

基于实例分割和显著性计算的人工视觉多目标优化处理

王静^{1,2}, 刘建云¹, 韩彦岭¹, 周汝雁¹, 沈晓晶¹

1. 上海海洋大学信息学院, 上海 201306; 2. 农业部渔业信息重点实验室, 上海 201306

【摘要】通过实例分割Swin-Transformer提取分割所有前景对象,融合亮度、大小和位置图像显著性特征,提出模拟人类视觉注意机制的多特征融合注意力层级计算模型,为不同级别的前景物体采用适合的光幻视分辨率和亮度表达,实现不同的刺激编码策略进行层级优化处理。通过人工假体视觉的仿真试验表明,在所提出的多目标层级优化表达策略下,试验被试完成多目标识别的准确率、识别时间表现具有一定的显著提升。利用深度学习实例分割技术,层级化光幻视编码以仿生人类视觉选择性注意,达到增强假体植入者在复杂场景下的多物体感知,为视觉假体图像信息编码和优化处理研究的发展与应用提供参考。

【关键词】视网膜假体; 多物体识别; Swin-Transformer; 层级优化表达

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2023)03-0320-08

Multi-object optimization approach for artificial vision based on instance segmentation and saliency detection

WANG Jing^{1,2}, LIU Jianyun¹, HAN Yanling¹, ZHOU Ruyan¹, SHEN Xiaojing¹

1. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; 2. Key Laboratory of Fisheries Information, Ministry of Agriculture, Shanghai 201306, China

Abstract: All foreground objects are extracted and segmented through instance segmentation using Swin-Transformer, and the image saliency features of luminance, size and location are combined to construct a multi-feature fusion attention hierarchical computational model for simulating human visual attention mechanism. The suitable phosphene resolution and luminance are adopted for foreground objects of different importance levels to realize the hierarchical optimization using different stimulus coding strategies. The simulation experiments of artificial prosthesis vision show that the experimental subjects will achieve higher recognition accuracy and take less time to complete multi-object recognition when they adopt the proposed multi-target hierarchical optimization approach. Instance segmentation technique is used to cascade the phosphene coding for mimicking human visual selective attention, thereby enhancing the multi-object perception in complex scenes for visual prosthesis. The study provides a reference for the development and application of image information coding and optimization strategies for visual prosthesis.

Keywords: retinal prosthesis; multi-object recognition; Swin-Transformer; hierarchical optimization expression

前言

视觉假体是对视网膜色素变性和老年黄斑变性失明患者进行视觉功能修复的一种可行技术,其中

因为视网膜假体的植入手术风险低,同时可以最大程度利用视觉通路处理视觉信息,目前发展最好^[1-2]。正常的人类视觉系统具有一个接近 $160^{\circ}\times 175^{\circ}$ 的视野^[3],但当前假体设备由于生理和物理因素的限制,可植入电极数目非常有限,如应用于临床的视网膜上视觉假体 Argus II 仅有 60 个刺激电极^[4]。因此所诱发的光幻视形成的人工视觉感受和正常视觉感受相比,分辨率较低且视野范围受限,同时丧失了大量的视觉特征信息(颜色、纹理、深度等),导致假体植入者在生活场景中完成基本的视觉任务存在一定难度。在此背景下,利用图像处理策略,检测提取场景中的重要信息,优化表达人工视觉感受,改善假体植

【收稿日期】2022-09-29

【基金项目】国家自然科学基金(61806123,41871325);国家重点研发计划(2019YFD0900805);上海市青年科技英才扬帆计划(16YF1415700)

【作者简介】王静,副教授,主要研究方向:计算机视觉理论及应用、医学图像处理, E-mail: wangjing@shou.edu.cn

【通信作者】沈晓晶,副教授,主要研究方向:图像处理、计算机视觉, E-mail: xjshen@shou.edu.cn

入者对视觉感知的理解,成为视网膜假体研究中的一个重要方向。然而,视觉假体目前还没有进行广泛的临床应用,参与试验的植入者数量受限,许多研究小组通过招募正常视力人群作为试验被试开展仿真假体视觉试验对运动物体感知识别^[5-6]、导航避障^[7-10]、场景、物体识别^[11-13]等视觉功能任务评估所提图像处理策略的有效性。由于物体识别是最基本的视觉任务,利用基于计算机视觉的图像处理算法提升假体视觉感知能力成为研究热点。从视觉自主选择机制角度,部分研究集中于显著性检测的图像处理策略,以突出前景信息,减少背景中的冗余信息。如Li等^[14]提出一种自上而下的基于颜色和亮度差异特征的全局显著性检测算法,以准确性、效率和头部运动范围指标进行评估眼手协调任务表现,所提出策略具有明显的优势。Li等^[15]利用基于GBVS视觉显著性模型的显著性分割方法和基于grabCut的自适应迭代框架来自动提取前景对象,并进一步采用单独低像素化和缩小背景像素两种图像处理策略,自动提取和增强日常生活场景中的目标,其仿真试验结果表明,有效的图像处理策略可以改善低分辨率下的人工视觉下的视觉感知。Guo等^[16]分别基于流形排序的显著性检测和基于多阈值与连通性分析的图像分割技术,提出两种优化假体视觉呈现的图像处理策略,显著提高被试在给定场景中物体定位和识别能力。王静等^[17]基于全局亮度对比度,结合图像颜色空间转换并提出仿生注意机制的高斯差分算子,以进一步优化显著物体信息表达。随着深度学习技术的发展,有效的目标检测和分割网络,如经典的Mask R-CNN^[18],在人工视觉信息处理领域得到广泛的关注和应用研究。Sanchez-Garcia等^[19]结合语义和实例分割,利用实例分割突出场景中的相关对象,用语义分割提取场景的结构边缘信

