

文章编号: 2095-2163(2020)06-0004-08

中图分类号: TP391

文献标志码: A

# 基于双通道卷积残差网络的人脸识别

吴涛, 蔡茂国, 沈冲冲, 周航

(深圳大学 电子与信息工程学院, 广东 深圳 518000)

**摘要:** 虽然人脸识别技术已经取得了很大的成就, 但是如何提高不同姿态、光照、表情识别的准确性也面临着很大的挑战。为了进一步提高精度, 神经网络模型被设计得越来越复杂, 这也直接导致在反向传播时会出现梯度消失等现象。为了缓解这些问题, 本文提出了一种复杂条件下参数较少的双通道卷积残差网络模型。该模型由多个输入通道组成, 共同学习输入图像的不同特征。将原始图像作为第一通道的输入, 然后利用 Sobel 算子提取原始图像的一阶导数特征, 并将其作为第二通道的输入, 来自两个通道的人脸特征信息经过融合后送入到一个残差模块, 经过平均池层, 最后用于识别。该网络模型结构简单、参数少、速度快、准确性高。模型在 FERET、AR 和 FEI 数据集上进行了训练和测试。实验结果表明, 本文的方法优于当前的一些先进方法。

**关键词:** 人脸识别; 双通道; 残差网络; sobel 算子

## Face recognition based on dual-channel residual convolution network

WU Tao, CAI Maoguo, SHEN Chongchong, ZHOU Hang

(College of Electronics and Information Engineering, Shenzhen University, Shenzhen Guangdong 518000, China)

**[Abstract]** Although face recognition technology has made great achievements, how to improve the accuracy of different pose, illumination and expression recognition is also facing great challenges. In order to further improve the accuracy, the neural network model is designed more and more complex, which directly leads to the phenomenon of gradient disappear in the back-propagation. In order to solve these problems, this paper proposes a two channel convolution residual network model with fewer parameters under complex conditions. The model consists of multiple input channels, which learn different features of input image together. The original image is taken as the input of the first channel, and then the first derivative feature of the original image is extracted by Sobel operator, which is taken as the input of the second channel. After fusion, the face feature information from the two channels converges in a residual module, then passes through the average pool layer, and finally is used for recognition. The network model has the advantages of simple structure, less parameters, fast speed and high accuracy. The model is trained and tested on FERET, AR and Fei datasets. The experimental results show that this method is superior to some advanced methods.

**[Key words]** Face recognition; Dual channel; Residual network; Sobel operator

### 0 引言

人脸识别是计算机视觉领域中的一个非常具有挑战性的课题。同一个人在不同的姿势、光照和表情 (PIE) 条件下, 看起来是不太一样的, 如图 1 所示。为了消除这些因素对人脸识别的影响, 本文提出的算法要求在增加个体间变化的同时减少个体内变化。为此, 研究人员提出了线性判别分析 (LDA)、主成分分析 (PCA)、贝叶斯分析等子空间人脸识别学习方法以及改进的方法。LDA 是 PCA 的增强, PCA 不使用类的概念, 而 LDA 使用。同一人的面部图像在 LDA 中被视为同一类。PCA 和 LDA 都做了降维, 并且它们都以矢量的形式将图像变换到新的空间和新的坐标轴。其他的改进还包括

将人脸图像投影到一个易于分类的子空间中。除此之外, 还有一些基于不同特征表示的方法, 例如: 局部二值模式、Gabor Filter、Scale-Sobel Filter 和 Scale-invariant feature transform (SIFT) 等手工图像算子。结合几种不同的算子, 可以获得更高的精度。



