·图形图像处理· 文章编号：1000-3428(2021)12-0266-08 文献标志码：A 中图分类号：TP391.41

基于CornerNet-Saccade的手部分割模型

林竞力，肖国庆，张 陶，文 鑫
（西华大学 电气与电子信息学院，成都 610039）

摘 要：手部分割技术受手部形态、分割背景等因素的影响，分割效率难以提高。在CornerNet-Saccade模型基础上构造一种基于扫视机制的分割模型。通过模拟人眼观察物体时先扫视再仔细观察的行为特征，降低待处理图像的像素数量并在初步判断手部位置后将掩码分支添加到不同尺度特征图中，完成精细分割任务。在此基础上，引入线性瓶颈结构完成模型轻量化操作以降低模型复杂度。实验结果表明，该模型在Egohands数据集上平均交并比高达88.4%，优于RefinNet、U-Net等主流模型，轻量化处理后其平均交并比虽降低了2.2个百分点，但参数量仅为原模型的44.9%。

关键词：手部分割；深度学习；CornerNet-Saccade模型；扫视机制；轻量化结构

开放科学（资源服务）标志码（OSID）：

中文引用格式：林竞力，肖国庆，张陶，等.基于CornerNet-Saccade的手部分割模型[J].计算机工程，2021，47(12)：266-273.

英文引用格式：LIN J L，XIAO G Q，ZHANG T，et al.Hand segmentation model based on CornerNet-Saccade[J].Computer Engineering，2021，47(12)：266-273.

Hand Segmentation Model Based on CornerNet-Saccade

LIN Jingli，XIAO Guoqing，ZHANG Tao，WEN Xin
（School of Electrical and Information Engineering，Xihua University，Chengdu 610039，China）

【Abstract】The existing hand segmentation technology is limited in the segmentation efficiency due to multiple factors，including various hand shapes and complex segmentation background.To address the problem，this paper optimizes the CornerNet-Saccade model，and on this basis constructs a hand segmentation model using saccade mechanism.This model simulates the action mode of human eyes，which scans a target first，and then observes it carefully.In this way，the model reduces the number of pixels in the to-be-processed image.After preliminary judgment of the hand position，mask branches are added to the feature maps of different scales to complete the fine segmentation task.Moreover，to reduce the complexity of the model，a linear bottleneck structure is introduced to make the model more lightweight.Experimental results show that the mIOU value of the model reaches 88.4% on the Egohands dataset，which is higher than that of mainstream methods such as RefinNet and U-Net.Additionally，the lightweight model further reduces the mIOU value by 2.2% compared with the original model，while its parameters are only 44.9% of the original model.

【Key words】hand segmentation；deep learning；CornerNet-Saccade model；saccade mechanism；lightweight structure

DOI：10.19678/j.issn.1000-3428.0060030

0 概述

随着人工智能和物联网时代的发展，以用户为中心的新型交互方式如姿态、手势、语音控制、表情等被广泛应用，其中手势交互由于具有直观性、自然性、丰富性等优点，成为目前的研究热点。手部分割作为手势交互技术研究的核心，具有极大的研究价值。

手部分割的目的是隔离手部数据与背景数据，目前已有诸多研究人员开展该方面的研究。文献[1-2]通过对图片的颜色、纹理、形状等特征建模完成分割，该类方法处理速度快但泛化性能差，受环境影响大，难以满足真实条件下的分割。文献[3]通过佩戴数据手套获取手部位置与形状，该类方法准

基金项目：国家自然科学基金(61571371)；国家自然科学基金青年科学基金项目(61901393)；教育部春晖计划合作科研项目(Z201405)。
作者简介：林竞力(1977—)，男，副教授、博士，主研方向为图形图像处理、机器学习；肖国庆、张 陶、文 鑫，硕士研究生。
收稿日期：2020-11-17 **修回日期：**2020-12-17 **E-mail：**jl.lin@qq.com

计算机工程 www.ecice06.com

确度高、鲁棒性好，但成本较高，便利性较低。文献[4-5]通过查找手部区域不规则光流模式分割手部数据，该类方法对亮度变化较为敏感，能识别到运动中的手部。

使用多模态图像如RGB-D等对手部进行分割也是发展趋势之一，文献[6-7]通过实验表明该类方法准确率高，使用方便，但对数据采集设备要求较高，且增加了成本。随着深度学习技术的发展，以神经网络为基础的手部分割方法成为主流，如文献[8]使用基于卷积神经网络（Convolutional Neural Network，CNN）的皮肤检测技术，文献[9]针对手部分割任务构造一种基于跃层连接的编解码网络，文献[10]提出基于密集注意力机制的DensAtentionSeg网络，文献[11]评估了基于RefineNet[12]的分割模型。这些方法均在手部分割任务中表现良好，但往往依赖较深层次和高分辨率的特征，因此，多数用于手部分割的网络模型存在参数量大、推理时间长的问题，难以满足实时性要求。

