

光学学报

VGG16 衍射光子神经网络及情境依赖处理

赵兴亚, 杨志伟, 戴键, 张天*, 徐坤

北京邮电大学信息光子学与光通信国家重点实验室, 北京 100876

摘要 针对电子计算中摩尔定律不断减慢、电子晶体管的规模接近物理极限等造成的计算速度难以进一步提高的问题, 提出了一种基于 VGG16 的衍射光子神经网络(VGG16-DONN)结构。该结构利用光衍射层作为 VGG16 的光学前端, 替换了 VGG16 中计算耗时占比最大的第一层电卷积层, 分别对 CelebA 数据集和猫狗数据集进行分类(分类精度分别达到 86.34% 和 88.53%), 实现了与电子神经网络相当的分类精度。此外, 基于此结构, 提出了一种面向情境依赖处理(CDP)的 VGG16-DONN 方法, 对 CelebA 数据集进行分类(平均精度为 83.10%), 同样达到了与电子神经网络相当的分类精度。不难看出, VGG16-DONN 以及其与 CDP 模块相结合的方式, 除了能够借助光计算速度快的优势克服电子神经网络计算慢的问题外, 还能够达到与电子神经网络相当的精度, 这对于图像处理、医疗、通信等领域都具有重要意义。

关键词 光计算; 衍射光子神经网络; 情境依赖处理; 正交权重修改算法

中图分类号 TP183

文献标志码 A

DOI: 10.3788/AOS202242.1920001

VGG16-Based Diffractive Optical Neural Network and Context-Dependent Processing

Zhao Xingya, Yang Zhiwei, Dai Jian, Zhang Tian*, Xu Kun

State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract As Moore's law slows down and the scale of electronic transistors approaches the physical limit, the computing speed is hard to be improved. In order to solve this problem, this paper proposes a VGG16-based diffractive optical neural network (VGG16-DONN). The structure uses the optical diffractive layer as the optical front end of VGG16 and thus replaces the first electronic convolutional layer which consumes the most time during the computing in VGG16. In addition, the CelebA dataset and cat and dog datasets are classified by the proposed structure, with high accuracy of 86.34% and 88.53%, respectively, which are equivalent to that of the electronic neural network. Furthermore, based on the proposed structure, the paper constructs a VGG16-DONN method for context-dependent processing (CDP) and classifies the CelebA dataset, with an average accuracy of 83.10%, which is equivalent to that of the electronic neural network. It should be noted that the VGG16-DONN and its combination with the CDP module can address the slow computation speed of the electronic neural network by taking advantage of the fast optical computation, and they can obtain a similarly high accuracy compared with the electronic neural network, which is of great significance to image processing, medical treatment, communication, and other fields.

Key words optical computation; diffractive optical neural network; context-dependent processing; orthogonal weight modification algorithm

1 引言

因为电子计算中摩尔定律正在不断减慢, 电子晶

体管的规模也正在接近物理极限^[1], 所以电子神经网络的计算速度难以进一步提高。而光计算具有大带宽、低功耗、低串扰等优势, 利用它有望极大地提高计

收稿日期: 2022-01-13; 修回日期: 2022-03-24; 录用日期: 2022-04-15

基金项目: 国家自然科学基金(62171055, 61705015, 61625104, 61821001, 62135009, 61971065)、国家重点研发计划(2019YFB1803504)、北京市科学技术委员会(Z181100008918011)、信息光子学与光通信国家重点实验室(北京邮电大学)基金(IP-OC2020ZT08)

通信作者: *ztian@bupt.edu.cn

算速度,因此光子神经网络应运而生^[2-5]。它的构建方式主要有两种:一种是采用马赫-曾德尔干涉仪^[6]来构建,以这种方式构建的光子神经网络具有集成度高、可重构性强等优势,但需要进行光电转换,并且受到制作工艺水平的限制,目前不适用于大规模神经网络;另一种是基于相位掩模版的衍射光子神经网络(DONN),它不需要进行光电转换处理,通过直接输入光信号构建大规模神经网络,进而实现更加复杂的功能。Lin等^[7]提出一种基于衍射的光子深度神经网络结构,实现全光机器学习,完成了MNIST手写数字数据集和Fashion-MNIST数据集的分类,其中每个相位掩模版的大小为 $8\text{ cm} \times 8\text{ cm}$,层间距为 3 cm ,掩模版由3D技术制作;Yan等^[8]提出一种傅里叶空间衍射深度神经网络,以光速执行显著性检测和高精度目标分类的任务,通过结合双 $2f$ (f 为焦距)光学系统来保持空间对应,有助于完成图像到图像映射的任务。但这些相位掩模版的实际结构过大,不利于集成,并且每个像素点一旦训练完成就不能改变,不适用于其他数据集。所以这种光计算架构不适用于构建网络权重可灵活调控的神经网络,而这一类神经网络是目前各个领域所亟需的。

为了构建网络权重灵活可调的神经网络,Chang等^[9]提出了一种具有优化衍射光学的混合光电卷积神经网络,通过在电子计算前加入一层光学卷积运算,即采用一层相位掩模版,在其前后分别加入凸透镜,构建 $4f$ 系统,从而实现卷积操作^[10],这样可以在保持网络性能的同时降低能耗,最终可以实现CIFAR-10数据集的分类;Colburn等^[11]提出了一种基于AlexNet的DONN(AlexNet-DONN),利用AlexNet卷积神经网络^[12]和仅进行相位调制的空间光调制器^[13]及复振幅调制方法,实现相位与振幅的可调,最后完成猫狗数据集分类的功能。对于AlexNet,其第一层卷积层和第二层卷积层的计算时间比随后的层更长,占总时间的62.7%,并且第一层处理以光作为载体的图像信息,不需要光电转换,因此利用光计算手段替换第一层电卷积操作具有重要意义。虽然它们都克服了全光衍射神经网络权重不可调的问题,但是都没有对实际环境中存在的情境信息做出灵活的响应,难以满足复杂多变的需求,即缺少情境依赖学习的能力,并且其中的AlexNet网络结构具有内存和时间消耗多、迭代次数多、判别力较弱、训练参数多的问题。因此Zeng等^[14]为提高人工智能系统对于复杂多变环境的适应性,受灵长类动物前额叶皮质的启发,提出了正交权重修改(OWM)算法和情境依赖处理(CDP)模块,使用ResNet^[15]网络结构对CelebA数据集^[16]的特征进行提取,但这种提取方式同样受到摩尔定律正在不断减慢的限制。因此,采用光电混合的方式对其进行加速具有重要意义。

本文提出了一种基于VGG16的DONN(VGG16-DONN),利用光学衍射层替换VGG16中计算耗时最多的第一层卷积层,不仅解决了传统的基于VGG16

的电子卷积神经网络(VGG16-ECNN)计算速度慢的问题,而且解决了AlexNet-DONN内存和时间消耗多、迭代次数多、判别力较弱、训练参数多的问题^[17]。本文方法实现了对CelebA数据集的分类,其分类精度与VGG16-ECNN得到的精度相当;本文方法对猫狗数据集进行分类时,分类精度高于AlexNet-DONN的精度。其次,将VGG16-DONN与CDP模块结合,采用混合光电的方式,借助光子计算速度快的优势,可将提取过程大大加快,实现对于相同输入不同情境下的不同输出。

