双通道 Faster R-CNN 在 RGB-D 手部检测中的应用

刘 壮[1,2,3,4] 柴秀娟[2,4] 陈熙霖[2,3,4]
(中国科学院上海微系统与信息技术研究所 上海 200050)[1]
(中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)[2]
(上海科技大学信息科学与技术学院 上海 201210)[3] (中国科学院大学 北京 100049)[4]

摘 要 在人机交互、手语识别等大量与人手有关的视觉任务中,手部检测是极为重要的一个预处理阶段。随着RGB-D数据采集设备的发展,额外提供的深度数据能够与传统使用的彩色数据互相补充以提供更强的特征表达。此外,传统的检测方法由于使用肤色、HOG等手工设计的特征,不能对手部进行很好的表达。而基于深度学习的检测方法通过从数据中自动学习有效的特征避免了这个问题。为了结合RGB-D数据和深度学习技术的优点,提出了一种融合彩色和深度数据的双通道 Faster R-CNN 检测框架。该方法在原有 Faster R-CNN 检测框架的基础上,增加了Depth通道信息,并在特征层面上将其与RGB通道信息进行融合。实验结果表明,所提方法在性能上比仅采用RGB或在数据层面上融合的 Faster R-CNN 框架有明显优势。因此,该方法能有效融合来自彩色和深度通道的数据,以提升手部检测性能。

关键词 手部检测,深度数据,深度学习,双通道 Faster R-CNN
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.05.040

Application of Two-stream Faster R-CNN in RGB-D Hand Detection

LIU Zhuang[1,2,3,4] CHAI Xiu-juan[2,4] CHEN Xi-lin[2,3,4]
(Shanghai Institute of Microsystm & Information Technology, Chinese Academy of Sciences, Shanghai 200050, China)[1]
(Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)[2]
(School of Information Science & Technology, Shanghai Tech University, Shanghai 201210, China)[3]
(University of Chinese Academy of Sciences, Beijing 100049, China)[4]

Abstract In most vision tasks related to human hands, such as human computer interaction and sign language recognition, hand detection is a distinctly important preprocessing phase. With the development of RGB-D data acquisition equipment, the extra depth data can complement the color data effectively, so they can provide more powerful feature representation. The traditional detection methods based on hand-crafted features (skin color or HOG) cannot form a well hand representation. While a lot of detection methods based on deep learning can avoid such weakness by learning effective features from data. To combine the advantages of RGB-D data and deep learning, a two-stream Faster R-CNN detection framework was proposed in this paper. The proposed method adds an extra depth stream information, and combines it with RGB stream information in the feature level. The experiment results show that the proposed method can achieve a higher detection precision than the Faster R-CNN framework which uses RGB or fuses the RGB and Depth in the data level. Thus, the proposed method can fuse the color and depth data effectively, and improve the performance of hand detection.

Keywords Hand detection, Depth data, Deep learning, Two-stream Faster R-CNN

1 引言

手部检测是指确定一幅图中是否存在人手,若存在则进一步确定其所处位置。在手语或手势识别、手势控制的人机交互及虚拟现实等大量与人手有关的应用中,手部检测通常作为预处理阶段。在获得手部区域的基础上,进一步提取相应的特征用于后续识别等任务。因此,手部检测精度的高低将直接影响手部特征的提取,进而影响最终任务的结果。而

到稿日期:2017-03-06 返修日期:2017-06-11 本文受大规模数据集3D手语识别的研究(61472398)资助。
刘 壮 男,硕士生,主要研究方向为计算机视觉、手部检测、手语评价,E-mail:zhuang.liu@vipl.ict.ac.cn;柴秀娟 女,博士,副研究员,主要研究方向为手语识别、手势交互、人机交互等;陈熙霖 男,博士,研究员,主要研究方向为计算机视觉、模式识别、多媒体技术、多模式人机交互等,E-mail:xlchen@ict.ac.cn(通信作者)。

随着深度摄像机的普及,越来越多的视觉任务开始采用带有深度捕获功能的采集设备。在获取彩色数据的同时也捕获了与之对应的深度数据。深度数据能够与彩色数据互相补充以提供更强的特征表达。因此,合理有效地使用深度数据能提高手部检测的精度。