为实现高精度且实时的手部分割，本文在CornerNet-Saccade模型[13]基础上利用多尺度特征，通过添加掩码分支构建一个基于扫视机制的手部分割模型，并使用以自我为中心的手部分割数据集Egohands进行训练。此外，利用MobileNet V2[14]中的线性瓶颈结构，构造实时性更高的轻量化模型，进一步提高模型效率。

1　CornerNet-Saccade模型

CornerNet-Saccade是一种基于扫视机制[15]的目标检测模型，在目标检测领域，以其高精度和低复杂度得到广泛的关注。扫视机制是指在推理期间选择性地裁剪和处理图像区域再进行后续操作。以输入大小为255×255×3、深度为54层的CornerNet-Saccade模型为例，该模型主要由Hourglass-54特征提取模块、注意力图模块和角点检测分支3个模块组成。

1.1　特征提取模块

CornerNet-Saccade模型中的Hourglass-54特征提取模块由3个三阶沙漏网络[16]组成，输入层大小为255×255×3。沙漏网络的主要结构为沙漏模块，图1所示为Hourglass-54网络所使用的最后一个沙漏模块，由多个残差模块[17]堆叠而成，输入张量经前2次降采样后变为64×64×256。

在图1中，方框内数据为输入输出的特征通道数，第1个残差模块对输入进行下采样并增加特征通道数，下采样后则进行特征提取，16×16、32×32、64×64均为为特征图尺寸。可以看到，左半边经过下采样后再进行上采样，逐步提取更深层次信息。右半边则直接在原尺度上进行，将两路结果相加，既提取到了较高层次信息又保留了原有层次信息。

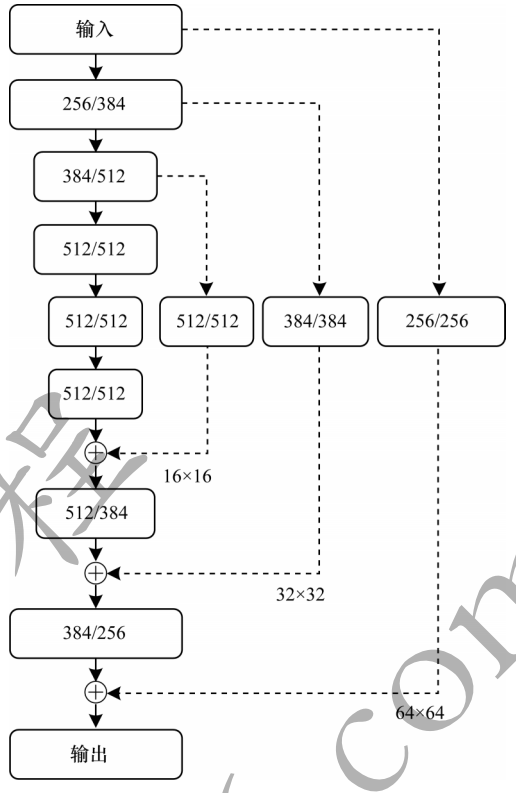


图1　沙漏模块结构

Fig.1　Structure of the hourglass module

1.2　注意力图模块

在所提网络中，每个沙漏模块都将生成3个不同尺度的特征图，如图1所示，在每次特征融合之前输出一个注意力图张量，这些张量具有不同的尺寸，能够反映不同层次的图片信息。使用这些不同尺度的张量来检测不同大小的手部图像。其中：尺寸为16×16的张量检测尺寸大于96的大物体；尺寸为32×32的张量检测尺寸大于32的中等大小物体；尺寸为64×64的张量检测尺寸小于32的小物体。

在训练过程中，将目标框的中心位置设定为正样本（y=1），其余为负样本（y=0），注意力图损失函数L_{att}如式（1）所示：

$$L_{\text{att}}=\begin{cases}\sum -\partial(1-y')^{\gamma}\times\log_{a}y', y=1\\ \sum(1-\partial)y'^{\gamma}\times\log_{a}(1-y'), y=0\end{cases}\tag{1}$$

其中：γ=2，α=4，用于平衡正负样本与难易样本之间的权重；y'表示预测到的值。

根据注意力图检测到关键点后将下采样后的图片进行缩放，根据物体尺寸选择缩放尺度，并在原图上进行裁剪，生成多个尺寸为255×255的图片，以进一步检测。

1.3　角点检测模块

如图2所示，特征提取模块获得特征图后进入角点检测模块以预测目标框。检测模块由结构相同的2组卷积组成，包括左上角和右下角预测。为提取角点特

计算机工程 www.ecice06.com

征，检测模块先对特征向量进行角点池化（Corner Pooling）[18]。对于左上角池化，从右侧观察边界框顶部水平方向，从下方观察边界框的最右边。因此，左上角的池化操作是分别对特征图中向量的左侧和顶部进行最大池化后相加的过程。将角点池化后得到的特征向量分别进行3组卷积得到目标框，产生的张量有热图（Heatmaps）、嵌入向量（Embeddings）和偏移（Offsets）。