2 VGG16-DONN

2.1 VGG16-ECNN

本文所提出的VGG16-DONN采用了文献[17]中所提出的VGG16-ECNN^[18-19]作为基本架构,如图1(a)所示。该网络数据的处理步骤如下。

1)输入图像预处理。输入大小为 224×224 的RGB图像,通过将在训练集中的每个像素减去RGB的均值进行预处理。

2)卷积(Conv)处理。卷积过程如图1(b)所示:首先输入一个 $n \times n \times c$ 的图像, n 和 c 分别代表图像的大小和输入通道,在卷积之前对输入图像进行填充操作,填充大小为 p ;然后使用 k 个 $f \times f \times c$ 的卷积核, f 代表卷积核的大小,将填充后的图像与卷积核进行卷积,即将图像每个通道元素与对应通道的卷积核进行卷积,并将通道的结果相加,得到输出图像的第二个元素;接着,将卷积窗口在图像上滑动,滑动步长为 s ,这样可以得到输出图像的第二个元素;随后依次滑动卷积窗口,最后可以得到 $o \times o \times k$ 的输出图像。本文所采用的VGG16-ECNN共有13层卷积层,采用 3×3 大小的卷积核($f=3, c=3$), $p=1, s=1$,卷积后利用ReLU激活函数实现非线性操作。

3)池化处理。池化层采用最大池化操作,大小为 2×2 ,步长为2,位于几个卷积层之后,总共有5层。

4)全连接层(FC)处理。三层全连接层中,前两层全连接层均有4096个通道,第三层全连接层有1000个通道。

5)输出处理。使用Softmax层作为输出层。由于文献[17]中处理的是ImageNet数据集,所以第三层全连接层和最后输出层均为1000个通道。

VGG16-ECNN共有5个卷积块,每个卷积块(Conv X 表示第 X 个卷积块, $X=1, 2, \dots, 5$)由2或3个卷积层组成,总共有13层卷积层(Conv X_Y 表示第 X 个卷积块的第 Y 个卷积层, $Y=1, 2, 3$)。VGG16-ECNN对一张图像进行推演时各卷积层的计算耗时如表1所示,第一层卷积层(Conv1_1)的计算耗时是最多的,占22.3%,其余卷积层都在8%及以下。测试获得该时间数据的电脑配置为:CPU为8核Intel E5,GPU为NVIDIA-GeForce GTX 1080 Ti,内存为125 G。

2.2 网络模型

图2(a)所示为VGG16-ECNN的简化图,对

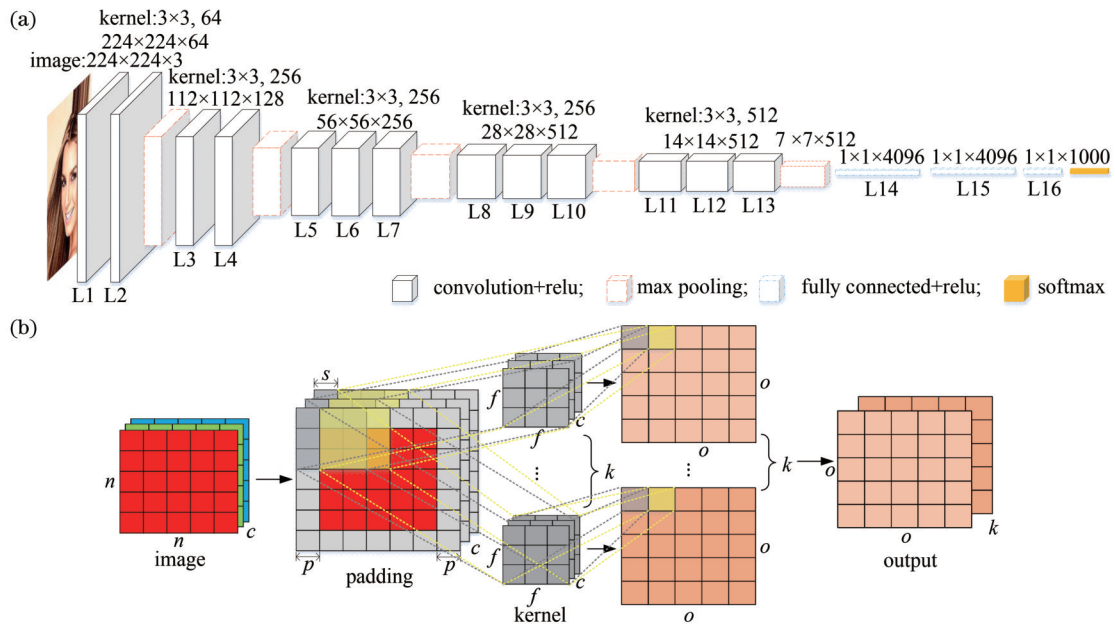


图 1 VGG16-ECNN 结构示意图和卷积过程。(a) VGG16-ECNN 结构示意图;(b) 卷积过程

Fig. 1 Schematic diagram of VGG16-ECNN, and convolution process. (a) Schematic diagram of VGG16-ECNN; (b) convolution process

表 1 VGG16-ECNN 各个卷积层的用时

Table 1 Time of each convolutional layer of VGG16-ECNN

Convolution layer	Time /s	Proportion of total time /%
Conv1_1	0.92	22.3
Conv1_2	0.33	8.0
Conv2_1	0.24	5.8
Conv2_2	0.27	6.6
Conv3_1	0.22	5.3
Conv3_2	0.25	6.1
Conv3_3	0.25	6.1
Conv4_1	0.24	5.8
Conv4_2	0.27	6.6
Conv4_3	0.27	6.6
Conv5_1	0.28	6.8
Conv5_2	0.29	7.0
Conv5_3	0.29	7.0
Total time	4.12	100.0

CelebA 数据集进行分类时,最后的全连接层只有 40 个输出,代表 CelebA 数据集的 40 种类别;第一层卷积层是处理图像的初始层,在 VGG16-ECNN 中,其计算耗时是最多的,所以使用如图 2(b)所示的 $4f$ 系统,让它作为光衍射层替换 VGG16-ECNN 结构中的第一层卷积层。 $4f$ 系统由物平面 O1,两个焦距为 f 的透镜 O2、O4,共焦平面 O3 和像平面 O5 组成,其中共焦平面又称变换平面,在此放置相位掩模版,其具体结构图如图 2(b)右半部分所示,它由 8×8 个 3×3 大小的卷积核平铺组成,灰度的深浅代表权重的不同,其部分具体数值

如图 3 所示,每个卷积核上共有 9 个权重,一共有 64 个卷积核。光在 $4f$ 系统中传播,首先通过图像和凸透镜 O2,可实现傅里叶变换,然后经过掩模版 O3,对输入图像频谱进行相位调制,最后经凸透镜 O4 实现傅里叶逆变换,还原出图像,整个过程采用快速傅里叶变换算法和角谱传播法模拟衍射过程,从而完成光学卷积^[10]。使用 $4f$ 系统替换 VGG16-ECNN 结构的第一层电卷积层,利用该 VGG16-DONN 可实现对图像等数据的快速分类。