息。人工视觉仿真试验结果表明,该策略对于室内场景的目标识别和房间识别具有较好的准确性。Han等^[12]利用基于深度学习的场景简化作为预处理,试验证明,与基于视觉显著性的模型相比,目标分割可以更好地支持场景理解,更好地提取高级场景信息。Jiang等^[20]提出一个基于Mask R-CNN层次化的图像处理策略,以优化表达场景中的多目标物体信息,增强场景感知能力。

以上研究工作均说明了基于视觉注意的图像处理算法能够改善人工视觉感知,有效提升低分辨率假体视觉下的物体识别能力。但此类研究中,面向复杂场景的多物体识别任务的信息处理相对较少。而且在上述大多数方法中,检测或分割出的前景物体在人工视觉下的信息表达形式大多都是一样的,易被混淆,不利于识别任务的开展。针对场景中的多目标物体,本文提出一种基于实例分割和显著性检测的多特征融合图像优化处理策略,通过有效的实例分割,根据不同视觉注意特征,层级化光幻视编码以仿生人类视觉选择性注意,达到增强假体植入者在复杂场景下的多物体感知的目的。

1 图像处理策略

在有限人工视觉条件下的物体识别任务中,应尽可能将场景中前景目标信息传递至假体植入者。本研究首先利用实例分割方法提取出场景中的多个前景物体。之后通过所提出的“多特征融合层级计算模型”,通过位置、大小及亮度显著性多特征,计算得出前景物体的重要性级别(Importance Level, IML)。根据IML,采用不同的光幻视分辨率和亮度等级对前景物体采用不同的刺激编码策略,实现低分辨率人工视觉条件下,真实场景中多目标物体的优化表达。本研究的图像处理流程如图1所示。

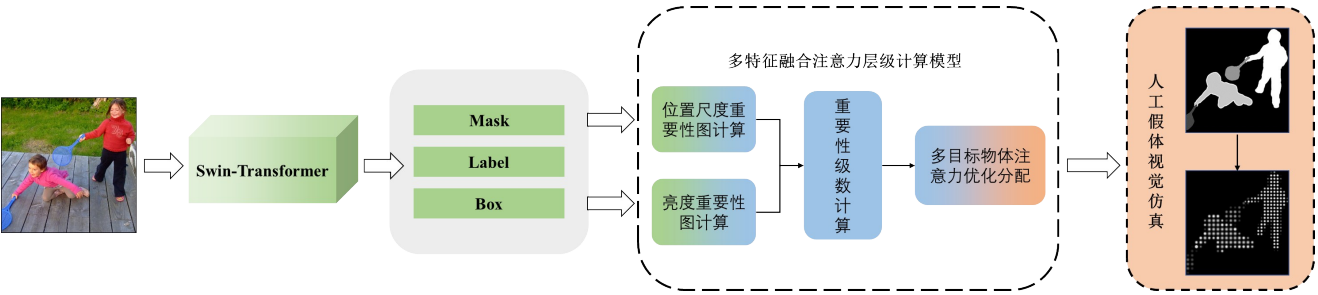


图1 基于实例分割和显著性检测的多特征融合图像优化处理流程图

Figure 1 Flowchart of multi-feature fusion image optimization based on instance segmentation and saliency detection

1.1 基于Swin-Transformer(Swin-T)的前景物体分割

考虑到实例分割结果的准确度和精度对后续的

图像优化表达算法具有重要影响,本研究首先对主流的实例分割模型Mask R-CNN^[18]、YOLACT^[21]、

SOLO^[22]以及Swin-Transformer-Segmentation (Swin-T-Seg)^[23]进行测试分析。综合选择了精度最高的Swin-T-Seg作为本研究实例分割预处理方法。

Swin-Transformer的结构一共有4个阶段,将图片输入网络之后,区块划分会将图片按照像素划分成不同的区块,并将区块在所有通道上拉伸为一维向量,然后将所有的向量组合成矩阵,经过4个阶段

之后进行分类和回归。第一阶段是区块组合在经过线性编码改变特征通道数,但不降低分辨率之后输入到Swin-Transformer Block中。除此之外的阶段都是由区块合并和Swin-Transformer Block组成,区块合并主要用来降低特征图分辨率。Swin-Transformer具体的网络架构如图2所示。