与手部检测相关的研究方向是人脸检测和通用物体检测。近年来,人脸检测和通用物体检测均取得了巨大的进展。手部检测的研究也开始借鉴人脸检测和通用物体检测的方法。然而,相对于人脸和通用物体,手部由于存在更多的姿态变化、手指间自遮挡严重且手部姿态在不确定的条件下没有明显的几何结构,面临着更大的挑战。

早期的手部检测方法主要利用手工设计的特征来获取图像中的手部区域,如一些基于肤色的检测方法[1-3]、基于轮廓的检测方法[4-5]、基于运动的检测方法[6]以及利用可变形部件模型(Deformable Part Model,DPM)的检测方法[7]等。具有代表性的工作是 Mittal 等人[8]提出的融合多个检测器的方法。该方法利用肤色信息得到肤色候选框,然后利用手的形状特征(采用方向梯度直方图(Histogram of Oriented Gradient,HOG)表示)、手的上下文信息(同样用 HOG 表示)分别获取形状候选框和上下文候选框。总的候选框数为 3 种候选框数量之和。接下来将每个候选框对应的肤色、形状和上下文特征进行融合,并采用支持向量机(Support Vector Machine,SVM)对融合后的特征进行分类,从而得到每个候选框的分类置信度。最后采用一种超像素非极大抑制方法对这些候选框进行处理以得到最终的检测结果。然而,基于手工设计特征的手部检测方法并不稳定,其特别容易受到光照条件、背景环境和手的姿态等因素的影响,故很难在实际系统中得到应用。

最近,随着深度学习技术的发展,通用物体检测领域取得了重大突破。其中具有代表性的是基于候选区域的卷积神经网络系列方法[9-14]。Girshick 等人[9]提出了 R-CNN 框架,该方法利用 Selective Search[15]或 Edge Boxes[16]等候选区域生成方法,预先找出图中待检测物体可能出现的区域。然后利用训练好的卷积神经网络对每个区域提取鲁棒的特征,并采用 SVM 对每个区域进行分类。与此同时,对每个候选区域进行边框回归以获得更加精准的检测结果。最后采用非极大抑制方法[17]合并信度较高的区域以得到最终检测结果。针对 R-CNN 中繁琐的训练过程以及提取区域特征时大量的向前传播计算,Girshick[11]对 R-CNN 框架进行了改进,提出了 Fast R-CNN 框架。与 R-CNN 不同,Fast R-CNN 仅对整张图进行一遍向前传播计算,并且采用感兴趣区域池化(Region of Interests Pooling,RoI Pooling)的方式来获得每个候选区域对应的特征。这种方式极大地缩短了候选区域特征提取的时间。除此之外,Fast R-CNN 还采用了多任务框架(分类+回归)来取代 R-CNN 中独立操作的 SVM 分类和边框回归,使得分类和边框回归结果更加准确。近期,Ren 等人[12]在 Fast R-CNN 的基础上提出了 Faster R-CNN 框架。该框架解决了之前 R-CNN 系列检测框架的瓶颈,即需要使用耗时的区域生成方法来获得候选区域。Faster R-CNN 首先采用一种候选区域网络(Region Proposal Network,RPN)来生成少量并且高质量的候选区域,之后如同 Fast R-CNN 一样进行检测。这种采用网络生成候选区域的方式极大地提高了检测的速度和精度。同时,由于是端到端的训练,整个操作过程也极为简洁。

此外,一些基于回归的方法使得实时检测趋于可能,如 YOLO[18,20]和 SSD[19]等方法。相较于基于候选区域的一些检测方法,基于回归的方法更加高效。其虽然在检测精度上稍显不足,但检测速度的提升效果十分明显。而另一种传承且经典的层级卷积神经网络采用滑动窗口的方式和从粗到细的分类方法在特定物体的检测上优势明显。

总结近几年的发展,相对于传统的检测方法,基于深度学习的检测方法由于能够学习到更加鲁棒的特征,从而能得到更精确的检测结果[21]。然而,目前基于深度学习的通用物体检测方法仅仅考虑了物体的彩色信息,而没有有效地利用其对应的深度信息。因此,本文提出了一种基于双通道 Faster R-CNN 的手部检测框架,通过对彩色和深度信息的有效融合,极大地提高了手部检测精度。本文第 2 节简单介绍了 Faster R-CNN 检测框架;第 3 节给出了本文提出的双通道 Faster R-CNN 手部检测框架;第 4 节给出了实验对比及结果分析;最后总结全文。