2.3 模型训练

训练采用的数据集分别是 CelebA 数据集和猫狗数据集,其中部分数据图如图 4 所示。CelebA 数据集包含 10177 个名人身份的 202599 张人脸图像,每张图像都做好了特征标记,包含人脸标注框、5 个人脸特征点坐标以及 40 个属性标记,此数据集由香港中文大学开放提供,广泛应用于人脸属性识别训练。猫狗数据集中训练集共有 25000 张图像,12500 张猫的图像,12500 张狗的图像。

训练过程如图 5 所示。输入 CelebA 数据集,其中训练集共有 182637 张图像,验证集共有 19962 张图像。采用 VGG16-DONN 结构,其前 13 层卷积层的权重不需要训练,使用文献[17]中对 ImageNet 数据集进行处理时所训练得到的网络权重,仅训练三层全连接的权重和偏置,共 119709696 个参数,最后有 40 个输出,40 个输出代表 CelebA 数据集的 40 个类别。训练采用 Adadelta 优化算法,学习率为 0.01,迭代次数为 20。图 5 中 N_{epoch} 代表 epoch 值。

分别向 VGG16-ECNN 和 VGG16-DONN 两种结构中输入一张人脸图像[图 6(a)],将人脸图像分别与第

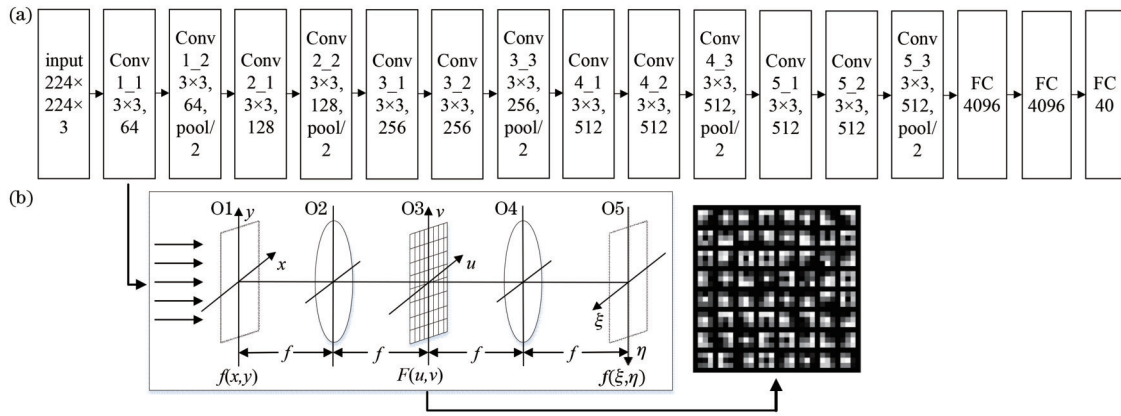


图 2 VGG16-DONN 结构示意图。(a) VGG16-ECNN 简化图；(b) 4f 系统 [$f(x,y)$ 为输入复振幅函数, $F(u,v)$ 为 $f(x,y)$ 的傅里叶变换, $f(\xi,\eta)$ 为经第二个凸透镜还原的改造信号]

Fig. 2 Schematic diagrams of VGG16-DONN structure. (a) Simplified diagram of VGG16-ECNN; (b) 4f system [$f(x,y)$ is input complex amplitude function, $F(u,v)$ is Fourier transform of $f(x,y)$, and $f(\xi,\eta)$ is reconstructed signal restored by second convex lens]

0.42947	0.37347	-0.06136		0.04134	0.11658	0.10312
0.27477	0.03868	-0.36722	...	-0.21259	-0.14905	0.03825
-0.05747	-0.26225	-0.35010		-0.21322	-0.18567	-0.00163
⋮						
0.04299	0.09854	0.22941		0.00758	-0.06304	-0.30051
-0.12696	-0.10365	0.18546	...	0.33994	0.37168	-0.23972
-0.01578	0.00424	0.15894		0.04556	0.05329	-0.26508

图 3 卷积核权重

Fig. 3 Weights of convolution kernels

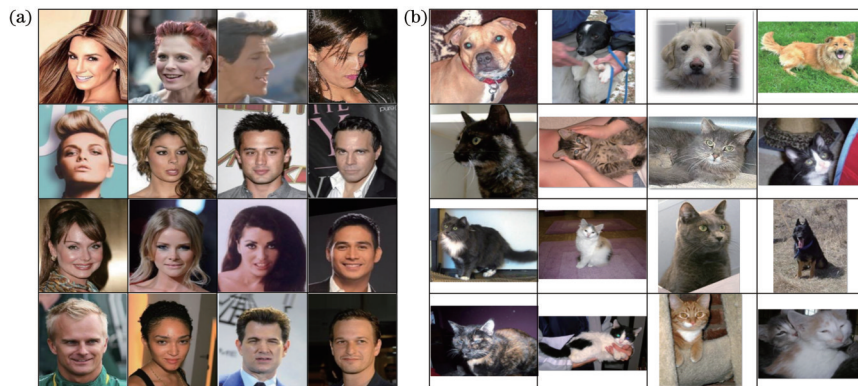


图 4 数据集示例。(a) CelebA 数据集；(b) 猫狗数据集

Fig. 4 Dataset examples. (a) CelebA dataset; (b) cat and dog dataset

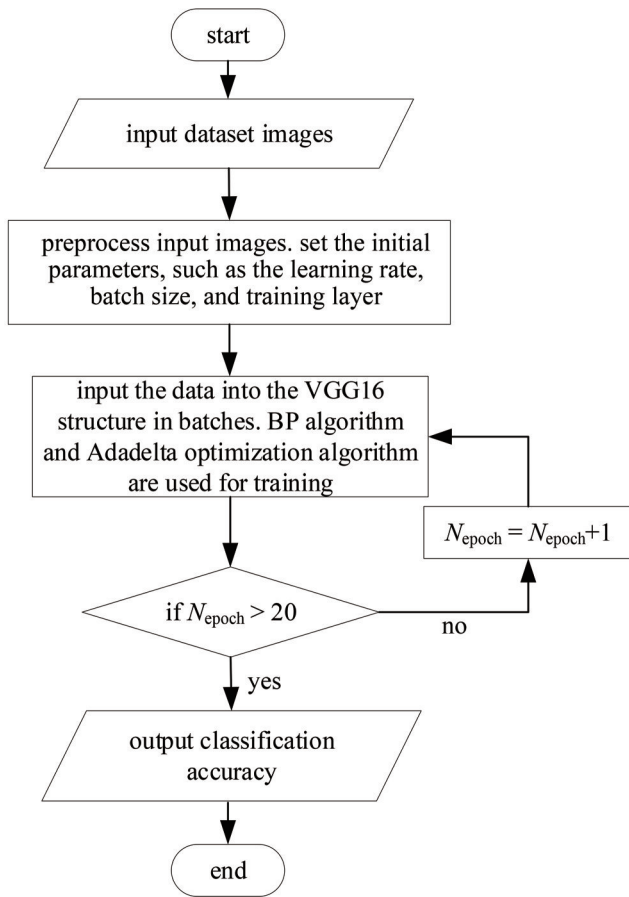


图 5 训练流程图

Fig. 5 Training flow chart

一个卷积层中的第 22、23、24 和 25 个卷积核进行卷积,经平方激活函数处理后的输出图如图 6(b)、(c)所示。图 6(b)、(c)分别为 VGG16-ECNN 结构和 VGG16-DONN 结构第一个卷积层中部分卷积核卷积后的输出图。比较图 6(b)、(c)可以看出:电卷积与光学卷积存在一定差异,输出图有差异是因为 VGG16-ECNN 第一层经卷积和激活函数处理时存在偏置参数,而 VGG16-DONN 没有偏置参数。图 6(a)中的人脸图像经 VGG16-DONN 第一个卷积层中 64 个卷积核的输出图如图 7 所示。图 6(a)中的人脸图像经 VGG16-DONN 中每个卷积块的输出图如图 8 所示,其中图 8(a)~(e)分别为人脸图像经 5 个卷积块(Conv1、Conv2、Conv3、Conv4、Conv5)的输出图,从图中可以看出图像的特征逐步简化并被逐步提取的过程。

