

DOI: 10.13382/j.jemi.B1801688

多尺度残差网络模型的研究及其应用^{*}

王飞¹ 张莹^{1,2} 邵豪¹ 张东波^{1,2} 牟清萍¹

(1.湘潭大学 信息工程学院 湘潭 411105; 机器人视觉感知与控制技术国家工程实验室 长沙 410082)

摘要: 针对传统的卷积神经网络不能充分利用图像的多尺度信息,以及网络层数的增加导致优化参数增加的问题。提出多尺度残差网络模型通过多尺度跨通道的卷积融合提高图像的特征表征能力;然后把大的卷积核分解成小的非对称卷积核降低网络的参数计算;接着利用残差网络原理来降低深层网络的梯度消失问题;最后将提出的多尺度卷积模块嵌入到 Lenet 网络中。在 Mnist 数据集上的测试结果证明分类准确率比原始 Lenet 网络提高了 0.24%,在 LFW 数据集上的测试结果表明分类准确率优于 DeepFace、WebFace 等传统算法。

关键词: 多尺度残差网络;卷积神经网络;跨通道卷积;核分解

中图分类号: TP391.73; TN911.41 文献标识码: A 国家标准学科分类代码: 520.60

Research and application of the multi-scale residual networks model

Wang Fei¹ Zhang Ying^{1,2} Shao Hao¹ Zhang Dongbo^{1,2} Mu Qingping¹

(1.College of Information Engineering, Xiangtan University, Xiangtan 411105, China;

2.National Engineering Laboratory for Robot Visual Perception and Control Technology, Changsha 410082, China)

Abstract: For the traditional convolutional neural network, the optimization parameter will increase as the number of network layer increases. Also, the multi-scale information of the image cannot be made full of use. To solve these problems, the multi-scale residual networks model is proposed to improve the feature characterization through convolution fusion of multiscale and cross channel. Then the large convolution kernel is decomposed into a small asymmetric convolution kernel in order to reduce network parameter calculation, the residual networks is used to reduce the impact of the gradient disappearance. Finally, The proposed multi-scale convolutional module is embedded in Lenet network for image classification. The experimental results on the Mnist dataset show that the classification accuracy is improved 0.24% than the original Lenet network. The experimental results on the LFW dataset show that the classification accuracy of the proposed model is better than DeepFace, WebFace and other traditional algorithms.

Keywords: multi-scale residual networks; convolutional neural networks; cross channel convolution; kernel decomposition

0 引言

随着人脸识别应用的广泛,人脸识别的准确性和稳定性要求也越来越高。传统的人脸识别方法主要依赖先验知识,通过人工方式提取浅层特征在进行分类,如 SIFT、HOG 算法^[1],但是浅层特征信息表达能力相对较差,容易受到光照、姿态和表情的干扰^[2-3]。2012年, Krizhevsky 等^[4]提出的卷积神经网络(AlexNet)显示了在

特征提取方面的优越性,因而被广泛应用于视觉领域,例如目标识别^[5-6]、目标检测^[7-10]、目标跟踪^[11-12]等。相对于传统的人工设计特征方法,卷积神经网络(convolutional neural networks, CNN)能够提取人脸更深层次特征^[13]。研究发现网络架构的深度对网络性能的提升至关重要^[14-15],但随着网络层数的增加,滤波器的尺寸选择将变得困难^[16]、网络的优化计算变得复杂^[17]。针对上述问题,文献[18-21]提出了具有固定组织的多个卷积层组的 Inception 模块,这些模块可与附加的自组织组

收稿日期: 2018-09-03 Received Date: 2018-09-03

^{*} 基金项目: 国家自然科学基金(61773330)、湖南省重点学科基金、湖南省自然科学基金(2017JJ2251)资助项目

合形成完整的网络,提高了网络的性能。

传统的 CNN 不能够充分利用人脸图像的多尺度信息^[22-23],每层只能提取一种尺度的特征信息,为了得到更多的特征,通常的做法是加深网络层数,容易出现过拟合,也意味着计算资源的增加,导致网络的训练优化变得复杂。针对这些问题,本文提出多尺度残差网络模型(multi-scale residual nets, MRResNet),网络每一层通过并联不同大小的卷积核,以充分提取不同尺度的人脸特征信息,同时采用残差模块的思想,使得网络更易于训练优化。

1 多尺度卷积网络

在图像特征提取中,单一尺寸的滤波器一次只能提取一种尺度的特征,如果在同一模块层同时使用多种尺寸的卷积核,可以增加对图像变化的鲁棒性。同时将一个大的卷积核分解为非对称小卷积核,可以有效降低需要学习的参数。根据上述设计思想,本文提出如下 4 种设计卷积网络的策略:1) 使用 1×1 的卷积核,构成跨通道卷积,压缩特征数量;2) 同一模块并联使用多种卷积核,构成多尺度模块来增加鲁棒性;3) 将大的卷积核分解成小卷积核,降低参数计算量;4) 使用跨层连接,缓解梯度消失问题。

1.1 多尺度跨通道卷积

图片数据的临近区域相关性极高,在同一空间位置不同通道的卷积核输出结果的相关性也极高^[24],因此可以用一个 1×1 ^[25] 的卷积核把这些相关性的特征、在同一空间位置但是不同通道的特征连接在一起,激活函数采用非线性修正单元(ReLU)^[26],网络结构如图 1 所示。其中 1×1 卷积得到的特征图可以按式(1)计算得到。