2　Faster R-CNN 检测框架

Faster R-CNN 检测流程如图 1 所示。

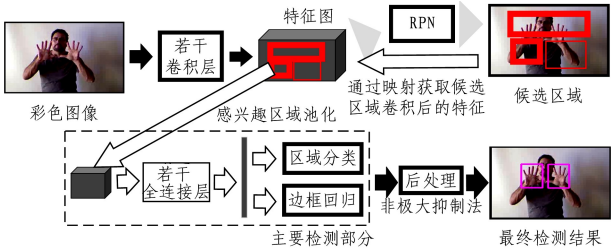


图 1　Faster R-CNN 检测流程图

Fig. 1　Detection flowchart of Faster R-CNN framework

Faster R-CNN 是近期由 Ren 等人[5]提出的一种端到端的通用物体检测框架。该框架主要包含两个模块:1)候选区域网络(RPN),该网络可以生成少量并且高质量的候选区域,其可用于之后的检测网络。RPN 是一种全卷积网络[22],它可以端到端地进行训练,以生成检测任务中的候选区域。2)对每一个候选区域进行检测处理的网络与FastR-CNN中的一样。在获取到生成的候选区域后,通过区域与特征图之间的映射关系得到每个候选区域卷积后的特征。之后通过感兴趣区域池化层将不同大小的区域特征归一化到同一尺度,并经过若干全连接层来获得每个区域的最终特征。每个区域的最终特征将经过两个并行的层,这两个并行的层一个用于判断区域所属的类别,另一个用于预测物体边框。在得到各个候选区域的检测结果后,通过非极大抑制法来得到最终的检测结果。

3　双通道 Faster R-CNN 检测框架

为有效利用不同通道提供的视觉信息,提出了双通道 Faster R-CNN 检测方法。本节主要对此检测框架进行介绍,整个检测流程如图 2 所示。

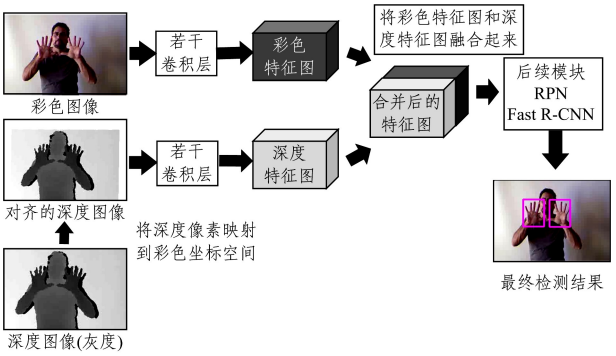


图 2 双通道 Faster R-CNN 的检测流程图

Fig. 2 Detection flowchart of two-stream Faster R-CNN framework

对于给定的待检测的彩色图像和其对应的深度图像(用灰度图表示),首先,通过彩色坐标空间和深度坐标空间的映射关系,将原始的深度像素对齐到对应的彩色坐标空间中;其次,分别对彩色图像和对齐后的深度图像进行各自的卷积等操作以提取特征,并将各自最后卷积层输出的特征图在通道维度上进行串联融合,构成 RGB-D 数据的最终特征;然后,在融合的特征图上利用 Faster R-CNN 中的 RPN 网络提取一定的候选区域,并对候选区域进行分类和边框回归;最后,通过非极大抑制法合并检测框,从而得到最终的检测结果。

整个框架主要分为 4 个部分:数据输入部分、特征提取部分、特征融合部分以及候选区域生成和检测处理部分。

3.1 数据输入

对于视觉任务而言,有效并充分地利用多模态间的信息十分重要。本文提出的双通道 Faster R-CNN 框架将彩色图像和深度图像作为输入,在输入的过程中,为了使彩色数据和深度数据保持一致,需要将深度图像上的像素对齐到彩色图像的坐标下。

采用相机标定的方法[23]将深度坐标下的像素对齐到彩色坐标下。在对齐的过程中,有些区域可能没有对应的深度信息,我们将这部分的像素值置为最大(灰度 255)。图 3 给出了对齐后的图像示例。

