JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2022
04
VOL.27

ISSN1006-8961
CN11-3758/TB

口罩人脸姿态分类 P1125

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第27卷第4期（总第312期）
2022年4月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@aircas.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

2022年4月 第27卷 第4期（总第312期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

残差密集结构的东巴画渐进式重建(第1084页)

结合双模板融合与孪生网络的鲁棒视觉目标跟踪(第1191页)

注意力机制的曲面沉浸式投影系统补偿(第1238页)

Volume 27, Number 4
Published April 16, 2022

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Gradual model reconstruction of Dongba painting based on residual dense structure(P1084)

Double template fusion based siamese network for robust visual object tracking(P1191)

An attention mechanism based inter-reflection compensation network for immersive projection system(P1238)

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2022)04-1125-12

论文引用格式:Chen S Q, Liu W B and Zhang G. 2022. A embedded dual-scale separable CBAM model for masked human face poses classification. Journal of Image and Graphics,27(04):1125-1136(陈森楸,刘文波,张弓. 2022. 嵌入双尺度分离式卷积块注意力模块的口罩人脸姿态分类. 中国图象图形学报,27(04):1125-1136)[DOI:10.11834/jig.200736]

嵌入双尺度分离式卷积块注意力模块的口罩人脸姿态分类

陈森楸[1,2*],刘文波[1,2],张弓[3]

1. 南京航空航天大学自动化学院,南京　211106; 2. 高速载运设施的无损检测监控技术工业和信息化部重点实验室,南京　211106;
3. 南京航空航天大学电子信息工程学院,南京　211106

摘　要:**目的**　针对口罩遮挡的人脸姿态分类新需求,为了提高基于卷积神经网络的人脸姿态分类效率和准确率,提出了一个轻量级卷积神经网络用于口罩人脸姿态分类。**方法**　本文设计的轻量级卷积神经网络的核心为双尺度可分离注意力卷积单元。该卷积单元由3 × 3和5 × 5两个尺度的深度可分离卷积并联而成,并且将卷积块注意力模块(convolutional block attention module,CBAM)的空间注意力模块(spatial attention module,SAM)和通道注意力模块(channel attention module,CAM)分别嵌入深度(depthwise,DW)卷积和点(pointwise,PW)卷积中,针对性地对DW卷积及PW卷积的特征图进行调整。同时对SAM模块补充1 × 1的点卷积挤压结果增强其对空间信息的利用,形成更加有效的注意力图。在保证模型性能的前提下,控制构建网络的卷积单元通道数和单元数,并丢弃全连接层,采用卷积层替代,进一步轻量化网络模型。**结果**　实验结果表明,本文模型的准确率较未改进SAM模块分离嵌入CBAM的模型、标准方式嵌入CBAM的模型和未嵌入注意力模块的模型分别提升了2.86%、6.41%和12.16%。采用双尺度卷积核丰富特征,在有限的卷积单元内增强特征提取能力。与经典卷积神经网络对比,本文设计的模型仅有1.02 MB的参数量和24.18 MB的每秒浮点运算次数(floating-point operations per second,FLOPs),大幅轻量化了模型并能达到98.57%的准确率。**结论**　本文设计了一个轻量高效的卷积单元构建网络模型,该模型具有较高的准确率和较低的参数量及计算复杂度,提高了口罩人脸姿态分类模型的效率和准确率。

关键词:轻量级卷积神经网络;口罩人脸姿态分类;深度可分离卷积;卷积块注意力模块(CBAM);深度学习;新冠肺炎(COVID-19)

A embedded dual-scale separable CBAM model for masked human face poses classification

Chen Senqiu[1,2*], Liu Wenbo[1,2], Zhang Gong[3]

1. College of Automation Engineering, Nanjing University of Aeronautics and Astronauts, Nanjing 211106, China;
2. Non-Destructive Testing and Monitoring Technology for High-Speed Transport Facilities Key Laboratory of Ministry of Industry and Information Technology, Nanjing 211106, China;

收稿日期:2020-12-11;**修回日期**:2021-01-06;**预印本日期**:2021-01-13

*通信作者:陈森楸　senqiuchen@nuaa.edu.cn

基金项目:国家重点研发计划资助(2018YFB2003304);国家自然科学基金项目(61871218);中央高校基本科研业务费专项资金资助(3082019NC2019002)

Supported by:National Key R&D Program of China(2018YFB2003304);National Natural Science Foundation of China(61871218);Fundamental Research Funds for the Central Universities (3082019NC2019002)

3. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronauts, Nanjing 211106, China

Abstract: **Objective** Human face poses classification is one of the key aspects of computer vision and intelligent analysis. It is a potential technology for human behavior analysis, human-computer interaction, motivation detection, fatigue driving monitoring, face recognition and virtual reality. Wearing masks is regarded as an intervened method during the outbreak of the corona virus disease 2019 (COVID-19) pandemic. It is a new challenge to achieve masked face poses classification. The convolutional neural network is widely applied to identify human face information and it is using in the face pose estimation. The convolutional neural network (CNN) research is to achieve face pose estimation based on low resolution, occlusion interference and complicated environment. In terms of the stronger capability of convolutional neural network and its successful application in face pose classification, we implement it to the masked face pose classification. Face pose estimation is one of the mediums of computer vision and intelligent analysis technology, and estimation results are used for subsequent analysis and decision. As an intermediate part of face pose estimation technology, the lightweight and efficient network structure can make the estimation to play a greater role within limited resources. Therefore, our research focuses on an efficient and lightweight convolutional neural network for masked face pose estimation. **Method** The core of the designed network is an efficient and lightweight dual-scale separable attention convolution (DSAC) unit and we construct the model based on 5 DSAC units stacking. The DSAC unit is constructed via two depthwise separating convolution with 3 × 3 and 5 × 5 kernel size in parallel, and we embed convolutional block attention module (CBAM) in these two convolutional ways. But, the embedding method that we proposed is different from the traditional CBAM embedding method. We split CBAM into spatial attention module (SAM) and channel attention module (CAM). We embed SAM following the depthwise (DW) convolution and embed CAM following the pointwise (PW) convolution respectively, and we cascade these two parts at final. In this way, the features of DW convolution and PW convolution can be matched. Meanwhile, we improve the SAM via the result of 1 × 1 pointwise convolution supplements, which can enhance the utilization of spatial information and organize a more effective attention map, we make use of 5 DSAC units to build the high accuracy network. The convolution channels and the number of units in this network are manipulated in rigor. The full connection layers are discarded because of their redundant number of parameters and the computational complexity, and a integration of point convolutional layer to global average pooling (GAP) layer is used to replace them. Therefore, these operations make the network more lightweight further. In addition, a large scale of human face data collection cannot be achieved temporarily because of the impact of COVID-19. We set the mask images that are properly scaled, rotated, and deformed on the common face pose dataset to construct a semisynthetic dataset. Simultaneously, a small amount of real masked face poses images are collected to construct a real masked face poses dataset. We use the transfer learning method to train the model under the lack of a massive real face poses dataset. **Result** The embedded demonstration results show that our proposed accuracy of the model is 2.86%, 6.41% and 12.16% higher than that of the model which is embedded separable CBAM without improved SAM module, the model embedded standard CBAM and the model without embedded CBAM. The results show that the model which is embedded separable CBAM with improved SAM module has an efficient and lightweight structure. It can effectively improve the performance of the model with little parameters and computational complexity. Depth wise separable convolution makes the model more compact and the CBAM attention module makes the model more efficient when the model is regressing. In addition, the dual scale convolution is used to tailor the features and enhance the feature extraction capability within the limited convolution unit. The method of extracting different scale features to enhance the performance of the model can avoid over-fitting and rapid growth of parameters derived of stacked convolutional layers. Compared with the classical convolutional neural network, such as AlexNet, Visual Geometry Group network (VGGNet), ResNet, GoogLeNet, the parameters and computational complexity of our designated model decrease significantly, and the accuracy is 3.57% ～30.71% higher than AlexNet, VGG16, ResNet18 and GoogLeNet. Compared with the classical lightweight convolutional neural network like SqueezeNet, MobileNet, ShuffleNet, EfficientNet, the demonstated model has the lowest parameters, computational complexity, and the highest accuracy, which only has 1.02 M parameters and 24.18 M floating-point operations per second (FLOPs), and achieves 98.57% accuracy. **Conclusion** A lightweight and efficient convolution unit is designed to construct the network, which has low parameters, computational complexity, and high accuracy.