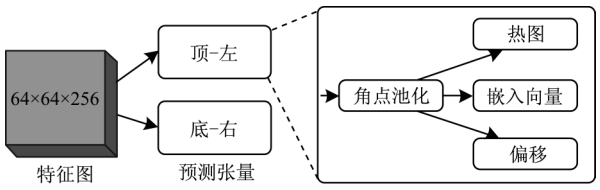


图2 角点检测模块

Fig.2 Corner detection module

热图的尺寸为64×64，表示对应位置是否为角点。由于使用离正样本近的负样本角点也能产生较好的边界框，因此在计算热图的损失函数时，使用高斯分布增强过的正样本标签处理损失函数，降低正样本附近圆周位置负样本的惩罚，权重由β参数控制。热图的损失函数L_{det}如式（2）所示：

$$L_{\text{det}}=\begin{cases}\sum(1-p_{ij})^{\alpha}\times\log_a(p_{ij}), y_{ij}=1\\ \sum(1-y_{ij})^{\beta}(p_{ij})^{\alpha}\times\log_a(1-p_{ij}), \text{其他}\end{cases} \tag{2}$$

其中：α参数控制难易样本的权重；p_{ij}为位置(i,j)预测到的值；y_{ij}为真实值。

在检测过程中可能会检测到多个角点，所以需要对角点进行配对，判断角点是否属于同一个目标框。因此，在CornerNet-Saccade中还使用了1-D嵌入向量，向量的距离决定了属于同一目标框的可能性。如式（3）和式（4）所示：

$$L_{\text{pull}}=\frac{1}{N}\sum_{k=1}^{N}[(\boldsymbol{e}_{tk}-\boldsymbol{e}_k)^2+(\boldsymbol{e}_{bk}-\boldsymbol{e}_k)^2] \tag{3}$$

其中，$\boldsymbol{e}_{\text{tk}}$代表目标k左上角点的嵌入向量；$\boldsymbol{e}_{\text{bk}}$代表其右下角点的嵌入向量；“pull”损失用于组合角点，“push”损失用于分离角点。

$$L_{\text{push}}=\frac{1}{N(N-1)}\sum_{k=1}^{N}\sum_{\substack{j=1\\ j\neq k}}^{N}\max(0,\Delta-|\boldsymbol{e}_k-\boldsymbol{e}_j|) \tag{4}$$

在预测角点的过程中，(x,y)位置的像素被映射为$(x/n,y/n)$而产生一定的精度误差，因此在映射回输入尺寸时，需对角点位置进行微调，偏移的损失函数L_{off}如式（5）所示：

$$L_{\text{off}}=\frac{1}{N}\sum_{k=1}^{N}\text{SmoothL1Loss}(o_k,\hat{o}_k) \tag{5}$$

其中：Smooth L1为损失函数；o表示偏移量。

2 改进手部分割算法

2.1 网络结构

针对手部分割效率较差的问题，在CornerNet-Saccade的基础上构造了基于扫视机制的手部分割模型。总体架构如图3所示，流程分为2步，先粗略检测（左），再精细分割（右）。

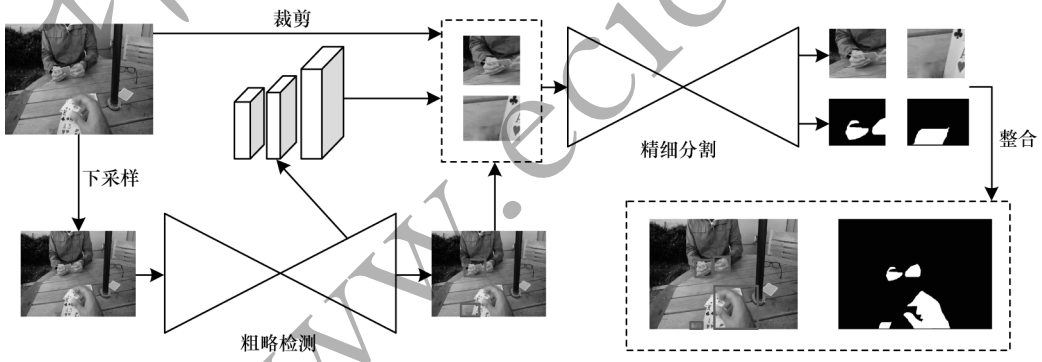


图3 手部分割总体架构

Fig.3 Overall architecture of hand segmentation

粗略检测阶段用下采样后的图片生成注意力图和特征图谱，估计目标在原图中的位置和粗略尺寸。精细分割阶段将检测到的位置映射回原图，对原图进行剪切后送入检测分割网络，精确定位目标位置并生成掩码。这种分两阶段检测的方法模拟了人观察物体的方式，即先检测物体大概位置，后对物体进行精细观察，使计算机在进行图像处理时无需对所有的像素均进行详细分析，进而大幅提高分割的精度和速度。