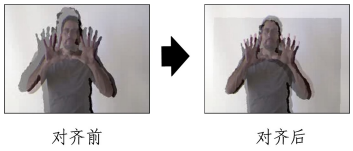


图 3 深度图像与彩色图像对齐示例

Fig. 3 Sample of RGB-D image pair before and after aligning depth pixel to RGB image space

在得到对齐的深度图像后,将彩色图像和对齐的深度图像作为输入,并分别对它们进行特征提取操作。

3.2 特征提取和融合

在卷积神经网络中,特征提取主要是指对图像进行卷积、激活、池化、全连接等操作,以获取表示图像的特征。对于给定的输入 X,经过卷积操作(线性变换)和激活函数(非线性变换)得到特征 Y:

$$Y=f(WX+b) \tag{1}$$

其中,W 为卷积权值,b 为偏置值,f 为非线性激活函数。而在使用卷积神经网络处理特定任务(例如手部检测)时,相对于直接采用特定任务数据(手部检测数据)来训练网络的方式,采用在通用数据(一般是用于分类的数据)上预训练网络并在特定数据上微调网络的方式更有优势。鉴于此,本文在彩色图像特征提取的卷积层结构中采取了已有预训练模型的卷积层结构;同时,在深度图像的特征提取过程中,采用了与彩色图像一样的卷积层结构。这样,我们便可以加载已有的预训练模型并在特定任务数据上对网络进行微调。对彩色图像和深度图像的特征进行提取,得到各自对应的彩色特征图和深度特征图。

在特征融合的过程中,本文采用在通道这一维度上串联特征图的方式将特征提取阶段得到的彩色特征图和深度特征图相融合。假设彩色特征图大小为 $h*w*c$,其中 h 代表特征图的高,w 代表特征图的宽,c 表示其通道数。由于深度特征提取的卷积结构与彩色相同,因此深度特征图的大小也为 $h*w*c$。在通道维度上对它们进行融合,融合后的特征图大小为 $h*w*(2c)$。采用矩阵形式表示,具体为:

$$Y_{Merge}(i,j,k)=\begin{cases}Y_{RGB}(i,j,k), & 0\leqslant k\leqslant c-1\\ Y_{Depth}(i,j,k-c), & c\leqslant k\leqslant 2c-1\end{cases} \tag{2}$$

其中,$i\sim[0,h-1]$,$j\sim[0,w-1]$,$k\sim[0,2c-1]$。

3.3 候选区域生成和检测处理

在候选区域生成和检测处理部分,我们沿用了 Faster R-CNN 中对应的部分。采用候选区域生成网络 RPN 在融合的特征图上进行卷积滑动操作,以生成相应的候选区域。由于融合后的特征既包含了彩色信息,也包含了深度信息,生成的候选区域更加有效。之后,我们采用 Faster R-CNN 中检测处理的网络,对每个候选区域进行感兴趣区域池化和全连接操作从而得到其特征,并将其用于分类和边框回归。这个过程中因为融合后的特征使结果更加精确。在得到分类信息和边框结果后,采用非极大抑制法来获取最终的检测结果。

3.4 具体实现

本文提出的双通道 Faster R-CNN 框架是在 Caffe[24]下搭建而成的。在数据的输入层中,我们将彩色通道(RGB)和深度通道(灰度图,DDD)串成一个 6 通道的数据对象并传入网络,之后通过 Caffe 中的分离层(Slice Layer)将其分离,然后对它们分别进行卷积等操作以提取特征。在彩色特征图和深度特征图的融合过程中,我们利用 Caffe 中的连接层(Concat Layer)将其串联起来。

为了达到更好的检测效果,采用加载预训练模型并进行微调的方式来训练网络,本文选取了 Faster R-CNN 中提到的 ZF 模型[25]。由于已有的预训练模型是在彩色图像(ImageNet[26])上训练得到的,因此其不能直接适应提出的双通道 Faster R-CNN 框架。我们将原预训练模型中的卷积参数分别拷贝到双通道 Faster R-CNN 中,并将其用于彩色图像特征提取的卷积层和深度图像特征提取的卷积层,然后存储成适应双通道 Faster R-CNN 框架的新预训练模型。