Key words: lightweight convolutional neural network; masked face poses classification; depthwise separable convolution; convolutional block attention module (CBAM); deep learning; corona virus disease 2019 (COVID-19)

0 引 言

人脸姿态估计是计算机视觉和智能分析领域的重要课题之一,是疲劳驾驶检测(庄员和戚湧,2021)、人机交互和虚拟现实等领域的关键技术并有着广泛应用。近年来,人脸姿态估计研究活跃,成果丰硕(卢洋 等,2015;董兰芳 等,2016;Borghi 等,2020;Dua 等,2019)。新型冠状病毒肺炎(corona virus disease 2019, COVID-19)的爆发严重影响了社会、经济和生产生活等各个方面。在新冠疫情防控的新形势下,佩戴口罩成为重要防控措施之一,实现口罩遮挡的人脸姿态估计具有重要的现实意义。

围绕人脸姿态估计任务,提出了许多技术路线(Murphy-Chutorian 和 Trivedi,2009)。其中,基于特征回归的方法具有突出的优越性,其思路为构建人脸图像的特征空间与姿态空间的映射关系。但由于口罩遮挡的人脸图像信息大量损失,传统方法不能获取丰富且鲁棒的特征,导致算法性能严重下降。随着深度学习的发展,基于卷积神经网络的人脸信息提取技术不断进步(LeCun 等,2015;吴从中 等,2021),卷积神经网络成功用于人脸姿态估计研究(Byungtae 等,2015;Patacchiola 和 Cangelosi,2017;Raza 等 2018;Ruiz 等,2018;Khan 等,2020)。卷积神经网络较传统方法具有更强的特征提取能力,研究者利用卷积神经网络在低分辨率、遮挡干扰和复杂环境等条件下实现了人脸姿态估计。鉴于卷积神经网络的特征提取能力及在人脸姿态分类中的成功应用,本文将其应用于口罩遮挡的人脸姿态分类。

卷积神经网络通过堆叠的卷积层和池化层对图像进行多重非线性映射,自动提取了浅层纹理、边缘等细节信息及高层语义信息,但高效的性能也造成了卷积神经网络结构的复杂。复杂的网络结构使模型获得高性能的同时在参数量和计算复杂度方面牺牲很多,导致实时性不佳,且计算、存储资源消耗大。而人脸姿态估计通常是计算机视觉和智能分析技术的中间环节之一,姿态估计结果用于后续的分析决策。作为中间环节的人脸姿态估计技术,轻量高效的网络模型能够使其在有限的资源范围内发挥高效的作用。因此,本文的研究重点为设计一个高效轻量的卷积神经网络用于口罩遮挡的人脸姿态估计。

为提高卷积神经网络的效率,针对轻量级卷积神经网络开展了大量研究(Denton 等,2014;Han 等,2015;Zhou 等,2020)。Iandola 等人(2016)设计 Fire 模块构建了 SqueezeNet,显著降低了参数量和计算复杂度。Howard 等人(2017)提出深度可分离卷积代替传统卷积,构建了 MobileNetV1,接着提出倒残差结构改进可分离卷积,构建了 MobileNetV2(Sandler 等,2018)。Zhang 等人(2018)采用 shuffle 模块解决了组卷积引起的通道间信息流不通的问题,提出了性能高效且轻量的 ShuffleNetV1 及改进的 ShuffleNetV2(Ma 等,2018)。EfficientNet(Tan 等,2019a)通过一个复合系数动态优化卷积网络的深度、宽度和分辨率,在降低参数量的同时优化了网络性能。然而,轻量化网络模型会在一定程度上造成模型性能下降。因此,需要在综合考虑模型的体量和准确率的基础上,设计一个轻量且高效的网络模型,以期较低的计算要求和较高的准确率。

本文采用深度可分离卷积分解传统卷积运算,引入并改进卷积块注意力模块(convolutional block attention module,CBAM)及其嵌入方式,利用双尺度卷积来优化该注意力模块的结构,形成双尺度分离嵌入 CBAM 的卷积单元,在保证较高模型性能的前提下,采用较少的卷积单元构建了一个轻量且高效的网络模型,同时利用卷积层替换全连接层,进一步轻量化模型。由于新冠疫情影响,暂时无法实现大规模的人脸数据采集。本文利用公开的人脸姿态图像叠加口罩图像制作半仿真口罩人脸姿态图像数据,同时采集了少量的真实口罩人脸姿态数据。采用迁移学习的方法,在半仿真数据上训练本文设计的模型,并将其迁移至真实数据集,在有限的真实口罩人脸姿态数据条件下有效地训练了网络模型,提高了模型泛化能力。

1 本文模型

1.1 深度可分离卷积

解耦卷积运算方式是降低计算量的重要措施之

一。深度可分离卷积是一种将传统卷积解耦为深度(depthwise,DW)卷积和点(pointwise,PW)卷积的特殊卷积方式,如图 1 所示。

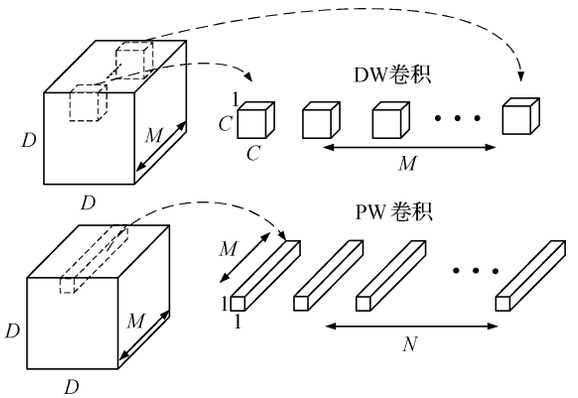

图 1　深度可分离卷积

Fig. 1　Depthwise separable convolution

传统卷积的计算量 $T_1 = M \times N \times C \times C \times D \times D$。而 1 个深度卷积的计算量为 $M \times C \times C \times D \times D$,1 个点卷积的计算量为 $M \times N \times D \times D$,深度可分离卷积的计算量 $T_2 = M \times C \times C \times D \times D + M \times N \times D \times D$。深度可分离卷积与传统卷积的计算量比例为

$$\frac{T_2}{T_1} = \frac{M \times C \times C \times D \times D + M \times N \times D \times D}{M \times C \times C \times D \times D \times N} = \frac{1}{N} + \frac{1}{C^2} \tag{1}$$

式中,M 为输入特征图的通道数,N 为卷积核个数,D 为特征图的尺寸,C 为卷积核的尺寸。卷积核个数 N 越多,计算量下降越大,所以采用深度可分离卷积可以大幅降低卷积计算量。