2.1.1 训练阶段

由于手部分割数据较少，因此需使用迁移学习方法得到模型参数。首先，在ImageNet数据集上对特征提取网络进行训练得到权重参数；然后在特征提取网络基础上添加角点检测网络、注意力图网络，使用COCO数据集对网络进行训练；最后将类别数降至1，添加掩码分支，利用训练集数据微调网络，得到模型参数。通过迁移学习的方法能够更快地训练出理想结果，并且在样本标签不够的情况下也能得

www.ecice06.com

到理想的结果。

本文要求模型既具有粗略检测又具有精细分割的能力，因此训练时，输入随机选取的下采样图片或裁剪图片，输入层大小为255×255×3。输入下采样图片时，将长边缩放至255，短边按比例缩放，其余位置补0；输入裁剪图片时，在目标框中心位置附近取中心点，根据物体大小随机决定缩放尺寸，裁剪出尺寸为255×255的图片。

2.1.2　推理阶段

推理阶段流程如图4所示，将下采样后大小为255×255×3的图片传入训练好的网络，提取到特征后，利用最后一个沙漏网络得到的注意力图和角点检测获得的位置对图片进行裁剪。

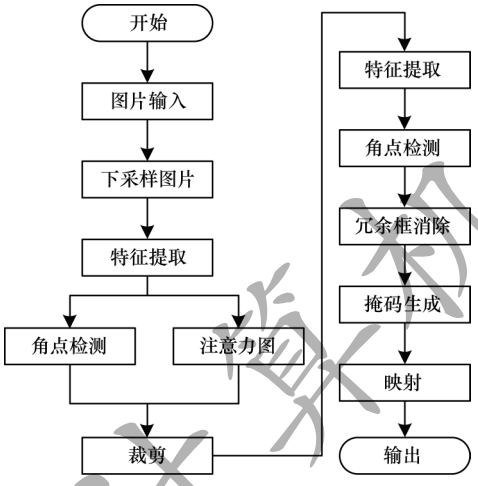


图4　推理阶段流程

Fig.4　Procedure of inference phase

在获得裁剪图片时，先选取得分大于阈值0.3的点作为图片中心位置，放大倍数则根据物体尺寸得出。物体尺寸越小，放大倍数S越大（S_{small}=4，S_{medium}=2，S_{large}=1）。根据角点检测结果获取裁剪图片时，通过边界框尺寸决定放大尺寸，边界框长边在放大后，小物体达到24，中物体达到64，大物体达到192。将裁剪后的图片再次送入网络，角点检测生成目标框，并使用Soft-NMS算法[19]消除冗余目标。

掩码结果与目标框结果相关，将目标框内区域作为感兴趣区域，选择置信度大于0.8的目标框，对目标框得分进行排序，得分由大到小排列。后续目标框与该目标框重合面积（IOU）较大时，则删除该目标框，最后将目标框与掩码图片映射至原图得到分割与检测结果。

2.2　掩码分支

在沙漏网络提取特征后，添加一组卷积网络预测手部图像掩码，结构参数如表1所示。输入128×128特征图，经最大池化后，与更深尺度特征融合，输出结果大小为64×64。将手部分割问题转化为二分类问题，并使用Sigmoid函数，大于0的位置认定为手部位置，小于0为背景。

表1　掩码分支结构参数

Table.1　Parameter of the mask branch structure

单元结构	类型	卷积核	大小/步长	输出
Input1	—	—	—	128×128
Conv1	卷积	256	(3×3)/2	64×64
	最大池化	—	—	
Input2	—	—	—	64×64
Add	特征融合	—	—	64×64
Conv2	卷积	256	(3×3)/2	32×32
	最大池化	—	—	
Conv3	卷积×3	256	(3×3)/2	32×32
TranConv	反卷积	256	(2×2)/2	64×64
Conv4	卷积	1	(3×3)/2	64×64

掩码分支在训练时使用Focal loss函数[20]，通过引入γ参数降低简单样本的权重，使训练时更加关注难以区分的样本。掩码分支损失函数L_{mask}如式(6)所示：

$$L_{mask}=\begin{cases}\sum -(1-p_{ij})^{\gamma}\times\log_a(p_{ij}), y_{ij}=1\\ \sum -(p_{ij})^{\gamma}\times\log_a(1-p_{ij}), y_{ij}=0\end{cases}\tag{6}$$

其中：γ=2，p_{ij}表示(i,j)位置掩码特征图的值；y_{ij}表示真实值；$y_{ij}=1$时该点为掩码区域。

在计算掩码的损失函数时，若将掩码图片转化为64×64大小后进行计算，则与输入图片对比，特征图上每个像素点均对应着裁剪后图片的16个像素，因此可通过改变掩码网络的尺寸来改变其分割精度。