整个网络是在 Ubuntu 平台下采用一块 Titan X GPU 训练完成的。由于是微调网络,一般采用较低的学习率。采用

Caffe 中“step”的方式对学习率进行动态调节。

4　实验及结果分析

本节主要介绍对所提出的双通道 Faster R-CNN 框架和一些基本方法在 ChaLearn 手部检测数据集上的对比实验。首先简单介绍一下实验所用数据集及相应的评测协议,然后详细对比不同方法之间的检测性能,并验证本文提出的双通道 Faster R-CNN 框架的有效性。

4.1　实验数据及测试协议

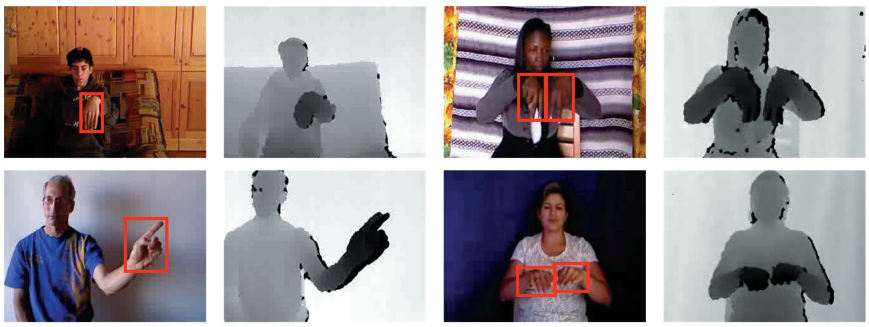
实验采用的是 ChaLearn 手部检测数据集。该数据集是从 2016 年 ChaLearn 手语竞赛[27]提供的彩色手语视频数据(包括彩色视频和深度视频)中随机挑选一定的帧,并对其进行人工标注而成。该数据集分为训练集和测试集两个部分,详细信息如表 1 所列。图 4 给出了 ChaLearn 手部检测数据集的部分示例。

表 1　ChaLearn 手部检测数据集

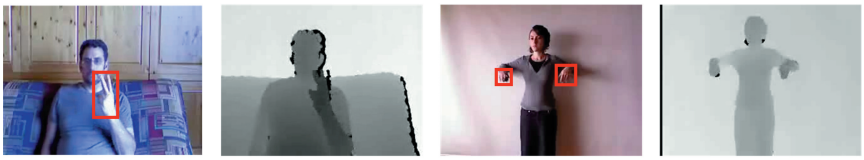
Table 1　ChaLearn hand detection dataset

集合	彩色图片数	深度图片数	手部框总数	手部框/每张图
训练集	50842	50842	83022	1.63
测试集	3155	3155	5006	1.59

由于划分的测试集图片中的手与训练集图片中的手来自于不同的人,该数据集相对来说更加具有挑战性。



(a)训练集中彩色和对应深度图像示例



(b)测试集中彩色和对应深度图像示例

图 4　ChaLearn 手部检测数据集示例

Fig. 4　Some examples of ChaLearn hand detection dataset

采用 2007 年的 PASCAL 视觉目标分类挑战(PASCAL Visual Object Classes Challenge 2007,VOC07)中的检测指标平均精度(Average Precision,AP)[28]来评价各个方法的性能。AP 代表查准率-查全率(Precision-Recall,PR)曲线下的面积。该值越大说明检测性能越好。

4.2　不同方法的实验对比

为了验证双通道 Faster R-CNN 框架的有效性,我们从数据的使用情况和数据间的融合策略两个方面进行了对比分析。选取了 4 种检测方式,具体为:采用彩色图像输入的 Faster R-CNN 网络、采用深度图像输入的 Faster R-CNN 网络、采用彩色图像和深度图像融合输入的 Faster R-CNN 网络以及分别输入彩色图像和深度图像的双通道 Faster R-CNN 网络。以下简称为单通道 RGB 网络、单通道 Depth 网络、单通道 RGB-D 网络和双通道 RGB-D 网络。图 5 给出了每种检测方式的大致流程。图 5(a)表示单通道 RGB 网络,该网络输入彩色图像,之后进行卷积等处理;图 5(b)表示单通道 Depth 网络,该网络输入的是深度图像,之后同样进行卷积等处理;图 5(c)表示单通道 RGB-D 网络,该网络输入的是彩色图像和深度图像融合的数据,即将 RGB 和 Depth 串联起来,之后对串联的数据进行卷积等处理;图 5(d)表示提出的双通道 RGB-D 网络,该网络输入彩色图像和深度图像,与图 5(c)不同的是,该网络对输入的彩色图像和深度图像分别进行卷积,然后在特征层面上进行融合。

