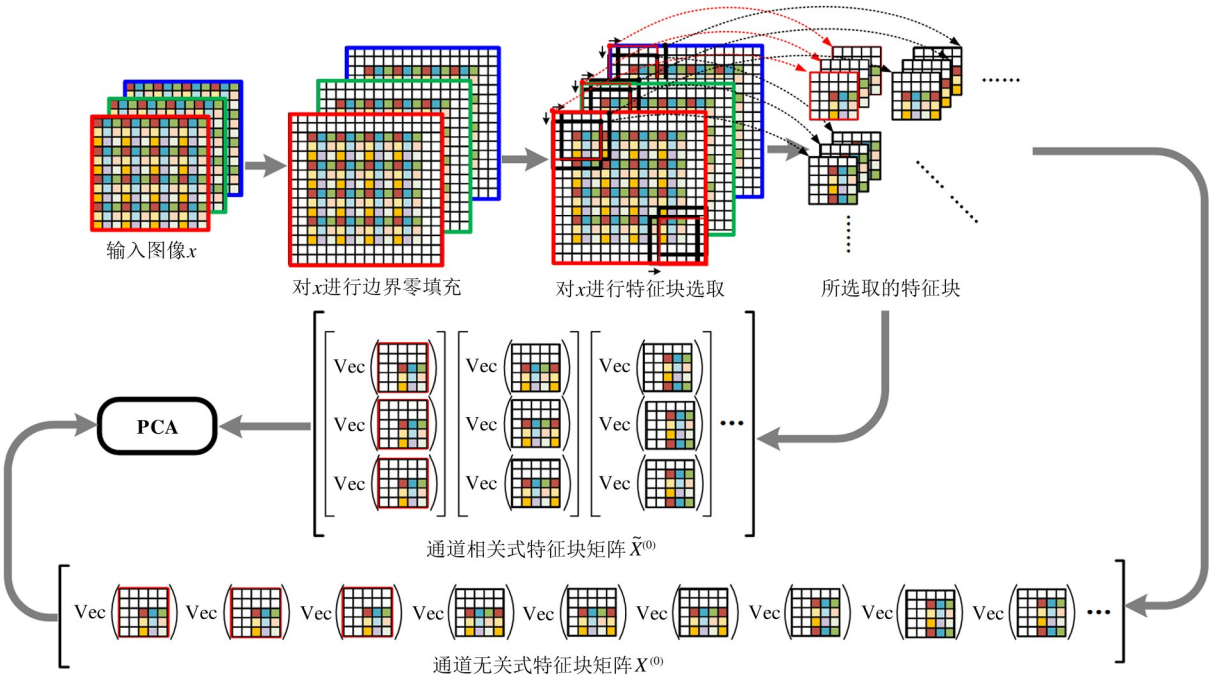
$$\boldsymbol{X}^{(l+1)} = CIC(\boldsymbol{X}^{(l)},\boldsymbol{W}^{(l)})$$

$$\widetilde{\boldsymbol{X}}^{(l+1)} = CDC(\widetilde{\boldsymbol{X}}^{(l)},\widetilde{\boldsymbol{W}}^{(l)})$$

注意，本文所提出的 DCH-PCANet 是将 CDC 和 CIC 联合使用，而非使用前者代替后者，这是因为 CDC 和 CIC 所生成的特征能够相互补偿。2.3 节将通过可视化比较，对 CDC 和 CIC 的特性进行进一步分析；另外，为了充分补偿低分辨图像所丢失的细节信息，DCH-PCANet 在特征提取阶段的 CIC 分支和 CDC 分支上同时使用了稠密连接，以形成更高维的特征空间。

2.2　通道相关式 PCA 滤波器

原始的 PCANet[26] 并未考虑特征图的“通道”，不加区分地从所有特征图中提取 PCA 滤波器，本文将基于这种方式获取的滤波器称为“通道无关式滤波器”。显然，使用通道无关式滤波器进行特征提取忽略了特征图的各通道之间的相关性。为了充分利用这种相关性，本文提出通道相关式 PCA 滤波器。图 3 给出了通道相关式和通道无关式 PCA 滤波器组的训练流程。下面给出第 l 层通道相关式 PCA 滤波器的生成(训练)方式。



沿着箭头所指向的方向依次为原始特征图、边界零填充、特征块选取、所选取的特征块、使用 Vec(·)向量化特征块并对特征块进行重组得到通道相关式/通道无关式矩阵、基于 PCA 提取滤波器。

图 3　通道相关式和通道无关式 PCA 滤波器组的训练流程

(1)获取通道相关式特征块 $\widetilde{\boldsymbol{X}}^{(l)}$

令 $\widetilde{\boldsymbol{X}}_i^{(l)} \in R^{h\times w\times c_l}$ 表示第 i 个训练样本在 DCH-PCANet 的通道相关式分支的第 l 层输出的特征图，$\tilde{\boldsymbol{x}}_{i,c,b}^{(l)} \in R^{k\times k}$ 表示从 $\widetilde{\boldsymbol{X}}_i^{(l)}$ 的第 $c \in \{1,2,\cdots,c_l\}$ 个通道中抽取的第 b 个大小为 $k\times k$ 的特征块，Vec(·)表示将矩阵拉伸为列向量的运算。如图 3 的第 1 行和第 2 行所示，$\widetilde{\boldsymbol{X}}_i^{(l)}$ 的第 b 个通道相关式特征块通过如下方式获取：

$$\tilde{\boldsymbol{x}}_{i,b}^{(l)} = [\mathrm{Vec}(\tilde{\boldsymbol{x}}_{i,1,b}^{(l)})^{\mathrm{T}},\ \mathrm{Vec}(\tilde{\boldsymbol{x}}_{i,2,b}^{(l)})^{\mathrm{T}},\cdots,\mathrm{Vec}(\tilde{\boldsymbol{x}}_{i,c_l,b}^{(l)})^{\mathrm{T}}]^{\mathrm{T}} \tag{3}$$

所有 N 个训练样本在第 l 层上形成的所有通道相关式特征块可通过如下方式拼接起来就构成了第 l 层的通道相关式矩阵 $\widetilde{\boldsymbol{X}}^{(l)} \in R^{k^2c_l \times NB}$：

$$\widetilde{\boldsymbol{X}}^{(l)} = [\underbrace{\tilde{\boldsymbol{x}}_{1,1}^{(l)},\cdots,\tilde{\boldsymbol{x}}_{1,B}^{(l)}}_{\text{第1个样本}},\underbrace{\tilde{\boldsymbol{x}}_{2,1}^{(l)},\cdots,\tilde{\boldsymbol{x}}_{1,B}^{(l)}}_{\text{第2个样本}},\cdots,\underbrace{\tilde{\boldsymbol{x}}_{N,1}^{(l)},\cdots,\tilde{\boldsymbol{x}}_{N,B}^{(l)}}_{\text{第}N\text{个样本}}] \tag{4}$$