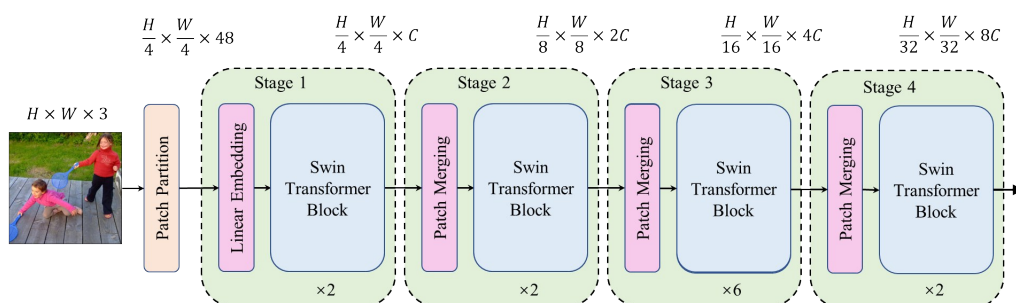


图2 Swin-Transformer的网络架构(引自文献[23])

Figure 2 Swin-Transformer network structure (cited from literature[23])

本研究中,通过Swin-T-Seg较为准确地获得采集图像中多个目标对象的mask。另外考虑到人工视觉仅能呈现有限视觉信息,小物体几乎不可能出现在植入者的视野中,因此将任何面积低于一定阈值的小对象剔除掉。参考文献[20],在本研究中,将阈值设置为2×2仿真光幻视点面积。

1.2 多特征融合注意力层级计算模型

Swin-Transformer实例分割获取多物体mask后,通过检测提取图像中人工视觉所缺乏的注意力相关特征,包括位置和尺度、亮度显著性,以此计算得出不同物体的IML,以实现后续的层级融合表达处理。具体的处理流程如图3所示。

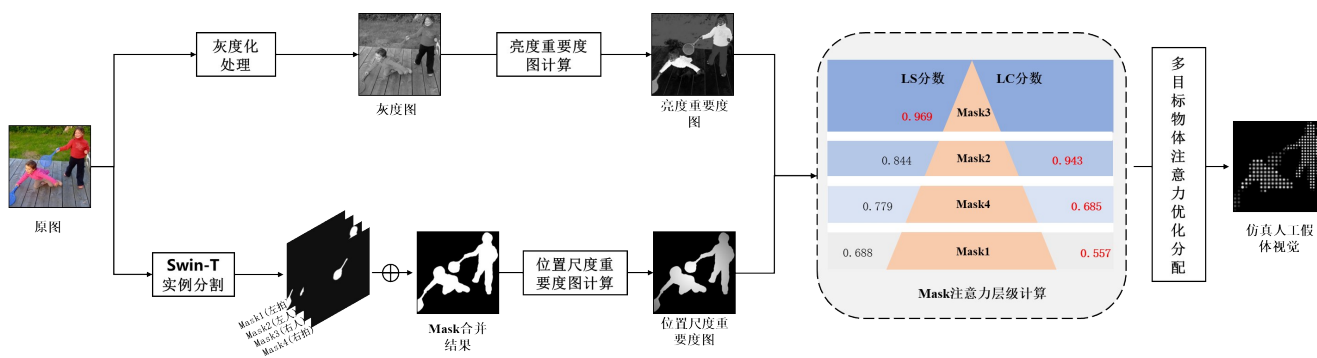


图3 多特征融合注意力层级计算模型

Figure 3 Multi-feature fusion attention hierarchical computational model

1.2.1 位置和尺度重要性计算 研究表明靠近图像中心的物体对人们更有吸引力,意味着靠近图像中心位置比远离中心位置更有可能是“重要”的^[24],也认为场景中物体的重要性是和其尺度和位置高度相关的^[25]。考虑到视觉注意的分布接近二维高斯模型,参考以上结论,本研究依据每个mask中像素点的位置和大小,采用式(1)进行像素点位置尺度重要性值的计算:

$$\text{LSimp}_{i,j} = \frac{1}{N} e^{-\frac{(i-c_x)^2 + (j-c_y)^2}{2\sigma^2}}, \quad (1)$$

$$(i,j) \in \text{mask}_k, k \in \{1, 2, 3, \dots, n\}$$

其中,mask_k是第k个mask区域,LSimp_{i,j}是mask_k中像素点(i,j)处的大小和位置重要性值,(c_x, c_y)是采集图像I的中心坐标。3σ=(w+h)/2,w和h分别是采集图像I的宽高。n是图像中mask的个数,N是图像像素

点数,用来归一化。其中第 k 个mask的尺度和位置的重要性的分数, $k \in \{1, 2, 3, \dots, n\}$,是通过式(2)计算的:

$$LS = \sum LSimp_{i,j}, (i,j) \in mask_k$$

(2)