图 1 跨通道卷积特征图

Fig.1 Cross-channel convolution feature map

$$f_{ijk} = \max(\mathbf{w}_k^T \cdot \mathbf{x}_{ijk} + b_k, 0) \quad (1)$$

式中: f_{ijk} 表示 1×1 卷积核对上—层特征图卷积得到并经过激活函数处理后的特征图; x_{ijk} 表示第 k 通道上第 i 行第 j 列对应的元素值; b 表示偏执项; k 表示通道的索引; \mathbf{w}_k^T 表示对应第 k 通道第 i 行第 j 列元素对应卷积核的权重。

为了在同一层能够得到多尺度的特征,通过同一模块层同时使用不同尺寸的卷积核,提取不同尺度的特征,提高对尺度变化的适应性,不同尺寸卷积核提取特征的过程如图 2 所示。为了验证设计原则 1) 和 2) 的有效性,本文对 Lenet 网络(https://github.com/BVLC/caffe/blob/master/examples/mnist/lenet_train_test.prototxt) 进行改进,并在 Mnist 数据集上(<http://yann.lecun.com/exdb/mnist/>) 进行实验。分别并联使用 1×1 、 3×3 、 5×5 三种不同的卷积核,网络模型的第一层卷积特征图可视化如图 3 所示,可知不同尺寸卷积核得到的特征图不一样,充分说明可以提取不同尺度的特征。网络结构和测试结果如表 1 所示,Lenet2 网络的准确率最高,故并联使用不同尺寸的卷积核的网络性能要优于原始使用单一尺寸网络的性能。

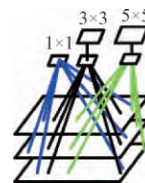
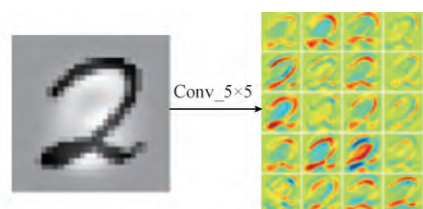
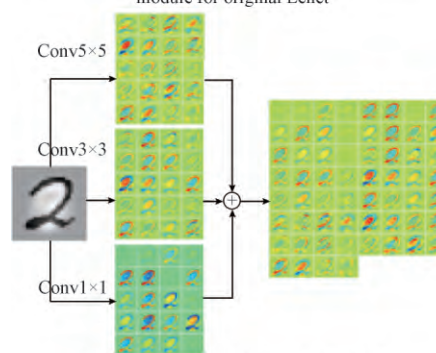


图 2 多尺度跨通道卷

Fig.2 Multiscale cross-channel convolution feature map



(a) 原始Lenet第一卷积模块可视化特征图
(a) The visual feature map of the first convolution module for original Lenet



(b) Lenet2第一卷积模块可视化特征图
(b) The visual feature map of the first convolution module for Lenet2

图 3 原始 Lenet 以及多尺度变体第一卷积模块层可视化特征图

Fig.3 The visual feature map of the first convolutional module layer for original Lenet and multi-scale convolution kernel variant

表1 网络结构和测试结果

Table 1 Structure and results for original Lenet and multiscale cross-channel variation convolutional network

Layers	Conv1	Conv2	准确率/%
Original Lenet	$\begin{bmatrix} k=5 \times 5 \\ p=0 \times 0 \end{bmatrix}$	$\begin{bmatrix} k=5 \times 5 \\ p=0 \times 0 \end{bmatrix}$	99.08
Lenet1	$\begin{bmatrix} k=5 \times 5 \\ p=0 \times 0 \end{bmatrix}$	$\begin{bmatrix} k=1 \times 1 \\ p=0 \times 0 \\ k=3 \times 3 \\ p=1 \times 1 \\ k=5 \times 5 \\ p=2 \times 2 \end{bmatrix} \times 1$	99.13
Lenet2	$\begin{bmatrix} k=1 \times 1 \\ p=0 \times 0 \\ k=3 \times 3 \\ p=1 \times 1 \\ k=5 \times 5 \\ p=2 \times 2 \end{bmatrix} \times 1$	$\begin{bmatrix} k=1 \times 1 \\ p=0 \times 0 \\ k=3 \times 3 \\ p=1 \times 1 \\ k=5 \times 5 \\ p=2 \times 2 \end{bmatrix} \times 1$	99.19

1.2 卷积核分解

卷积操作是卷积核和特征图进行的全连接操作(图4)。3×3 卷积核的级联可以获得 5×5 卷积核同样大小的感受野,但需要计算的参数是原来的 9/25=0.36,显著降低了需要学习的参数。如果在两个 3×3 卷积核之间再使用一次激活函数,其特征表达能力比单个 5×5 的卷积核更好。本文提出 3×3 的卷积核分解为 3×1 和 1×3 的卷积核级联,这样将卷积核需要计算 9 个参数降为只需要计算 6 个参数,而且比 3×3 卷积核多了一层非线性处理,可以增强细小特征的表达能力^[27]。3×3 卷积核分解如图5所示,5×5 卷积核的分解如图6所示。

为了研究非对称大卷积核、对称小卷积核、非对称小卷积核对网络性能的作用,按照前面原则3)的设计思想,对原始的 Lenet 网络进行了改进,将 5×5 卷积核进行



图4 卷积操作示意图

Fig.4 The schematic diagram of the convolution operation



图5 3×3 卷积核分解示意图

Fig.5 3×3 convolution kernel decomposition schematic

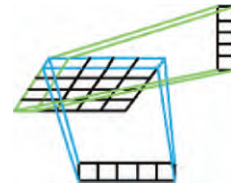


图6 5×5 卷积核分解示意图

Fig.6 5×5 convolution kernel decomposition schematic

了多种形式的分解,第一卷积模块层的可视化特征图如图7所示,图7(b)表示直接原始 Lenet 中的 5×5 的卷积核分解成 5×1 和 1×5 卷积核,图7(c)为直接将原始的 Lenet 中的每个 5×5 卷积核均分解成两个 3×3 的卷积核对应的特征图,图7(d)是将图7(c)中的卷积核进一步分解成非对称小卷积核得到的特征图,网络参数和测试结果如表2所示。