其中，B 为从每个训练样本的特征图 $\tilde{X}_i^{(l)}$ 中所能抽取的最大特征块的个数。B 的大小取决于用于分割特征块的滑动窗口在水平方向和垂直方向上的步长(stride)，当在 2 个方向上的步长均为 1 且采用了零填充时，$B = hw$。

(2)零中心化各通道相关式特征块

为了对 $\widetilde{\boldsymbol{X}}^{(l)}$ 进行主成分分析，还需要中心化 $\widetilde{\boldsymbol{X}}^{(l)}$ 中的各通道相关式特征块。令 $\tilde{\boldsymbol{\mu}}^{(l)}$ 为所有 $\tilde{\boldsymbol{x}}_{i,b}^{(l)}(1 \leqslant i \leqslant N, 1 \leqslant b \leqslant B)$ 的均值(也就是 $\widetilde{\boldsymbol{X}}^{(l)}$ 的各列向量的均值)：

$$\tilde{\boldsymbol{\mu}}^{(l)} = \frac{1}{NB}\sum_{i=1}^{N}\cdot\sum_{b=1}^{B}\tilde{\boldsymbol{x}}_{i,b}^{(l)} \tag{5}$$

中心化后的 $\tilde{\boldsymbol{x}}_{i,b}^{(l)}$ 为：$\bar{\tilde{\boldsymbol{x}}}_{i,b}^{(l)} = \tilde{\boldsymbol{x}}_{i,b}^{(l)} - \tilde{\boldsymbol{\mu}}^{(l)}$；相应地，中心化后的通道相关式矩阵 $\widetilde{\boldsymbol{X}}^{(l)}$ 为

$$\bar{\widetilde{\boldsymbol{X}}}^{(l)} = \left[\underbrace{\bar{\tilde{\boldsymbol{x}}}_{1,1}^{(l)},\cdots,\bar{\tilde{\boldsymbol{x}}}_{1,B}^{(l)}}_{\text{第1个样本}},\underbrace{\bar{\tilde{\boldsymbol{x}}}_{2,1}^{(l)},\cdots,\bar{\tilde{\boldsymbol{x}}}_{1,B}^{(l)}}_{\text{第2个样本}},\cdots,\underbrace{\bar{\tilde{\boldsymbol{x}}}_{N,1}^{(l)},\cdots,\bar{\tilde{\boldsymbol{x}}}_{N,B}^{(l)}}_{\text{第}N\text{个样本}}\right] \tag{6}$$

(3)基于奇异值分解计算 $\bar{\widetilde{\boldsymbol{X}}}^{(l)}$ 的 c_{l+1} 个主方向(通道相关式 PCA 滤波器)

从 $\bar{\widetilde{\boldsymbol{X}}}^{(l)}$ 中提取 c_{l+1} 个 PCA 主方向，也就是求解如下目标式：

$$\min_{\widetilde{W}^{(l)}} \| \bar{\widetilde{\boldsymbol{X}}}^{(l)} - \widetilde{\boldsymbol{W}}^{(l)}(\widetilde{\boldsymbol{W}}^{(l)})^{\mathrm{T}}\bar{\widetilde{\boldsymbol{X}}}^{(l)} \|_F^2$$
$$\text{s.t.}\quad (\widetilde{\boldsymbol{W}}^{(l)})^{\mathrm{T}}\widetilde{\boldsymbol{W}}^{(l)} = \boldsymbol{I}_{c_{l+1}\times c_{l+1}} \tag{7}$$

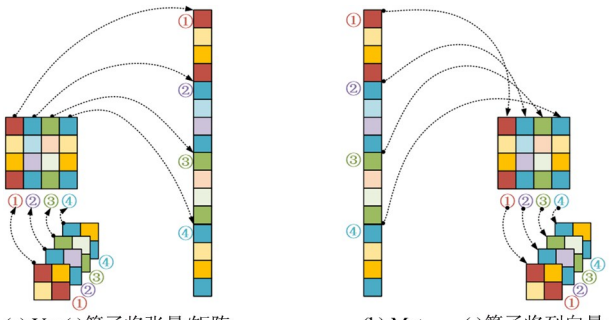
其中，$\boldsymbol{W}^{(l)} \in R^{k^2c_l \times c_{l+1}}$ 为 c_{l+1} 个主方向构成的矩阵，$\boldsymbol{I}_{c_{l+1}\times c_{l+1}}$ 表示大小为 $c_{l+1} \times c_{l+1}$ 的单位矩阵。满足优化式(7)的 $W^{(l)}$ 的解为 $\bar{\widetilde{\boldsymbol{X}}}^{(l)}$ 的协方差矩阵 $\bar{\widetilde{\boldsymbol{X}}}^{(l)}(\bar{\widetilde{\boldsymbol{X}}}^{(l)})^{\mathrm{T}}$ 的前 c_{l+1} 个最大特征值对应的特征向量(主方向)。令 $\widetilde{\boldsymbol{w}}_d^{(l)}$ 表示 $\bar{\widetilde{\boldsymbol{X}}}^{(l)}(\bar{\widetilde{\boldsymbol{X}}}^{(l)})^{\mathrm{T}}$ 的第 d 个主方向，则

$$\widetilde{\boldsymbol{W}}^{(l)} = [\widetilde{\boldsymbol{w}}_1^{(l)},\widetilde{\boldsymbol{w}}_2^{(l)},\cdots,\widetilde{\boldsymbol{w}}_{c_{l+1}}^{(l)}] \tag{8}$$

进一步地，每个主方向 $\widetilde{\boldsymbol{W}}_d^{(l)}$ 可通过从列向量到张量/矩阵的变换算子 $\mathrm{Mat}_{a\times b\times c}(\cdot)$ 被重置为多通道的滤波器核，如式(9)所示。

$$\widetilde{\boldsymbol{w}}_d^{(l)} = \mathrm{Mat}_{k\times k\times c_l}(\widetilde{\boldsymbol{w}}_d^{(l)}) \tag{9}$$

这里，从列向量到张量/矩阵的变换算子 $\mathrm{Mat}_{a\times b\times c}(\cdot)$ 是从张量/矩阵到列向量的变换算子 $\mathrm{Vec}(\cdot)$ 的逆运算。为清晰起见，图 4 给出了 $\mathrm{Vec}(\cdot)$ 算子和 $\mathrm{Mat}_{a\times b\times c}(\cdot)$ 算子的具体示例。



(a) Vec(·)算子将张量/矩阵拉伸为列向量的过程　　(b) $\mathrm{Mat}_{a\times b\times c}(\cdot)$算子将列向量恢复为张量/矩阵的过程