输入猫狗数据集,其中训练集共有 17500 张图像,验证集共有 7500 张图像。同样使用 VGG16-DONN 结构,仅训练三层全连接层中的权重,使用 Adam 优化算法,学习率为 0.00001,迭代次数为 10。分别向 VGG16-ECNN 和 VGG16-DONN 两种结构中输入一张狗的图像[图 9(a)],将狗的图像分别与第一个卷积层中的第 42、43、44 和 45 个卷积核进行卷积,经平方激活函数处理后的输出图如图 9(b)、(c)所示。图 9(b)、(c)分别为 VGG16-ECNN 结构和 VGG16-DONN 结构第一个卷积层中部分卷积核卷积后的输出图,比较图 9(b)、(c)可以看出电卷积与光学卷积同样存在一定差异,其存在差异的原因与图 6(b)、(c)相同。

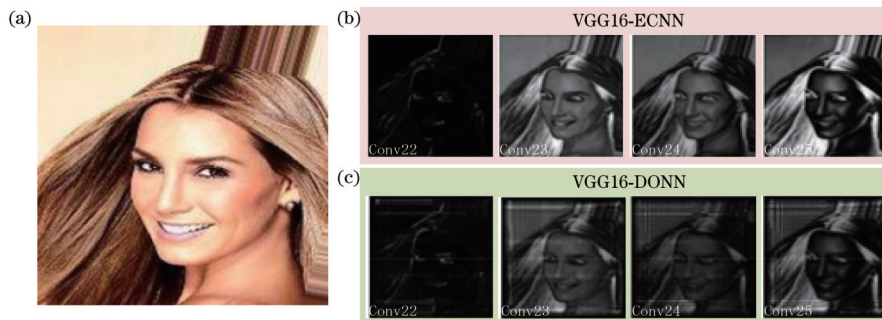


图 6 原图像和两种网络结构经第一层卷积层中部分卷积核的输出图。(a)一张人脸图像;(b) VGG16-ECNN;(c) VGG16-DONN
Fig. 6 Original image and output images of two network structures through partial convolution kernel in the first convolution layer.

(a) Face image; (b) VGG16-ECNN; (c) VGG16-DONN

3 面向 CDP 的 VGG16-DONN

3.1 网络模型

为了解决“灾难性遗忘”问题,在上述 VGG16-DONN 的基础上加入 CDP 模块,其结构如图 10 所示,该结构主要由两部分组成:特征提取器和 CDP 模块。其中:第一部分为特征提取器,利用 VGG16-DONN 的前 13 层卷积层实现;第二部分为 CDP 模块,它受大脑前额叶皮层启发,可以有效整合情境信息,调制神经网络的信息处理过程,此处的情境信息指的是 VGG16-

DONN 所分出的类别信息。VGG16-DONN 将多个类别同时进行分类,而加入 CDP 模块的网络可以针对其中一个类别进行学习,采用在时间上连续的方式学习多个类别,而不需要同时对其进行学习。CDP 模块包括两个子模块:1)编码子模块,其负责将情境信息编码为适当的控制信号;2)旋转子模块,该模块结合编码子模块的控制信号来处理特征提取器所提取的特征信号。

通过特征提取器进行图像的特征提取,将特征输入到 CDP 模块,CDP 模块的输出为

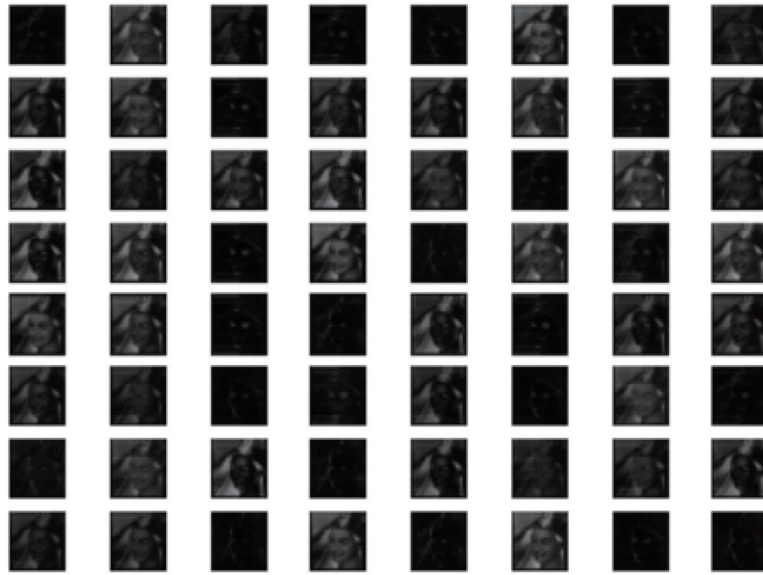


图 7 人脸图像与第一层卷积层中所有卷积核进行卷积后的输出

Fig. 7 Output of face image convolved with all convolution kernels in the first convolution layer

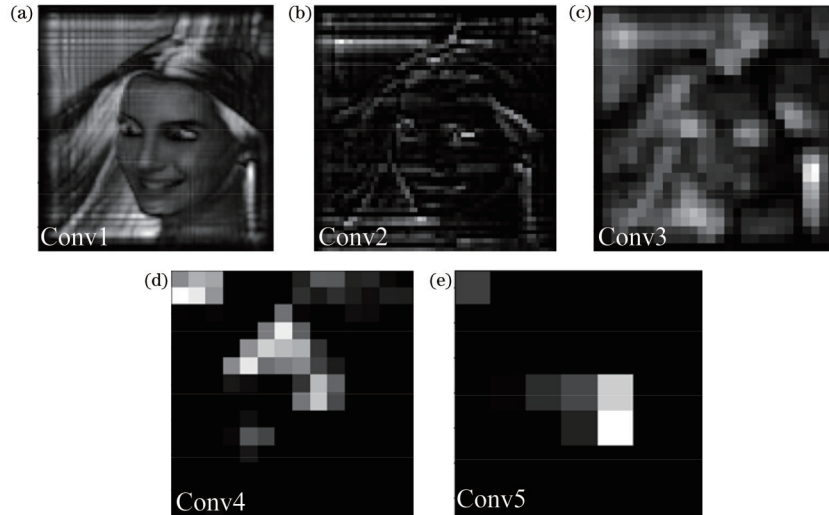


图 8 向 VGG16-DONN 结构输入一张人脸图像后各卷积块的输出图。(a) Conv1; (b) Conv2; (c) Conv3; (d) Conv4; (e) Conv5

