

基于眼动数据的分类视觉注意模型

王凤娇 田 媚 黄雅平 艾丽华

(北京交通大学计算机与信息技术学院 北京 100044)

摘 要 视觉注意是人类视觉系统中的重要部分,现有的视觉注意模型大多强调基于自底向上的注意,较少考虑自顶向下的语义,也鲜有针对不同类别图像的特定注意模型。眼动追踪技术可以客观、准确地捕捉到被试的注意焦点,但在视觉注意模型中的应用还比较少见。因此,提出了一种自底向上和自顶向下注意相结合的分类视觉注意模型 CMVA,该模型针对不同类别的图像,在眼动数据的基础上训练分类视觉注意模型来进行视觉显著性预测。实验结果表明:与现有的其它 8 个视觉注意模型相比,该模型的性能最优。

关键词 视觉注意,视觉显著性,分类模型,自底向上,自顶向下

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2016.1.020

Classification Model of Visual Attention Based on Eye Movement Data

WANG Feng-jiao TIAN Mei HUANG Ya-ping AI Li-hua

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract Visual attention is a very important part of the human visual system. Most of the existing visual attention models emphasize bottom-up attention, considering less top-down semantic. There is few specific attention model for different categories of images. Eye tracking technology can capture the focus of attention objectively and accurately, but its application in visual attention model is still relatively rare. Therefore, we proposed a classification model of visual attention (CMVA) combining bottom-up with top-down factors, which trains classification models for different categories of images on the basis of eye movement data so as to predict visual saliency. Our model was compared with other existing eight models, proving its superior performance than other models.

Keywords Visual attention, Visual saliency, Classification model, Bottom-up, Top-down

1 引言

人类的信息加工在很大程度上依赖于视觉,来自外界的信息约有 80%~90%是通过人眼获得的,进而通过人眼进行高层次的认知和复杂处理,如物体认知或者场景理解,这都依赖于视觉注意机制。注意机制研究在生物视觉处理过程中是一个难以解释清楚但又非常重要的问题,因为其涵盖了计算机视觉、心理学、神经生物学等多个领域,如基于注意机制的显著技术生产的仿造视网膜假体——仿视网膜眼镜是部分失明者重见光明的第一步。目前,注意机制问题已经得到了国内外的广泛关注。

眼动追踪技术可以记录被试的视线停留位置及浏览路径,判断出被试的真实意愿,近年来逐渐成为国内外注意机制研究领域的热点课题。但由于眼动设备非常昂贵且很难在处理数据时都配备实时交互的眼动平台,因此迫切需要视觉注意模型(visual attention model)代替眼动设备来准确预测人眼的注视焦点。视觉注意模型依据人类视觉系统感知理论模拟人眼观察图像的过程,提取一幅图像中人眼所能观察到的

引人注意的焦点,也即该图像的显著性区域。

很多注意模型的基础能够追溯到 1980 年 Treisman 和 Gelade^[1]提出的“特征整合理论”,它指出哪些视觉特征是最重要的,以及它们如何引导人们的注意力。其中能够吸引注意的因素主要有两方面:自底向上的特征,主要包括与周围环境有较大差异的因素,如颜色、亮度、方向等,是数据驱动的;自顶向下的语义理解,包括图像中容易引起人们视觉注意的人脸、动物、汽车等目标,是任务驱动的。

按照处理方式的不同,注意机制模型分为自底向上(bottom-up)的基于数据驱动的计算模型和自顶向下(top-down)的基于任务驱动的计算模型。大多数注意模型基于自底向上的计算模型,其中以 Itti^[2]模型最具代表性。Itti 模型主要包括 3 个方面:特征提取、显著图生成和注意焦点转移。Itti 和 Koch 通过对已知图像提取多个方面的底层特征,如方向、颜色、亮度等通道,建立金字塔式结构,并借助中央周边差获得图像的多通道上的特征映射图。近年来,有关注意模型的研究涌现了很多新的具有代表性的研究成果。其中, J. Harel 等人^[3]提出的 GB 模型对图像的兴趣区域的提取与金字塔式

到稿日期:2015-03-28 返修日期:2015-05-20 本文受北京市高等学校“青年英才计划”(YETP0541, YETP0546),中央高校基本科研业务费专项资金(2015JBM036),国家自然科学基金(61273364, 61473031, 61472029)资助。

王凤娇(1989—),女,硕士生,主要研究方向为计算机视觉、模式识别、图像处理, E-mail: 15300262021@163.com; 田 媚(1980—),女,博士,讲师,主要研究方向为视觉感知的智能计算; 黄雅平(1974—),女,博士,教授,主要研究方向为图像处理、模式识别和机器学习; 艾丽华(1964—),女,博士,教授,主要研究方向为并行与分布式计算。