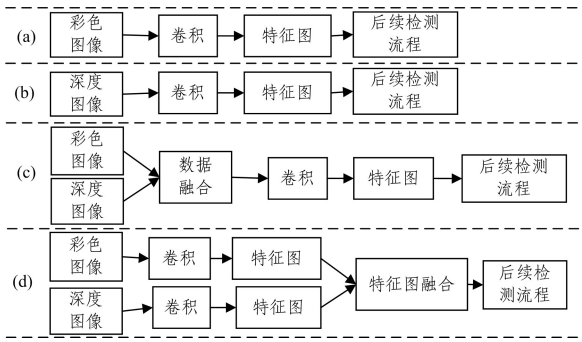


图 5　4 种用于手部检测的网络框架结构

Fig. 5　Four different frameworks for hand detection

根据 3.4 节中的具体实现,采用 4.1 节中的 ChaLearn 手部检测训练集对 4 种检测方法对应的网络分别进行训练。每个网络的基本训练参数设置都一致,如学习率、迭代次数等。对每个网络进行训练,设初始学习率为 0.001,并且在每 7 万次迭代后将学习率降低 10 倍,训练迭代的总次数为 35 万。在网络训练完毕后,分别对 4 个网络在 4.1 节中的 ChaLearn 手部检测测试集上进行性能测试。表 2 描述了随着迭代次数的增加,不同检测方式在 ChaLearn 手部检测测试集上的性能变化。

表 2 4 种网络在测试集上平均精度的变化

Table 2 Changes in average precision on test set of four different frameworks

网络	网络迭代次数						
	5W	10W	15W	20W	25W	30W	35W
单通道 RGB	0.462	0.538	0.536	0.542	0.542	0.543	0.542
单通道 Depth	0.253	0.323	0.324	0.322	0.323	0.323	0.325
单通道 RGB-D	0.431	0.528	0.545	0.542	0.546	0.550	0.549
双通道 RGB-D	0.541	0.596	0.602	0.606	0.606	0.606	0.607
* 单通道 RGB	0.575	0.601	0.598	0.598	0.599	0.592	0.598
* 单通道 Depth	0.277	0.336	0.333	0.340	0.341	0.342	0.343
* 双通道 RGB-D	0.583	0.627	0.622	0.628	0.626	0.627	0.626

其中 * 表示网络采用加载预训练模型并微调的方式进行训练。由于单通道 RGB-D 网络更改了网络输入数据的通道数(由 3 到 6),导致第一层卷积的通道也产生变化,因此不能很好地使用预训练模型,故没有采用加载预训练并微调的方式对其进行比较。

4.2.1 不同数据输入的比较

由表 2 可知,单通道 RGB 网络较单通道 Depth 网络更优。这是因为彩色数据包含了更多刻画手部表观的信息,相较于深度数据(灰度图)更加有判别力;同时,由于深度数据反映了手部空间的位置信息,能在一定程度上区分背景区域,相对于彩色数据,能更容易也更清晰地刻画手的形状,是彩色数据的有益补充,因此能有效帮助检测任务。而单通道 RGB-D 网络和双通道 RGB-D 网络的性能则都优于前两者。由此可见,融合使用两种数据后,网络能够学习到更强的特征,从而提升检测性能。

4.2.2 数据融合策略的比较

为了进一步探讨如何有效地融合彩色和深度数据,本文进行了单通道 RGB-D 网络和双通道 RGB-D 网络的比较实验。两种网络结构如图 5(c)、图 5(d)所示。二者都使用了彩色和深度数据。不同的是,单通道 RGB-D 网络是在输入中对两种数据进行融合,而双通道 RGB-D 网络则首先分别对两种数据进行特征提取,然后在特征上进行融合。从结果可以看出,特征层面上的融合更加有效。若只是将两种数据相融合来进行学习,由于彩色数据更加具有判别力,网络会倾向于利用彩色数据,而不能有效利用深度数据。