Fig. 8 Output graph of each convolution block after inputting face image to VGG16-DONN structure. (a) Conv1; (b) Conv2; (c) Conv3; (d) Conv4; (e) Conv5

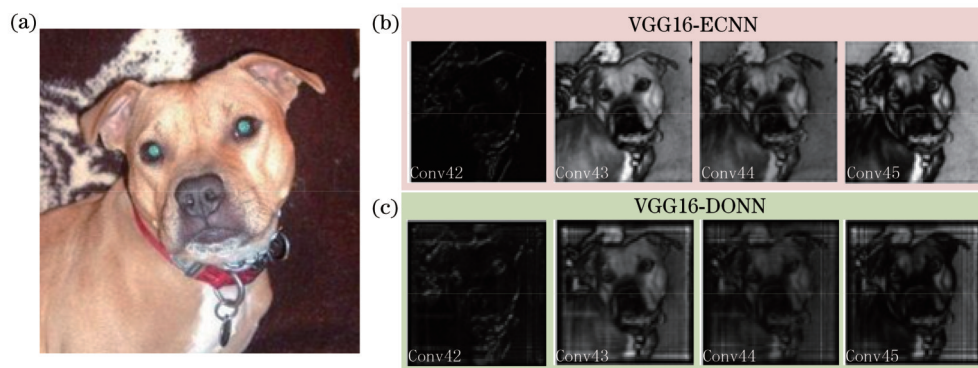


图 9 原图像和两种网络结构经第一层卷积层中部分卷积核的输出图。(a) 一张狗的图像; (b) VGG16-ECNN; (c) VGG16-DONN

Fig. 9 Original image and output graphs of two network structures through partial convolution kernel in the first convolution layer.

(a) Dog image; (b) VGG16-ECNN; (c) VGG16-DONN

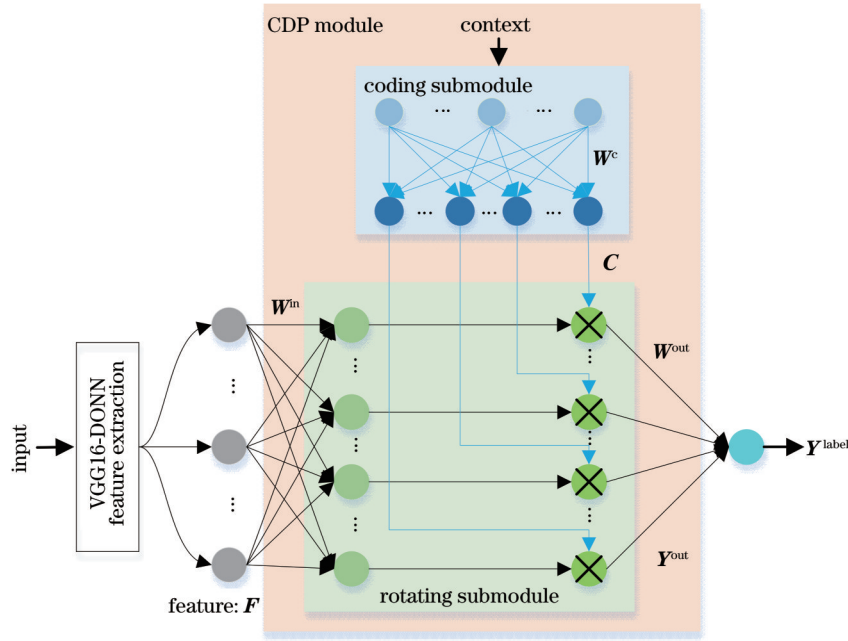


图 10 面向 CDP 的 VGG16-DONN

Fig. 10 VGG16-DONN for CDP

$$\begin{aligned}
 Y^{\text{out}} &= g\left[(W^{\text{in}})^{\text{T}} F\right] \odot C = g\left([\mathbf{w}_1^{\text{in}}, \mathbf{w}_2^{\text{in}}, \dots, \mathbf{w}_m^{\text{in}}]^{\text{T}} F\right) \odot C = \\
 &g\left[\left[c_1 \| F \| \|\mathbf{w}_1^{\text{in}}\| \cos \theta_1, c_2 \| F \| \|\mathbf{w}_2^{\text{in}}\| \cos \theta_2, \dots, c_m \| F \| \|\mathbf{w}_m^{\text{in}}\| \cos \theta_m\right]^{\text{T}}\right) = \\
 &g\left(\left[c_1 \|\mathbf{w}_1^{\text{in}}\| \cos \theta_1, c_2 \|\mathbf{w}_2^{\text{in}}\| \cos \theta_2, \dots, c_m \|\mathbf{w}_m^{\text{in}}\| \cos \theta_m\right]^{\text{T}}\right) \| F \|, \quad (1)
 \end{aligned}$$

式中： $g(\cdot)$ 为 ReLU 激活函数； $W^{\text{in}} = [\mathbf{w}_1^{\text{in}}, \mathbf{w}_2^{\text{in}}, \dots, \mathbf{w}_m^{\text{in}}] \in \mathbb{R}^{k \times m}$ 为特征输入到 CDP 模块时的输入权重； $F = [f_1, f_2, \dots, f_k]^{\text{T}} \in \mathbb{R}^k$ 为特征提取器提取到的特征向量，作为 CDP 模块的输入； \odot 为按元素相乘操作； $C = [c_1, c_2, \dots, c_m]^{\text{T}} \in \mathbb{R}^m$ 为控制信号，代表编码后的情境信息； $Y^{\text{out}} = [y_1, y_2, \dots, y_m]^{\text{T}} \in \mathbb{R}^m$ 为 CDP 模块的输出； θ_i 为 F 和 \mathbf{w}_i^{in} 之间的角度，其中 $i = 1, 2, \dots, m$ 。对于一张人脸图像，其特征向量 F 、输入权重 W^{in} 和 $\cos \theta_1, \dots, \cos \theta_m$ 均固定不变，通过不同的控制信号 C 来完成不同的任务（VGG16-DONN 所分出的类别）。

3.2 模型训练

根据图 10 所示的网络模型，使用正交权值修正 (OWM) 算法对编码子模块中的权重 W^c 和输出权重 W^{out} 进行训练。OWM 算法是为了避免“灾难性遗忘”，即防止先前学习过的映射被随后的训练擦除、为实现连续学习而提出的一种方法，通过正交投影算子

与误差反向传播算法得到权重增量可实现该目的。

假设一个层数为 $l(0, 1, \dots, L)$ 的前馈神经网络，输入层和输出层分别为 $l=0$ 和 $l=L$ ，第 $l-1$ 层和第 l 层之间的权重为 $W_l \in \mathbb{R}^{s \times m}$ ，激活函数为 $g(\cdot)$ ，第 l 层的输入和输出分别为 y_l 和 x_l ，其中 $y_l = W_l^{\text{T}} x_{l-1}$ ， $x_l = g(y_l)$ ($y_l \in \mathbb{R}^m$, $x_{l-1} \in \mathbb{R}^s$)。连续学习的分类任务数为 j ，其最大值为数据集的类别数 C ，对于 CelebA 数据集， C 为 40，第 j 个任务的批次 i 的取值范围为 $(1, \dots, n_j)$ ，本文 n_j 为 500。对于 OWM 算法， P_l 为学习任务的第 l 层的正交投影算子，它是解决“灾难性遗忘”问题的关键。OWM 算法训练过程如图 11 所示，具体过程为：