1.2.2 亮度重要性计算 视觉显著性是指视觉刺激所具有的一类显著的客观属性,如颜色、亮度、边缘等,能够使具有该类特征的物体和周围其他物体形成强烈的对比,从而迅速得到人们的注意。这里考虑计算某个像素在整个图像上的全局亮度对比度,即该像素与图像中其他所有像素在亮度上的差异之和作为该像素的亮度显著值,采用加速计算的全局亮度对比显著性检测方法,生成全局显著图^[26],如式(3)所示:

$$S(a_m) = \sum_{n=0}^{255} f_n \|a_m - a_n\|$$

(3)

其中, f_n 表示灰度图中灰度级 a_n 出现的次数, $\|a_m - a_n\|$ 是灰度值 a_m 和灰度值 a_n 的欧式距离, $S(a_m)$ 是灰度值为 a_m 的像素的全局亮度重要性值。基于此,每个mask的亮度重要性分数通过式(4)计算:

$$LC = \frac{1}{M} \sum S(I_k), I_k \in mask_k, k \in \{1, 2, 3, \dots, n\}$$

(4)

其中, $mask_k$ 是第 k 个mask区域, n 是mask的个数, $S(I_k)$ 是像素点 I_k 的亮度重要性值, LC 是整个mask的亮度重要性分数, M 是像素点数,用来归一化。

1.2.3 多目标注意力优化分配 以往研究提出利用二值化和边缘提取增强来抑制背景信息^[27],或者前景背景采用不同的灰度表达以及背景边缘化来提升前景灰度级动态范围的图像处理策略增强前景对象^[28],对于单个物体识别表现均有显著的提升。参考以上工作思想,本研究在实例分割及每个mask的位置和尺度重要性分数和亮度重要性分数计算后,

计算不同物体的IML,目的是根据各mask的IML,采用提高亮度级别、边缘信息增强、降低光幻视点分辨率几种表达处理对场景中的多目标物体进行层级优化表达,提高各前景目标间的对比,以模拟正常视觉下的视觉注意分配。IML评价准则如式(5)所示:

$$ID_{mask_k} = \begin{cases} 1, LS(mask_k) = \operatorname{argmax}\{LS(mask_k), k \in [1, n]\} \\ \operatorname{index}\{\operatorname{sorted}(LC(mask_k), k \in [1, n])\} + 1 \end{cases}$$

(5)

其中, ID_{mask_k} 是 $mask_k$ 的IML, $\operatorname{argmax}\{LS(mask_k), k \in [1, n]\}$ 代表尺度和位置分数最高, $\operatorname{index}\{\operatorname{sorted}(LC(mask_k), k \in [1, n])\}$ 代表除尺度和位置重要性分数最高的mask外,其余各mask的LC分数排序索引。即依据mask的LS分数,筛选出场景中最重要物体,其余mask的重要性程度由它们的LC分数决定。根据每个mask的IML,对每个mask以及其边缘根据所提的多目标注意力优化分配策略分配不同的灰度等级及不同的仿真假体视觉的分辨率,具体策略如表1所示。其中第一等级的mask分配最高的灰度,第二重要的mask分配第二等级的灰度,并且对其边缘增强,在利用Swin-T-Seg提取出的mask的基础上,采用5×5的Canny算子基于mask提取边缘信息,并进行膨胀增强,对其分配最高级别亮度,其余的mask依据其重要性程度分配相应等级的灰度。考虑到像素数的增加会提高图像的平均识别准确率^[29],并局限于目前可植入的电极数,本研究以物体的IML为参考,将场景中的多个物体的mask采用局部平均法^[14],分别降采样至32×32或24×24的分辨率。其中第一重要的mask采用32×32的分辨率,其余物体采用24×24的分辨率。

表1 多目标注意力优化分配策略

Table 1 Multi-object attention optimization allocation strategy

Mask重要性级别优化策略	第一等级	第二等级	第三等级	第N等级(N>3)
亮度	最高亮度	次高亮度	相应等级亮度	相应等级亮度
边缘增强	否	是	否	否
分辨率	32×32	24×24	24×24	24×24

1.3 人工假体视觉仿真处理

为了验证本文提出的图像优化表达策略,主要对比以下两种处理方法:(1)Direct Pixelization(DP):输入图像直接灰度化,降采样为高斯点,无其它的图像处理;(2)SM:通过Swin-T-Seg提取前景对象,然后使用像素化,生成高斯点。

图4为几种处理方法的部分结果图示。从DP结果可以看出,一定的光幻视点密度下的人工视觉仍能感知部分场景信息。但是对观察场景中物体的数量、内容的理解仍然存在较大的困难。文献[30]表明,植入者即使在视网膜假体设备的帮助下,识别一个字母也需要花几十秒或者几百秒,或将袜子按照