1.2　CBAM 注意力模块及改进

CBAM 注意力模块(Woo 等,2018)是一个低参数量、可灵活嵌入基础网络中即插即用的模块,由通道注意力模块(channel attention module,CAM)和空间注意力模块(spatial attention module,SAM)级联而成。

通道注意力模块 $\boldsymbol{M}_{\mathrm{C}} \in \mathbf{R}^{C\times1\times1}$,如图 2 所示,⊕表示元素相加。假设输入特征图 $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_C]$($\boldsymbol{f}_i \in \mathbf{R}^{H\times W}$),首先通过最大值池化和均值池化将 $\boldsymbol{F}$ 进行挤压,结果为 $[q_1, q_2, \cdots, q_C]$($q_i \in \mathbf{R}$),每个通道的 2 维特征图由一个实数表示,代表该通道的整体信息。接着将这两组描述符送入一个包含隐含层的共享网络(shared multi-layer perceptron, shared MLP)学习得到不同通道间的注意力权值。CAM 计算过程为

$$\begin{aligned}\boldsymbol{M}_{\mathrm{C}}(\boldsymbol{F}) &= \sigma(\boldsymbol{f}_{\mathrm{mlp}}([\boldsymbol{f}_{\mathrm{a}}(\boldsymbol{F})) + \boldsymbol{f}_{\mathrm{mlp}}(\boldsymbol{f}_{\mathrm{m}}(\boldsymbol{F})) = \\ &\quad \sigma(\boldsymbol{W}_1(\boldsymbol{W}_0(\boldsymbol{F}_{\mathrm{a}}^{c})) + [\boldsymbol{W}_1(\boldsymbol{W}_0(\boldsymbol{F}_{\mathrm{m}}^{c})))\end{aligned} \tag{2}$$

式中,σ 为 sigmoid 函数,$\boldsymbol{W}_0 \in \mathbf{R}^{C/r\times C}$ 和 $\boldsymbol{W}_1 \in \mathbf{R}^{C\times C/r}$ 为 shared MLP($\boldsymbol{f}_{\mathrm{mlp}}$)的权重,$r$ 为控制 shared MLP 节点数的衰减系数,$\boldsymbol{F}$ 为输入特征图,$\boldsymbol{F}_{\mathrm{a}}^{c}$ 和 $\boldsymbol{F}_{\mathrm{m}}^{c}$ 为 $\boldsymbol{F}$ 求取全局均值池化 $\boldsymbol{f}_{\mathrm{a}}$ 和全局最大值池化 $\boldsymbol{f}_{\mathrm{m}}$ 的特征图。

空间注意力模块通过学习人脸图像中不同空间位置的重要性,生成空间注意力图。传统的 SAM 通过对特征图沿通道维度分别进行均值池化和最大值池化挤压图像空间信息,但这种挤压方式对图像的空间信息利用并不充分。

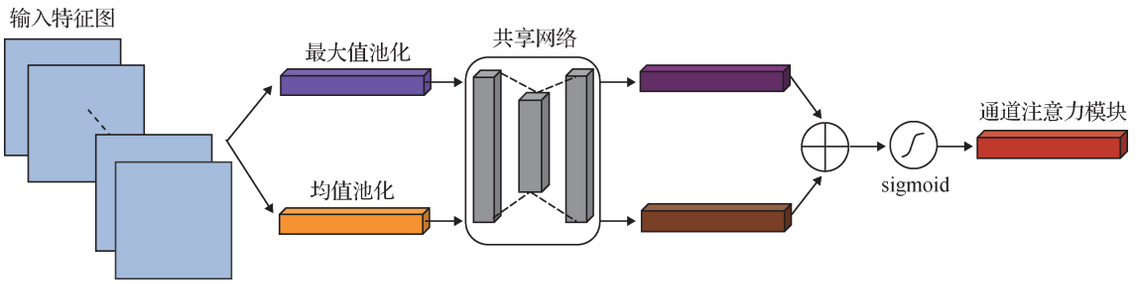

图 2　通道注意力模块示意图

Fig. 2　Schematic diagram of channel attention module

本文在利用均值池化和最大值池化挤压图像空间信息的基础上,增加对特征通道的挤压,丰富模块挤压的空间信息。具体操作为采用 1 × 1 的点卷积逐像素点地对各通道进行挤压。通过补充 1 × 1 点卷积结果,挤压操作能够获取更丰富的信息,形成更有效的注意力图,从而更好地把握空间信息。如图 3 所示,改进空间注意力模块采用 1 × 1 的点卷积将输入特征图挤压为 1 维,接着将 3 个特征描述符串联并采用 3 × 3 的卷积核进行运算得到空间注意力图。改进 SAM 模块计算过程为

$$\begin{aligned}\boldsymbol{M}_{\mathrm{S}}(\boldsymbol{F}) &= \sigma(\boldsymbol{f}^{3\times3}([\boldsymbol{f}_{\mathrm{a}}(\boldsymbol{F}); \boldsymbol{f}_{\mathrm{m}}(\boldsymbol{F}); \boldsymbol{f}_{\mathrm{p}}(\boldsymbol{F})])) = \\ &\quad \sigma(\boldsymbol{f}^{3\times3}([\boldsymbol{F}_{\mathrm{a}}^{\mathrm{S}}; \boldsymbol{F}_{\mathrm{m}}^{\mathrm{S}}; \boldsymbol{F}_{\mathrm{p}}^{\mathrm{S}}]))\end{aligned} \tag{3}$$

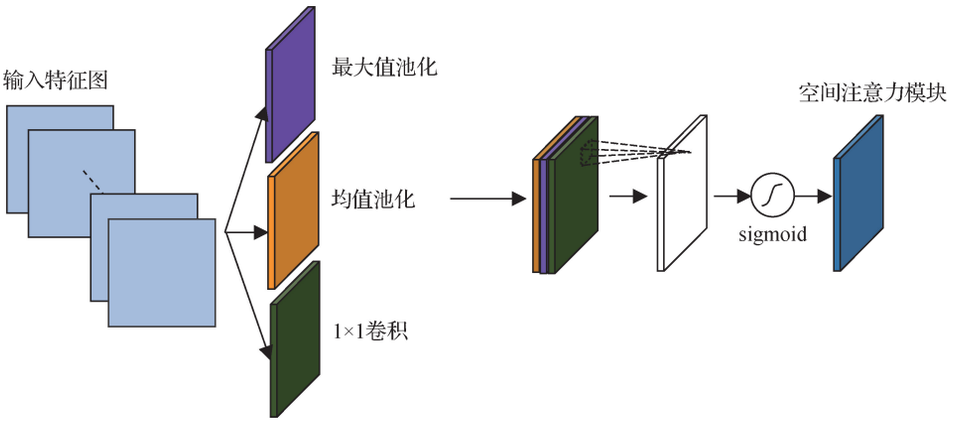

图 3　改进空间注意力模块

Fig. 3　Improved spatial attention module

式中，σ 为 sigmoid 函数，$\boldsymbol{f}^{3\times3}$ 代表卷积核为 3 × 3 的卷积操作，$\boldsymbol{F}_{\mathrm{a}}^{\mathrm{S}}$，$\boldsymbol{F}_{\mathrm{m}}^{\mathrm{S}}$ 和 $\boldsymbol{F}_{\mathrm{p}}^{\mathrm{S}}$ 为均值池化 $\boldsymbol{f}_{\mathrm{a}}$ 、最大池化 $\boldsymbol{f}_{\mathrm{m}}$ 和 1 × 1 点卷积 $\boldsymbol{f}_{\mathrm{p}}$ 挤压后的特征图。