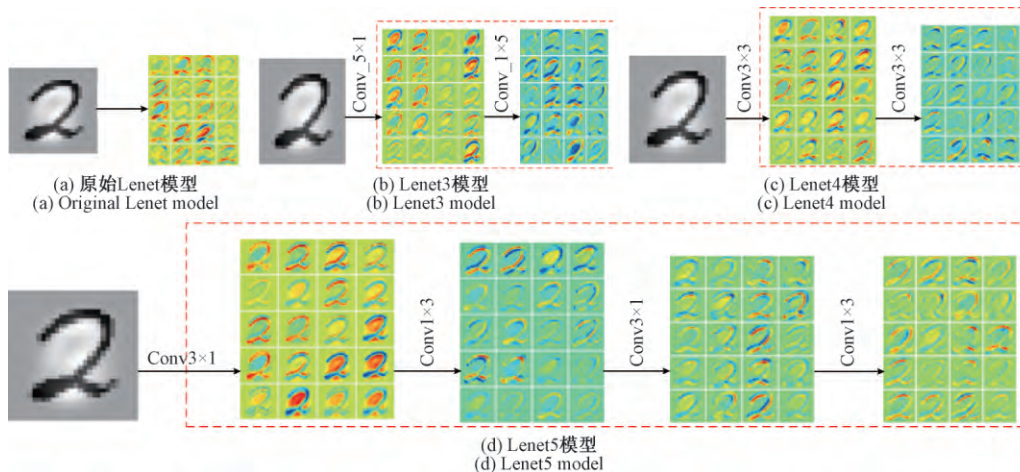


图7 原始 Lenet 以及分解卷积核变体第一卷积模块层可视化特征图

Fig.7 The visual feature map of the first convolutional module layer for original Lenet and decomposed convolution kernel variant

表 2 网络参数和测试结果

Table 2 Parameters and results for original Lenet and multiscale cross-channel variation convolutional network

Layers	Conv1	Conv2	准确率/%
Original Lenet	$\begin{bmatrix} k = 5 \times 5 \\ p = 0 \times 0 \end{bmatrix}$	$\begin{bmatrix} k = 5 \times 5 \\ p = 0 \times 0 \end{bmatrix}$	99.08
Lenet3	$\left[\begin{bmatrix} k = 5 \times 1 \\ p = 0 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 5 \\ p = 0 \times 0 \end{bmatrix} \right] \times 1$	$\left[\begin{bmatrix} k = 5 \times 1 \\ p = 0 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 5 \\ p = 0 \times 0 \end{bmatrix} \right] \times 1$	99.07
Lenet4	$\begin{bmatrix} k = 3 \times 3 \\ p = 0 \times 0 \end{bmatrix} \times 2$	$\begin{bmatrix} k = 3 \times 3 \\ p = 0 \times 0 \end{bmatrix} \times 2$	99.16
Lenet5	$\left[\begin{bmatrix} k = 3 \times 1 \\ p = 0 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 3 \\ p = 0 \times 0 \end{bmatrix} \right] \times 2$	$\left[\begin{bmatrix} k = 3 \times 1 \\ p = 0 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 3 \\ p = 0 \times 0 \end{bmatrix} \right] \times 2$	99.22

对比表 2 的 4 种模型以及图 7 的可视化特征图可以发现, 将 5×5 的卷积核直接分解成 5×1 和 1×5 卷积核, 特征图的噪声增强了, 模型性能低于原始 Lenet 模型。将 5×5 卷积核分解成两个 3×3 卷积核级联得到的特征图比原始 Lenet 模型的第一层卷积模块得到的特征图更平滑, 有更低的噪声, 分类准确率有提高。如果进一步将 3×3 的卷积核分解成两个 3×1 和 1×3 的卷积核级联, 可以得到噪声更小、更平滑的特征, 分类准确率最高。

1.3 多尺度小卷积核

同一卷积模块同时并联 1×1 、 3×3 、 5×5 三种感受野大小的卷积核和把 5×5 的卷积核分解成两个 3×3 或者分解成两对 3×1 、 1×3 的卷积核均有有助于网络模型性能的提升, 因而将 Lenet2 网络中的 5×5 的卷积核分解成

两个 3×3 的卷积核得到 Lenet7 网络, 将 Lenet2 中的 3×3 、 5×5 分解成非对称的 3×1 和 1×3 、 5×1 和 1×5 得到 Lenet8, 进而将 Lenet7 中的 3×3 卷积核分解成 3×1 和 1×3 得到 Lenet9, 对应的第一卷积模块可视化特征如图 8 所示, 图中虚线框的感受野同 5×5 卷积核等效, 但比单个 5×5 卷积核的特征表达能力更强, 同时需要学习的参数大幅度降低。图 8(d) 同时并联使用不同尺寸的卷积核能提取的特征数量更多, 而且最后得到的特征层次相对更抽象。网络架构参数和测试结果如表 3 所示。Lenet9 模型使用的多尺度非对称小卷积核得到的特征图噪声最低, 分类准确率最高, 故并联使用 1×1 、 3×1 、 1×3 非对称的小卷积核的性能最优。