的模型不同,它在生成显著图的过程中采用的是纯数学的方法,借助马尔科夫链的平稳分布得到显著区域。M. M. Cheng 等人^[4]提出了基于区域对比度的显著区域检测算法 HC 方法和 RC 方法。HC 方法基于颜色通道上的直方图,图中的每一个像素的显著值都是由它和其它所有像素的颜色值差异决定的;RC 方法基于空间信息来增强对比度,区域的显著性值由每个区域和其他区域的对比度加权而来表示。该算法简单、高效,并且能够产生全分辨率的高质量显著性图。J. M. Zhang^[5]提出了一个新的基于布尔图的显著性模型,这个模型利用了全局拓扑测度,这一拓扑测度已被证明有助于感知的图形-背景分割。

但是大多数视觉注意模型没有考虑自顶向下的语义,尤其是容易引起人们注意的目标检测这一因素;虽然 Judd^[6]提出的少数自顶向下的模型^[7,8]融合了目标检测这一特征,但是类似这样的注意模型训练时融合的特征没有针对性,人脸、动物、汽车等目标同时加入到模型中需要耗费的时间长并且有的目标特征并不涉及,没有与实际的眼动数据相匹配。

针对以上问题,本文主要工作有两点:(1)从现有的几个图像库中挑选出动物、植物、人物和交通工具 4 类图像,对每一类图像进行眼动实验,得到每一幅图像的眼动数据,产生一个“ground truth”显著图来记录被试的实际注视焦点,构建 4 类图像的眼动数据集。(2)在 Judd^[6]的注意模型的基础上,对每一类图像引入不同的高层特征,并提取出一系列的底层特征来定义显著位置,用一个线性支持向量机来训练一个注意模型,称之为分类视觉注意模型(Classification Model of Visual Attention, CMVA)。CMVA 模型结果如图 1 所示。我们可以针对不同类别的图像选择相应的分类模型,通过实验得到不同类别的图像需要加入哪些高层特征来得到性能最优的视觉注意模型,并与其他模型进行比较分析。



图1 CMVA 模型得到的视觉显著性图

本文第 2 节阐述了眼动实验的相关工作;第 3 节具体描述了 CMVA 模型;第 4 节给出了实验结果;最后对本文研究进行了总结和展望。

2 眼动实验

通过眼动实验,我们得到有人眼注视路径的眼动数据集,进而产生“ground truth”显著图,再由阈值得到正负样本点。由正负样本训练产生视觉注意模型。

2.1 眼动数据集

我们收集的每类图像包含 120 幅图像,图像的尺寸范围为 400×400 到 1280×1280 。

动物类主要来自网络图像和 NUSEF^[9],包括背景简单、背景复杂和单目标、多目标图像,动物种类有狗、猫、鸟、马等;人物类主要来自 MSRA^[10]和 MIT^[11],包括单人图与多人图;植物类主要来自 MSRA^[10]和 MIT^[11],包括背景简单和背景复杂两类,种类包括花和绿色植物;交通工具类主要来自 PASCAL VOC 2007^[12],包括自行车、汽车、公交车等。

2.2 眼动实验过程

一共有 14 位 20~30 岁的被试进行眼动实验。采用的眼动仪是型号为 SMI iView X RED 的遥测式眼动仪,采样频率为 120Hz,可视屏幕分辨率为 1280×1024 。进行眼动实验时,每一位被试坐在可视屏幕前以全分辨率观看每幅图,如图 2 所示。每幅图的显示时间是 4s~5s,每两幅之间有 1s 的灰色屏幕。为了确保高质量的追踪结果,每一位被试在看图之前要对眼睛进行校准。4 类图分开进行实验,以免被试产生视觉疲劳。通过眼动实验,我们可以得到一幅图像的多种类型的实验结果图。



图2 一名被试正在进行眼动实验

为了从被试的眼动数据中获取图像的连续的显著图,我们对 14 位被试的叠加的焦点图进行高斯卷积,得到连续平滑的“ground truth”显著图。同时每位被试对每幅图像的注视位置形成一个文本文件,汇总所有被试对每幅图像的注视位置,并对汇总的文本文件进行处理,得到每一幅图像的所有被试所看的注视点图。

2.3 眼动实验结果分析

通过眼动实验结果可以得到 14 位被试关注的区域。所有被试关注的主要区域是相同的,但也有些不同的视线点。

针对这些结果,我们发现不管是哪一类图像,人们主要注意的是目标明显的前景区域,比如对于动物类,人们的视线集中在动物身上;对于人物类,视线集中在人身上;对于人物和动物五官清晰的图像,人们的视线会集中在五官上,而不是看整张脸;同时人们的视线偏向于图像中心的区域;而对于背景比较杂乱的图像,颜色鲜艳的区域也会引起人们的关注。