1.3　双尺度分离嵌入注意力单元及整体网络设计

由于轻量化模型会导致模型性能受损，本文通过嵌入注意力模块改善轻量化模型的性能。注意力模块是一个即插即用的小参数量插件，能够以增加较少参数量的代价提升模型性能。

本文采用深度可分离卷积结合 CBAM 构建了一个轻量高效的卷积结构。注意力模块作为一个独立的组件嵌入在主干网络中，Woo 等人（2018）研究了不同的 CAM 和 SAM 组合嵌入方式（串联/并联）且确立了先 CAM 后 SAM 的串联组合方式。假设输入特征图为 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$，首先经过 CAM 对通道特征进行处理，然后经过 SAM 对空间特征进行处理。具体为

$$\begin{aligned}\boldsymbol{F}' &= \boldsymbol{M}_{\mathrm{C}}(\boldsymbol{F})\otimes\boldsymbol{F}\\ \boldsymbol{F}'' &= \boldsymbol{M}_{\mathrm{S}}(\boldsymbol{F}')\otimes\boldsymbol{F}'\end{aligned}\tag{4}$$

式中，$\boldsymbol{F}'$ 为通道注意力的结果，$\boldsymbol{M}_{\mathrm{C}}\in\mathbf{R}^{C\times1\times1}$ 为通道注意力模块，$\boldsymbol{F}''$ 为空间注意力的结果，$\boldsymbol{M}_{\mathrm{S}}\in\mathbf{R}^{1\times H\times W}$ 为空间注意力模块，⊗ 代表元素的乘法。

CBAM 以标准方式（Woo 等，2018）嵌入深度可分离卷积的效果是次优的。深度可分离卷积由 DW 卷积和 PW 卷积组成。在一个 DW 卷积中，卷积核数与输入特征通道数一致，单个卷积核仅对一个特征通道进行运算，所以各通道间的信息不流通。PW 卷积以 1 × 1 的点卷积核逐点地对 DW 卷积结果进行处理，融合不同通道间的特征。特征图经过 DW 卷积后仅能获取各特征通道的空间信息，而经过 PW 卷积后才能获取特征图的空间及通道的混合信息。传统卷积则是一步获取空间及通道的混合信息，后接 CBAM 模块对包含混合信息的特征图进行处理。然而，按标准方式对深度可分离卷积嵌入 CBAM 模块的效果并未能达到最佳。本文将 CBAM 模块拆分，在 DW 卷积后嵌入 SAM，对仅包含空间信息的特征图进行空间注意力调整，而后将处理过的特征图送入 PW 卷积获取包含空间及通道特征的混合信息，且在其后嵌入 CAM 对特征图进行调整。具体为

$$\begin{aligned}\boldsymbol{F}' &= \boldsymbol{M}_{\mathrm{S}}(\boldsymbol{F}_{\mathrm{DW}})\otimes\boldsymbol{F}_{\mathrm{DW}} = \boldsymbol{M}_{\mathrm{S}}(\boldsymbol{f}_{\mathrm{DW}}(\boldsymbol{F}))\otimes\boldsymbol{f}_{\mathrm{DW}}(\boldsymbol{F})\\ \boldsymbol{F}'' &= \boldsymbol{M}_{\mathrm{C}}(\boldsymbol{F}_{\mathrm{PW}})\otimes\boldsymbol{F}_{\mathrm{PW}} =\\ &\quad \boldsymbol{M}_{\mathrm{C}}(\boldsymbol{f}_{\mathrm{PW}}(\boldsymbol{F}'))\otimes\boldsymbol{f}_{\mathrm{PW}}(\boldsymbol{F}')\end{aligned}\tag{5}$$

式中，$\boldsymbol{F}_{\mathrm{DW}}$ 为 DW 卷积结果，$\boldsymbol{f}_{\mathrm{DW}}$ 为 DW 卷积，$\boldsymbol{F}_{\mathrm{PW}}$ 为 PW 卷积结果，$\boldsymbol{f}_{\mathrm{PW}}$ 为 PW 卷积。

本文将 CBAM 分离嵌入深度可分离卷积，所提卷积结构 DW-SAM-PW-CAM 如图 4 所示（⊗表示元素相乘），其效果优于 CBAM 以标准方式嵌入的结构 DW-PW-CAM-SAM。该结构能够更有效地将注意力模块应用于卷积运算。

卷积神经网络通过不断堆叠卷积层或者扩宽卷积通道数可以在一定程度上增强模型性能，但是这样的操作会增大模型的参数量和计算复杂度。在 Inception（Szegedy 等，2015）结构启发下，本文采用不同尺度的卷积核分担单个卷积通道数，提取不同尺度的特征信息，丰富模型获得的图像特征（Tan 和 Le，2019b）。本文采用 3 × 3 和 5 × 5 两种尺度的卷积核替换单一尺度的卷积核，以牺牲较少的模型

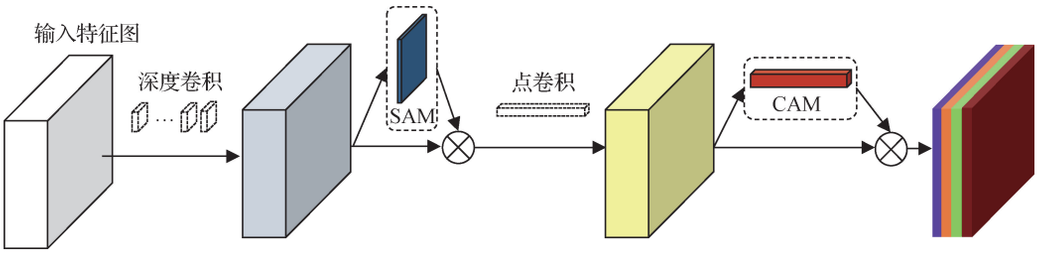

图 4　DW-SAM-PW-CAM 结构

Fig. 4　The structure of DW-SAM-PW-CAM

参数量有效提升模型性能。结合 DW-SAM-PW-CAM 结构,首先分别采用 3 × 3 和 5 × 5 的 DW 卷积和 SAM 处理输入特征图,接着将结果送入 PW 卷积和 CAM 进行处理,最后将计算后不同尺度的特征图连接起来作为整个卷积块单元的输出。本文将该单元称为双尺度分离注意力卷积(dual-scale separable attention convolution, DSAC)单元,如图 5 所示。本文利用设计的 DSAC 单元搭建网络模型,但简单地堆叠卷积单元不仅造成模型参数量和计算复杂度激增,还容易导致模型过拟合,性能下降。因此在保证模型准确率的前提下,将每个 DSAC 单元以较少的通道数构建为一个仅包含 5 个 DSAC 单元的轻量级网络模型。此外,模型参数大量集中在网络的全连接层部分。因此本文丢弃全连接层,并在最后一个 DSAC 卷积单元添加新卷积层,其输入通道数为最后一个轻量卷积块单元提取的特征图通道数,输出则为 n 个特征映射,对应 n 个目标的高维特征,然后经过 softmax 得到最终输出结果。本文设计的模型在保证准确率的前提下,大幅降低了参数量和计算复杂度,整体网络结构如图 6 所示。