图 4　变换过程示意图

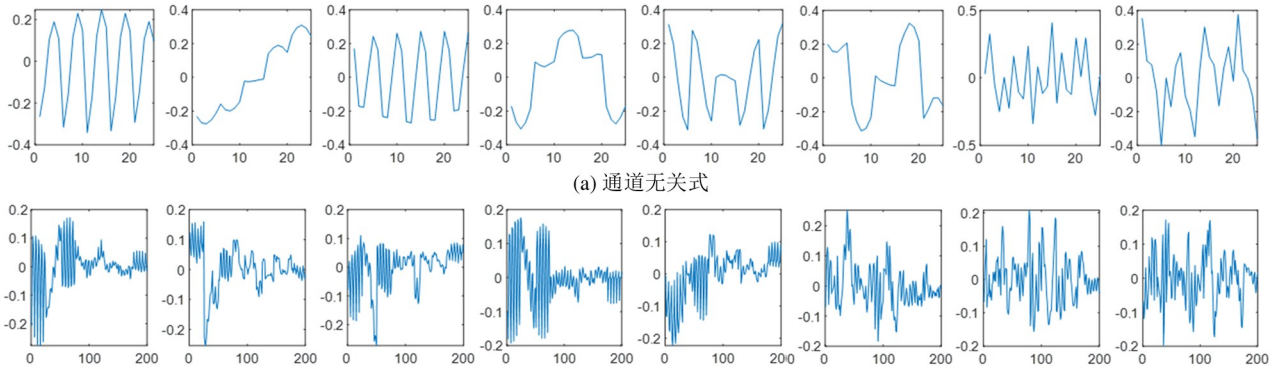
2.3　通道相关式卷积和通道无关式卷积的比较

本文所提出的 DCH-PCANet 是将通道相关式卷积(CDC)和通道无关式卷积(CIC)联合使用，而非使用前者代替后者。这是因为，首先，CIC 在特征图的生成方式上有自身的优势，CIC 以将各滤波器与各输入的特征图相组合的方式生成特征图，此种方式在本质上是以一种比较经济的方式产生尽可能丰富的特征图；其次，CDC 所生成的特征和 CIC 所生成的特征具有各自的优缺点，将两者联合起来可以形成较好的互补性。

CDC 不仅描述通道内部的特征变化，而且描述通道之间的特征变化，因此其所描述的信息更具有全局性，而其所捕捉的特征也更容易受到局部变化的干扰，从而能够刻画较为敏感的“细节”信息。然而，CDC 忽略了不同通道上的特征图在主方向上的差异性；而 CIC 由于只描述通道内部的特征变化，所描述的特征更具有局部性和稳定性(在局部范围内的特征变化不如 CDC 显著)，能够最大限度地从各个通道上提取特征，具有更为明显的方向性，但在各个通道上获取的滤波器较为相似，不利于增强特征的丰富性。

图 5、图 6、图 7 和图 8 分别对 CIC 和 CDC 2 种滤波器的一维展示、二维展示、所生成的特征图、所生成的模式图进行了比较。从图 5和图 6 对 2 种滤

波器的一维展示和二维展示来看,通道相关式 PCA 滤波器比通道无关式 PCA 滤波器具有更大的局部扰动和变化。这导致基于 CDC 所提取的特征图和所生成的模式图都比基于 CIC 所提取的特征图和所生成的模式图具有更为均匀和丰富的特征(如图 7和图 8 所示)。这一点从对高分辨率图像的处理结果可以看得更为清楚。如图 9 所示,CDC 对纹理信息(如围巾的纹理信息)的刻画更为细致。正如李小薪和梁荣华[29]所指出的那样,更为均匀和丰富的特征对于因遮挡和光照变化等造成的特征损失可以进行更好的特征补偿,同样,这样的特征补偿也适用于因分辨率过低而所损失的特征,详见本文实验部分。

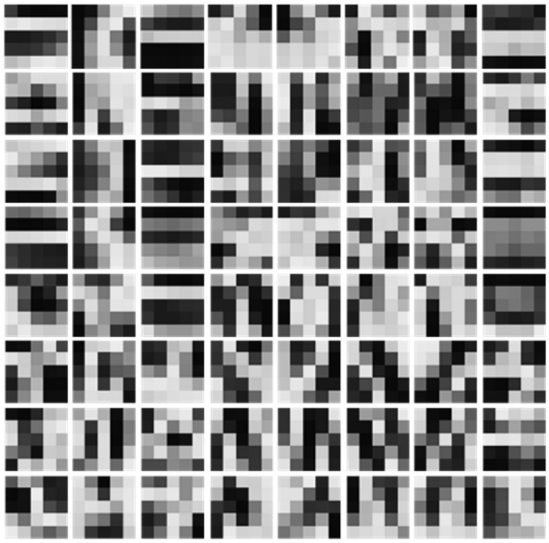


(a) 通道无关式

(b) 通道相关式

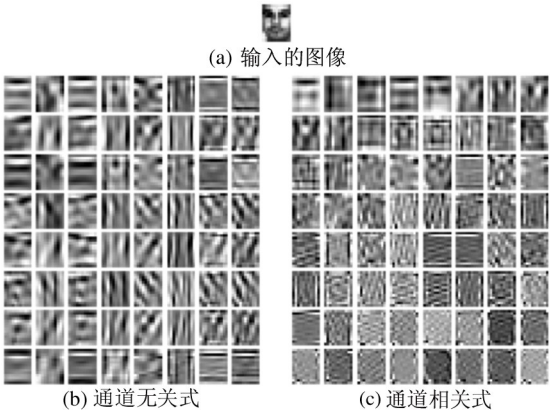
图 5　PCA 滤波器((5×5)@8)的一维展示

(a) 通道无关式



(b) 通道相关式

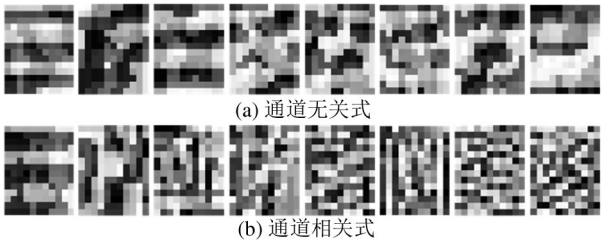
图 6　PCA 滤波器((5×5)@8)的二维展示



(a) 输入的图像

(b) 通道无关式　　(c) 通道相关式

图 7　PCANet 所生成的特征图



(a) 通道无关式

(b) 通道相关式

图 8　PCANet 所生成的模式图

从上面的分析(特别是图 8 和图 9)也可以看出,将 CIC 和 CDC 结合起来可以达到多尺度分析的效果,对于从图像尤其是低分辨率图像中提取丰富的特征以补偿丢失的细节信息具有明显的作用。

2.4　特征图的稠密连接

李小薪和梁荣华[29]指出,PCANet 的各卷积层在本质上是对输入图像进行了多尺度(不同卷积层

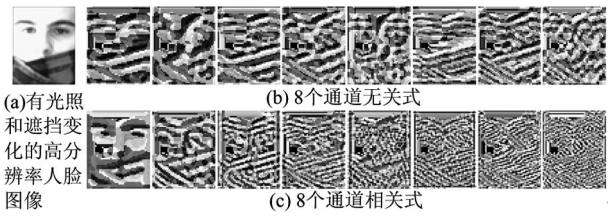


图 9　PCANet 所生成的模式图

的感受野不同)和多方向(在不同的主方向上的投影)分解。然而,PCANet 丢弃了低层卷积层(较小尺度)的特征图输出,只使用了最后一个卷积层的特征图。对于有偏移图像的识别而言,低层卷积层产生的特征图往往比高层卷积层产生的特征图更具有判别性,这是因为高层卷积层由于具有较大的感受野,更容易将偏移所导致的局部坏特征在更大的范围内传播到邻域特征点。