图 1 不同条件下的同一人脸

Fig. 1 The same person under a different illumination environment

在过去的十年中, 卷积神经网络 (convolutional

**基金项目:** 国家自然科学基金 (61872244)。

**作者简介:** 吴涛 (1995-), 男, 硕士研究生, 主要研究方向: 图像处理、深度学习; 蔡茂国 (1965-), 男, 博士, 教授, 主要研究方向: 图像处理;

沈冲冲 (1995-), 男, 硕士研究生, 主要研究方向: 目标检测; 周航 (1997-), 男, 硕士研究生, 主要研究方向: 图像处理、活体检测。

**通讯作者:** 吴涛 Email: 997969986@qq.com

收稿日期: 2020-03-25

neural network, CNN) 已经成为计算机视觉中一种非常流行的技术, 并且取得了显著的成功。与传统的图像处理方法相比, 卷积神经网络在目标检测、图像分类、语义分割等领域取得了很大的成就。基于卷积神经网络的人脸识别也取得了重大突破。DeepID、DeepID2、DeepID3、FaceNet 和 VGG-Face 是需要大量数据训练的人脸识别模型, 这些模型可以达到较高的准确率。为了提高训练速度, Xiang Wu 等人提出了一个轻量型的卷积神经网络<sup>[1]</sup>, 其计算成本达到了 VGG 模型的 9 倍以上。Xi Yin 等人认为人脸识别不是单一任务, 他们将人脸识别视为一个多任务问题。虽然这些卷积神经网络模型已经达到了很高的精度, 但是它们需要非常庞大的数据集来训练, 由于训练过程中会产生大量的参数, 这些方法对计算机硬件有较高的要求。为了充分利用图像的信息, 减轻网络模型的负担, 本文提出了一种参数较少的双通道卷积残差网络用于人脸识别。人脸识别系统的结构图如图 2 所示。首先, 利用人脸检测和对齐来消除背景的影响。其次, 通过数据增容来获得更多的训练数据。最后, 利用双通道残差网络进行训练识别原始图像, 经过 Sobel 算子处理过的图像能够表征人脸的不同信息, 因此, 利用多通道输入获得不同的人脸特征, 将两个通道得到的特征进行融合, 经过残差网络进一步的特征提取, 最后用于识别。提取特征之后, 使用平均池化层而不是全连接层, 这样可以使得网络更加轻盈。本文提出的模型在 FERET、AR 和 FEI 数据集上进行了训练和测试。实验结果表明, 该人脸识别算法更有优势。

## 1 相关工作

传统的 CNN 模型如 AlexNet、VGGNet 以及新提出的 InceptionNet、ResNet 等在图像分类方面取得了显著的效果。这些方法也可以应用于人脸识别。AlexNet 由 Alex Krizhevsky、Geoffrey Hinton 和 Ilya Sutskever 组成的研究小组所设计, 它比 LeNet 更深。AlexNet 首先使用线性整流单元 (ReLU)、Dropout 层和局部响应归一化 (LRN) 来提高 CNN 的性能。在 2012 年的全球最权威的计算机视觉识别挑战 (ILSVRC) 中, 该网络获得了 15.3% 的 top-5 错误, 领先亚军 10.8 个百分点。VGGNet 是由牛津大学的 K.Simonyan 和 A.Zisserman 提出的 CNN 模型。整个网络只使用  $3 \times 3$  卷积核和  $2 \times 2$  池化层, VGGNet 表明网络的深度对于特征的提取起着重要作用, 网络越深, 效果越好。但是随着网络模型的加深, 网络变得越来越难训练, 这主要是由于梯度弥散和梯度爆炸导致的, 为了解决这个问题, 何凯明提出了 ResNet, 并于 2015 年在 ILSVRC 中获得冠军。ResNet 可以训练数百甚至数千层网络, 并且仍然可以获得非常好的效果, 该模型比 VGGNet 参数少, 收敛速度快。

DeepFace 是 Facebook 的一个研究小组创建的人脸识别的深度神经网络, 是第一个在 440 万张人脸数据库中训练卷积神经网络, 采用人脸对齐进行数据预处理, 网络参数超过 1.2 亿。据称, 该模型在 Wild (LFW) 数据集上的人脸识别准确率高达 97%。然而, 由于使用局部连接层, 模型的参数出现了指数增加。为了进一步提高精度, Sun Y、Wang X 和 Tang X 采用了多尺度模型。DeepID 通过深度神经网络来提取人脸深层次的抽象特征, 这种特征可以区分出两张不同的人脸, 在提取特征后, 输入到联合贝叶斯模型, 进行人脸验证, DeepID 在 LFW 数据集上的准确率达到 97.53%。Florian Schroff 等人提出了一个名为 FaceNet 的系统, 该系统可以将原始图片降维到欧式距离空间的映射, 其中距离与人脸相似性相关, FaceNet 首次引入三重态损失。Visual Geometry Group 提出了一个名为 VGG-Face 的模型, 该模型结合了 VGGNet 和三重态损失。VGG-Face 和 FaceNet 使用类似的度量学习方法, 并在 260 万人脸数据集进行训练, 在 LFW 上准确率达到 98.95%, 在 YTF 上准确率达到 97.3%。虽然这些方法在 LFW 上达到了很高的精度, 但是他们都是通过设计复杂的网络模型来提取人脸特征, 都需要大量的数据来训练模型, 对计算机硬件的要求比较高。

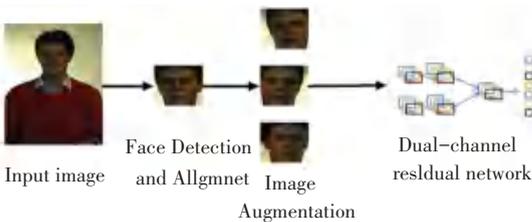


图 2 人脸识别系统的结构图

Fig. 2 The structure of face recognition system

本文的主要贡献如下: (1) 使用原始图像和 Sobel 算子处理过的图像作为网络的输入。(2) 提出了一种用于人脸识别的双通道卷积残差网络。该模型将多通道卷积神经网络与残差网络相结合, 使用平均池化层代替全连接层。(3) 网络模型参数较少, 不到 300 K, 网络模型文件大小约为 20 M。通过与其他模型的相比, 本文提出的模型不仅参数少、网络简单, 还加快了人脸识别的处理时间。

因此,本文主要针对一种轻量级的 CNN 人脸识别模型进行研究。

本文提出的人脸识别模型如图 3 所示。它是由两个输入通道构成,包括 6 个卷积层,6 个 max-pooling 层,1 个残差模块,1 个 average-pooling 层和