1) 初始化网络参数。随机初始化 $W_l(i, j)$ ，设 $P_l(i, j) = I_l / \beta$ ，其中 I_l 是单位矩阵， β 是正则化常数。

2) 第 j 个任务的第 i 批输入需要经过前向传播，然后利用反向传播算法计算权重增量 $\Delta W_l^{\text{BP}}(i, j)$ 并更新每一层的权重：

$$\begin{cases} W_l(i, j) = W_l(i-1, j) + \alpha(i, j) \Delta W_l^{\text{BP}}(i, j) & j=1 \\ W_l(i, j) = W_l(i-1, j) + \alpha(i, j) P_l(n_{j-1}, j-1) \Delta W_l^{\text{BP}}(i, j) & j=2, 3, \dots \end{cases}, \quad (2)$$

式中： $\alpha(i, j)$ 为学习率。

3) 下一批输入重复步骤 2)。

4) 如果第 j 个任务完成, 则依次向前传播第 j 个任务中每一批输入的平均值。随后, 更新 $W_l(i, j)$ 中的 $P_l(n_j, j)$, 利用以下公式迭代 n_j 次可以得到:

$$P_l(i, j) = P_l(i-1, j) + k_l(i, j) \bar{x}_{l-1}^T(i, j) P_l(i-1, j), \quad (3)$$

式中: $k_l(i, j) = P_l(i-1, j) \bar{x}_{l-1}(i, j) / [1 + \bar{x}_{l-1}^T(i, j) P_l(i-1, j) \bar{x}_{l-1}(i, j)]$; $x_{l-1}(i, j)$ 为第 j 个任务第 i 批输入均值的 $l-1$ 层输出。

5) 下一个任务重复步骤 2) 到步骤 4)。

利用 VGG16-DONN 的前 13 层卷积层对 CelebA 数据集进行特征提取, 提高特征提取的速度。随后将所提取的特征数据(数据长度为 25088)输入到 CDP 模

块, 此时仅使用 500 张图像的特征作为训练集, 50 张图像的特征作为测试集。一方面, CDP 模块首先利用 W^m 权重矩阵和神经元所执行的 ReLU 操作对特征数据进行处理; 另一方面, CDP 模块的另一输入为情境信息, 该情境信息需要首先通过用于自然语言处理的 Gensim 模块来处理 40 类人脸属性名的词向量, 随后利用编码器模块对词向量进行编码。其中, 每一类人脸属性名所对应的词向量长度和经过编码后的长度分别为 200 和 4000。利用 OWM 算法对权重 W^c 和 W^{out} 进行训练。为了加快训练速度和保证样本的多样性, 采用小批量梯度下降法对权重等参数进行更新, 从而仅利用一个分类器就可以连续学习不同的任务。最后, 将经过处理的特征数据和情境信息相结合, 完成最终的分类功能。

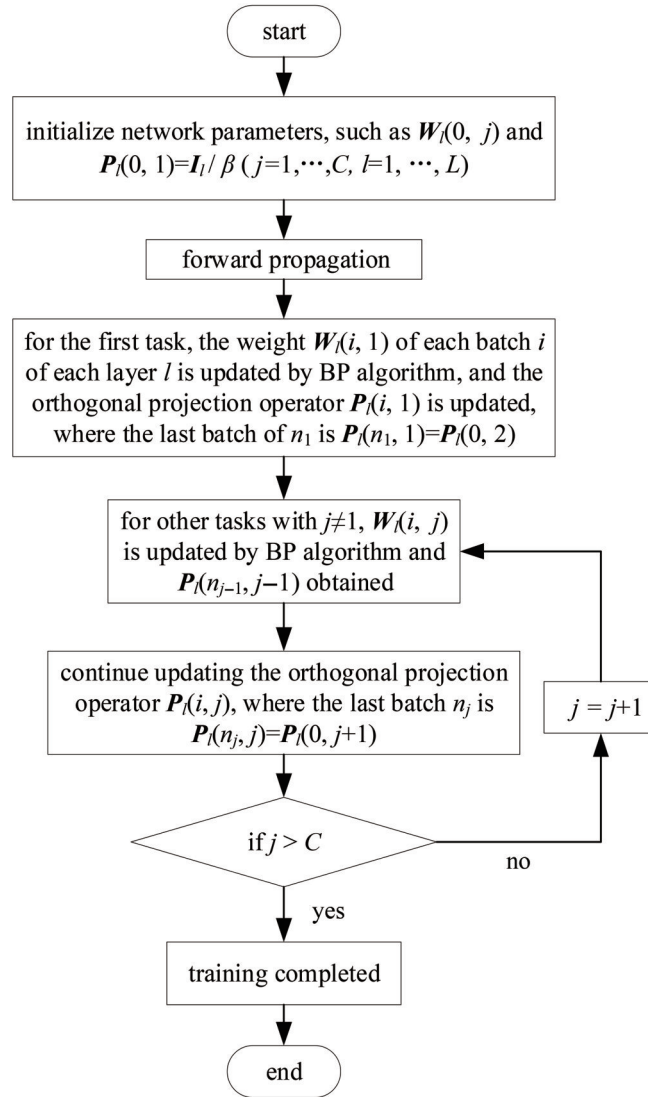


图 11 OWM 算法流程图

Fig. 11 OWM algorithm flow chart

4 分析与讨论

4.1 VGG16-DONN

对 CelebA 数据集进行分类, 如图 12 所示: VGG16-

DONN 进行 20 次迭代训练后得到的训练精度为 87.06%, 验证精度为 86.34%, VGG16-ECNN 进行 20 次迭代训练后得到的训练精度为 87.17%, 验证精度为 86.77%。经过对比可看出 VGG16-DONN 的精度与

VGG16-ECNN 的精度相当,由此可说明 VGG16-DONN 可对 CelebA 数据集进行有效分类。从图 12(a) 可看出分类精度受迭代次数的影响,随着迭代次数的增加,分类精度也随之提高,当迭代次数为 20 时,分类精度可达到 86.34%。虽然迭代次数继续增加可以进一步提高分类精度,但是提高速度极慢,因此,综合权衡所需迭代时间和分类精度,本文后续研究中将迭代次数固定为 20 次。此外,从图 12(a) 可以发现, VGG16-

DONN 的精度没有达到 VGG16-ECNN 的精度,造成这一现象的原因是光卷积过程存在一定的误差,没有实现与电卷积完全一致的操作,这一点通过对比图 6 (b)、(c) 和图 9(b)、(c) 可以看出,但是这里光卷积造成的误差并不大,训练精度和验证精度与 VGG16-ECNN 相比分别只差 0.11% 和 0.43%,这些误差可以通过在后续的工作中将光卷积也加入训练来消除,同时采用这种方式也可以消除光卷积的噪声带来的影响。