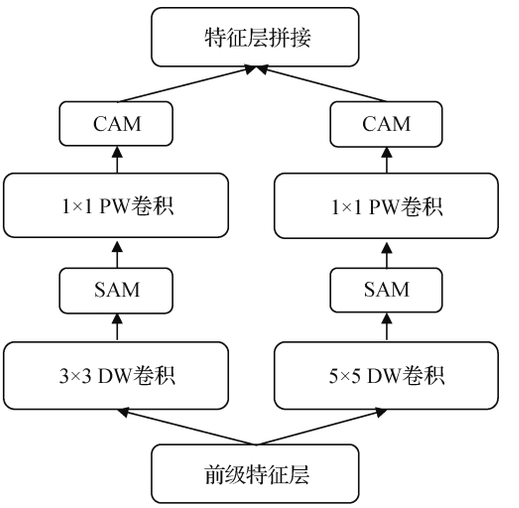

图 5　DSAC 卷积块单元

Fig. 5　DSAC convolution block unit

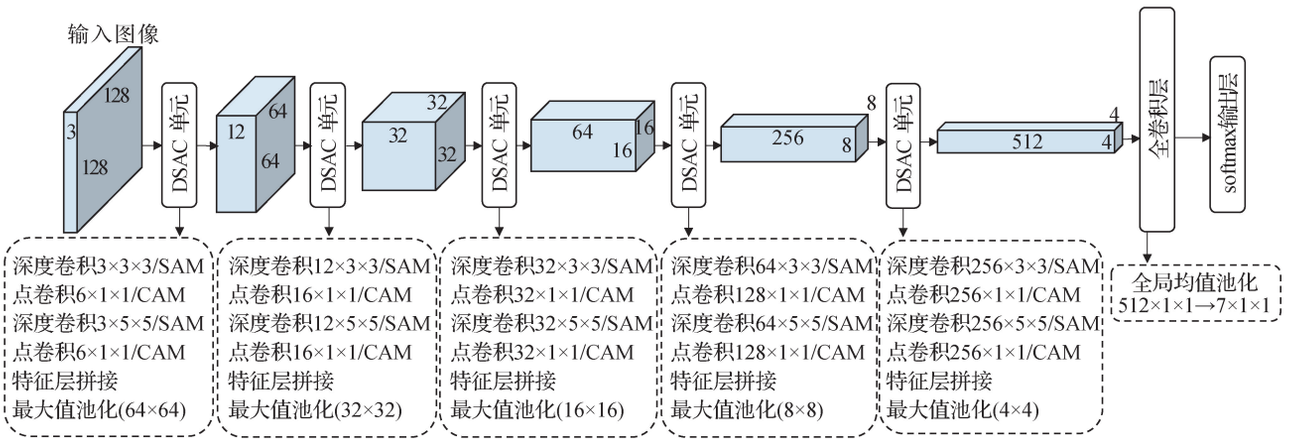

图 6　本文所提卷积神经网络模型

Fig. 6　The overall structure of the lightweight convolution neural network proposed in this paper

2 实验与结果分析

2.1 数据集准备

受新冠疫情影响，本实验暂时无法实现大规模人脸采集。在 CAS-PEAL-R1（张晓华 等，2005）人脸姿态数据集基础上，与经过缩放、旋转和变形操作的口罩图像叠加，制作了一个半仿真口罩人脸姿态数据集。同时采集少量真实口罩人脸姿态图像，构建了一个真实口罩人脸姿态数据集。数据集样例如图 7 所示，第 1 行是半合成样本，第 2、3 行是真实样本，包含偏航（Yaw）方向 ±67°、±45°、±22°和 0°共 7 种姿态类别。半合成数据集包括 1 040 个人在 7 种不同姿态下的口罩人脸姿态图像 7 280 幅，其中随机选取每个姿态 740 幅共 5 180 幅作为训练样本，剩余 2 100 幅作为测试样本。真实数据集为 57 个人在相同 7 个姿态下的真实口罩人脸姿态图像 798 幅，其中随机选取每个姿态 94 幅共 658 幅作为训练样本，剩余 140 幅作为测试样本。将图像尺寸统一缩放为 128 × 128 像素以符合网络输入要求，同时为了增强模型的泛化能力，随机对数据采取了亮度变换、加噪声和模糊等数据增强，其中噪声为椒盐噪声和均值为 0、方差为 0.002 的高斯噪声；亮度变换为原来的 0.5 倍和 1 倍；图像模糊采用均值模糊滤波器处理。

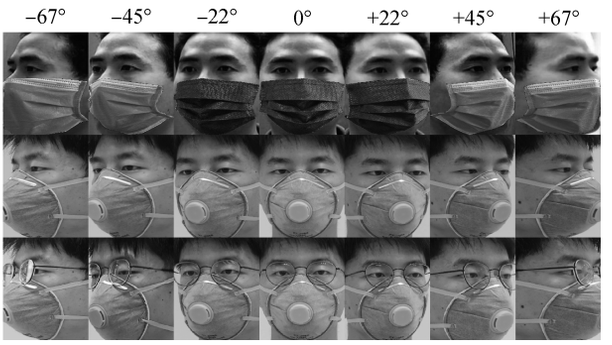

图 7 半合成和真实口罩人脸姿态图像

Fig. 7 Semisynthetic and real masked face poses images

2.2 训练方法及参数设置

本文采用迁移学习的训练方法，将在半仿真数据集上预训练的模型迁移至真实数据集中。因为源域和目标域具有类似的数据分布和相同的任务，所以本文采用微调的迁移方式，将迁移网络在目标域中进行非冻结训练（Yosinski 等，2014）。实验采用随机梯度下降算法，其动量设置为 0.9，权重衰减设置为 0.000 5；采用变学习率的训练方式，当迭代到训练次数的 2/3 时，学习率降低为原来的 1/10，以使误差收敛更加平稳，设置批大小为 16，损失函数选择交叉熵函数。在模拟数据集上从头训练的迭代次数设置为 50，初始学习率设置为 0.005。在真实口罩人脸姿态数据集上迁移训练的迭代次数设置为 10，初始学习率设置为 0.000 8。实验的软硬件平台为 PC 端，Windows10 操作系统，8 GB 内存的 Core i7-9750H CPU 处理器，4 GB 显存的 NVIDIA GeForce GTX 1650GPU 显卡，Pytorch 深度学习框架。

2.3 结果与分析

2.3.1 CBAM 模块嵌入方式对比实验

为了分析本文所提嵌入方式的性能，将未引入注意力模块的 DW-PW 方法与以标准 CBAM 嵌入方式的 DW-PW-CAM-SAM、采用分离嵌入注意力模块的 DW-SAM-PW-CAM 和改进了 SAM 模块的 DW-SAM(+)-PW-CAM 方法通过可视化方法 Grad-CAM（Selvaraju 等，2017）进行对比，结果如图 8 所示。Grad-CAM 可以清楚地显示网络在学习中重点关注的区域，通过观察网络认为对预测类重要的区域，从而试图去查看网络如何充分利用图像信息。从图 8 可以看出，由于 DW-PW 方法未引入注意力模块，网络对图像信息利用不充分。将 CBAM 模块以标准方式嵌入深度可分离卷积的 DW-PW-CAM-SAM 方法覆盖目标区域较基线增多，有效提升了网络对图像信息的利用程度。采用分离嵌入注意力模块的 DW-SAM-PW-CAM 方法对目标覆盖区域较 DW-PW-CAM-SAM 方法增大，说明采用 DW-SAM-PW-CAM 方法有效改进了 CBAM 注意力模块的嵌入方式。本文所提 DW-SAM(+)-PW-CAM 方法的目标区域覆盖程度在 DW-SAM-PW-CAM 方法的基础上进一步增大，表明改进 SAM 模块能进一步提升图像利用程度。此外，从图 8 可以清楚地看到网络对未遮挡人脸部分的信息利用程度较大。实验结果表明，采用 DW-SAM(+)-PW-CAM 方法构建的网络对目标区域信息利用程度最高，模型能有效获取图像特征。