1 个 SoftMax 层,具体每层的参数数量如表 1 所示,在 TensorBoard 上运行的数据流图如图 4 所示,残差模块如图 5 所示,该模块中总共有三层卷积构成,卷积核分别为  $1 \times 1$ ,  $3 \times 3$  和  $1 \times 1$ 。残差模块解决了网络训练过程中出现的梯度弥散和梯度爆炸等现象。

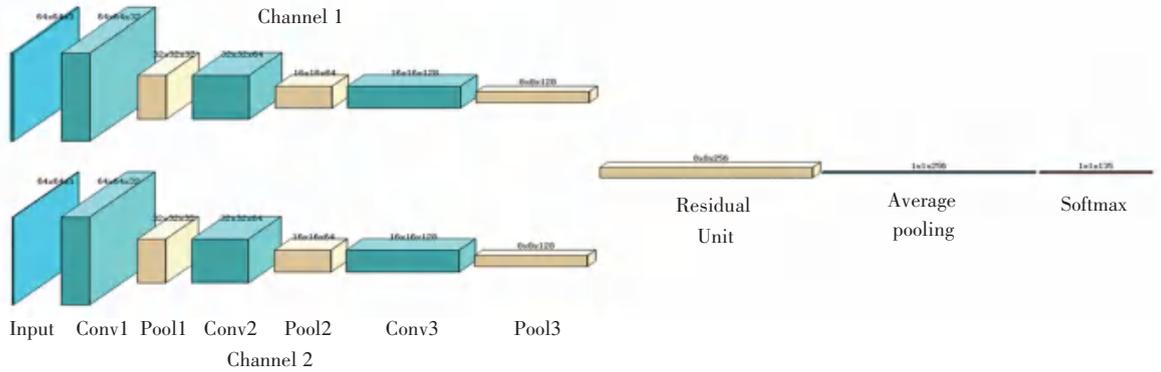


图 3 双通道卷积残差网络结构图

Fig. 3 The architecture of dual-channel residual network

表 1 双通道卷积残差网络的参数数量统计表

Tab. 1 The total params about dual-channel residual network

Channel1				Channel2			
Name	Filter Size/Stride	Output Size	#param	Name	Filter Size/Stride	Output Size	#param
input	-	$64 * 64 * 3$	-	Input	-	$64 * 64 * 1$	-
Conv1	$5 * 5/1$	$64 * 64 * 32$	2.4k	Conv1	$3 * 3/1$	$64 * 64 * 32$	0.28k
Pool1	$2 * 2/2$	$32 * 32 * 32$	-	Pool1	$2 * 2/2$	$32 * 32 * 32$	-
Conv2	$5 * 5/1$	$32 * 32 * 64$	51k	Conv2	$3 * 3/1$	$32 * 32 * 64$	18.5k
Pool2	$2 * 2/2$	$16 * 16 * 64$	-	Pool2	$2 * 2/2$	$16 * 16 * 64$	-
Conv3	$3 * 3/1$	$16 * 16 * 128$	73.7k	Conv3	$3 * 3/1$	$16 * 16 * 128$	73.7k
Pool3	$2 * 2/2$	$8 * 8 * 128$	-	Pool3	$2 * 2/2$	$8 * 8 * 128$	-
Concatenate: $8 * 8 * 256$							
Name	Filter Size/Stride	Output Size	#param				
Res Unit	$1 * 1/1$	$8 * 8 * 64$	16.3k				
	$3 * 3/1$	$8 * 8 * 64$	36.8k				
	$1 * 1/1$	$8 * 8 * 256$	16.3k				
Ave-pool	$8 * 8/8$	256	-				
SoftMax		Num	-				
total			288.98k				

将大小为  $64 \times 64$  的人脸图像作为两个并行通道的双通道卷积残差网络的输入,把原始图像作为第一通道的输入,利用 Sobel 算子提取原始图像的一阶导数特征,并将其作为第二通道的输入。每个通道都有 3 个卷积层,每个卷积层后面都有一个最大值池化层。这两个通道之间唯一的区别是卷积核的大小不同。考虑到 Sobel 算子处理的图像对轮廓更敏感,在第二通道中,统一选择尺寸为  $3 \times 3$ ,步长

为 1 的卷积核,这样做的目的是提取图像小区域的特征。在第一通道中,除最后一层采用尺寸为  $3 \times 3$  的卷积核,其余两个采用  $5 \times 5$  的卷积核。来自两个通道的特征融合在一个残差模块中,然后经过平均池化层,最后用于分类。本文将多通道卷积神经网络与残差模块相结合,充分利用融合后的特征,克服了反向传播时造成的梯度消失等现象,这样不仅能提高人脸识别的准确率,还能加快网络的收敛速度。