不同的颜色完成分类,准确率也达不到80%。特别对假体佩戴者来说,在实际中经常面临多种物体图像

信息被采集、编码并以刺激呈现时,能够区分出多个物体并能找到感兴趣目标,具有更大的挑战性。

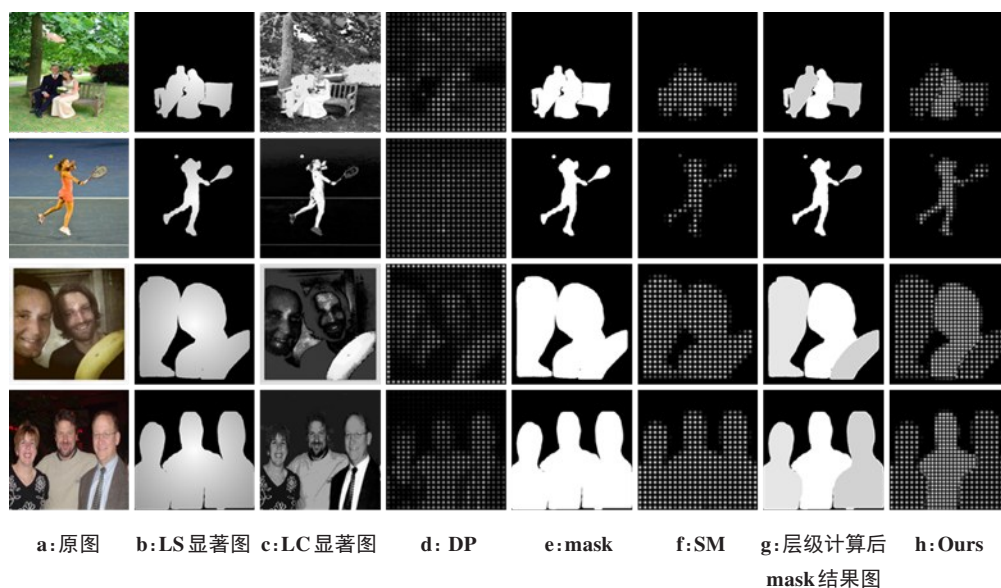


图4 不同图像处理策略下仿真人工假体视觉结果图对比

Figure 4 Comparison of visual results of artificial prosthesis with different image processing strategies

SM方法能够准确提取分割多个前景物体,便于对多物体数量的感知。但是当物体相互之间有遮挡时,由于每个前景对象的分辨率和亮度都是同级的,不能很好被区分开。如图4的第3行,SM处理之后不能直观地看出右边的人手里还拿着一根香蕉。

通过仿生视觉注意力分配,所提出的层级计算模型融合了注意力偏移程度、目标尺度及亮度显著性等多特征,评价多目标的重要性程度,据此采用不同的刺激编码策略,以优化分配人工视觉自下而上的视觉注意。以图4的第1行结果为例,中间的女士与背景的差异最大且占据了图像的中心,层级模型中该区域计算得出的LS和LC重要性分数都是最高的,因而对其赋予最高的亮度等级和最多的信息表达,使其获得更多的视觉注意。

1.4 统计学方法

采用SPSS 25.0软件对数据进行处理,符合正态分布的计量资料用均数±标准差表示,组间比较采用 t 检验, $P<0.05$ 为差异有统计学意义。

2 人工假体视觉仿真试验

2.1 试验被试

36名被试参与试验,年龄20~30岁,均来自上海海洋大学,视力正常或矫正正常。他们均未参加过假体视觉仿真试验,不具备先验知识。正式试验之前,被试首先参加一个预试验,目的是帮助被试熟悉试验环境,

了解试验步骤,以顺利进行后续的正式试验。

2.2 试验平台和环境

试验平台包括便携式电脑和头戴式显示设备HMD(HTC vive)。其中一台移动工作站(Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz, 16.0 GB, RAM, Dell, China)用来进行图像处理和显示,并且和HMD串流,将视觉刺激呈现给被试,另外一台供试验人员记录数据。本研究提出的图像处理方法基于Python3.8, CUDA 10.1,在Pytorch1.2开源框架下实现。同时数据记录程序在Visual Studio 2017下基于C++语言开发。

2.3 图像获取和预处理

所有的试验场景和物体对象是经过流调所得日常生活中常用或者常见的。但对于被试试验中的图片都是陌生的。通过摄像头采集图像,然后根据Argus II视网膜植入系统只能提供近似 20° 的视野,试验中所用的所有图片都裁剪至 480×480 的分辨率,覆盖 20° 对角视角,传递至计算机进行图像处理。

2.4 仿真试验设计

试验具体的过程是:呈现视觉刺激后,要求被试尽力描述HMD上呈现场景中的对象数量和场景内容,被试可以进行适当的猜测。试验主要分为两个阶段任务。任务(1)为识别场景中的物体数量,如果被试能够准确,说出图片中物体的数量,得1分,否则不得分。任务(2)为识别场景内容,要求被试描述场景中全部物体的名

称或者场景中的活动,如果被试正确描述场景中每个物体的名称或者正确描述场景,他们可以获得3分;如果被试不确定物体是什么,他们可以描述物体的形状或者外观,如果场景中所有对象的名称或者形状都回答正确,可以获得3分;如果一大半对象的形状或者名称回答正确,可以获得2分;如果少于一半的物体名称或者形状回答正确,获得1分;如果全部物体的形状或名称都没有回答正确,不得分。