为了充分利用 PCANet 的多尺度多方向特征,本文将稠密连接[27,30]引入其中,如图 1 所示。这里,稠密连接通过特征拼接的方式实现,即将浅层特征图与深层特征图通过拼接的方式连接在一起。值得注意的是,本文在 PCANet 中加入的“稠密连接”与 DenseNet[27]中所使用的“稠密连接”有如下几点不同。

首先,稠密连接的范围不同。为了防止高层网络参数规模的指数级增长,DenseNet 中的稠密连接局限于单个的稠密块(dense block)中;而本文所使用的稠密连接局限于同一类型的卷积块中。

其次,稠密程度不同。DenseNet 中的稠密连接将所有的低层特征图与其后所有的高层特征图进行拼接,因此,在一个 L 层的稠密块中将发生 $L(L+1)/2$ 次连接;而本文所使用的稠密连接仅将低层特征指向其后紧邻的下一层特征和最终的特征拼接层,也就是除了最后一个特征块,每个特征图只有 2 个后向连接,因此,一个 L 层的卷积块中只有 $2L-1$ 次连接。容易看出,$L(L+1)/2 \geqslant 2L-1$,且当 $L>2$ 时,$L(L+1)/2>2L-1$。之所以出现这样的连接稠密性差异主要是由于 DenseNet 所采用的稠密连接更适用于面向多通道特征图的通道相关式卷积,而在 PCANet 中只使用了通道无关式卷积,由于没有用到通道方向上的特征融合,因此无需将低层特征图与其后续的所有高层特征图拼接。另外,PCANet 通过卷积将蕴含在像素域中的丰富特征逐渐在各个主方向上分离出来,形成更具有判别性的特征空间。然而,如果将第 $l-1$ 次和第 l 次的主方向分解的结果混合起来进行第 $l+1$ 次主方向分解,就会将已经分离的特征再次混合起来(此时所用到的主方向会与第 l 次分解所用到的主方向混叠在一起),这样就不利于特征的充分分解。

最后,主要作用不同。正如 Huang 等人[27]所指出的,在 DenseNet 中引入稠密连接的作用主要有 3 个:(1)可以极大地减少网络的参数;(2)特征重用;(3)强化特征在各网络层之间的传递,从而可以有效地消除梯度消失问题。然而,本文在 PCANet 中引入的稠密连接主要是为了充分利用低层卷积层的特征图以增强特征的多样性。

2.5　具体算法

算法 1 对本文所提出的基于稠密连接的通道混合式 PCANet 的特征提取过程进行了详细总结。

算法 1　DCH-PCANet 的特征提取算法

输入:待提取特征的人脸图像 $\boldsymbol{X}^{(0)}=\boldsymbol{x}/B_{h\times w}(\boldsymbol{y})\in R^{h\times w\times c}$,通道无关式滤波器组 $\{\boldsymbol{W}^{(l)}\in R^{k^2\times c_{l+1}}\}_{l=0}^{L-1}$ 及其均值向量 $\{\boldsymbol{\mu}^{(l)}\in R^{k^2}\}_{l=0}^{L-1}$ 和通道相关式滤波器组 $\{\widetilde{\boldsymbol{W}}^{(l)}\in R^{k^2c_l\times c_{l+1}}\}_{l=0}^{L-1}$ 及其均值向量 $\{\widetilde{\boldsymbol{\mu}}^{(l)}\in R^{k^2c_l}\}_{l=0}^{L-1}$;

1. **输出**:$\boldsymbol{X}^{(0)}$ 的 DCH-PCANet 特征 H 初始化:令 $l=0$,$\widetilde{\boldsymbol{X}}^{(0)}=\boldsymbol{X}^{(0)}$;
2. 从 $X^{(l)}$ 和 $\widetilde{X}^{(l)}$ 中分别提取通道无关式特征块 $\boldsymbol{X}^{(l)}\in R^{k^2\times c_lB}$ 和通道相关式特征块 $\widetilde{X}^{(l)}\in R^{k^2c_l\times B}$,且 $B=hw$(采用步长为 1 提取特征块);
3. 中心化 $\boldsymbol{X}^{(l)}$ 和 $\widetilde{\boldsymbol{X}}^{(l)}$:$\overline{\boldsymbol{X}}^{(l)}=\boldsymbol{X}^{(l)}-\mu^{(l)}\times\mathbf{1}$,$\overline{\widetilde{\boldsymbol{X}}}^{(l)}=\widetilde{\boldsymbol{X}}^{(l)}-\widetilde{\boldsymbol{\mu}}^{(l)}\times\mathbf{1}$;这里,$\mathbf{1}$ 为全 1 的行向量:$\mathbf{1}=[1,1,\cdots,1]\in R^{1\times B}$;
4. 将 $\overline{\boldsymbol{X}}^{(l)}$ 和 $\overline{\widetilde{\boldsymbol{X}}}^{(l)}$ 分别投影到 $\boldsymbol{W}^{(l)}$ 和 $\widetilde{\boldsymbol{W}}^{(l)}$:$\widehat{\boldsymbol{X}}^{(l+1)}=(\overline{\boldsymbol{X}}^{(l)})^H\boldsymbol{W}^{(l)}\in R^{c_lB\times c_{l+1}}$,$\widehat{\widetilde{\boldsymbol{X}}}^{(l+1)}=(\overline{\widetilde{\boldsymbol{X}}}^{(l)})^H\widetilde{\boldsymbol{W}}^{(l)}\in R^{B\times c_{l+1}}$;
5. 将 $\widehat{\boldsymbol{X}}^{(l+1)}$ 中的元素重新组织为特征图 $\boldsymbol{X}^{(l+1)}$:$\boldsymbol{X}^{(l+1)}=Cat(\{\breve{\boldsymbol{x}}_{i,j}^{(l+1)}\}_{i=1,\cdots,c_l;\,j=1,\cdots,c_{l+1}})$,其中,$\breve{\boldsymbol{x}}_{i,j}^{(l+1)}=\mathrm{Mat}_{h\times w}(\widehat{X}_{r+1:r+B,\,j}^{(l+1)})\in R^{h\times w}$,$r=(i-1)B$;
6. 将 $\widehat{\widetilde{\boldsymbol{X}}}^{(l+1)}$ 中的元素重新组织为特征图:$\widetilde{\boldsymbol{X}}^{(l+1)}=Cat(\{\breve{\widetilde{\boldsymbol{x}}}_j^{(l+1)}\}_{j=1,\cdots,c_{l+1}})$,其中,$\breve{\widetilde{\boldsymbol{x}}}_j^{(l+1)}=\mathrm{Mat}_{h\times w}(\widehat{\widetilde{X}}_{:,j}^{(l+1)})\in R^{h\times w}$;