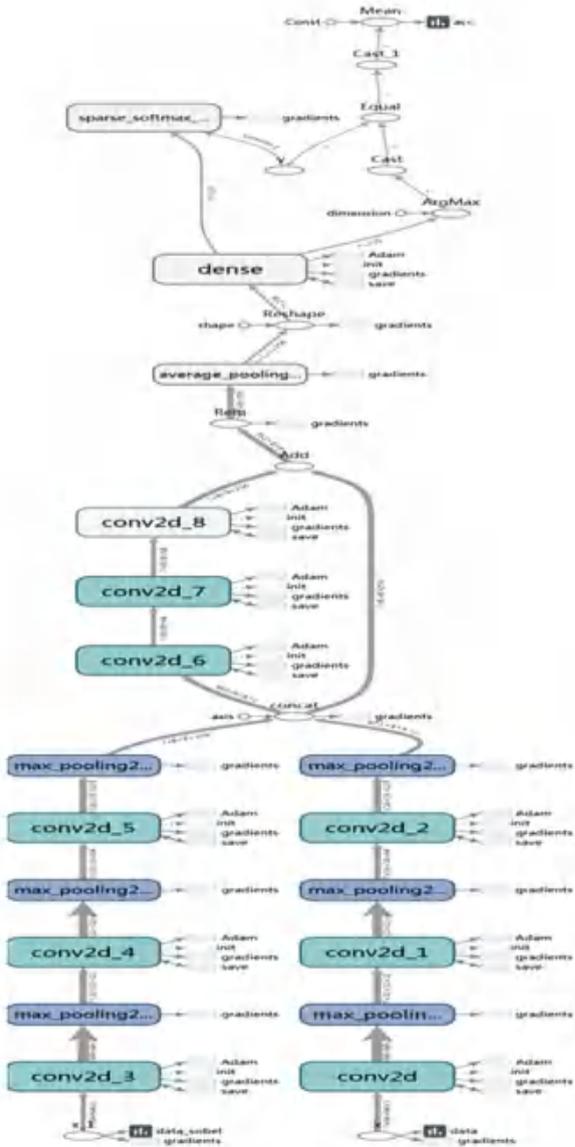


图 4 Tensorboard 可视化模型数据流图  
Fig. 4 The tensor graph of tensorboard

AR、FERET 和 FEI。AR 数据集是由美国广播公司计算机视觉中心 (CVC) 的 Alex Martinez 和 Robert Benavente 创建的,共包含了 4 000 多张彩色的人脸图像。图像以具有不同面部表情、照明条件和遮挡 (太阳眼镜和围巾) 的正面视图为特征。这些照片是在 CVC 严格控制的环境下拍摄的,参与者不受穿戴 (衣服、眼镜等)、化妆、发型等的限制,每个人参加两次拍摄,间隔 14 天。图 6 是来自 AR 数据库的部分图像。FEI 是巴西的一个人脸数据库,包含了一组在 2005 年 6 月到 2006 年 3 月之间在巴西圣保罗的圣贝尔纳·杜坎波的 FEI 人工智能实验室拍摄的人脸图像,共有 200 个人,每人有 14 张图像,总共 2800 张。所有的图片都是彩色的,在白色均匀的背景下拍摄,正面直立,侧面旋转可达 180 度,比例可能会变化约 10%,每个图像的原始大小是 640×480 像素。所有面孔主要由 FEI 的学生和工作人员代表,年龄在 19 到 40 岁之间,有着独特的外貌、发型和装饰。男性和女性受试者的数量完全相同,分别都为 100 张。图 7 显示了来自 FEI 数据库的部分图像。

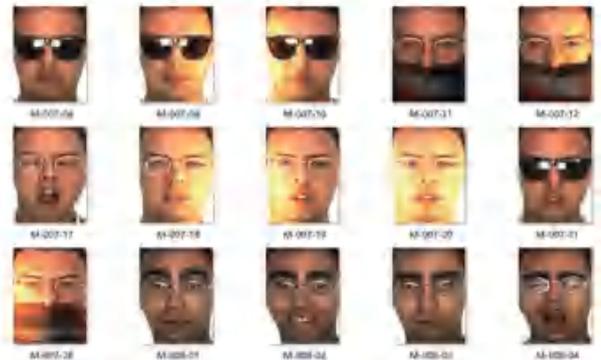


图 6 AR 数据集  
Fig. 6 AR dataset

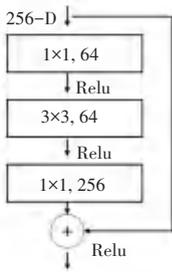


图 5 残差模块结构图  
Fig. 5 The architecture of residual block



图 7 FEI 数据集  
Fig. 7 FEI dataset

## 2 前期准备

### 2.1 数据集

本文将采用 3 种不同的数据集进行训练和测试,将数据集以 8 : 1 : 1 的比例分别划分为训练集、测试集以及验证集,本次使用到的数据集分别为:

## 2.2 人脸检测

目前实现人脸检测的方法有很多,本文采用2016年中国科学院深圳研究院提出的多任务级联卷积神经网络模型<sup>[2]</sup>来对数据集进行人脸检测,该模型主要采用了3个级联的卷积神经网络,3个网络分别为P-net、R-net、O-net。首先,将图片以0.709的比例进行缩小,形成一个图像金字塔,P-net会对输入的图片进行第一次判断,把人脸置信度高的图片留下来,去除置信度低的图像;其次,进行非极大值抑制(NMS),将所有图片的置信度进行排序,将置信度最大的人脸图像和其他图像做iou处理,粗略筛除重复框出的人脸,将输出的图像进一步放进R-net里面,进一步筛选人脸图像,做非极大值抑制,筛除重复的人脸框。最后,经过O-net网络,这是最后一层筛选,也是最精确的一次筛选,最终得到的最精确的人脸框,将人脸的5个关键的特征点标记出来,其过程如下图8所示。

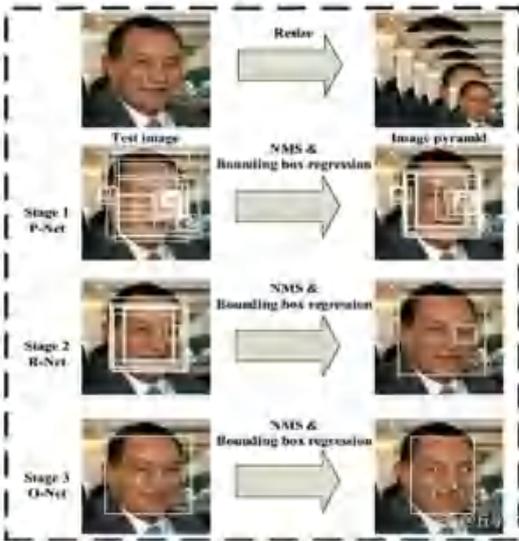


图8 mtcnn人脸检测流程图

Fig. 8 The architecture of mtcnn

## 2.3 数据增容

随着网络的加深,训练批次的增加,在数据集不够大的时候,网络训练可能会出现过拟合,除了给网络添加正则化和 dropout 以外,还可以进行数据增容,提高网络的泛化能力,常见数据增容的方式有以下几种:(1)图像翻转。将图像镜面翻转,一般选择水平翻转。(2)图像旋转。将图像以一定的角度进行旋转。(3)图像剪裁。对原始图像进行一定尺度的裁剪。(4)图像增强。将图像变暗或者变明,增强图像的对比度,经过图像增容后的数据集如下图9所示。



图9 数据增容后的部分数据

Fig. 9 The data after data augmentation

## 2.4 sobel 边缘检测算法

Sobel 算子,有时称为 Sobel 滤波器,在图像处理和计算机视觉中,用于图像中的边缘检测<sup>[3]</sup>。sobel 算子主要应用于边缘检测,在技术上,实际就是一个离散微分算子,用来运算图像亮度函数的梯度的近似值。在图像的任何一点使用此算子,将会产生对应的梯度矢量或是其法矢量。最终获取新图像的灰度值,其过程如下:

假设原始图像为  $I$ ,分别在水平方向和垂直方向求导:

水平方向。将图像  $I$  和一个卷积核为奇数的矩阵相乘,一般将卷积核的大小设为 3,如公式(1)所示。

$$G_x = \begin{bmatrix} \hat{e} & -1 & 0 & 1 & \hat{u} \\ \hat{e} & -2 & 0 & 2 & \hat{u} \cdot I \\ \hat{e} & -1 & 0 & 1 & \hat{u} \end{bmatrix} \quad (1)$$

垂直方向。将图像  $I$  和一个卷积核为奇数的矩阵相乘,同样,卷积核的大小和水平方向保持一致,如公式(2)所示:

$$G_y = \begin{bmatrix} \hat{e} & -1 & -2 & -1 & \hat{u} \\ \hat{e} & 0 & 0 & 0 & \hat{u} \cdot I \\ \hat{e} & 1 & 2 & 1 & \hat{u} \end{bmatrix} \quad (2)$$

接下来,将水平方向的  $G_x$  和垂直方向的  $G_y$  进行平方和再开方,得到近似梯度  $G$ ,如公式(3)所示:

$$G = \sqrt{G_x^2 + G_y^2} \quad (3)$$

为了计算的简便,通常采用公式(4)简单代替。

$$G = |G_x| + |G_y| \quad (4)$$

图10为通过 sobel 算子边缘检测过的图像。

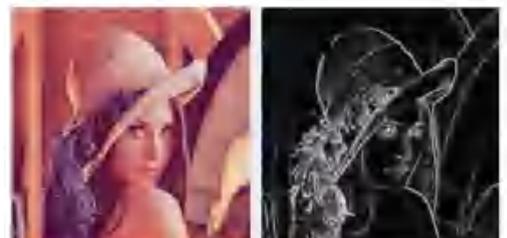


图10 sobel 算子处理过的图像

Fig. 10 Images processed by Sobel operator

### 3 实验

使用 MTCNN 人脸检测器来精确定位原始图像中的人脸,利用旋转、裁剪、翻转、缩放等数据增容方法,得到更多的训练图像,将得到的图像归一化为 64×64,利用 Sobel 算子获取图像的 Sobel 边缘映射,最后将这些图像输入到网络。采用留一法对模型进行评估,首先打乱所有数据的顺序,随机选择其中的 80%的人脸图像作为双通道残差网络训练集,剩下的 20%用来做测试集和验证集,验证集主要用来调节神经网络的参数。使用开源的深度学习框架 TensorFlow 对模型进行训练。双通道卷积残差网络的第一通道输入为 64×64×3 的 RGB 人脸图像,第二通道的输入为 64×64×1 的 sobel 算子处理后的图像。为了防止网络过拟合,采用 L2 正则化,正则化中的权重衰减设置为 0.003,初始学习速率设置为  $e^{-4}$ ,如果模型不收敛则降低学习率,采用交叉熵损失函数来计算损失,批量大小首先设置为 256,然后减少到 32。设置 200 个 epoch 训练本文的模型,并在必要时增加。卷积的参数初始化采用截断的正态分布。本文的模型使用 GeForce GTX1050TI-A4G GPU 进行训练。

表 2 展示了所提出的模型与 FERET 数据集上其他现有方法的人脸识别的准确性的比较,表 3 展示了 AR 数据集上其他现有方法的人脸识别的准确性的比较。在 FERET 数据集中,本文模型达到了 99.83%的准确性,而 CFV 达到了 97.55%,而 PCANet在 FERET 数据集上的准确性是97.26%。在 AR 数据集中,本文模型达到了 99.70%的准确性,而 LCSPT 和 CFV + LRA 的精度为 99.16%。通过比较发现本文提出的方法优于其他方法。

表 2 在 FERET 数据集上的各种方法的准确性比较

Tab. 2 Comparison of recognition accuracies on FERET

Method	Accuracy
CFV	0.9755
MDSP+KNN	0.854 4
LBPNet <sup>[4]</sup>	0.968 5
DCP <sup>[5]</sup>	0.928 0
Extended PLS	0.916 2
PCANet <sup>[6]</sup>	0.972 6
ESRC <sup>[7]</sup>	0.947 0
ANN	0.940 0
The Proposed	0.998 3

表 3 在 AR 数据集上的各种方法的准确性比较

Tab.3 Comparision of recognition accuracies on AR

Method	Accuracy
CFV + LRA	0.991 6
WSRC <sup>[8]</sup>	0.944 3
ASRC <sup>[9]</sup>	0.947 1
LCSPT <sup>[10]</sup>	0.991 6
TWR + PCA <sup>[11]</sup>	0.991 0
DRR + KNN <sup>[12]</sup>	0.990 0
MFL <sup>[13]</sup>	0.918 6
NUPDWT	0.982 1
The Proposed	0.997 0

接下来,利用已有的 VGG16 和 ResNet18 模型来验证双通道卷积残差网络的有效性。VGG16 是一个 16 层的卷积神经网络,有 13 个卷积层和 3 个全连接层。ResNet18 是一个残差网络,共有 18 层,其中 1 个卷积层,4 个残差模块,2 个残差单元。这两个模型的架构如图 11 所示。在相同的配置下实现了 VGG16 和 ResNet18 两个模型,并在 FERET、AR、FEI 人脸数据集上与本文提出的模型进行了比较。表 4、表 5 和表 6 分别为 FERET 数据集、AR 数据集和 FEI 数据集的人脸识别准确率,可以看出双通道残差网络的精度有了明显的提高,GPU 时间比 VGG16 快 3 倍,比 ResNet18 快 1.5 倍。结果表明,本文的模型性能优于以前的模型。