4.2.3 预训练模型对性能的影响

由于提出的双通道 RGB-D 网络能够很好地利用已有的预训练模型,因此对加载预训练模型后的情况也进行了对比实验。由表 2 不难发现,加载预训练模型并微调的训练方式能够给检测带来 2%~4%的性能提升。同样,无论加载预训练模型与否,提出的双通道 RGB-D 网络性能都明显优于其他网络。

结束语 在诸如手势识别等应用中,通过摄像头获取到的手部数据越来越多样化,如何合理、有效地利用不同模态的手部数据是目前研究的热点。针对手部检测任务,本文在传统 Faster R-CNN 网络结构的基础上,提出了双通道 Faster R-CNN 网络结构。网络结构以彩色数据和深度数据为输入,并对两种数据分别进行卷积等操作以提取各自的特征。然后融合两种特征用于检测。这种网络结构能够充分利用彩色信息和深度信息,极大地提升了检测性能,但是目前我们只是较简单地在最后的特征图上进行融合,如何提出更加有效的特征融合方式将是我们下一步的工作重点;同时,由于目前的方法在 CPU 环境下尚不能进行实时检测,因此如何提升算法的处理速度也要需要纳入考虑。

参 考 文 献

[1] KAKUMANU P,MAKROGIANNIS S,BOURBAKIS N. A survey of skin-color modeling and detection methods[J]. Pattern recognition,2007,40(3):1106-1122.

[2] DAWOD A Y,ABDULLAH J,ALAM M J. Adaptive skin color model for hand segmentation[C]// 2010 International Conference on Computer Applications and Industrial Electronics(ICCAIE). IEEE,2010:486-489.

[3] KÖLSCH M,TURK M. Robust Hand Detection[C]//FGR. 2004:614-619.

[4] SHOTTON J,BLAKE A,CIPOLLA R. Contour-based learning for object detection[C]// Tenth IEEE International Conference on Computer Vision,2005(ICCV 2005). IEEE,2005:503-510.

[5] ONG E J,BOWDEN R. A boosted classifier tree for hand shape detection[C]// Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. IEEE, 2004: 889-894.

[6] SHEIKH Y,JAVED O,KANADE T. Background subtraction for freely moving cameras[C]// 2009 IEEE 12th International Conference on Computer Vision. IEEE,2009:1219-1225.

[7] FELZENSZWALB P F,GIRSHICK R B,MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Nnalysis and Machine Intelligence,2010,32(9):1627-1645.

[8] MITTAL A,ZISSERMAN A,TORR P H S. Hand detection using multiple proposals[C]// Proceedings of British Machine Vision Conference. 2011:1-11.

[9] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2014:580-587.

[10] HE K,ZHANG X,REN S,et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]// European Conference on Computer Vision. Springer International Publishing,2014:346-361.

[11] GIRSHICK R. Fast r-cnn[C]// IEEE International Conference on Computer Vision. 2015:1440-1448.

[12] 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. 2015:91-99.

[13] DAI J,LI Y,HE K,et al. R-FCN:Object Detection via Region-based Fully Convolutional Networks[J]. arXiv preprint arXiv: 1605.06409,2016.

[14] ZHANG L,LIN L,LIANG X,et al. Is Faster R-CNN Doing Well for Pedestrian Detection? [C]// European Conference on Computer Vision. Springer International Publishing,2016:443-457.

[15] UIJLINGS J R R,VAN DE SANDE K E A,GEVERS T,et al. Selective search for object recognition[J]. International Journal of Computer Vision,2013,104(2):154-171.

[16] ZITNICK C L,DOLLÁR P. Edge boxes:Locating object proposals from edges[C]// European Conference on Computer Vision. Springer International Publishing,2014:391-405.

[17] NEUBECK A,VAN GOOL L. Efficient non-maximum suppression[C]// 18th International Conference on Pattern Recognition,2006(ICPR 2006). IEEE,2006,3:850-855.

[18] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once:Unified,real-time object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2016:779-788.

[19] LIU W,ANGUELOV D,ERHAN D,et al. SSD:Single shot multibox detector[C]// European Conference on Computer Vision. Springer International Publishing,2016:21-37.

[20] REDMON J,FARHADI A. YOLO9000:Better,Faster,Stronger [J]. arXiv preprint arXiv:1612.08242,2016.

[21] SHARIF RAZAVIAN A,AZIZPOUR H,SULLIVAN J,et al. CNN features off-the-shelf:an astounding baseline for recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014:806-813.