正式试验共有72张图片,为了最大限度地减少学习效应,将36名受试者分为6组,每组观看36张照片,要求6组受试者按不同顺序用3种方法(DP、SM和Ours)完成试验。每3组的试验顺序都呈拉丁方形式,具体顺序如表2所示。

表2 不同被试组的图像处理策略呈现顺序
Table 2 Presentation order of image processing strategies for different groups of subjects

被试组	第1组 图像处理策略	第2组 图像处理策略	第3组 图像处理策略
第1组	DP	SM	Ours
第2组	SM	Ours	DP
第3组	Ours	DP	SM

2.5 仿真假体视觉试验结果分析

本文采用识别时间(reaction time)、识别准确率(Recognition Accuracy, RA)对几种图像处理策略进行评估。识别时间是由每名被试识别每幅图像所花的时间平均计算所得。根据被试在两项任务中的表现得分,分别计算完成任务的RA。同时计算场景中有相互遮挡对象的图像(41/72)的RA。

从反应时间结果来看,被试在DP下的平均识别时间为(14.3±1.36) s。应用SM策略和本文的优化表达策略之后,识别时间分别下降到(12.67±1.1) s和(11.8±0.97) s,如图5所示。在DP策略下,由于看到

零星的黑点、白点,许多被试反映无法感知对象的轮廓。任务过程中大部分时间停留在观察感知阶段,描述上大多为不确定、犹豫的状态。由于SM策略和本文的策略都能提取出前景物体,大多数被试在识别过程中是放松的,在看到视觉刺激时,能够迅速地描述场景中对象的数目,较为肯定地说出场景中对象的名称或者描述出场景对象的活动或者形状和特征。因此被试在SM和本文的策略下识别场景中的多物体对象的时间差异并不显著。

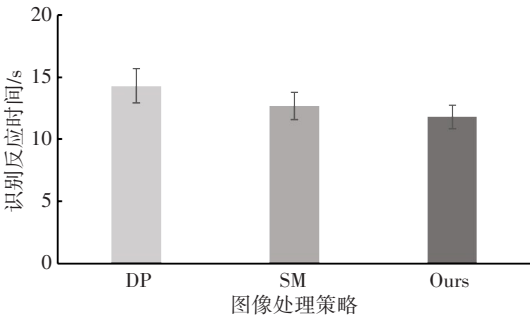


图5 3种图像处理策略下的识别时间
Figure 5 Time required for recognition with 3 image processing strategies

如图6所示,在识别场景中的对象数目任务(1)中,DP策略下被试平均识别正确率值为15.56%±2.38%。应用SM策略和本文的策略后,分别上升到55.09%±4.73%和89.35%±1.23%(图6a)。在使用本文的策略和SM策略时,被试的平均识别正确率显著提高。同时,本文的策略相对于SM策略也具有明显的优势。对于任务(1)有对象相互遮挡的场景中,DP、SM和本文策略下的均识别正确率值分别为13.74%±2.92%、45.20%±6.30%和89.43%±2.45%(图6b)。从图6中可以看出,当场景中出现对象遮挡时,本文的策略有更显著的提升,因为层级的图像优化表达策略显著提高了对象之间的区分能力。

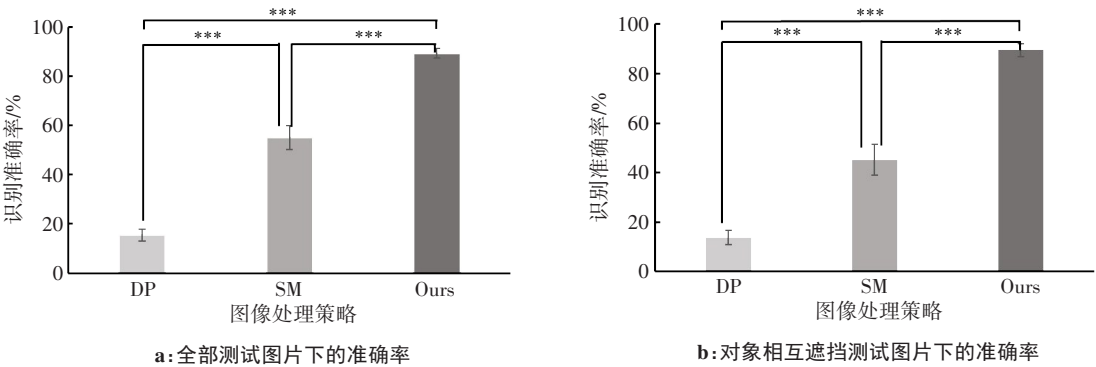


图6 任务(1)在3种处理策略下的识别准确率(***)
Figure 6 Recognition accuracies of different strategies for task 1 (***)

如图7所示,在描述场景中对象的名称或者形状和特征的任务(2)中,DP策略下被试的平均识别正确率值为 $12.42\% \pm 2.13\%$,而SM策略和本文的策略被试的平均识别正确率显著提高至 $56.94\% \pm$