3　线性瓶颈结构

本文使用Hourglass-54作为特征提取网络，该网络由许多残差块堆叠而成，因此大部分计算资源将消耗在残差模块上，残差模块如图5所示，分为直接映射和残差部分。虽然残差模块的使用减少了计算资源的消耗及缓解了梯度消失/爆炸的问题，但就参数数量和推理时间而言，其代价依旧昂贵。

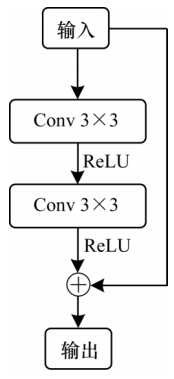
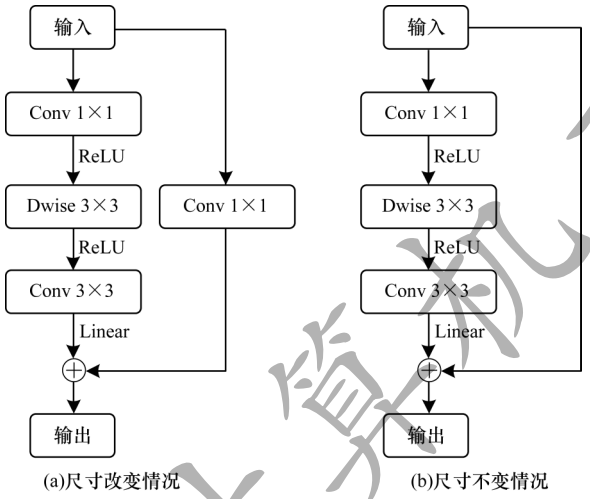


图5　残差块

Fig.5　Residual block

计算机工程 www.ecice06.com

为更近一步减少模型复杂度，提高模型效率，使用线性瓶颈结构对网络模型进行优化。线性瓶颈结构如图6所示，图6(a)为输入和输出尺寸改变的残差模块，图6(b)为输入和输出尺寸相同的残差模块。为匹配特征维度，将输入和输出改变的残差模块在直接映射部分添加1×1卷积。为减少计算资源的消耗，使用分离卷积替代原来残差模块中的普通卷积。为提高特征提取的准确率，先对特征向量进行扩张操作。为防止破坏特征，保证模型的表达能力，模块的最后一个卷积使用线性激活函数(Linear：Linear activation function)代替线性整流函数(Rectified Linear Unit，ReLU)。



(a)尺寸改变情况 (b)尺寸不变情况

图6 线性瓶颈结构

Fig.6 Linear bottleneck structure

4 实验

4.1 数据集准备

实验使用的Egohands数据集是以自我为中心背景复杂下的手分割数据，共由4 800张720×1 280像素的JPEG图片组成。这些图片取自48个不同环境和活动，将其随机分为训练集4 400张和测试集400张，并生成掩码图片和角点信息。部分数据集如图7所示。

(a)原图 (b)掩码图

图7 Egohands数据集

Fig.7 Egohands dataset

4.2 实验环境

实验在Linux系统上进行，显卡为11 GB的Geforce RTX 2080 Ti，CPU型号为i5-8500，使用Pytorch深度学习框架。

4.3 网络训练

根据2.1.1节中所描述方法，训练改进手部分割网络。在完成预训练后，使用批梯度下降法和Adam梯度下降法更新权重，Batch_Size设为8，共进行200次迭代，每次从训练集中随机选取3 000张图片进行迭代。

迭代时的损失函数如式(7)所示，由3部分组成，包括注意力图损失L_{att}，角点检测损失L_{cor}和掩码分支损失L_{mask}。其中角点检测损失由上文中的热图损失(L_{det})、嵌入向量损失($L_{push}+L_{pull}$)和偏移量损失(L_{off})组成。

$$L_{Loss}=L_{att}+L_{cor}+L_{mask} \tag{7}$$

在训练时，3种损失函数的损失值曲线如图8所示，cor_loss为角点损失值，att_loss为注意力图损失值，mask_loss为掩码分支损失值。由图8可知，3种损失函数所占比重基本相同，且随着迭代次数的增加，损失函数能够很好的收敛，网络参数逐渐平衡，因此可以判断该损失函数具有良好的稳定性和有效性。

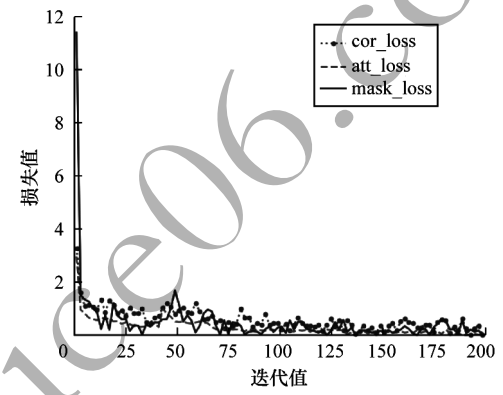


图8 3种损失函数的损失值曲线

Fig.8 Loss values of three loss functions

图9所示为训练过程中训练集总损失值(Train_Loss)和测试集总损失值(Vaild_Loss)。可以看到，测试集上损失值比训练集上小，随着迭代次数增加，两者均收敛良好，且从损失函数上看，没有产生过拟合现象，因此可以判断该网络训练成功。