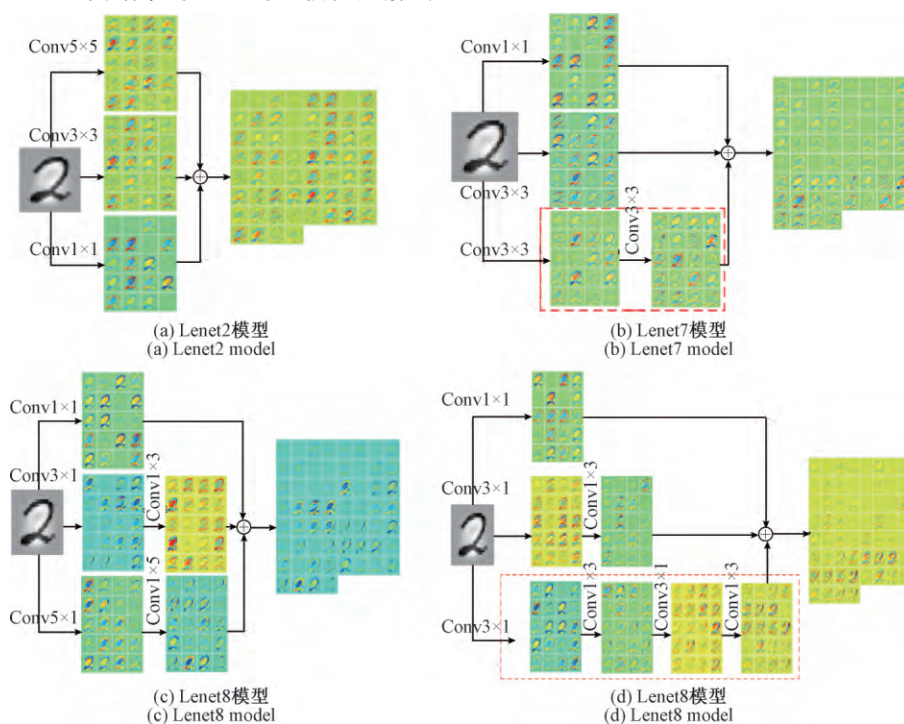


图 8 原始 Lenet 以及多尺度小卷积核变体第一卷积模块层可视化特征图

Fig.8 Visual feature map of the first convolutional module of the original Lenet and multi-scale small convolution kernel

表 3 多尺度小卷积核变体网络架构与结果

Table 3 Network architecture and results for the multi-scale convolution kernel variant

layers	Conv1	Conv2	准确率/%
Lenet2	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 3 \\ p = 1 \times 1 \end{bmatrix} \\ \begin{bmatrix} k = 5 \times 5 \\ p = 2 \times 2 \end{bmatrix} \end{bmatrix} \times 1$	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 3 \\ p = 1 \times 1 \end{bmatrix} \\ \begin{bmatrix} k = 5 \times 5 \\ p = 2 \times 2 \end{bmatrix} \end{bmatrix} \times 1$	99.08
Lenet6	$\begin{bmatrix} k = 5 \times 5 \\ p = 0 \times 0 \end{bmatrix}$	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 3 \\ p = 1 \times 1 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 3 \\ p = 1 \times 1 \end{bmatrix} \times 2 \end{bmatrix} \times 1$	99.15
Lenet7	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 3 \\ p = 1 \times 1 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 3 \\ p = 1 \times 1 \end{bmatrix} \times 2 \end{bmatrix} \times 1$	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 3 \\ p = 1 \times 1 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 3 \\ p = 1 \times 1 \end{bmatrix} \times 2 \end{bmatrix} \times 1$	99.2
Lenet8	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 1 \\ p = 1 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 3 \\ p = 0 \times 1 \end{bmatrix} \\ \begin{bmatrix} k = 5 \times 1 \\ p = 2 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 5 \\ p = 0 \times 2 \end{bmatrix} \end{bmatrix} \times 1$	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 1 \\ p = 1 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 3 \\ p = 0 \times 1 \end{bmatrix} \\ \begin{bmatrix} k = 5 \times 1 \\ p = 2 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 5 \\ p = 0 \times 2 \end{bmatrix} \end{bmatrix} \times 1$	99.15
Lenet9	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 1 \\ p = 1 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 3 \\ p = 0 \times 1 \end{bmatrix} \\ \begin{bmatrix} \begin{bmatrix} k = 3 \times 1 \\ p = 1 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 3 \\ p = 0 \times 1 \end{bmatrix} \end{bmatrix} \times 2 \end{bmatrix} \times 1$	$\begin{bmatrix} \begin{bmatrix} k = 1 \times 1 \\ p = 0 \times 0 \end{bmatrix} \\ \begin{bmatrix} k = 3 \times 1 \\ p = 1 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 3 \\ p = 0 \times 1 \end{bmatrix} \\ \begin{bmatrix} \begin{bmatrix} k = 3 \times 1 \\ p = 1 \times 0 \end{bmatrix} \rightarrow \begin{bmatrix} k = 1 \times 3 \\ p = 0 \times 1 \end{bmatrix} \end{bmatrix} \times 2 \end{bmatrix} \times 1$	99.32

比较表 1 的 Lenet1 和表 3 的 Lenet6 可以发现,将 5×5 的卷积核分解成 3×3 的卷积核提高了分类准确率,从表 3 中的 Lenet7 和 Lenet8 以及图 8 对应的可视化特征图可知,Lenet7 卷积模块得到的特征图噪声低于 Lenet8 的卷积模块,且特征图更平滑。