为进一步验证不同嵌入方法的性能，对上述 4 种方法进行定量对比分析。首先给定以下几个评价参数。总体准确率（overall accuracy，OA）代表着一种方法的总体性能，是所有类别中分类正确的样本数占总样本数的比例。模型体量评价指标采用常用的模型参数量和每秒浮点运算次数（floating-point

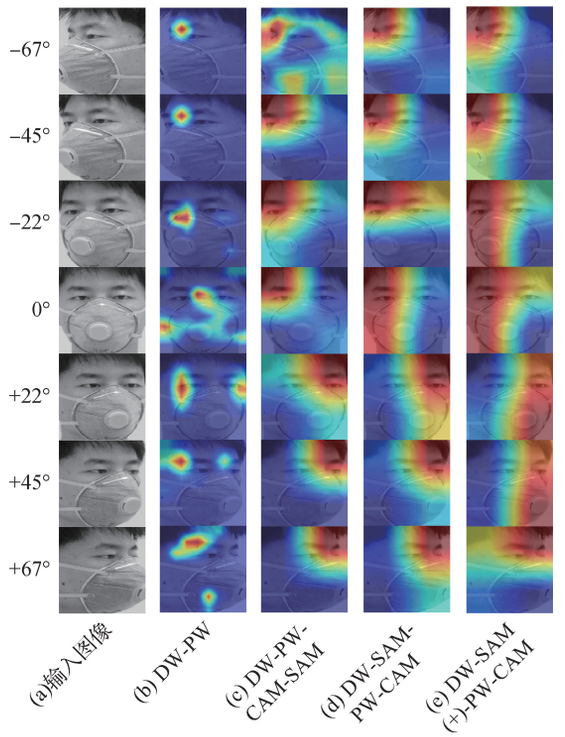

图 8　Grad-CAM 可视化结果

Fig. 8　Grad-CAM visualization results((a) input images; (b) DW-PW; (c) DW-PW-CAM-SAM; (d) DW-SAM-PW-CAM; (e) DW-SAM(+)-PW-CAM)

operations per second, FLOPs)。此外,通过嵌入注意力模块提升准确率会导致参数量和 FLOPs 的增加,本文希望在增加较少量参数量和 FLOPs 的前提下尽可能地提升准确率。通过与基准方法对比,提升的 OA 分别与增加的参数量和 FLOPs 做比值,将其定义为得分 S。提升的 OA 与增加的参数量的比值为 S_1,提升的 OA 与增加的 FLOPs 的比值为 S_2。显然,S 越大表示该方法较基准方法而言,能够牺牲较少参数量/计算复杂度以达到更高的准确率。S_1 和 S_2 的具体计算为

$$S_1 = \frac{O_n - O_b}{p_n - p_b} \tag{6}$$

$$S_2 = \frac{O_n - O_b}{R_n - R_b} \tag{7}$$

式中,O_n 和 O_b 分别代表当前方法和基准方法的总体准确率,p_n 和 p_b 分别代表当前方法和基准方法的参数量,R_n 和 R_b 分别代表当前方法和基准方法的 FLOPs。

实验结果如表 1 所示。可以看出:1)嵌入注意力模块可以有效提升 OA。CBAM 以标准方式嵌入网络的 DW-PW-CAM-SAM 方法较未引入注意力模块的方法(基线)提升了 5.75%。将 CBAM 分离嵌入深度可分离卷积的 DW-SAM-PW-CAM 方法较基线提升了 9.3%。在 DW-SAM-PW-CAM 方法上改进 SAM 模块的 DW-SAM(+)-PW-CAM 方法较基线提升了 12.16%。2)在不增加参数量和 FLOPs 的前提下,DW-SAM-PW-CAM 方法的 OA 较 DW-PW-CAM-SAM 方法有明显提升,表明分离 CBAM 的嵌入方式比标准嵌入的方式更具优势,合理地嵌入 CBAM 能有效提升模型性能。3)DW- SAM(+)-PW-CAM 方法的 OA 较 DW-SAM-PW-CA 方法有明显提升。DW-SAM(+)-PW-CAM 方法将通过对 SAM 模块增加 1 × 1 点卷积的结果作为补充信息,有效提升了 OA,但同时也导致模型的参数量和 FLOPs 小幅增加。4)DW-SAM(+)-PW-CAM 方法的 S_1 和 S_2 均为最高,表明该方法通过牺牲较少的模型参数量和计算复杂度获得了较高的模型准确率。

表 1　不同嵌入方法的性能对比

Table 1　Performance comparison of different embedding methods

方法	参数量/MB	FLOPs/MB	OA/%	S_1	S_2
DW-PW(基线)	**0.85**	**18.24**	86.41	–	–
DW-PW-CAM-SAM(CBAM)	1.01	22.81	92.16	0.36	0.012 6
DW-SAM-PW-CAM	1.01	22.81	95.71	0.58	0.020 4
DW-SAM(+) – PW-CAM(本文)	1.02	24.18	**98.57**	**0.72**	**0.020 5**

注:加粗字体表示各列最优结果;“–”表示此处数据为空。

4 种方法的网络训练收敛过程如图 9 所示。可以看出,DW-SAM(+)-PW-CAM 方法具有较快的收敛速度,并且最终能够获得较高的准确率。

2.3.2　不同尺度卷积核对比实验

为了验证本文采用多尺度卷积核的效果,对不同尺寸的卷积核组合进行对比实验。3 × 3 卷积核

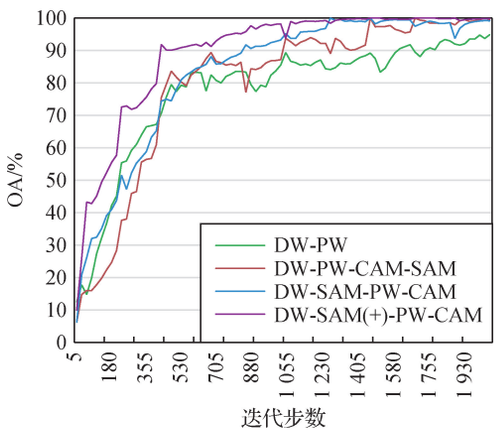

图 9　网络训练收敛过程

Fig. 9　Convergence of networks training process

已经广泛应用于各种网络模型，因此将仅包含 3×3 卷积核的网络作为基准，采用 5×5、7×7 和 9×9 这 3 种尺寸的卷积核组合方式分担卷积通道进行对比实验。不采用 11×11 及以上的卷积核的原因在于：1）除了 AlexNet 采用 11×11 卷积核作为前置特征提取层外，少有网络采用大卷积核；2）大尺寸卷积核会造成模型参数量激增；3）过大尺寸的卷积核可能造成提取特征冗余并影响模型性能。

本文设置{3×3,5×5}、{3×3,7×7}、{3×3,9×9}、{3×3,5×5,7×7}、{3×3,5×5,9×9}、{3×3,7×7,9×9}和{3×3,5×5,7×7,9×9}等网络结构作为对比实验组。实验结果如由表 2 所示。可以看出：1）多尺度卷积核构建的网络的 OA 都高于单一尺度的网络。在不增加卷积深度和宽度的前提下，采用多尺度卷积核能够有效提升模型的准确率。2）采用三尺度及四尺度卷积核构建的网络的 OA 高于采用双尺度卷积核的网络，但双尺度卷积核构建的网络的 S_1 和 S_2 普遍高于采用三尺度或四尺度卷积核的网络。即采用多尺度卷积核能够有效提升模型的 OA，但会增加模型的参数量和计算复杂度。而双尺度卷积核构建的网络能够通过增加较少参数量及计算复杂度有效提升 OA。3）包含 9×9 卷积核的网络的 S 均较低，表明采用过大尺寸的卷积核的网络不能很好地在提升 OA 和牺牲参数量及计算复杂度之间取得平衡。网络采用不同尺度的卷积核能够获得丰富的特征信息，但过大尺寸的卷积核会造成模型参数量和计算复杂度的大幅增加。4）采用{3×3,5×5}双尺度卷积核的网络的准确率为 98.57%，较基准网络仅增加了 0.03 MB 的参数量和 3.68 MB 的 FLOPs，构建的模型具有最高的 S_1 和 S_2，即在保持较少参数量和计算复杂度增加的前提下有效提升了网络的准确率。