3 分类视觉注意模型 CMVA

我们对分好类的图像进行特征提取,对每一类图像引入不同的特征并用线性支持向量机 LSVM 来训练产生分类视觉注意力模型 CMVA。生成流程如图 3 所示。

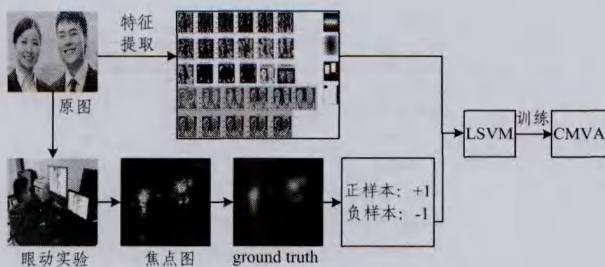


图3 CMVA模型流程

3.1 特征提取

在CMVA模型中,我们首先需要将每幅图像尺寸调整为 200×200 ,提取图像中的底层和高层特征,如图3所示,用这些特征来训练模型。而模型之间的一个主要的差别是它们是否依赖自底向上的因素、自顶向下的因素(也就是底层特征和高层特征)或者是两者的组合。



A为特征(1);B为特征(2);C为特征(3)(4);D和E为特征(5);F为水平线检测;G为偏向中心;H为检测人;I为检测人脸;J为汽车、动物类检测;K为原图

图4

1. 底层(自底向上)特征

自底向上的因素主要基于视觉场景的特性,是激励驱动的。这种方式中,吸引我们注意力的感兴趣区域必须充分地不同于其周围特征,这种注意力机制叫做外在的、自动的、灵活的或者周边的因素。一般来说,方向、颜色和亮度特征经常被用于图像的显著性检测。除此之外,还有其他的特征,如深度、大小、光流(optical flow)等。这里,我们用到的底层特征有:

(1)在4个方向和3个尺度上的13个可控金字塔的子带特征;

(2)由Itti和Koch的显著模型计算得到的方向、颜色和亮度3个对比通道的特征;

(3)红色、绿色、蓝色3个通道值以及这3个特征对应的各颜色通道的概率值;

(4)在6个缩放尺度上,图像的三维颜色直方图用中值滤波器滤波之后计算得到的上述颜色通道的5个概率;

(5)Torralla^[13]和GBVS^[3]模型得到的显著性特征。

如图4所示,A—E,一共29个底层特征。

2. 高层(自顶向下)特征

自顶向下的因素是由认知现象如知识、期望、奖励和当前任务决定的,是任务驱动的。因为大多数目标位于地球表面,所以水平线自然是人们关注的显著性特征,水平线检测是我们需要的高层特征。人们在拍摄一幅图像时,会自然而然地将目标置于中心区域,同时人们在看一幅图像时,也倾向于先看图像的中心,因此我们在训练的模型中加入了中心偏向这一高层特征。

为了得到4个分类视觉注意模型,提高模型的时间效率和功能效率,除了水平线检测和中心偏向,不同类别的模型需要加入的高层特征不尽相同。对于不同种类的图像,我们加入的特征具有各自类别图像目标,如动物类需要的是各种动物的检测,但是这样需要耗费的时间比较长,因为仅仅是其中一个物种的种类就很多,如动物有狗、猫、马、鸟、牛、猪等。经过实验验证,发现对于动物类模型,只引入狗、猫、鸟这种常见的动物目标检测特征的检测性能就不亚于加入所有动物类特征,并且大大节省了时间成本。同样地,对于人物类模型,只需要加入检测人以及检测人脸的特征;对于交通工具类模型,只需要加入自行车和汽车两类检测的特征即可,不必将摩托车、公交车、火车等所有车类目标引入。不同于以上3类明显的目标特征,对于植物类模型,这类图像的显著性主要体现在其底层特征上,不需要加入目标检测的特征。

3.2 训练模型

训练的系统平台是64位的Ubuntu14.04,所用工具是Linux下的MATLAB R2013a。

为了训练所提出的模型,从每类图像中随机选出20幅作为训练集,10幅作为测试集。在训练之前,我们对20幅图进行特征收集,每幅图已经缩放到 200×200 像素,因此一共收集 $200 \times 200 \times 20 = 400000$ 个样本点的特征。在每幅图所对应的“ground truth”显著图中,按照显著性从最引人注目的前30%的显著位置中选10个正样本,标记为1;同样从最不引人注目的后30%的区域里选10个负样本,标记为-1。从而得到训练集样本数为 $20 \times 20 = 400$ 个,测试集样本数为 $20 \times 10 = 200$ 个。然后使用线性支持向量机(Linear Support Vector Machine,LSVM)进行训练。