2 多尺度残差网络

第 2 节验证了多尺度小卷积模块对 Lenet 网络性能提升的有效性,然而浅层网络模型对人脸图像的特征表达能力远远不够,但是网络深度的增加会导致网络参数

的求解变得困难^[28],残差网络针对此问题,提出将上一层的网络特征直接跨过多层卷积模块映射到下层网络,能有效缓解梯度消失问题。因此首先根据 Lenet2 的设计思想和原则 4) 设计原则,设计了图 9(a) 的多尺度残差模块,标有 Maxout 的表示使用 Maxout^[29] 激活函数,未标出的表示使用 ReLU 激活函数。为了降低特征损失,采用步长为 2 的卷积核和步长为 2 的最大池化层对网络进行降维操作,模块如图 9(b) 所示。图 9 的 n 值表示相应卷积核得到的特征图个数。本文将这种网络架构命名为 MRResNet-A,其网络模型如图 10 所示,其中的模块 A 和模块 C 的架构如图 9(a) 和 (b) 所示,各个模块卷积核输出特

征图个数以及每个模块的参数如表 4 所示,其中 n_i ($i=1, 2, \dots, 7$) 表示网络中使用的第 m 个多尺度残差模块, C_m ($m=1, 2, \dots, 7$) 表示图 9 中相应卷积核的输出特征图的个数, A_m ($m=1, 2, 3, 4$) 表示网络中使用的第 n 个降维模块。

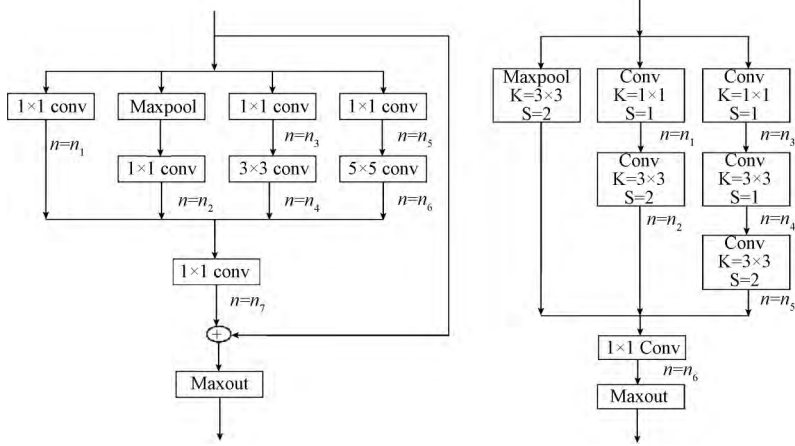


图 9 MResNet-A 的多尺度残差模块和降维模块

Fig.9 The multi-scale residual module and dimension reduction module for the MResNet-A

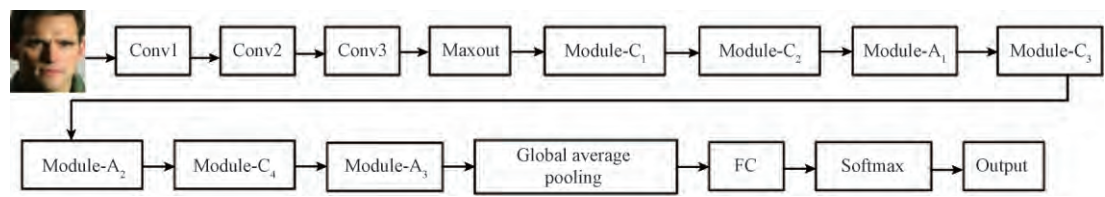


图 10 MResNet-A 网络模型架构示意图

Fig.10 The schematic diagram of MResNet-A model

表 4 MResNet-A 各模块卷积核的输出特征图个数及需要学习参数

Table 4 The number of output feature map and the parameters that need to be learned of the convolution kernel for MResNet-A model

Model	n_1	n_2	n_3	n_4	n_5	n_6	n_7	参数($\times 10^3$)
C_1	32	64	32	64	128	192	—	159.744
C_2	96	192	96	128	192	384	—	571.392
A_1	96	96	96	128	96	192	384	841.728
C_3	96	256	96	128	256	512	—	1 024
A_2	128	128	128	192	128	256	512	1 531.904
C_4	192	384	192	256	512	1024	—	3 563.52
A_3	256	256	256	384	256	512	1024	6 127.616

表 5 MResNet-B 各模块卷积核的输出特征图个数及需要学习参数

Table 5 The number of output feature map and the parameters that need to be learned of the convolution kernel for MResNet-B

Model	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	n_{11}	Parameter
C_1	32	64	32	64	128	192	—	—	—	—	—	159.744
C_2	96	192	96	128	192	384	—	—	—	—	—	571.392
B_1	96	96	96	128	128	96	128	128	192	192	384	1 290.688
C_3	96	256	96	128	256	512	—	—	—	—	—	1 024
B_2	128	128	128	192	192	128	192	192	256	256	512	1 204.224
C_4	192	384	192	256	512	1024	—	—	—	—	—	3 563.52
B_3	256	256	256	384	384	256	384	384	512	512	1 024	4 816.896

从表4可以看出,每个模块的参数数量很大,为了进一步降低 MResNet-A 网络架构的参数,本文综合 1)~4) 的设计原则,按照 Lenet9 模型架构的设计方法,设计了如图 11 所示的多尺度小卷积残差模块 MResNet-B,降维模块仍然使用图 9(b) 所示的降维模块架构。图 11 中的多尺度小卷积残差模块各卷积核对应的输出特征个数及参数如表 5 所示。可以看出,虽然模块 B_1 、 B_2 、 B_3 比对应的模块 A_1 、 A_2 、 A_3 使用了更多的卷积核,但是参数量均要小于 A_1 、 A_2 、 A_3 的参数量,图 9(a) 和图 11 中的 1×1 的卷积核步长均为 1, 3×3 、 5×5 、 1×3 、 3×1 采用零填充策略。