表 4 在 FERET 数据集上不同算法的准确性比

Tab. 4 Comparison with different architecture on FERET

Method	Accuracy
VGGNet16	0.993 0
ResNet18	0.970 7
The Proposed	0.998 3

表 5 在 AR 数据集上不同算法的准确性比较

Tab. 5 Comparision with different architecture on AR

Method	Accuracy
VGGNet16	0.992 4
ResNet18	0.997 5
The Proposed	0.996 7

表 6 在 FEI 数据集上不同算法的准确性比较

Tab. 6 Comparision with different architecture on FEI

Method	Accuracy
VGGNet16	0.921 2
ResNet18	0.952 2
The Proposed	0.992 0

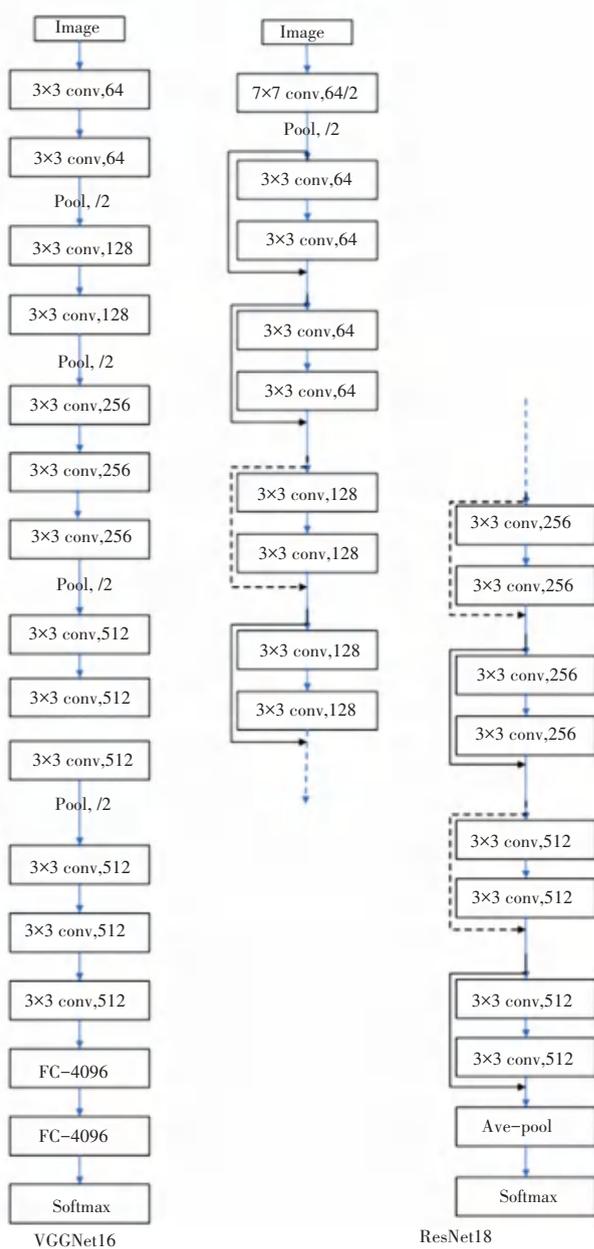


图 11 VGGNet16 和 ResNet18 网络结构图

Fig. 11 The architecture of VGGNet16 and ResNet18

为了进一步比较这些模型的性能,分析了训练和测试中的精度和损失等参数。AR 数据库中不同模型的结果如图 12 所示,比较中可以看出,本文的模型比其他模型收敛更快,获得了更高的精度。本文提出的模型在训练过程中训练集和验证集的收敛情况如图 13 所示,从图中可以看出,在训练 100 个 epoch 时,模型基本上已经收敛,准确性已经达到一个比较好的结果。

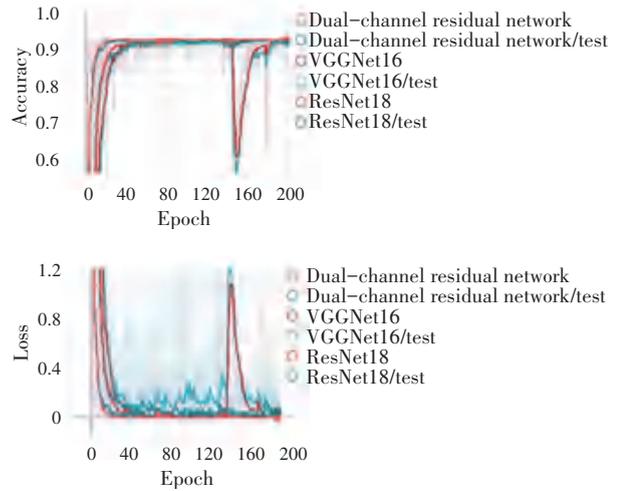


图 12 不同模型在 AR 数据集上的损失和准确率

Fig. 12 The results of different models on AR database

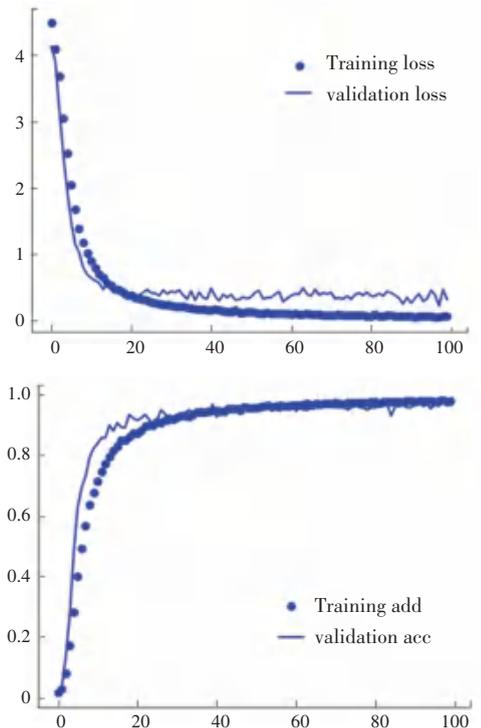


图 13 本文提出的模型在 AR 数据集上的收敛情况

Fig. 13 The results of different models on AR database

### 4 结束语

本文提出了一种双通道卷积残差网络算法用来做人脸识别,利用双通道输入,将人脸图像的特征充分的利用,sobel 算子能对提取人脸的细节特征更有鲁棒性。除此之外,该算法还利用到了残差网络,因此整个网络具有卷积网络和残差网络的优点,可以使模型收敛更快,达到更高的精度,还可以防止训练过程中梯度消失,导致网络无法正常训练。本文的模型在 FERET、AR、FEI 数据集上分别达到了 99.83%、99.67%、99.20% 的准确率,不仅超过了传统方法的性能,