7. 令 $l = l + 1$, 执行上述步骤 2 ~ 6,直至 $l = L$, 这里 L 表示预先给定的最大卷积层数;
8. 将上面步骤中所生成的特征图全部拼接起来,形成新的特征图集 Y:

$$\boldsymbol{Y} = Cat(X^{(1)}, X^{(2)}, \cdots, X^{(L)}, \tilde{X}^{(1)}, \tilde{X}^{(2)}, \cdots, \tilde{X}^{(L)})$$

9. 对特征图 Y 进行模式图编码,得到模式图 $\boldsymbol{P}$:$\boldsymbol{P} = Cat(\{\boldsymbol{p}_\gamma\}_{\gamma=1,\cdots,\Gamma})$, 其中, 第 γ 个模式图 $\boldsymbol{p}_\gamma = \sum_{t=1}^{T} 2^{t-1}\mathrm{USF}(\boldsymbol{Y}_{\cdot,\cdot,(\gamma-1)T+t})$, $\Gamma = \lfloor N(Y)/T \rfloor$, $\boldsymbol{Y}_{\cdot,\cdot,z}$ 表示 Y 在第 z 个通道上的特征图, T 表示参与单个模式图编码的通道数, $N(Y)$ 表示 Y 在通道方向上的维度,USF(·)表示单位阶跃函数(unit step function, USF),将输入数值通过与 0 比较进行二值化,即:

$$\mathrm{USF}(x) = \begin{cases} 1 & x \geqslant 0 \\ 0 & x < 0 \end{cases};$$

10. 从模式图 P 提取柱状图特征 $\boldsymbol{H}$:$\boldsymbol{H} = [H_1^{\mathrm{T}}, \cdots, H_\Gamma^{\mathrm{T}}]^{\mathrm{T}}$, 其中, $\boldsymbol{H}_\gamma = Q\mathrm{hist}(\boldsymbol{p}_\gamma)$, $Q\mathrm{hist}(\boldsymbol{p}_\gamma)$ 表示将模式图 $\boldsymbol{p}_\gamma$ 划分为 Q 块,从每一块中提取柱状图,每个柱状图使用 2^T 个分组,也就是,统计模式图的编码值在每一特征块的 2^T 个分组中出现的频率。

3　实　验

3.1　实验设置

数据集。为了验证本文方法的有效性和通用性,针对受控和非受控 2 种数据采集环境,模拟低分辨和真实低分辨率 2 种低分辨率情形,使用含有真实遮挡的人脸图像数据集进行测试。首先,使用 AR 人脸数据集[31],在受控环境、真实遮挡和模拟低分辨率的情景下进行识别性能评估,详见 3.2 节;3.3 节使用 MFR2[32] 和 PKU-Masked-Face[33] 2 个数据集,在非受控环境、真实遮挡和模拟低分辨率的情景下进行识别性能评估;3.4 节使用作者自己采集的人脸数据库,在非受控环境、真实遮挡和真实低分辨率的情景下进行识别性能评估。另外,为了对网络模型进行训练,使用 WebFace[34] 作为训练集。WebFace 是一个大规模的人脸数据集,由来自10 575 个人的 494 414 张人脸图像组成。由于 WebFace 中大部分图像不含遮挡,从 WebFace 中随机选取 50% 张不含遮挡的人脸图像,使用遮挡生成工具[32]对其添加模拟遮挡,并将其与 WebFace 中剩下的不含遮挡的人脸图像混合在一起,形成一个新的混合了模拟遮挡的人脸数据集,称为“WebFace-Occ”。

与前沿方法的对比。为了评估本文方法的有效性,针对受控环境和非受控环境下的数据集,采用了不同的对比方法。针对受控环境数据集,主要对比了传统的机器学习方法基于核范数的矩阵回归(nuclear norm based matrix regression, NMR)[35] 以及 2 个有代表性的卷积神经网络模型,即 VGG-Face[36] 和 LCNN(Lightened CNN)[37]。针对非受控环境数据集,与前沿的 CNN 模型进行对比。首先,以 ResNet28 作为主干网络,训练了 2 个 CNN 模型 ResNet28 和 ResNet28,其中,ResNet28 在原始的 WebFace 数据集上训练,而 ResNet28 在 WebFace-Occ 数据集上训练,并使用了 ArcFace 损失[38]。另外,选取 FaceNet[2]、InterpretFR[39]、LPD[33]、ECBAM[40] 作为前沿的对比方法。这里,InterpretFR 和 LPD 均提供了公开可获取的预训练模型。ECBAM 主要基于 CBAM[41],本文基于 CBAM 的官方代码实现了 ECBAM。为了使本文所提出的方法能更好地适用于非受控环境,将 PCANet 和本文提出的 DCH-PCANet 追加在训练好的 ECBAM 模型的最后一个卷积层之后,并分别用 + PCANet 和 + DCH-PCANet 进行表示。对不同类型的特征采用不同的分类器:由于 PCANet 最终生成的是柱状图特征,采用基于卡方距离[26]的最近邻分类器;对其他方法统一采用基于协同表示的分类器[42]。

所有实验在 Win10 64 位操作系统下进行部署,主要硬件配置为:8 GB 内存(RAM),Intel i7-4600 双核处理器(主频分别为 2.10 GHz 和 2.69 GHz)。所有方法均使用 Python 语言实现,并使用 PyTorch 作为深度学习框架。

3.2　受控环境下在模拟低分辨率及有实际遮挡的数据集上的识别性能

本节使用受控环境下采集的 AR 人脸数据集[31]来验证本文提出的 DCH-PCANet 的有效性。为探究稠密连接对 PCANet 识别性能的影响,使用 DPCANet(dense PCANet)表示在原始的 PCANet 上只施加了稠密连接的 PCANet;为探究通道相关式卷积(CDC)对 DCH-PCANet 的性能的影响,本节考虑 DCH-PCANet 的 3 个版本,即 DCH-PCANet-I、DCH-PCANet-II 和 DCH-PCANet-III。3 个版本的 DCH-

PCANet 均采用两层卷积,且所采用的通道无关式卷积核的个数均为 8(卷积层 1) + 8(卷积层 2),它们的区别主要体现在对通道相关式卷积的使用上:DCH-PCANet-I 所采用的通道相关式卷积核的个数分别为 8 和 24,DCH-PCANet-II 所采用的通道相关式卷积核的个数分别为 8 和 32,DCH-PCANet-III 所采用的通道相关式卷积核的个数分别为 8 和 40。

针对有实际遮挡和极端光照变化的情形,测试 DCH- PCANet 的识别性能。从 AR 库中选取 119 人(男性 65 人和女性 54 人)的人脸图像作为实验数据。训练集选取 119 × 8 = 952 张不含遮挡和光照变化的人脸图像,如图 10(d)所示;测试集 I 选取 119 × 3 = 357 张有左侧、右侧和正面等 3 种强光变化的人脸图像,如图 10(a)所示;测试集 II 选取 357 张有太阳镜遮挡且分别包含了 3 种强光变化的人脸图像,如图 10(b)所示;测试集 III 选取 357 张有围巾遮挡且混合了 3 种强光变化的人脸图像,如图 10(c)所示。AR 库中的强光光照与遮挡类似,造成了大面积的判别特征的损失,具有更大的识别难度。本文重点关注低分辨率情境下的识别性能,因此,所有图像首先被裁剪和对齐为 56 × 46 的像素矩阵,在此基础上,对 3 个测试集中的图像进一步下采样,令其分辨率为 14 × 11 = 154;为了测试各方法的识别性能随着训练图像分辨率的变化情况,对所有训练图像进行了 4 倍、3 倍、2 倍下采样,分别对应于 14 × 11 = 154、18 × 15 = 270 和 28 × 23 = 644 等像素维数。