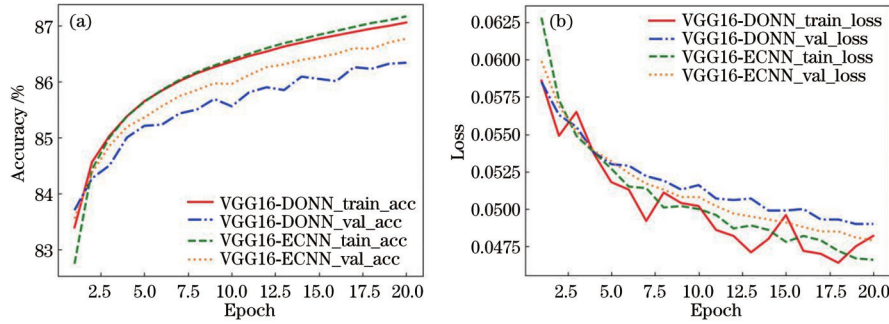


图 12 VGG16-DONN 和 VGG16-ECNN 的训练结果。(a) VGG16-DONN 和 VGG16-ECNN 的训练精度和验证精度;(b) VGG16-DONN 和 VGG16-ECNN 的训练损失和验证损失
Fig. 12 Training results of VGG16-DONN and VGG16-ECNN. (a) Training accuracy and validation accuracy of VGG16-DONN and VGG16-ECNN; (b) training loss and validation loss of VGG16-DONN and VGG16-ECNN

为进一步提高对 CelebA 数据集进行分类的精度,可以通过改变 VGG16-DONN 结构需要训练的层、优化器算法、学习率等方法来实现。针对需要训练的层,本文选择三种代表性情况进行分析,分别为训练全连接层(FC)、训练全连接层和第一层卷积层(FC+Conv1_1),以及训练全部层(All)。使用 Adadelta 优化

算法、0.01 的学习率,通过训练不同层的权重来提高分类精度,如图 13(a)、(b) 所示。图 13(a) 为 VGG16-ECNN 在迭代 20 代时得到的分类精度和时间,图 13 (b) 为 VGG16-DONN 在迭代 20 代时得到的分类精度和时间。当选择训练全部层的权重时得到的分类精度最高,为 88.8%。对比图 13(a)、(b) 可看出 VGG16-

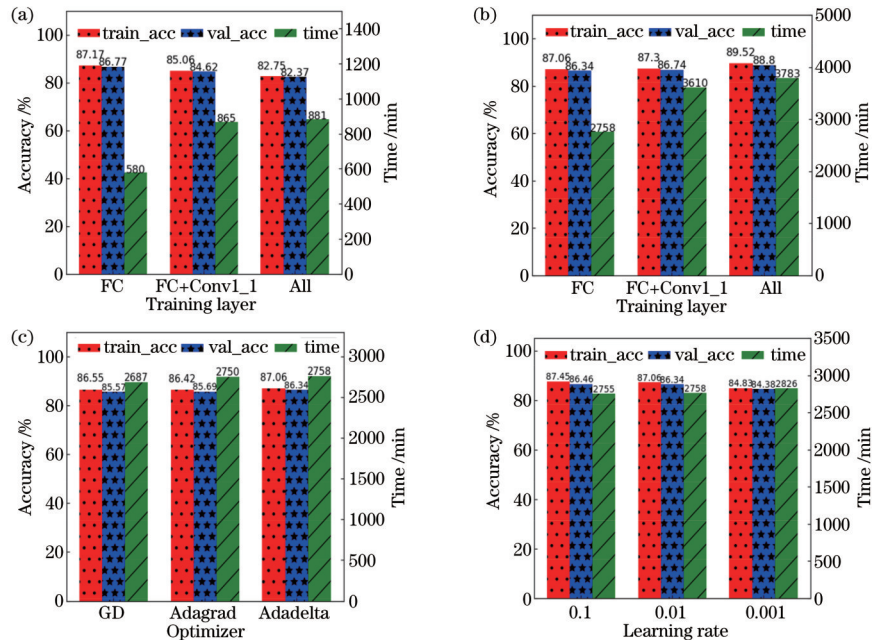


图 13 通过改变不同参数得到的训练/验证精度和时间。(a) VGG16-ECNN 结构训练不同层的权重;(b) VGG16-DONN 结构训练不同层的权重;(c) VGG16-DONN 结构改变优化器;(d) VGG16-DONN 结构改变学习率
Fig. 13 Training/validation accuracy and time obtained by varying different parameters. (a) Weights of different layers trained by VGG16-ECNN structure; (b) weights of different layers trained by VGG16-DONN structure; (c) VGG16-DONN structure changes optimizer; (d) VGG16-DONN structure changes learning rate

ECNN 和 VGG16-DONN 在训练全连接层权重时具有相当的分精度。对于 VGG16-DONN, 所有网络层都参与训练时的准确度比只有全连接层参与训练时的准确度只提高了 2.46%, 说明卷积层参与训练对于最终的准确度影响并不大[图 13(b)], 所以, 在本方案的后续研究中都采用了只训练全连接层的方式。另外, 通过改变优化算法也可以提高分类精度, 如图 13(c) 所示。优化器算法有 GD、Adadelta 和 Adagrad, 其中 Adadelta 优化算法得到的分类精度最高(86.34%)。如图 13(d) 所示, 学习率为 0.1 时得到的分类精度最高, 为 86.46%。除了采用改变迭代次数、需要训练的层、优化器算法、学习率的方法外, 还可以通过增加光卷积层数、掩模版上的神经元个数、神经元之间的连接数等方式提高分类精度。

另外, 利用 VGG16-DONN 结构对猫狗数据集进行分类, 可以实现比 AlexNet-DONN 结构^[11]更好的效果。如表 2 所示, VGG16-DONN 结构得到的分类精度高于 AlexNet-DONN 结构, 并且训练所用的时间也比 AlexNet-DONN 少, 由此可以说明 VGG16-DONN 对猫狗数据集分类比 AlexNet-DONN 更有效。

表 3 VGG16-DONN 和 VGG16-ECNN 结合 CDP 模块下 40 种人脸属性的精度

Table 3 Accuracy of 40 kinds of face attributes obtained by VGG16-DONN and VGG16-ECNN combined with CDP module unit: %

Attribute	VGG16-DONN	VGG16-ECNN	Attribute	VGG16-DONN	VGG16-ECNN
5_o_Clock_Shadow	90.20	90.70	Mouth_Slightly_Open	62.00	65.30
Arched_Eyebrows	70.80	71.70	Mustache	96.80	96.30
Attractive	68.00	71.60	Narrow_Eyes	87.00	86.10
Bags_Under_Eyes	79.60	78.60	No_Beard	86.00	86.00
Bald	98.20	98.20	Oval_Face	71.00	73.40
Bangs	83.20	86.00	Pale_Skin	95.80	95.70
Big_Lips	67.20	66.20	Pointy_Nose	71.80	71.00
Big_Nose	81.80	80.30	Receding_Hairline	92.00	92.90
Black_Hair	75.80	76.00	Rosy_Cheeks	93.20	93.40
Blond_Hair	87.40	86.90	Sideburns	94.60	95.30
Blurry	96.20	94.60	Smiling	67.00	75.00
Brown_Hair	81.20	81.10	Straight_Hair	79.20	77.50
Bushy_Eyebrows	85.00	86.40	Wavy_Hair	66.60	73.60
Chubby	95.00	95.90	Wearing_Earrings	79.40	78.10
Double_Chin	97.40	95.70	Wearing_Hat	96.00	95.80
Eyeglasses	94.60	93.70	Wearing_Lipstick	75.20	81.60
Goatee	94.80	95.60	Wearing_Necklace	84.20	85.70
Gray_Hair	97.20	97.50	Wearing_Necktie	91.00	93.30
Heavy_Makeup	73.80	79.00	Young	77.60	76.90
High_Cheekbones	62.80	71.20	Male	77.60	84.90