4.10% 和 $81.37\% \pm 2.69\%$ (图7a)。在评估任务(2)下的对象相互遮挡的场景时,DP、SM和本文策略下的平均准确率分别为 $10.70\% \pm 2.50\%$ 、 $50.68\% \pm 5.04\%$ 和 $84.01\% \pm 3.20\%$ (图7b)。

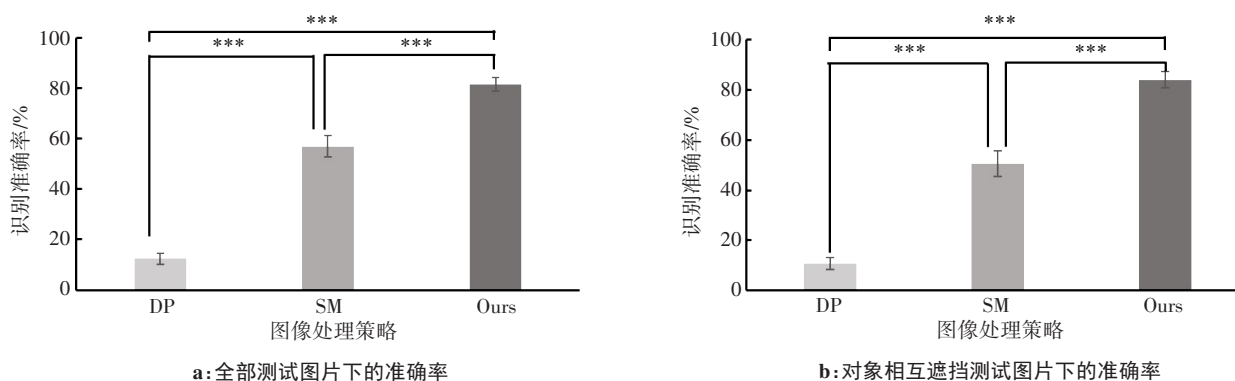


图7 任务(2)在不同策略下的识别准确率(*** $P < 0.01$)

Figure 7 Recognition accuracies of different strategies for task 2 (*** $P < 0.01$)

3 讨论

本文所提策略下被试可以较好地地区分不同物体,但是有些对象仍然不能正确识别。分析图像后发现,这类图像刺激中Swin-T-Seg提取部分对象轮廓较为粗糙模糊,相对于物体本身边缘会有缺失和扭曲。同时由于获得的mask缺乏物体内部细节,进一步缺少了利于识别的重要特征。如两人撑伞的图像中,因为实例分割处理没有提取出雨伞的伞柄,而且缺乏内部的伞骨细节,除了个别被试认为是雨伞,还有避雨斗篷,甚至是公交站牌等其他描述。

为了区分不同对象,本研究还对较显著的物体采用边缘增强策略,即边缘较亮,但是内部的像素点相对其边缘暗一些。尽管对边缘进行了膨胀处理,但是在最后处理效果中,由于边缘占据光幻视数比例较低,边缘仍然是整个物体区域的最暗部分。如图4中的第3行对应的g和h列,可以看出香蕉的边缘比内部亮,但是低像素化以后香蕉的边缘反而没有内部亮。在以后的工作中,会进一步解决边缘增强的处理方式,使人工视觉下不同对象间的边界更明确。

4 结论

本研究面向人工假体视觉下的多物体识别任务,提出了基于实例分割和显著性计算的多特征融合的层级优化表达策略。利用分割效果较好的Swin-T-Seg提取场景中的所有前景对象,提出仿生视觉注意的层级优化表达模型,对不同前景对象分配不同的分辨率和亮度等级。开展的仿真试验结果表

明,本文所提策略可以提高被试对复杂场景下的多目标物体的感知识别能力,特别地,当场景中出现有相互遮挡对象时,本文所提策略有更显著的效果。本研究为假体设备的图像处理单元性能提升提供一种可行的图像处理策略,也为植入者在日常生活中更有效地开展物体识别任务提供试验参考。

【参考文献】

- [1] Kramme R, Hoffmann KP, Pozos RS. Springer handbook of medical technology[M]. Heidelberg: Springer, 2011: 821-834.
- [2] 闫妍, 柴新高, 陈焱, 等. 视觉假体的研究进展[J]. 生理学报, 2016, 68(5): 628-636.
Yan Y, Chai XY, Chen Y, et al. Research progress of visual prosthesis[J]. Acta Physiologica Sinica, 2016, 68(5): 628-636.
- [3] 刘勇. 干细胞移植治疗视网膜变性疾病临床试验现状及前景[J]. 第三军医大学学报, 2019, 41(3): 187-191.
Liu Y. Stem cell transplantation for retinal degeneration: current status and future prospects[J]. Journal of Army Medical University, 2019, 41(3): 187-191.
- [4] Humayun MS, Dorn JD, Cruz LD, et al. Interim results from the international trial of second sight's visual prosthesis[J]. Ophthalmology, 2012, 119(4): 779-788.
- [5] Guo F, Yang Y, Xiao Y, et al. Recognition of moving object in high dynamic scene for visual prosthesis[J]. IEICE Trans Inf Syst, 2019, E102-D(7): 1321-1331.
- [6] 姜广森, 赵瑛, 王铁, 等. 仿真假体视觉下的运动感知研究[J]. 中国医学物理学杂志, 2022, 39(9): 1105-1113.
Jiang GM, Zhao Y, Wang T, et al. Motion perception in simulated prosthetic vision[J]. Chinese Journal of Medical Physics, 2022, 39(9): 1105-1113.
- [7] Vergnien V, Macé MJ, Jouffrais C. Simplification of visual rendering in simulated prosthetic vision facilitates navigation[J]. Artif Organ, 2017, 41(9): 852-861.
- [8] White J, Kameneva T, McCarthy C. Deep reinforcement learning for task-based feature learning in prosthetic vision[C]//2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Berlin: IEEE, 2019: 2809-2812.
- [9] Perez-Yus A, Bermudez-Cameo J, Lopez-Nicolas G, et al. Depth and motion cues with phosphene patterns for prosthetic vision[C]//Proceedings of the IEEE International Conference on Computer Vision