图 10　来自 AR 人脸数据库的低分辨率测试集和高分辨率训练集样本

表 1 比较了 DCH-PCANet 的 3 个版本(DCH-PCANet-I ~ III)与现有方法(NMR[35]、VGG-Face[36]、LCNN[37]、PCANet[26]、DPCANet-III)在 AR 的 3 个测试子集上的识别率。可以看出,DCH-PCANet-I ~ III 都表现出了比 PCANet 更优的识别性能,尤其是当待识别图像的分辨率较低的时候,这一优势更为显著;另外,也可看出,从 DCH-PCANet-I 到 DCH-PCANet-III,随着特征维度的提升,各版本的 DCH-PCANet 都能够更好地利用训练集的高分辨率信息来提升低分辨率测试集的识别性能。从表 1 的实验结果也可以看出,与现有各方法相比,对于实物遮挡(测试集 II 和 III),当分辨率较低时,LCNN 具有较好的识别性能,然而,对于强光光照所引发的遮挡(测试集 I),LCNN 在各分辨率下的识别性能都远弱于其他方法。而 VGG-Face 的识别性能正好相反,对于实物遮挡,不管是对高分辨率的训练集图像还是对低分辨率的训练集图像,其识别性能几乎都是最低的;而对于强光光照遮挡,其识别性能与 NMR 和 PCANet 几乎相当。造成这一现象的原因,在很大程度上与网络结构和所使用的训练集图像有关。而本文提出的 DCH-PCANet 模型较好地解决了这一问题,在各个维度下和各种遮挡条件下,DCH-PCANet 都能取得较为稳定且优越的识别性能。

为什么 DCH-PCANet 在低分辨率情形下具有更好的识别性能?究其原因,主要是 DCH-PCANet 所采用的通道相关式卷积(CDC)分支对通道无关式卷积(CIC)分支所提取的特征起到了较好的补充作用。图 11 对此进行了说明。通过对比图 11(b)和(c)可以发现,通道无关式卷积由于没有考虑各特征图在通道方向上的相关性及相互抑制作用,可以更好地凸显局部纹理特征的作用,这一点对低分辨率图像的特征提取尤为重要,但同样的,通道无关式卷积也会强化遮挡区域的特征;而通道相关式卷积由于充分考虑了特征图在各通道方向上的相关性,从而可以较好地抑制坏特征的作用,从图 11(c)可以看出,通道相关式卷积的特征响应更为稀疏。因

表 1　在 AR 人脸库的 3 个测试集上的识别率(%)比较

像素维度	测试集 I			测试集 II			测试集 III		
	154	270	644	154	270	644	154	270	644
NMR	48.50	50.41	53.27	23.41	27.32	30.79	33.49	36.58	37.14
VGG-Face	43.25	47.63	50.72	5.32	7.51	9.58	9.52	10.82	12.79
LCNN	10.36	12.72	15.23	21.57	22.43	26.69	42.30	45.13	47.78
PCANet	44.82	48.37	49.18	19.61	26.25	31.76	31.37	37.22	39.50
DPCANet	54.00	57.70	59.20	30.50	32.20	35.30	39.20	42.91	47.40
DCH-PCANet-I	56.22	58.19	62.07	24.65	29.02	33.51	39.69	41.98	42.87
DCH-PCANet-II	58.18	61.75	63.07	24.70	31.54	35.35	40.81	42.15	45.27
DCH-PCANet-III	59.86	62.85	67.63	25.49	32.98	36.19	43.89	46.06	49.11

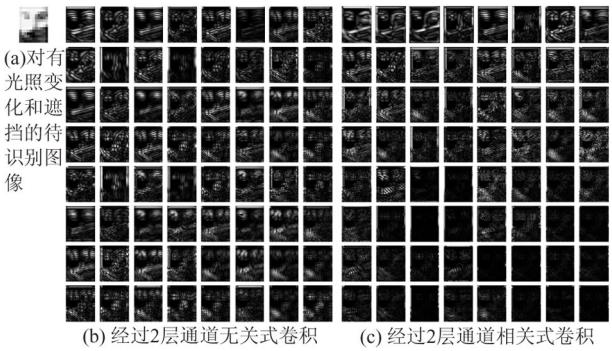


图 11　DCH-PCANet 所生成的特征图的模值对比

此,通道相关式卷积和通道无关式卷积可以相互补偿,从而有效提取低分辨率图像的特征,达到更优的识别性能。

3.3　非受控环境下在模拟低分辨及真实遮挡数据上的识别性能比较

本节及下一节评估本文方法在非受控环境下对低分辨率及有真实遮挡的人脸图像的识别性能。本节针对 MFR2[32] 和 PKU-Masked-Face[33] 2 个有真实遮挡的人脸数据集,测试不同方法在不同分辨率下的识别性能。在 U-O 和 H-L 情境下测试各模型的性能。这里,U-O 指的是:相比对的 2 幅图像中,库存人脸图像(也就是在人脸识别系统中事先注册的人脸图像)不含遮挡(unoccluded),待比对的人脸图像含有遮挡(occluded);H-L 指的是:相比对的 2 幅图像中,库存人脸图像为高清的(high resolution),待比对的人脸图像是低分辨率的(low resolution)。也就是说,这里限定库存图像是无遮挡的高清图像,待比对图像是低分辨率的有遮挡图像。为此,首先按照 U-O 协议分别在 MFR2 和 PKU-Masked-Face 上选取样本对。在 MFR2 上,随机选取 800 个图像对,包括 400 个正样本对和 400 个负样本对;在 PKU-Masked-Face 上,随机选取 1 890 个图像对,包括 945 个正样本对和 945 个负样本对。按照 U-O 协议,所选取的样本对中,有一张是不含遮挡的,将其作为库存样本,有一张是含有遮挡的,将其作为待比对样本。并且将所有选出的人脸图像裁剪、缩放为 240×240 大小,然后再将待比对图像进行 1 倍、2 倍、4 倍和 6 倍下采样,分别得到分辨率依次为 240×240、120×120、60×60 和 40×40 的待比对的人脸图像。图 12 给出了本节实验所用到的来自于 MFR2(第 1 行)和 PKU-Masked-Face(第 2 行)的各分辨率下的有实际遮挡的人脸图像样本。