LSVM^[14]是台湾林智仁(Chih-Jen Lin)教授于2001年开发的一套支持向量机的库,这套库运算速度快,可以很方便地对数据做分类或回归。

SVM分类模式主要应用于对两类模式的识别。它的思路是:给定一组训练样本集 X 和 Y ,其中 X 为 $\{x_1, x_2, x_3, \dots, x_i, \dots, x_n\}$, Y 为 $\{y_1, y_2, y_3, \dots, y_i, \dots, y_n\}$ 。 x_i 表示输入向量,即待分类的样本。 $y_i \in \{-1, 1\}$ 表示输出样本,即所有训练样本的实际类别,指定两类的其中一类为正类,另一类为负类。

我们将400个样本点的标记和各个特征值输入到LSVM中,进行训练,可得到需要的分类视觉注意模型。采用线性核

的原因是它与非线性的多项式核和 RBF 核的性能一样好,并且它执行速度更快,效率更高。

4 实验结果

我们在分类数据集上进行测试来进一步验证 CMVA 模型,测试的系统平台和训练的是一样的,都是 64 位的 Ubuntu14.04,所用工具同样是 Linux 下的 MATLAB R2013a。

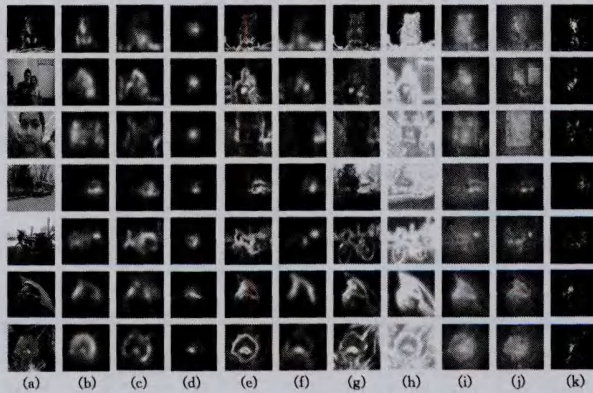
4.1 评价指标

训练出的模型 CMVA 的性能需要通过评价指标来与其他模型进行比较来评价,这里我们用 AUC(Area Under the ROC Curve)、NSS(Normalized Scanpath Saliency)和 SIM(Similarity)这 3 个评价指标,值越大,则该模型性能越好。

AUC 是指 ROC 曲线下的面积,NSS 称为归一化扫描路径分析法,用来衡量每一次扫视时视觉显著图与人实际看的视觉注意图之间的关联程度。SIM 用来计算归一化后的视觉显著图与人实际看的视觉注意图之间的相似度。

4.2 性能分析

对于 4 种分类注意模型,我们将其与现有的 8 种显著性算法进行了比较,分别是:GBVS^[3],Itti^[2],CovSal(with covariance and means)^[6],Saliency for Image Manipulation^[16],Image Signature (with LAB color space)^[17],SUN^[18],Torralba^[13],Judd^[6]。图 5 是 9 种注意模型的视觉显著图。



(a) 原图; (b) GBVS; (c) Itti; (d) CovSal (with covariance and means); (e) Saliency for Image Manipulation; (f) Image Signature (with LAB color space); (g) SUN; (h) Torralba; (i) Judd; (j) CMVA; (k) 实际的注视点

图 5 9 种显著性算法结果图

同时,我们使用 AUC、NSS、SIM 比较 9 种模型的性能。通过表 1,我们看到:

(1) 对于动物类和交通工具类图像,CMVA 模型的 AUC、NSS 和 SIM 值都是最高的;对于人物类图像,CMVA 模型的 AUC、NSS 值最高,SIM 值接近 Judd 模型;对于植物类图像,CMVA 模型的 AUC 和 SIM 值最高,NSS 值接近 Judd 模型。

(2) Judd 模型的性能与我们的模型接近,但是 Judd 模型在收集特征阶段,需要检测多种目标的特征,耗费的时间长。

(3) 大多数模型是自底向上的注意模型,没有考虑自顶向下的因素,因此虽然它们的执行时间相对较少,3 种评价指标中 SIM 值相差不大,但是 AUC 和 NSS 与我们的模型相差较大。