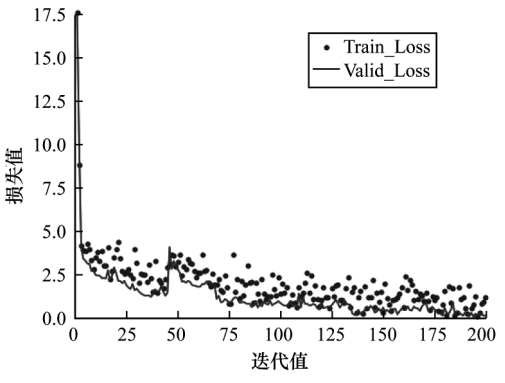


图9 总损失值曲线

Fig.9 Total loss curve

计算机工程 www.ecice06.com

4.4　实验结果

将训练好的网络在划分好的Egohands测试集上进行验证，使用平均交并比（mIOU）、平均精确率（mean average Precision，mPrec）和平均召回率（mean average Recall，mRec）等图片分割常用评价指标衡量模型分割效果，分割结果如图10所示。

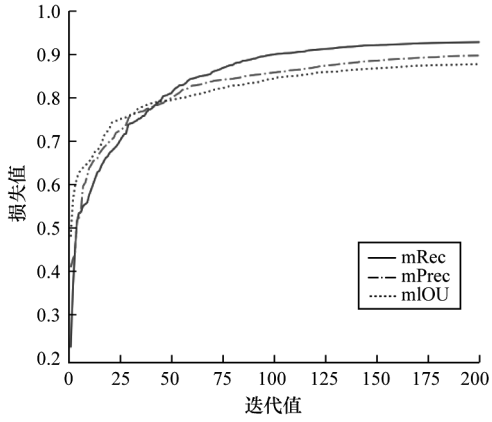
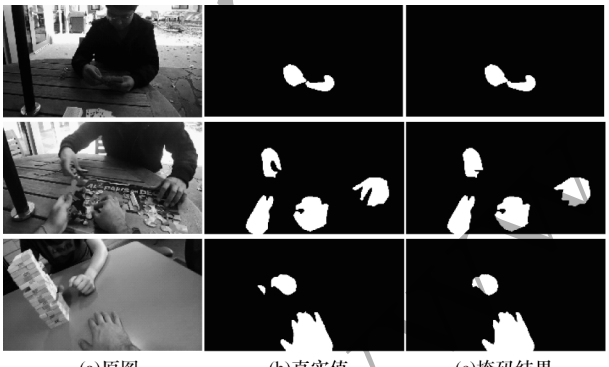


图10　本文模型分割性能

Fig.10　Segmentation performance of model in this paper

从图10可以看出，随着迭代次数的增加，平均交并比、平均精确率和平均召回率稳步上升，且在接近100次迭代后三者均趋于稳定，没有太大改变。此外，平均交并比、平均精确率、平均召回率分别高达88.4%、90.6%、91.2%。

手部分割结果如图11所示，图11（a）为原始图片，图11（b）是图片掩码的真实值，图11（c）为测试数据产生的掩码结果。



(a)原图　(b)真实值　(c)掩码结果

图11　分割效果对比

Fig.11　Segmentation effect comparison

从图11中可知，在复杂背景下，手部分割结果基本正确，手部形态的变化对实验结果没有产生很大影响。为进一步评估模型性能，分别选取了基于全局外观混合模型、基于GrabCut模型和基于CNN模型的方法进行对比，对比结果如表2所示，其中，CNN模型包括RefinNet、U-Net[21]和Deeplab V3+[22]。

表2　不同模型分割性能对比

Table 2　Comparison of segmentation performance of different models

模型	平均交并比	平均精确率	平均召回率
混合模型	0.478	—	—
GrabCut	0.556	—	—
RefinNet	0.814	0.879	0.919
U-Net	0.845	0.875	0.898
Deeplab V3+	0.870	0.909	0.958
本文模型	0.884	0.906	0.912

基于全局外观混合模型的方法，利用颜色、纹理和渐变直方图的稀疏组合完成分割，基于GrabCut的模型采用图分割和最大流技术，两者均属于传统的图像处理方法。基于CNN模型的方法中，RefinNet模型利用递归方式融合不同层特征完成语义分割，U-Net模型使用包含压缩路径和扩展路径的对称U形结构，Deeplab V3+模型利用空洞卷积[23]和Xception[24]提取丰富的语义信息，三者在语义分割领域均取得了较好的成果。本文所提模型具有最高平均交并比，且精确率和召回率也仅次于Deeplab V3+模型，具有较高的准确率。虽然如此，本文所提模型就参数数量和推理时间而言仍然昂贵。为适应实时图像处理任务，使用MobileNet V2模型中的线性瓶颈结构对网络模型进行优化，构造轻量级手部分割模型。轻量化手部模型分割结果如图12所示。其中，图12（a）为原模型的分割结果，图12（b）为轻量化模型的分割结果。