表 2　不同尺寸卷积核组合的性能对比

Table 2　Performance comparison of combinations with different kernel sizes

卷积核组合方式	参数量/MB	FLOPs/MB	OA/%	S_1	S_2
{3×3}（基线）	**0.99**	**20.50**	95.14	–	–
{3×3,7×7}	1.05	26.27	97.29	0.36	0.003 73
{3×3,9×9}	1.09	30.89	96.28	0.12	0.001 11
{3×3,5×5,7×7}	1.18	32.24	98.15	0.16	0.002 56
{3×3,5×5,9×9}	1.23	36.86	98.43	0.14	0.002 01
{3×3,7×7,9×9}	1.26	40.32	99.14	0.15	0.002 02
{3×3,5×5,7×7,9×9}	1.13	46.22	**99.29**	0.30	0.001 61
{3×3,5×5}（本文）	1.02	24.18	98.57	**1.14**	**0.009 32**

注：加粗字体表示各列最优结果；“–”表示此处数据为空。

2.3.3　不同网络模型对比实验

为了验证本文模型的性能，与 AlexNet、VGGNet（Visual Geometry Group network）、ResNet（residual neural network）和 GoogLeNet 等经典卷积神经网络模型以及 SqueezeNet、MobileNet、ShuffleNet 和 EfficientNet 等优秀的轻量级卷积神经网络进行比较，采用模型参数量、FLOPs 和 OA 作为评价指标，实验结果如表 3 所示。

表 3　不同网络模型的性能对比

Table 3　Performance comparison of different models

网络模型	参数量/MB	FLOPs/MB	OA/%
AlexNet	61.60	715.54	67.86
VGG16	138.36	15 620.12	83.57
VGG19	143.67	19 793.67	77.86
ResNet18	11.69	1 582.81	95.00
ResNet50	25.56	3 530.02	**99.14**
GoogLeNet	6.80	1 550.65	94.25
SqueezeNet	1.25	829.88	73.57
MobileNetV1	4.24	569.66	90.00
MobileNetV2	3.40	320.24	96.43
ShuffleNetV1	2.45	140.62	93.57
ShuffleNetV2	2.28	49.28	95.00
EfficientNet	5.30	390.69	96.43
本文	**1.02**	**24.18**	98.57

注:加粗字体为各列最优结果。

从表 3 可以看出,1)经典卷积神经网络中的 AlexNet 网络结构相对简单,性能较差。更深更复杂的 VGGNet、ResNet 和 GoogLeNet 网络能够有效提升模型准确率。2)VGG19 网络在理论上较 VGG16 网络更加复杂,性能更强,但实验结果显示其性能反而下降。这是因为简单叠加卷积层在提升网络非线性映射能力的同时,更加容易导致网络梯度弥散或爆炸,容易造成过拟合,反而导致 OA 下降。3)具有残差结构的 ResNet 和具有 Inception 结构的 GoogLeNet 网络能够解决简单叠加网络带来的缺陷,大幅提升网络性能,其中 ResNet50 的 OA 达到了 99.14%。4)虽然采用 ResNet 或 GoogLeNet 能够提高模型的准确率,但传统网络的模型参数量和 FLOPs 均较庞大,其中 VGG19 的参数量达到了 143.67 MB,FLOPs 为 19 793.67 MB。5)采用轻量级卷积神经网络可以在降低模型参数量和计算复杂度的同时具有较高的准确率。MobileNet 的倒残差结构、ShuffleNet 的通道混洗策略以及 EfficientNet 多维特征混合方法都使模型以轻量的体量和较低的计算复杂度获得了较高的准确率。本文模型的 DSAC 单元同样在轻量化模型的同时具有较高的准确率。6)与经典的轻量级卷积神经网络相比,本文模型以更少的参数量和计算复杂度获得了更高的准确率,较 SqueezeNet、MobileNetV1、MobileNetV2、ShuffleNetV1、ShuffleNetV2 和 EfficientNet 提升了 2.14%～25%,参数量降低了 0.23 ～ 4.28 MB,FLOPs 降低了 25.1 ～ 805.7 MB。其中,较 SqueezeNet 提升了 25% 且 FLOPs 降低了 805.7 MB。7)本文模型以 1.02 MB 的参数量和 24.18 MB 的 FLOPs 获得了 98.57% 的准确率,设计的网络模型结构紧凑、轻盈且高效。在具有较高分类准确率的同时,大幅降低了参数量和计算复杂度,提升了计算效率。

2.3.4　不同训练方法对比实验

由于真实场景中的口罩人脸姿态数据较少,采取一种有效的小样本学习方法是成功训练模型的关键。本文设计了两种方案解决数据缺乏问题,其一是通过混合制作的半仿真数据和真实数据,将模型在混合数据集中进行训练;其二是根据半仿真数据具有与真实数据相似数据分布的特点,采用迁移学习的方法能够有效地训练模型。所以将在半仿真数据集上训练的网络模型迁移至真实数据集中,提升模型的准确率。

不同训练方法的实验结果如表 4 所示。可以看出:1)仅在半仿真数据集中训练的模型缺乏在真实数据下的泛化能力,直接在真实数据集中测试则准确率不高。2)仅在真实数据集中训练的网络模型的准确率也较低,这是因为真实数据集过小,网络容易过拟合,导致测试准确率下降。3)在真实数据和半仿真数据混合的数据集上训练的模型的准确率能够达到 90.2%,通过迁移学习训练的模型的准确率能够达到 98.57%。实验表明,采用迁移学习方法能够在有限的真实口罩人脸姿态数据条件下有效训练网络模型,且具有较高的模型准确率。

表 4　不同训练方法的 OA

Table 4　OA of different training methods

训练样本源	OA/%
半合成数据集	67.86
真实数据集	74.29
混合数据集	90.02
本文方法	**98.57**

注:加粗字体为最优结果。

3　结　论

本文设计了一个轻量级的卷积神经网络模型用

于口罩人脸姿态分类。将通过深度可分离卷积解耦传统卷积、采用卷积层替代全连接层、缩减网络深度及卷积通道数等作为网络轻量化的主要手段，并引入注意力机制提升轻量化模型的性能。

首先，创新性地将 CBAM 注意力模块分离嵌入 DW 卷积和 PW 卷积，针对性地对特征图的空间信息和通道信息进行调整。其次，对 SAM 模块补充 1×1 的点卷积特征图，使 SAM 模块能够获取更丰富的空间信息，更好地把握了感受域的信息。然后，采用双尺度卷积核优化 DW-SAM(+)-PW-CAM 卷积结构，构建了 DSAC 模块，仅利用 5 个 DSAC 模块搭建了本文轻量高效的卷积神经网络模型。最后，将设计的网络模型在构建的半仿真口罩人脸姿态数据集上进行预训练后迁移至真实数据集中微调训练。