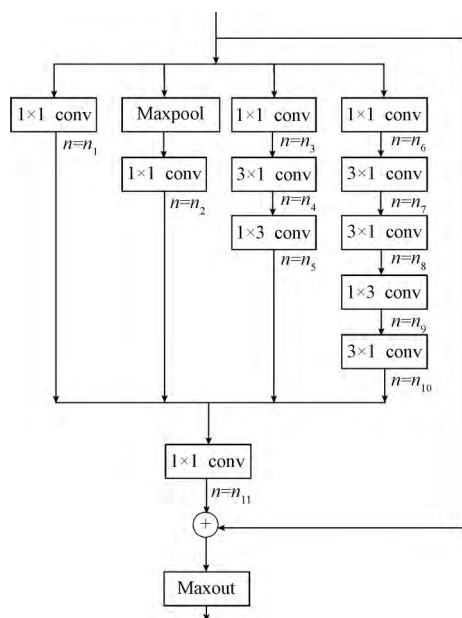


图 11 MResNet-B 多尺度小卷积残差模块

Fig.11 The multiscale small convolution residual module for MResNet-B

3 实验及数据分析

为了验证本文提出的算法性能,采用 MS-Celeb-1Mv1^[30] 人脸数据集来训练 MResNet-A 和 MResNet-B 模型,MS-Celeb-1Mv1 数据集包含了 99 891 个人,每个人有多张不同的图片,一共大约 10 000 000 张图片。然后在 LFW^[31] 数据集上来评估模型的性能,LFW 数据集包含 5 749 个人,一共 13 233 张图片。

3.1 图像预处理

首先利用 MTCNN^[32] 算法对 MS-Celeb-1Mv1 和 LFW 数据集检测面部的关键点,包括 2 个眼部中点、鼻尖、两个嘴角,以及每张人脸的边框信息,然后利用检测到的边框信息将人脸的面部图像裁剪出来(图 12)。实验将图像裁剪成 165×165 ,作为网络的输入训练 MResNet-A 和

MResNet-B 网络。



图 12 人脸图像预处理

Fig.12 Face image preprocessing

3.2 模型训练

随机选取 MS-Celeb-1Mv1 数据集的 75% 作为训练,剩下的 25% 作为验证,并将输入的 165×165 大小的图像进行随机裁剪 149×149 的图像作为第一个卷积层的输入。初始学习率从 0.01 开始,按照步长逐渐降低学习率。

3.3 性能评估

本文将 LFW 数据集分割成 6 000 对人脸图像,每对包含两张图片,直接使用在 MS-Celeb-1Mv1 数据集上训练的模型文件来提取每张图片的特征,为了同现有算法进行性能比较,采用 10 倍交叉验证计算这两个特征的余玄距离的相似性。表 6 所示为本文算法以及现有算法在 LFW 上的识别效果,图 13 所示为本文算法在 LFW 数据集上的 ROC 曲线。

表 6 各算法在 LFW 上的评估性能

Table 6 The evaluation performance of each algorithm on LFW

算法	准确率/%
MResNet-A	98.26
MResNet-B	99.13
DeepFace ^[33]	98.95
Webface ^[34]	97.73
Web-scale ^[35]	98.37
VIPLFaceNet ^[36]	98.62

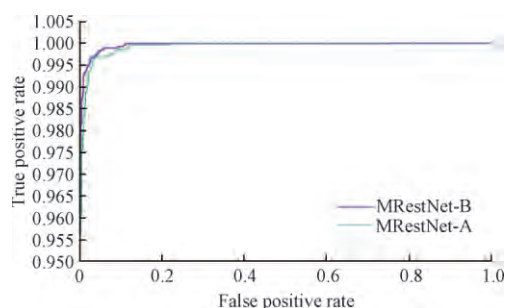


图 13 MResNet-A 和 MResNet-B 在 LFW 上的 ROC 曲线

Fig.13 The ROC curves for MResNet-A and MResNet-B on the LFW

3.4 实验结果分析

通过图 9 和 11 可以发现,模型 MResNet-B 中的每个残差模块比模型 MResNet-A 的每个残差模块使用了更多的卷积核,级联了更深的层数,从而提高了特征的表达能力。从表 6 可知,MResNet-B 的性能要比 MResNet-A 提高近 1%,而对比表 4 和 5 可以发现,MResNet-B 的参数计算量要小于模型 MResNet-A,充分说明将大卷积核分解成小卷积核,不仅能够降低参数计算,而且能够提高网络架构的特征表达能力。由图 13 可以发现,MResNet-B 模型的 ROC 曲线与横坐标所围成的面积要高于 MResNet-A 模型的 ROC 曲线与横坐标围成的面积,间接说明了 MResNet-B 网络模型的诊断准确性高于 MResNet-A 模型的准确性。

4 结 论

本文提出在每一卷积模块并联使用多个不同大小的卷积核提取不同尺度的特征,增加了模型对复杂环境变化的适用性,并通过将大卷积核分解成小卷积核级联降低模型参数的计算量,且根据残差网络的思想直接将上一模块层的特征直接跨过多尺度模块映射到下一层来减少梯度消失问题带来的影响,提高了图像分类的准确率。最近谷歌团队针对人脸识别问题直接使用 Triplets^[37] 损失函数直接将人脸图象映射到欧几里得空间,也得到了很好的效果。在未来的研究工作中,将针对此问题进行研究,进一步提高人脸识别的精度。