- Workshops. Venice: IEEE, 2017: 1516-1525.
- [10] Ge C, Kasabov N, Liu Z, et al. A spiking neural network model for obstacle avoidance in simulated prosthetic vision[J]. *Inf Sci*, 2017, 399: 30-42.
- [11] Mallya A, Lazebnik S. Learning informative edge maps for indoor scene layout prediction [C]//IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 936-944.
- [12] Han N, Srivastava S, Xu A, et al. Deep learning-based scene simplification for bionic vision [C]//Augmented Humans Conference 2021. New York: Association for Computing Machinery, 2021: 45-54.
- [13] Elnabawy RH, Abdennadher S, Hellwich O, et al. A yolo-based object simplification approach for visual prostheses [C]//35th International Symposium on Computer-Based Medical Systems. Shenzhen: IEEE, 2022: 183-186.
- [14] Li H, Han TT, Wang J, et al. A real-time image optimization strategy based on global saliency detection for artificial retinal prostheses[J]. *Inf Sci*, 2017, 415: 1-18.
- [15] Li H, Su X, Wang J, et al. Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision[J]. *Artif Intell Med*, 2018, 84: 64-78.
- [16] Guo F, Yang Y, Gao Y, et al. Optimization of visual information presentation for visual prosthesis[J]. *Int J Biomed Imaging*, 2018, 2018: 1-12.
- [17] 王静, 张羽婷, 张云, 等. 基于图像显著性的人工视觉图像处理策略[J]. *中国医学物理学杂志*, 2019, 36(11): 1277-1283.
- Wang J, Zhang YT, Zhang Y, et al. Image processing strategy based on saliency detection for artificial vision[J]. *Chinese Journal of Medical Physics*, 2019, 36(11): 1277-1283.
- [18] Huang Z, Huang L, Gong Y, et al. Mask scoring r-cnn [C]//Conference on Computer Vision and Pattern Recognition. Long Beach: Computer Vision Foundation/IEEE, 2019: 6409-6418.
- [19] Sanchez-Garcia M, Martinez-Cantin R, Guerrero JJ. Semantic and structural image segmentation for prosthetic vision[J]. *PLoS One*, 2020, 15(1): e0227677.
- [20] Jiang H, Li H, Liang J, et al. A hierarchical image processing strategy for artificial retinal prostheses [C]//2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE). Beijing: IEEE, 2020: 359-362.
- [21] Bolya D, Zhou C, Xiao F, et al. Yolact: real-time instance segmentation [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 9156-9165.
- [22] Wang X, Kong T, Shen C, et al. Solo: segmenting objects by locations [C]//European Conference on Computer Vision. Glasgow: Springer, 2020: 649-665.
- [23] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 9992-10002.
- [24] Judd T, Ehinger K, Durand F, et al. Learning to predict where humans look [C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto: IEEE Computer Society, 2009: 2106-2113.
- [25] Goyal A. Understanding and predicting importance in images [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. RI: IEEE Computer Society, 2012: 3562-3569.
- [26] Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues [C]//Proceedings of the 14th ACM International Conference on Multimedia. CA: ACM, 2006: 815-824.
- [27] Lu YY, Kan H, Liu J, et al. Optimizing chinese character displays improves recognition and reading performance of simulated irregular phosphene maps[J]. *Invest Ophthalmol Vis*, 2013, 54(4): 2918-2926.
- [28] Wang J, Lu YY, Gu L, et al. Moving object recognition under simulated prosthetic vision using background-subtraction-based image processing strategies[J]. *Inf Sci*, 2014, 277: 512-524.
- [29] Zhao Y, Lu YY, Tian Y, et al. Image processing based recognition of images with a limited number of pixels using simulated prosthetic vision[J]. *Inf Sci*, 2010, 180(16): 2915-2924.
- [30] Dagnelie G, Christopher P, Arditi A, et al. Performance of real world functional vision tasks by blind subjects improves after implantation with the Argus® II retinal prosthesis system[J]. *Clin Exp Ophthalmol*, 2017, 45(2): 152-159.

(编辑:陈丽霞)