(a)原模型　(b)轻量化模型

图12　模型轻量化前后分割效果对比

Fig.12　Comparison of segmentation effects before and after model lightweight

对比图12发现，轻量化模型损失了部分精度，对于模糊物体而言存在少部分误检现象，但对于大部分图片而言无明显差异。为进一步比较分割结果的准确率和实时性，统计了原模型和轻量化模型的平均交并比、平均精确率、平均召回率、网络参数量和计算时间参数。对比结果如表3所示。

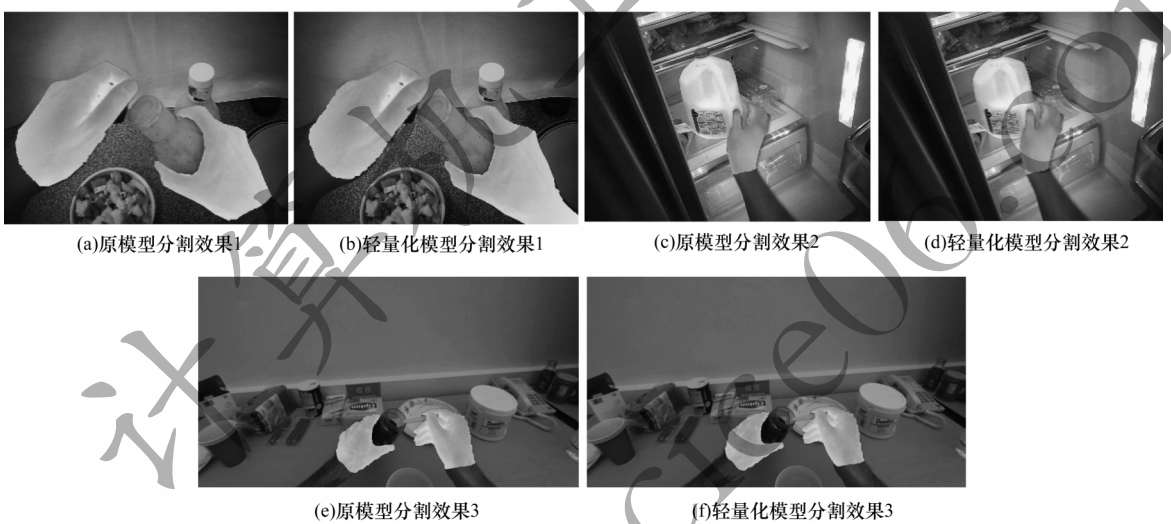
表3 模型轻量化前后分割性能对比

Table 3 Comparison of segmentation performance before and after model lightweight

模型	计算时间/ms	网络参数量/10^6	平均交并比	平均精确率	平均召回率
原模型	102	118	0.884	0.906	0.912
轻量化模型	44	53	0.862	0.896	0.896

可以看到，与原模型相比，轻量化模型平均交并比、精确率、召回率分别下降2.2、1、1.6个百分点，但参数减少了65×10^6，计算时间也降低至44 ms，在移动端部署时将比原模型具有更大的优势。

由于实际的应用场景更加复杂，因此训练出的模型应具有很好的泛化能力。为进一步测试模型在其他场景下的应用，引入新的图像数据[25]对训练好的模型进行测试，部分图片的分割效果如图13所示，图13(a)、图13(c)、图13(e)为本文基于CornerNet-Saccade网络所构造的模型，图13(b)、图13(d)、图13(f)为轻量化处理后的模型。可以看到，本文基于CornerNet-Saccade网络所构造的手部分割网络模型具有良好的泛化性能，在其他以自我为中心的数据集中也能很好地分割出手部掩码。



(a)原模型分割效果1 (b)轻量化模型分割效果1 (c)原模型分割效果2 (d)轻量化模型分割效果2

(e)原模型分割效果3 (f)轻量化模型分割效果3

图13 模型轻量化前后在新数据集下的分割效果对比

Fig.13 Comparison of segmentation effect under new date set before and after model lightweight

5 结束语

本文基于CornerNet-Saccade构造一种高效实时的手部分割模型。应用扫视机制避免对所有像素均进行精细处理，通过在不同尺度特征图中添加掩码分支完成多尺度分割任务，同时使用线性瓶颈结构对残差块进行改进以降低模型参数量，使模型满足实时处理的要求。实验结果表明，该模型在Egohands数据集上平均交并比为88.4%，高于RefinNet、U-Net、Deeplab V3+等常用手部分割模型。轻量化模型虽然在准确率上有所降低，但计算时间和参数复杂度均大幅减少。下一步将通过融合多尺度特征的方法优化分割细节，并针对严重遮挡情况下的分割问题进行研究，以拓展模型的应用范围。

参考文献

[1] LI C, KITANI K M. Pixel-level hand detection in ego-centric videos [C]// Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2013: 3570-3577.