5 结 论

提出了 VGG16-DONN 结构, 即将 VGG16 结构中计算耗时最多的第一层卷积层替换为光学衍射层, 对 CelebA 数据集进行分类, 分类精度与 VGG16-ECNN 得到的分类精度相当; 对猫狗数据集进行分类, 分类精

表 2 VGG16-DONN 和 AlexNet-DONN 对猫狗数据集的分类精度和用时

Table 2 Classification accuracy and time of VGG16-DONN and AlexNet-DONN for cat and dog dataset

Network structure	Classification accuracy / %	Time / min
VGG16-DONN	88.53	320
AlexNet-DONN	83.81	1117

4.2 面向 CDP 的 VGG16-DONN

本文将 VGG16-DONN 结构的前 13 层卷积层作为特征提取器, 并且使用文献[17]提供的卷积核权重参数, 不需要训练卷积层部分的权重, 再结合 CDP 模块, 对 CelebA 数据集进行分类, 所得分类精度如表 3 所示。由表 3 可看出 VGG16-DONN 和 VGG16-ECNN 在处理情境信息时所得到的各属性精度及平均精度均相当, 由此得到的平均精度分别约为 83.10% 和 84.37%。但是光学衍射层替换了 VGG16 网络中计算占比最大的第一层卷积层, 能够大大提升该网络的计算速度, 因此, 该方案具有重要意义。

度高于 AlexNet-DONN 结构^[11]得到的分类精度。另外, 针对 CelebA 数据集的分类任务, 对 VGG16-DONN 结构的参数进行优化。仿真结果表明, 卷积层是否参与训练对于最终的准确度影响并不大, 只对全连接层权重进行训练得到的分类精度为 86.34%。此外, 利用 VGG16-DONN 的前 13 层卷积层对图像进行

特征提取,将该网络与CDP模块进行结合,可以在不同情境下对CelebA数据集的分类,与VGG16-ECNN结合CDP模块相比能达到相当的分类精度;同时,由于采用光学衍射层对第一层卷积层进行了替代,本文方法的计算速度大大提升。以混合光电的方式对数据集进行图像分类和特征提取,实现了网络权重灵活可调,大大提高了运算速度,有望解决电子计算的速度瓶颈,对于图像处理、医疗、通信等领域都具有重要意义。

参 考 文 献

- [1] Shainline J M, Buckley S M, Mirin R P, et al. Superconducting optoelectronic circuits for neuromorphic computing[J]. *Physical Review Applied*, 2017, 7(3): 034013.
- [2] Zhang T, Wang J, Dan Y H, et al. Efficient training and design of photonic neural network through neuroevolution [J]. *Optics Express*, 2019, 27(26): 37150-37163.
- [3] 陈宏伟, 于振明, 张天, 等. 光子神经网络发展与挑战 [J]. *中国激光*, 2020, 47(5): 0500004.
Chen H W, Yu Z M, Zhang T, et al. Advances and challenges of optical neural networks[J]. *Chinese Journal of Lasers*, 2020, 47(5): 0500004.
- [4] 项水英, 宋紫薇, 高爽, 等. 光神经形态计算研究进展与展望(特邀)[J]. *光子学报*, 2021, 50(10): 1020001.
Xiang S Y, Song Z W, Gao S, et al. Progress and prospects of photonic neuromorphic computing (invited) [J]. *Acta Photonica Sinica*, 2021, 50(10): 1020001.
- [5] 栾海涛, 陈希, 张启明, 等. 人工智能纳米光子学: 光学神经网络与纳米光子学[J]. *光学学报*, 2021, 41(8): 0823005.
Luan H T, Chen X, Zhang Q M, et al. Artificial intelligence nanophotonics: optical neural networks and nanophotonics[J]. *Acta Optica Sinica*, 2021, 41(8): 0823005.
- [6] Reck M, Zeilinger A, Bernstein H J, et al. Experimental realization of any discrete unitary operator[J]. *Physical Review Letters*, 1994, 73(1): 58-61.
- [7] Lin X, Rivenson Y, Yardimci N T, et al. All-optical machine learning using diffractive deep neural networks [J]. *Science*, 2018, 361(6406): 1004-1008.
- [8] Yan T, Wu J M, Zhou T K, et al. Fourier-space diffractive deep neural network[J]. *Physical Review Letters*, 2019, 123(2): 023901.
- [9] Chang J L, Sitzmann V, Dun X, et al. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification[J]. *Scientific Reports*, 2018, 8(1): 12324.
- [10] 马锁冬, 曾春梅, 许峰. 基于4f系统的光学图像加密与解密仿真教学[J]. *大学物理实验*, 2018, 31(6): 39-45.
Ma S D, Zeng C M, Xu F. Teaching of 4f system with optical image encryption and decryption simulations[J]. *Physical Experiment of College*, 2018, 31(6): 39-45.
- [11] Colburn S, Chu Y, Shilzerman E, et al. Optical frontend for a convolutional neural network[J]. *Applied Optics*, 2019, 58(12): 3179-3186.
- [12] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [13] 宋菲君, Jutamulia S. 近代光学信息处理[M]. 北京: 北京大学出版社, 2014.
Song F J, Jutamulia S. Modern optical information processing[M]. Beijing: Peking University Press, 2014.
- [14] Zeng G, Chen Y, Cui B, et al. Continual learning of context-dependent processing in neural networks[J]. *Nature Machine Intelligence*, 2019, 1(8): 364-372.
- [15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [16] Liu Z W, Luo P, Wang X G, et al. Large-scale celebfaces attributes (CelebA) dataset[EB/OL]. [2021-02-05]. <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>.
- [17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04) [2021-05-06]. <https://arxiv.org/abs/1409.1556>.
- [18] 邸江磊, 唐唯, 吴计, 等. 卷积神经网络在光学信息处理中的应用研究进展[J]. *激光与光电子学进展*, 2021, 58(16): 1600001.
Di J L, Tang J, Wu J, et al. Research progress in the applications of convolutional neural networks in optical information processing[J]. *Laser & Optoelectronics Progress*, 2021, 58(16): 1600001.
- [19] 国强, 彭龙. 基于三维卷积神经网络与超像素分割的高光谱分类[J]. *光学学报*, 2021, 41(22): 2210001.
Guo Q, Peng L. Hyperspectral classification based on 3D convolutional neural network and super pixel segmentation[J]. *Acta Optica Sinica*, 2021, 41(22): 2210001.