本文设计的网络模型具有紧凑轻盈的结构，大幅缩减了参数量和计算复杂度，具有较高的分类准确率。采用迁移学习的方法在缺乏真实口罩遮挡人脸姿态数据集的条件下成功训练了模型，提高了模型的泛化能力和准确率。与经典卷积神经网络对比，本文设计的模型仅有 1.02 MB 的参数量和 24.18 MB 的 FLOPs，而准确率达到了 98.57%。然而，本文研究受限于人脸姿态类别数量，未能实现较精细化的人脸姿态估计。未来的工作中，将构建更加完备的口罩人脸姿态数据集，考虑更多细分的人脸姿态，设计能够估计更加细分姿态类别的模型。

参考文献(References)

Byungtae A, Park J and Kweon I S. 2015. Real-time head orientation from a monocular camera using deep neural network//Cremers D, Reid I, Saito H and Yang M H, eds. Lecture Notes in Computer Science. Cham: Springer: 82-96 [DOI: 10.1007/978-3-319-16811-1_6]

Borghi G, Fabbri M, Vezzani R, Calderara S and Cucchiara R. 2020. Face-from-depth for head pose estimation on depth images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(3): 596-609 [DOI: 10.1109/TPAMI.2018.2885472]

Denton E, Zaremba W, Bruna J, LeCun Y and Fergus R. 2014. Exploiting linear structure within convolutional networks for efficient evaluation//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: ACM: 1269-1277 [DOI: 10.5555/2968826.2968968]

Dong L F, Ren L L and Dong Y D. 2016. Complex illumination face pose estimation. Journal of Chinese Computer Systems, 37(3): 598-602 (董兰芳, 任乐乐, 董玉德. 2016. 复杂光照下的人脸姿态估计. 小型微型计算机系统, 37(3): 598-602)

Dua I, Nambi A U, Jawahar C V and Padmanabhan V. 2019. AutoRate: how attentive is the driver? //Proceedings of the 14th IEEE International Conference on Automatic Face & Gesture Recognition. Lille, France: IEEE: 1-8 [DOI: 10.1109/FG.2019.8756620]

Han S, Pool J, Tran J and Dally W J. 2015. Learning both weights and connections for efficient Neural Networks//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: ACM: 1135-1143 [DOI: 10.5555/2969239.2969366]

Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2020-11-11]. http://arxiv.org/pdf/1704.04861.pdf

Iandola F N, Han S, Moskewicz M W, Ashraf K, Dally W J and Keutzer K. 2016. Squeezenet: alexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size [EB/OL]. [2020-11-11]. http://arxiv.org/pdf/1602.07360.pdf

Khan S D, Ali Y, Zafar B and Noorwali A. 2020. Robust head detection in complex videos using two-stage deep convolution framework. IEEE Access, 8: 98679-98692 [DOI: 10.1109/ACCESS.2020.2995764]

LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. Nature, 521(7553): 436-444 [DOI: 10.1038/nature14539]

Lu Y, Wang S G, Zhao W T and Wu W. 2015. Technology of virtual eyeglasses try-on system based on face pose estimation. Chinese Optics, 8(4): 582-588 (卢洋, 王世刚, 赵文婷, 武伟. 2015. 基于人脸姿态估计的虚拟眼镜试戴技术. 中国光学, 8(4): 582-588) [DOI: 10.3788/CO.20150804.0582]

Ma N N, Zhang X Y, Zheng H T and Sun J. 2018. Shufflenetv2: practical guidelines for efficient CNN architecture design//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 122-138 [DOI: 10.1007/978-3-030-01264-9_8]

Murphy-Chutorian E and Trivedi M M. 2009. Head pose estimation in computer vision: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(4): 607-626 [DOI: 10.1109/TPAMI.2008.106]

Patacchiola M and Cangelosi A. 2017. Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. Pattern Recognition, 71: 132-143 [DOI: 10.1016/j.patcog.2017.06.009]

Raza M, Chen Z H, Rehman S U, Wang P and Bao P. 2018. Appearance based pedestrians' head pose and body orientation estimation using deep learning. Neurocomputing, 272: 647-659 [DOI: 10.1016/j.neucom.2017.07.029]

Ruiz N, Chong E and Rehg J M. 2018. Fine-grained head pose estimation without keypoints//Proceedings of 2018 IEEE/CVF Conference

on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, USA: IEEE: 2155-215509 [DOI: 10.1109/CVPRW.2018.00281]

Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetv2: inverted residuals and linear bottlenecks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]

Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//Proceedings of 2017 International Conference on Computer Vision. Venice, Italy: IEEE: 618-626 [DOI: 10.1109/ICCV.2017.74]

Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 1-9 [DOI: 10.1109/CVPR.2015.7298594]

Tan M X and Le Q V. 2019a. Efficientnet: rethinking model scaling for convolutional neural networks [EB/OL]. [2020-11-11]. https://arxiv.org/pdf/1905.11946.pdf

Tan M X and Le Q V. 2019b. MixConv: mixed depthwise convolutional kernels [EB/OL]. [2020-11-11]. http://arxiv.org/pdf/1907.09595.pdf

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Wu C Z, Zheng R S, Zang H J, Liu M W, Xu J J and Zhan S. 2021. Face pose correction based on morphable model and image inpainting. Journal of Image and Graphics, 26(4): 828-836 (吴从中, 郑荣生, 臧怀娟, 刘明威, 徐甲甲, 詹曙. 2021. 结合形变模型与图像修复的人脸姿态矫正. 中国图象图形学报, 26(4): 828-836) [DOI: 10.11834/jig.200011]

Yosinski J, Clune J, Bengio Y and Lipson H. 2014. How transferable are features in deep neural networks?//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: ACM: 3320-3328 [DOI: 10.5555/2969033.2969197]

Zhang X, Zhou X, Lin M and Sun J. 2018. ShuffleNet: an extremely efficient convolutional neural network for mobile devices//Proceedings of 2018 Computer Society Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6848-6856 [DOI: 10.1109/CVPR.2018.00716]

Zhang X H, Shan S G, Cao B, Gao W, Zhou D L and Zhao D B. 2005. CAS-PEAL: a large-scale Chinese face database and some primary evaluations. Journal of Computer-Aided Design & Computer Graphics, 17(1): 9-17 (张晓华, 山世光, 曹波, 高文, 周德龙, 赵德斌. 2005. CAS-PEAL 大规模中国人脸图像数据库及其基本评测介绍. 计算机辅助设计与图形学学报, 17(1): 9-17) [DOI: 10.3321/j.issn:1003-9775.2005.01.002]

Zhou Y, Chen S C, Wang Y M and Huan W M. 2020. Review of research on lightweight convolutional neural networks//Proceedings of the 5th Information Technology and Mechatronics Engineering Conference (ITOEC). Chongqing, China: IEEE: 1713-1720 [DOI: 10.1109/ITOEC49072.2020.9141847]

Zhuang Y and Qi Y. 2021. Driving fatigue detection based on pseudo 3D convolutional neural network and attention mechanisms. Journal of Image and Graphics, 26(1): 143-153(庄员, 戚湧. 2021. 伪3D卷积神经网络与注意力机制结合的疲劳驾驶检测. 中国图象图形学报, 26(1): 143-153) [DOI: 10.11834/jig.200079]

作者简介

陈森楸,1997 年生,男,硕士研究生,主要研究方向为图像处理、深度学习。
E-mail: senqiuchen@nuaa.edu.cn

刘文波,女,教授,主要研究方向为信号处理、计算机测控技术。E-mail: wenboliu@nuaa.edu.cn

张弓,男,教授,主要研究方向为图像处理与分析、目标探测与识别、雷达信号处理。E-mail: gzhang@nuaa.edu.cn