[2] BAMBACH S, LEE S, CRANDALL D J, et al. Lending a hand: detecting hands and recognizing activities in complex egocentric interactions [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 1949-1957.

[3] ZHANG X, CHEN X, LI Y, et al. A framework for hand gesture recognition based on accelerometer and emg sensors[J]. IEEE Transactions on Systems, 2011, 41(6): 1064-1076.

[4] FATHI A, REN X, REHG J M. Learning to recognize objects in egocentric activities[C]//Proceedings of IEEE

Conference on Computer Vision & Pattern Recognition. Washington D. C. ，USA：IEEE Press，2011：3281-3288.

[5] LI M，SUN L，HUO Q. Flow-guided feature propagation with occlusion aware detail enhancement for hand segmentation in egocentric videos[J]. Computer Vision and Image Understanding，2019，187：1-11.

[6] MIRSU R，SIMION G，CALEANU C D，et al. A pointnet-based solution for 3d hand gesture recognition[J]. Sensors，2020，20(11)：3226.

[7] SHARMA P，ANAND R S. Depth data and fusion of feature descriptors for static gesture recognition[J]. IET Image Processing，2020，14(5)：909-920.

[8] ROY K，MOHANTY A，SAHAY R R. Deep learning based hand detection in cluttered environment using skin segmentation [C]//Proceedings of IEEE International Conference on Computer Vision Workshops. Washington D. C. ，USA：IEEE Press，2017：640-649.

[9] WANG W，YU K，HUGONOT J，et al. Recurrent U-Net for resource-constrained segmentation [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2019：2142-2151.

[10] BO Z H，ZHANG H，YONG J H，et al. DenseAttentionSeg：segment hands from interacted objects using depth input [EB/OL]. [2020-10-03]. https://www.researchgate.net/publication/332110106_DenseAttentionSeg_Segment_Hands_from_Interacted_Objects_Using_Depth_Input.

[11] UROOJ A，BORJI A. Analysis of hand segmentation in the wild[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2018：4710-4719.

[12] LIN G，MILAN A，SHEN C，et al. RefineNet：multi-path refinement networks for high-resolution semantic segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2017：1925-1934.

[13] LAW H，TENG Y，RUSSAKOVSKY O，et al. CornerNet-Lite：efficient keypoint based object detection[EB/OL]. [2020-10-03]. arXiv preprint arXiv：1904.08900，2019.

[14] SANDLER M，HOWARD A，ZHU M，et al. MobileNetv2：inverted residuals and linear bottlenecks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2018：4510-4520.

[15] 秦升，张晓林，陈利利，等. 基于人类视觉机制的层级偏移式目标检测[J]. 计算机工程，2018，44(6)：253-258.
QIN S，ZHANG X L，CHEN L L，et al. Hierarchical offset object detection based on humanvisual mechanism[J]. Computer Engineering，2018，44(6)：253-258. (in Chinese)

[16] NEWELL A，YANG K，DENG J. Stacked hourglass networks for human pose estimation[C]//Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2016：483-499.

[17] HE K，ZHANG X，REN S，et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2016：770-778.

[18] LAW H，DENG J. Cornernet：detecting objects as paired keypoints[C]//Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2018：734-750.

[19] BODLA N，SINGH B，CHELLAPPA R，et al. Soft-NMS--improving object detection with one line of code[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2017：5561-5569.

[20] LIN T Y，GOYAL P，GIRSHICK R，et al. Focal loss for dense object detection[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. ，USA：IEEE Press，2017：2980-2988.

[21] 朱辉，秦品乐. 基于多尺度特征结构的U-Net肺结节检测算法[J]. 计算机工程，2019，45(4)：254-261.
ZHU H，QIN P L. U-Net pulmonary nodule detection algorithm based on multi-scale feature structure [J]. Computer Engineering，2019，45(4)：254-261. (in Chinese)

[22] CHEN L C，ZHU Y，PAPANDREOU G，et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2018：801-818.

[23] 宦海，陈逸飞，张琳，等. 一种改进的BR-YOLOv3目标检测网络[J]. 计算机工程，2021，47(10)：186-193.
HUAN H，CHEN Y F，ZHANG L，et al. An Improved BR-YOLOv3 object detection network [J]. Computer Engineering，2021，47(10)：186-193. (in Chinese)

[24] CHOLLET F. Xception：deep learning with depthwise separable convolutions[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2017：234-241.

[25] LI Y，YE Z，REHG J M. Delving into egocentric actions [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. ，USA：IEEE Press，2015：287-295.

编辑　赖玉玲

计算机工程 www.ecice06.com