图 12　本节实验所用到的非受控环境下的来自于 MFR2(第 1 行)和 PKU-Masked-Face(第 2 行)的模拟低分辨率和实际遮挡的人脸图像样本

表 2 报告了各对比方法在本节实验环境下的识别性能。可以看出,随着待比对图像的分辨率的下

降,各对比方法的识别性能都有了显著下降,从 240 ×240 到 40 ×40,分辨率下降了 6 倍,而识别率下降了将近 50% ,几乎呈指数级下降。产生这一现象的一个主要原因是待比对图像和库存图像的分辨率严重不一致所致。由于所比对的各网络模型都要求输入图像的大小是固定的 240 ×240,所以,所有图像,不论其分辨率大小,都被重新调整到了 240 ×240 大小,再进行匹配识别。而低分辨率图像由于要进行像素插值,可能会引入人工痕迹,这会导致输入图像在进入网络时的特征就有了一定程度的破坏,而这种人工痕迹会随着输入图像分辨率的降低而越发严重,从而导致识别率显著下降。显然,通过对输入图像进行有效的超分辨率重建可能是提升低分辨率人脸图像识别性能的最直接有效的方法。然而,由于本文重在构建鲁棒的人脸识别模型,因此,如何提升输入图像的分辨率将是进一步的研究课题。尽管如此,从表 2 的最后 3 行来看,本文提出的方法 DCH-PCANet 可以作为现有模型的有效补充,并稳定提升其识别性能,尽管这种提升并非是数量级上的提升。

表 2　在 MFR2 和 PKU-Masked-Face 2 个真实遮挡人脸数据集上的最大识别率(%)

模型	MFR2 (U-O)				PKU-Masked-Face (U-O)			
	240^2	120^2	60^2	40^2	240^2	120^2	60^2	40^2
ResNet28	91.14	89.74	62.15	42.85	88.37	83.87	63.35	40.18
ResNet28	90.63	89.96	63.90	43.50	90.00	87.61	64.38	42.16
FaceNet	92.01	89.50	62.20	44.17	89.58	87.71	66.38	40.12
InterpretFR	90.63	88.43	61.36	41.20	92.85	83.42	62.22	39.15
LPD	88.76	87.54	63.92	43.67	91.96	84.13	64.59	39.19
ECBAM	92.38	89.56	63.15	45.78	93.97	87.33	62.28	41.46
+ PCANet	92.42	89.62	64.51	46.74	94.15	87.73	64.03	43.37
+ DCH-PCANet	**93.56**	**89.92**	**65.56**	**46.75**	**94.52**	**89.20**	**67.60**	**46.85**

3.4　非受控环境下在真实低分辨及真实遮挡数据上的识别性能比较

本节进一步在非受控环境下测试本文方法对真实低分辨、真实遮挡的人脸图像的识别性能。3.2 节和 3.3 节通过对高清人脸图像进行下采样得到低分辨率人脸图像。然而,在现实世界中,低分辨率图像的产生可能是一个复杂的过程,例如,可能是由于光照条件差、人体运动、相机抖动等众多因素造成的。因此,对真实场景中产生的低分辨率人脸图像进行识别往往具有更大的挑战性。

为了测试现有方法对真实场景中的低分辨率人脸图像的识别性能,本文自行采集了一个小规模的低分辨率人脸图像数据集。该数据集包括 80 个参与者(30 个女生,50 个男生)的人脸图像。首先,由参与者自行提供一张高清的人脸图像作为库存图像;其次,使用 MI 8 SE 手机对参与者进行拍摄,让参与者随意行走在光线并不充足的走廊上,在距离参与者大约 20 m 时进行拍摄,每个参与者来回走动 5 次,拍摄 10 张照片。对采集到的 80 ×10 张人脸图像,采用 MTCNN[43] 检测和对齐人脸图像,将每张人脸图像裁剪为 86 ×90 大小,从每个人的 10 张图像中挑选 5 张照片进行识别,如图 13 所示。

表 3 报告了相关方法的识别结果。可以看出,在真实的低分辨率人脸图像数据集上,各方法的识别率都在 60.00% 以下,本文提出的方法达到了最高 57.68% 的识别率。显然,在现实环境下,低分辨的人脸图像对现有的人脸识别方法构成了极大的挑战。正如本文开篇所指出的,相对于其他生物识别技术,人脸识别最大的优势在于其更自然,不需要被识别对象的配合。然而,从表 3 报告的识别结果来看,现有方法对于真正达到这一目标还有不小的距离。

本文也对这些人脸图像进行了人工识别测试:选取 5 个受试者,任意选取 10 个参与人的库存照片

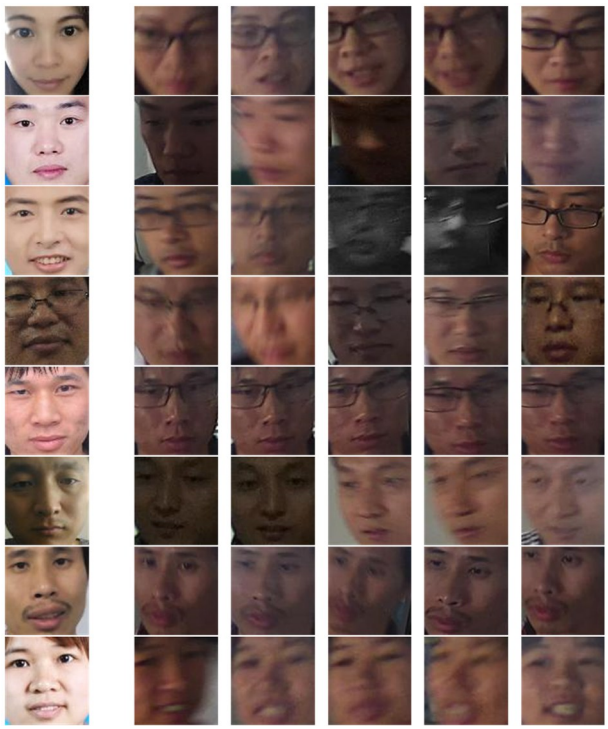


图 13　本节实验所用到的非受控环境下的真实低分辨率和真实遮挡人脸数据集的样本示例

表 3　在自建的低分辨率人脸数据集上的识别率

模型	识别率/%
ResNet28	52.34
ResNet28	50.52
FaceNet	52.73
InterpretFR	52.35
LPD	48.67
ECBAM	54.98
+ PCANet	55.32
+ DCH-PCANet	**57.68**

和 5 张低分辨照片,让受试者指出低分辨率照片与哪张高清库存图片匹配。实验结果表明,人工识别的准确率一般在 80% 以上。可见,尽管现有的研究工作已经很多次声称所构建的模型超过了人眼的分类准确度,但这并不代表在所有情况下,人工训练的模型都能超过人眼。深度学习模型与人类视觉机制的内在联系还有待于进一步探讨。

4　结　论

针对低分辨率下有遮挡人脸识别问题,本文提出了基于稠密连接的混合通道 PCANet(DCH-PCANet)模型。实验结果表明,DCH-PCANet 能够在一定程度上弥补图像分辨率不足和遮挡造成的信息损失,是现有的深度网络模型的有益补充。本文对低分辨率图像只是进行了简单的双三次插值处理,而简单的插值处理往往会在低分辨图像中引入很多人工痕迹,从而破坏输入图像的局部判别信息。如何在保留判别性的同时提升低分辨率图像的分辨率是未来进一步探讨的课题。本文研究也表明人眼不仅有很好的遮挡鲁棒性,也有很好的“低分辨率鲁棒性”,即使不加训练(无须熟悉待识别图像),也可以在同时含有遮挡和低分辨率的人脸图像上达到很好的泛化能力。探讨深度学习模型与人类视觉机制的内在联系也将是进一步的研究目标。

参考文献

[1] 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, USA: IEEE, 2014:1701-1708.