而且与深度学习方法具有一定的竞争力。以上结果表明,本文的模型可以显著提高人脸识别率。此外,本文模型的每个 epoch 的时间大约比 VGG16 快 3 倍,比 ResNet18 快 1.5 倍,模型在硬盘上的存储空间仅占 19.7 MB 左右,不仅准确性高,速度快,需要的内存小,明显优于其他现有的算法,更具有实用价值。

### 参考文献

- [1] WU X, HE R, SUN Z. A lightened cnn for deep face representation [C]//2015 IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR). 2015, 4.
- [2] 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.
- [3] SOBEL I, FELDMAN G. A 3×3 isotropic gradient operator for image processing [J]. a talk at the Stanford Artificial Project in, 1968: 271–272.
- [4] XI M, CHEN L, POLAJNAR D, et al. Local binary pattern network; a deep learning approach for face recognition [C]//Image processing (ICIP), 2016 IEEE international conference on. IEEE, 2016: 3224–3228.
- [5] DING C, CHOI J, TAO D, et al. Multi-directional multi-level dual-cross patterns for robust face recognition [J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(3): 518–531.

- [6] 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.
- [7] DENG W, HU J, GUO J. Extended SRC: Undersampled face recognition via intraclass variant dictionary [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(9): 1864–1870.
- [8] LU C Y, MIN H, GUI J, et al. Face recognition via weighted sparse representation [J]. Journal of Visual Communication and Image Representation, 2013, 24(2): 111–116.
- [9] WANG J, LU C, WANG M, et al. Robust face recognition via adaptive sparse representation [J]. IEEE transactions on cybernetics, 2014, 44(12): 2368–2378.
- [10] UÇAR A. Color face recognition based on steerable pyramid transform and extreme learning machines [J]. The Scientific World Journal, 2014, 2014.
- [11] SHI X, GUO Z, NIE F, et al. Two-dimensional whitening reconstruction for enhancing robustness of principal component analysis [J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(10): 2130–2136.
- [12] LI H, SUEN C Y. Robust face recognition based on dynamic rank representation [J]. Pattern Recognition, 2016, 60: 13–24.
- [13] YANG M, ZHANG L, YANG J, et al. Metaface learning for sparse representation based face recognition [C]//Image Processing (ICIP), 2010 17th IEEE International Conference on. IEEE, 2010: 1601–1604.

(上转第 3 页)

CTARC-E 数据集上训练 Stanford AR 模型, Two-column finetuning 代表使用微调做迁移学习, Two-column PNN 代表使用原始的渐进神经网络做迁移学习, Two-column PNN+CRA 代表使用融合交叉注意力适配器的渐进神经网络做迁移学习。在应用渐进神经网络的设置时,使用 RACE 作为第一个任务,使用 CTARC-E 作为第二个任务。

表 3 实验结果

Tab. 3 Main results

实验设置	Dev	Test
RACE single	45.82	43.3
CTARC single in Chinese	30.95	27.78
CTARC single in English	40.48	31.48
Two-column finetuning	30.95	33.33
Two-column PNN	35.71	29.63
<b>Two-column PNN+CRA</b>	<b>38.01</b>	<b>37.04</b>

Stanford AR 模型在 RACE 数据集上的性能为 43.3%, 远低于人类上限性能 94.5%, 这意味着 RACE 是一个具有挑战性的数据集, 现有的模型可能会从数据集中学习到有限的知识。然而, 所有的迁移学习方法, 都有助于模型在单一中文数据上集提高性能。在这些方法中, 本文提出的框架获得了 10% 的显著提高。在迁移学习的三种设置中, Two-column

finetuning 优于 Two-column PNN 3.7%, 这表明使用原始的渐进神经网络对于处理阅读理解问题并不是一种有效的方法。本文所提出的方法的取得了最好的效果, 显示出了交叉注意力适配器的有效性。

### 3 结束语

基于迁移学习的方法利用从一个大型的英语数据集中转移的知识来帮助模型更有效地解决中文高考中的阅读理解任务。本文提出了一种具有交叉注意力适配器的渐进神经网络模型, 用于解决阅读理解任务。该方法在渐进神经网络模型的基础上进行了改进, 提供了一种跨越不同任务的注意力机制, 并使用长短记忆神经网络来表示所有注意力层的组合输出。实验结果表明, 与所有的基线模型相比, 该模型在解决目标中文阅读理解任务上的表现更好。

### 参考文献

- [1] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 22(10): 1345–1359.
- [2] RUSU A A, RABINOWITZ N C, DESJARDINS G, et al. Progressive neural networks [J]. arXiv preprint arXiv: 1606.04671, 2016.
- [3] SEO M, KEMHAVI A, FARHADI A, et al. Bidirectional attention flow for machine comprehension [J]. arXiv preprint arXiv: 1611.01603, 2016.