参考文献

- [1] 祝军,赵杰煜,董振宇. 融合显著信息的层次特征学习图像分类[J]. 计算机研究与发展, 2014, 51(9): 1919-1928.
ZHU J, ZHAO J Y, DONG ZH Y. Image classification using hierarchical feature learning method combined with image saliency [J]. Journal of Computer Research and Development 2014, 51(9): 1919-1928.
- [2] 张盼,练秋生. 融合整体与局部特征的低秩松弛协作表示[J]. 计算机研究与发展, 2014, 51(12): 2663-2670.
ZHANG P, LIAN Q SH. Low-rank relaxed collaborative representation combined with global and local features for face recognition [J]. Journal of Computer Research and Development 2014, 51(12): 2663-2670.
- [3] HU C, LU X, YE M, et al. Singular value decomposition and local near neighbors for face recognition under varying illumination [J]. Pattern Recognition, 2017(64): 60-83.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]. Advances in Neural Information Processing Systems. New York: Curran Associates Inc, 2012: 1097-1105.
- [5] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition [C]. European Conference on Computer Vision. Berlin: Springer, 2016: 499-515.
- [6] 随婷婷,王晓峰. 一种基于 CLMF 的深度卷积神经网络模型[J]. 自动化学报, 2016, 42(6): 875-882.
SUI T T, WANG X F. Convolutional neural networks with candidate location and multi-feature fusion [J]. Acta Automatica Sinica 2016, 42(6): 875-882.
- [7] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn [C]. IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]. European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.
- [9] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [C]. Advances in Neural Information Processing Systems. New York: Curran Associates Inc, 2015: 91-99.
- [10] 尹宏鹏,陈波,柴毅,等. 基于视觉的目标检测与跟踪综述[J]. 自动化学报, 2016, 42(10): 1466-1489.
YIN H P, CHEN B, CHAI Y, et al. Vision-based object detection and tracking: A review [J]. Acta Automatica Sinica 2016, 42(10): 1466-1489.
- [11] WANG L, OUYANG W, WANG X, et al. Visual tracking with fully convolutional networks [C]. International Conference on Computer Vision. Piscataway, IEEE 2015: 3119-3127.
- [12] SUN C, LU H, YANG M H. Learning spatial-aware regressions for visual tracking [J]. Computer Vision and Pattern Recognition, 2017, arXiv: 1706.07457.
- [13] 王飞,张莹,张东波,邵豪,成超鹏. 基于捷径的卷积神经网络在人脸识别中的应用研究[J]. 电子测量与仪器学报, 2018, 32(4): 80-86.
WANG F, ZHANG Y, ZHANG D B, et al. Research on application of convolutional neural networks in face recognition based on shortcut connection [J]. Journal of Electronic Measurement and Instrumentation, 2018, 32(4): 80-86.
- [14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. Computer Vision and Pattern Recognition, 2014, arXiv: 1409.1556.
- [15] 王佩琪,高原,刘振宇,等. 深度卷积神经网络的数

- 据表示方法分析与实践[J]. 计算机研究与发展, 2017, 54(6): 1348-1356.
- WANG P Q, GAO Y, LIU ZH Y, et al. A comparison among different numeric representation in deep convolution neural networks [J]. Journal of Computer Research and Development 2017, 54(6): 1348-1356.
- [16] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size [J]. Computer Vision and Pattern Recognition, 2016, arXiv: 1602.07360.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. IEEE conf on Computer Vision and Pattern Recognition. Piscataway, 2016: 770-778.
- [18] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]. IEEE conf on Computer Vision and Pattern Recognition. Piscataway 2015: 1-9.
- [19] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]. International Conference on Machine Learning, 2015: 448-456.
- [20] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]. IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, 2016: 2818-2826.
- [21] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]. AAAI, 2017: 12.
- [22] 周晔, 张军平. 基于多尺度深度学习的商品图像检索[J]. 计算机研究与发展, 2017, 54(8): 1824-1832.
- ZHOU Y, ZHANG J P. Multi-scale deep learning for product image search [J]. Journal of Computer Research and Development 2017, 54(8): 1824-1832.
- [23] ZHU C, ZHENG Y, LUU K, et al. CMS-RCNN: Contextual multi-scale region-based CNN for unconstrained face detection [C]. Deep Learning for Biometrics. Cham: Springer, 2017: 57-79.
- [24] ARORA S, BHASKARA A, GE R, et al. Provable bounds for learning some deep representations [C]. International Conference on Machine Learning, 2014: 584-592.
- [25] LIN M, CHEN Q, YAN S. Network in network [J]. Neural and Evolutionary Computing, 2013, arXiv: 1312.4400.
- [26] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [C]. Fourteenth International Conference on Artificial Intelligence and Statistics, 2011: 315-323.
- [27] MA J, WANG W, WANG L. Irregular convolutional neural networks [J]. Computer Vision and Pattern Recognition, 2017, arXiv: 1706.07966.
- [28] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Highway networks [J]. Machine Learning, 2015, arXiv: 1505.00387.
- [29] GOODFELLOW I J, WARDE-FARLEY D, MIRZA M, et al. Maxout networks [J]. Machine Learning, 2013, arXiv: 1302.4389.
- [30] GUO Y, ZHANG L, HU Y, et al. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition [C]. European Conference on Computer Vision. Berlin: Springer, 2016: 87-102.
- [31] HUANG G B, RAMESH M, BERG T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments [R]. Technical Report 07-49, Amherst: University of Massachusetts, 2007.
- [32] 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.
- [33] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep face recognition [C]. BMVC. 2015, 1(3): 6.
- [34] YI D, LEI Z, LIAO S, et al. Learning face representation from scratch [J]. Computer Vision and Pattern Recognition 2014, arXiv: 1411.7923.
- [35] TAIGMAN Y, YANG M, RANZATO M A, et al. Web-scale training for face identification [C]. IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, 2015: 2746-2754.
- [36] LIU X, KAN M, WU W, et al. VIPLFaceNet: An open source deep face recognition SDK [J]. Frontiers of Computer Science, 2017, 11(2): 208-218.
- [37] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering [C]. IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, 2015: 815-823.