[2] 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, USA: IEEE, 2015:815-823.

[3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016:770-778.

[4] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification[C]//Proceedings of the International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:1026-1034.

[5] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//Proceedings of the International Conference on Machine Learning. Lille, France: ACM, 2015:448-456.

[6] QUIONERO-CANDELA J, SUGIYAMA M, SCHWAIGHOFER A, et al. Dataset shift in machine learning[M]. Massachusetts: MIT Press, 2009.

[7] SHIMODAIRA H. Improving predictive inference under covariate shift by weighting the log-likelihood function [J]. Journal of Statistical Planning and Inference, 2000, 90(2):227-244.

[8] 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.

[9] WRIGHT J, YANG A, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009,31(2):210-227.

[10] ZHOU 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, USA: IEEE, 2009:1050-1057.

[11] 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.

[12] 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, USA: IEEE Press, 2017:3795-3799.

[13] 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, USA: IEEE, 2019:773-782.

[14] 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.

[15] 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.

[16] 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.

[17] ZHANG W, SHAN S, GAO W, et al. Local Gabor binary pattern histogram sequence: a novel non-statistical model for face representation and recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. Beijing, China:IEEE, 2005:786-791.

[18] AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: application to face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006,28(12):2037-2041.

[19] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[EB/OL]. (2013-12-21)[2023-09-01]. https://arxiv.org/pdf/1312.6199.pdf.

[20] NGOC-SON V, CAPLIER A. Enhanced patterns of oriented edge magnitudes for face recognition and image matching[J]. IEEE Transactions on Image Processing, 2012, 21(3):1352-1365.

[21] NGOC-SON V. Exploring patterns of gradient orientations and magnitudes for face recognition[J]. IEEE Transactions on Information Forensics and Security, 2013,8(2): 295-304.

[22] AHONEN T, HADID A, PIETIKÄINEN M. Face recognition with local binary patterns[C]//Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2004:469-481.

[23] ZHANG B, GAO Y, ZHAO S, et al. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor[J]. IEEE Transactions on Image Processing, 2009,19(2):533-544.

[24] YANG M, ZHANG L, SHIU S C-K, et al. Monogenic binary coding: an efficient local feature extraction approach to face recognition[J]. IEEE Transactions on Information Forensics and Security, 2012,7(6):1738-1751.

[25] YANG M, ZHANG L, SHIU S C-K, et al. Robust kernel representation with statistical local features for face recognition[J]. IEEE Transactions on Neural Networks and Learning Systems, 2013,24(6):900-912.

[26] 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.

[27] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017:4700-4708.

[28] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE, 2017:136-144.

[29] 李小薪, 梁荣华. 有遮挡人脸识别综述: 从子空间回归到深度学习[J]. 计算机学报, 2018,41(1):177-207.

[30] TAI Y, YANG J, LIU X, et al. Memnet: a persistent memory network for image restoration[C]//Proceedings of the IEEE International Conference on Computer Vision. Honolulu, USA: IEEE, 2017:4539-4547.

[31] MARTINEZ A M. The AR Face Database: CVC[EB/OL], [2023-10-01]. https://www.researchgate.net/profile/Robert-Benavente/publication/243651904_The_AR_face_database/links/5486f00f0cf289302e2ea746/The-AR-face-database.pdf.

[32] ANWAR A, RAYCHOWDHURY A. Masked face recognition for secure authentication[EB/OL]. (2020-08-25)[2023-09-01]. https://arxiv.org/pdf/2008.11104.pdf.

[33] DING F, PENG P, HUANG Y, et al. Masked face recognition with latent part detection[C]//Proceedings of

the ACM International Conference on Multimedia. Seattle, USA: ACM, 2020:2281-2289.

[34] YI D, LEI Z, LIAO S, et al. Learning face representation from scratch [EB/OL]. (2014-11-28) [2023-09-01]. https://arxiv.org/pdf/1411.7923.pdf.

[35] YANG J, LUO L, QIAN J, et al. Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(1):156-171.

[36] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep face recognition[C]//Proceedings of the British Machine Vision Conference. Durham, UK: BMVA, 2015: 6.

[37] WU X, HE R, SUN Z. A lightened CNN for deep face representation [EB/OL]. (2015-11-09) [2023-09-01]. https://arxiv.org/pdf/1511.02683.pdf.

[38] 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, USA: IEEE, 2019: 4690-4699.

[39] 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, USA: IEEE, 2019:9348-9357.

[40] 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.

[41] WOO S, PARK J, LEE J-Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018:3-19.

[42] ZHANG L, YANG M, FENG X. Sparse representation or collaborative representation: which helps face recognition? [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Barcelona, Spain: IEEE, 2011:471-478.

[43] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10):1499-1503.

Dense channel-hybrid PCANet for low-resolution and occluded face recognition

QIN E*, HE Jiayao*, LIU Yinwei*, ZHU Yani**, LI Xiaoxin*

(*College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023)

(**School of Computer Science, Hanzhou Dianzi University, Hangzhou 310018)

Abstract

A dense channel-hybrid PCANet (DCH-PCANet) is proposed to recognize low-resolution and occluded face images. Only channel-independent convolutions (CIC) are used in the convolutional layer of the existing principal component analysis network(PCANet) model. Since CIC does not consider the correlation of the feature maps in the channel direction, it can better highlight the local texture features of a single feature map, which is of great significance for compensating the feature loss caused by low resolution and occlusion. However, CIC will also strengthen the occlusion features, hence enlarging the influence range of bad features. The channel-dependent convolution (CDC) fully considers the correlation of all feature maps in the channel direction, which can better suppress the effect of bad features and form a sparse feature map. A CDC-based feature-map extraction branch is added to PCANet to form a channel-hybrid PCANet. And dense connections are also introduced to make full use of low-level features to improve the robustness of occluded image recognition. Experiments are conducted on the following four datasets: AR face dataset, where face images with real occlusions and simulated low-resolutions are acquired in controlled environment; MFR2 and PKU-Masked-Face, where face images with real occlusions and simulated low-resolutions are acquired in uncontrolled environment; our own dataset, where face images with real occlusion and real low-resolution are acquired in uncontrolled environment. Experimental results show that compared with the existing methods, the proposed DCH-PCANet has better occlusion and low-resolution robustness, which can be used as an effective supplement to the cutting-edge methods to improve their recognition performance.

Key words: face recognition with occlusion, principal component analysis network(PCANet), channel-dependent convolution (CDC), dense connection