[22] LONG J,SHELHAMER E,DARRELL T. Fully convolutional networks for semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2015:3431-3440.

[23] BRADSKI G,KAEHLER A. Learning OpenCV:Computer vision with the OpenCV library[M]. Sebastopol:O'Reilly Media,Inc.,2008.

[24] JIA Y,SHELHAMER E,DONAHUE J,et al. Caffe:Convolutional architecture for fast feature embedding[C]// 22nd ACM International Conference on Multimedia. ACM,2014:675-678.

[25] ZEILER M D,FERGUS R. Visualizing and understanding convolutional networks[C]// European Conference on Computer Vision. Springer International Publishing,2014:818-833.

[26] DENG J,DONG W,SOCHER R,et al. Imagenet:A large-scale hierarchical image database[C]// IEEE Conference on Computer Vision and Pattern Recognition,2009(CVPR 2009). IEEE,2009:248-255.

[27] WAN J,ZHAO Y,ZHOU S,et al. Chalearn looking at people rgb-d isolated and continuous datasets for gesture recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016:56-64.

[28] EVERINGHAM M,ZISSERMAN A,WILLIAMS C K I,et al. The PASCAL visual object classes(VOC) challenge[J]. International Journal of Computer Vision,2010,88(2):303-338.

(上接第 223 页)

[7] PLATT J C,CRISTIANINI N,SHAWE-TAYLOR J. Large margin DAGs for multiclass classification[C]// 12th International Conference on Neural Information Processing Systems,MIT Press. 1999:547-553.

[8] KIJSIRIKUL B,USSIVAKUL N. Multiclass support vector machines using adaptive directed acyclic graph[C]// 2002 International Joint Conference on Neural Networks. IEEE,2002:980-985.

[9] BENNETT K P,BLUE J A. A support vector machine approach to decision trees[C]// 1998 IEEE International Joint Conference on Neural Networks. IEEE,1998:2396-2401.

[10] FEI B,LIU J. Binary tree of SVM:a new fast multiclass training and classification algorithm[J]. IEEE Transactions on Neural Networks,2006,17(3):696-704.

[11] CHEONG S,SANG H,LEE S Y. Support vector machines with binary tree architecture for multi-class classification[J]. Neural Information Processing Letters and Reviews,2004,2(3):47-51.

[12] KANG S,CHO S,KANG P. Multi-class classification via heterogeneous ensemble of one-class classifiers[J]. Engineering Applications of Artificial Intelligence,2015,43(C):35-43.

[13] TOMAR D,AGARWAL S. A comparison on multi-class classification methods based on least squares twin support vector machine[J]. Knowledge-Based Systems,2015,81(C):131-147.

[14] YANG Z,WU H,LI C,et al. Least squares recursive projection twin support vector machine for multi-class classification[J]. International Journal of Machine Learning and Cybernetics,2016,7(3):411-426.

[15] LIU M,ZHANG D,CHEN S,et al. Joint binary classifier learning for ecoc-based multi-class classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,38(11):2335-2341.

[16] SONG Q,XIAO X,JIANG H,et al. A new multi-class classification method based on minimum enclosing balls[J]. Journal of Mechanical Science and Technology,2015,29(8):3467-3473.

[17] KOSTIN A. A simple and fast multi-class piecewise linear pattern classifier[J]. Pattern Recognition,2006,39(11):1949-1962.

[18] ARONSZAJN N. Theory of reproducing kernels[J]. Transactions of the American Mathematical Society,1950,68(3):337-404.

[19] SNYDER W E,TANG D A. Finding the extrema of a region [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1980,2(3):266-269.

[20] BRAZDIL P,GAMA J. Statlog datasets [OL/DB]. [2016-10-25]. http://www.liacc.up.pt/ml/old/statlog/datasets.html.

[21] FRANK A,ASUNCION A. UCI machine learning repository [OL/DB]. [2016-10-20]. http://archive.ics.uci.edu/ml.

[22] CHANG C C,LIN C J. Libsvm:a library for support vector machines [OL]. [2016-10-26]. http://www.csie.ntu.edu.tw/cjlin/libsvm.

[23] HSU C W,LIN C J. A comparison of methods for multiclass support vector machines[J]. IEEE Transactions on Neural Networks,2002,13(2):415-425.