作者简介



王飞, 2018 年于湘潭大学获硕士学位, 现为中国科学院博士研究生, 主要研究方向为深度学习、模式识别与智能信息处理。

E-mail: sdauwangfei2010@163.com

Wang Fei received his M.Sc. degree from Xiangtan University in 2018. Now he is a Ph.D. candidate in Chinese Academy of Sciences. His main research interests include deep learning, pattern recognition and intelligent information processing.



张莹,2010 年于湖南大学获得博士学位,现为湘潭大学副教授,主要研究方向为模式识别、图像处理与视觉导航。

E-mail: 122880704@qq.com

Zhang Ying received his Ph.D. degree from Hunan University in 2010. Now he is an associate professor in Xiangtan University. His main research interests include pattern recognition, image processing and visual navigation.



邵豪,2018 年于湘潭大学获硕士学位,现为广州 LOTES 公司工程师,主要研究方向为机器学习、模式识别与智能信息处理。

Shao Hao received her M.Sc. degree from Xiangtan University in 2018. Now she is an engineer in LOTES Guangzhou Co., Ltd. Her main research interests include machine learning, pattern recognition and intelligent information processing.



张东波,2007 年于湖南大学获博士学位,现为湘潭大学教授,主要研究方向为机器学习、机器视觉感知。

Zhang Dongbo received his Ph. D. degree from Hunan University in 2007. Now he is a professor in Xiangtan University. His main research interests include machine learning, machine vision and machine perception.



牟清萍,2017 年于湘潭大学学士学位,现为湘潭大学硕士研究生,主要研究方向为深度学习,目标检测与视觉跟踪。

Mu Qingping received her B.Sc. degree from Xiangtan University in 2017. Now she is a M.Sc. candidate in Xiangtan University. Her main research interests include deep learning, object detection, and visual tracking.

是德科技与国家信息光电子创新中心成立硅基光电子测试测量共建实验室

是德科技(NYSE: KEYS)近日宣布与国家信息光电子创新中心(NOEIC)在武汉光谷成立硅基光电子测试测量共建实验室。这也是国内首家拥有最高级别光通信测试能力的实验室,目前该实验室具备 400G 硅光,超 100GBaud 复杂相干调制等热门光通信领域测试能力,必将会为光谷产业前沿科技的加速应用带来显著影响力,并成为国内一流的信息光电子产业创新平台。

2019 年 4 月 24 日,国家信息光电子创新中心在东湖高新区烽火科技园揭牌。国家信息光电子创新中心采用“公司+联盟”模式运营,汇聚了光迅科技、烽火通信、亨通光电、武汉光电工业技术研究院等 30 余家行业骨干企业和院所高校资源。国家信息光电子创新中心主要面向新一代网络、数据中心光互联、5G 等信息光电子应用领域,力争建成国际一流的信息光电子制造业创新平台。目标在高端材料生长、核心芯片工艺、先进封装集成等方面突破关键技术和共性技术瓶颈,打通先进制造技术从基础研究到应用研究、首次商业化和规模化生产的创新链条,促进行业关键共性技术向规模化、经济高效的制造能力转化。

是德科技全球销售高级副总裁 Mark Wallace 说:“很高兴能见证是德科技与国家信息光电子创新中心这一合作,客户的成功一直都是是德科技的成功,这一平台的建立也会进一步推动我们双方在此领域更广阔更紧密的合作。”

是德科技网络与数据中心解决方案副总裁 Joachim Peerlings 博士说:“光通信领域的创新在连接世界、实现

5G、数据中心和电信服务货币化的今天和未来都至关重要,凭借我们在行业内领先的光通信和高速数字测试解决方案以及专业的测试和测量知识,是德科技必将与国家信息光电子创新中心共同加速光通信领域的创新发展。”

国家信息光电子创新中心常务副总经理肖希博士说:“硅基光电子技术是未来光电子技术的重要发展方向,能够有效突破微电子和光电子目前面临的发展瓶颈,也是国家信息光电子创新中心的重点攻关方向。此次创新中心与是德科技成立的联合实验室将瞄准硅光芯片测试测量所面临的共性技术问题,共同探索更加标准化和自动化的光电芯片测试测量方案。此次合作是创新中心构建公共开放的技术服务平台的重要举措,必将为加快信息光电子技术成果转化、构建成熟完整硅光产业链条作出重要贡献。”

关于是德科技

是德科技(NYSE: KEYS)是一家领先的技术公司,致力于帮助企业、服务提供商和政府客户加速创新,创造一个安全互联的世界。从设计仿真、原型验证、生产测试到网络和云环境的优化,是德科技提供了全方位的测试与分析解决方案,帮助客户深入优化网络,进而将其电子产品以更低的成本、更快地推向市场。我们的客户遍及全球通信生态系统、航空航天与国防、汽车、能源、半导体和通用电子终端市场。2018 财年,是德科技收入达 39 亿美元。更多信息,请访问 www.keysight.com。