表 1 9 种显著性算法性能

	模型	AUC	NSS	SIM
动物组	Human	0.92	1.97	1
	CMVA(Ours)	0.89	1.56	0.49
	Judd	0.88	1.49	0.46
	CovSal	0.72	1.21	0.47
	GBVS	0.67	1.14	0.41
	Manipulation	0.65	1.10	0.39
	Signature	0.63	1.06	0.40
	Torralba	0.61	0.78	0.39
	SUN	0.61	0.74	0.34
人物组	Itti	0.60	0.68	0.20
	Human	0.91	2.04	1
	CMVA(Ours)	0.88	1.88	0.46
	Judd	0.87	1.63	0.47
	CovSal	0.88	1.43	0.44
	GBVS	0.87	1.23	0.47
	Manipulation	0.81	1.34	0.33
	Signature	0.80	0.86	0.44
	Torralba	0.80	0.89	0.47
交通工具组	SUN	0.77	0.93	0.35
	Itti	0.77	0.82	0.40
	Human	0.92	2.57	1
	CMVA(Ours)	0.89	1.30	0.50
	Judd	0.88	1.16	0.45
	CovSal	0.81	1.21	0.49
	GBVS	0.79	1.05	0.43
	Manipulation	0.80	1.02	0.42
	Signature	0.78	0.96	0.39
植物组	Torralba	0.72	0.73	0.40
	SUN	0.74	0.67	0.38
	Itti	0.73	0.58	0.33
	Human	0.91	2.45	1
	CMVA(Ours)	0.90	1.23	0.55
	Judd	0.90	1.36	0.46
	CovSal	0.73	1.04	0.49
	GBVS	0.67	0.99	0.45
	Manipulation	0.69	1.13	0.43
Signature	0.68	0.79	0.47	
Torralba	0.69	0.67	0.49	
SUN	0.65	0.62	0.39	
Itti	0.66	0.57	0.29	

因此,CMVA 模型视觉显著性预测性能更好。

结束语 本文建立了一个眼动数据集(数据集包含 14 个人对 4 类图像的眼动数据,每类图像包含 120 幅图),可以用它来进行眼动的分析与研究。同时,我们使用机器学习中的 SVM 算法训练了 4 类结合自底向上和自顶向下的特征的视觉显著模型并用 3 种衡量指标进行评价,通过比较,CMVA 模型优于现有的大多数模型,与人类视觉感知最匹配。

在未来的工作中,我们需要对分类模型进行更深一步的研究,比如图像自动匹配或者选择合适的分类模型,一类模型针对同一幅图的黑白图和彩色图的性能是否相同,眼动仪分析得到的“ground truth”和人工手动标记的“ground truth”训练得到的模型有何不同等。

参考文献

- [1] Treisman A M, Gelade G. A feature-integration theory of attention[J]. Cognitive psychology, 1980, 12(1): 97-136
- [2] Itti L, Koch C, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[J]. IEEE Transactions on PAMI, 1998, 20(11): 1254-1259

(下转第 115 页)

- [2] Signals and information Group [EB/OL]. <http://www.cspl.umd.edu/sig>
- [3] Wang B B, Wu Y L, Liu K J. Game theory for cognitive radio networks: An overview[J]. *Computer Networks*, 2010, 54(14): 2537-2561
- [4] Park J, van der Schaar M. The Theory of Intervention Games for Resource Sharing in Wireless Communications[J]. *IEEE Journal on Selected Areas in Communications*, 2012, 30(1): 165-175
- [5] Elias J, Martignon F, Capone A, et al. Non-cooperative spectrum access in cognitive radio networks: A game theoretical model [J]. *Computer Networks*, 2011, 55(17): 3832-3846
- [6] Akyildiz I F, Lo B F, Balakrishnan R. Cooperative spectrum sensing in cognitive radio networks: A survey [J]. *Physical Communication*, 2011, 4(1): 40-62
- [7] Wang Xing-wei, Wang Qing-jun, Lai Wu-wen, et al. Static Traffic Grooming Scheme Based on Predatory Search and Gaming [J]. *Computer Science*, 2012, 39(12): 30-32(in Chinese)
王兴伟, 王庆军, 赖武文, 等. 基于捕食搜索和博弈的静态通信量疏导模式[J]. *计算机科学*, 2012, 39(12): 30-32
- [8] Wang Xing-wei, Wang Qi, Huang Min, et al. A Fuzzy Integral and Game Theory Based QoS Multicast Routing Scheme [J]. *Journal of Software*, 2008, 19(7): 1743-1752(in Chinese)
王兴伟, 王琦, 黄敏, 等. 基于模糊积分和博弈论的 QoS 组播路由机制[J]. *软件学报*, 2008, 19(7): 1743-1752
- [9] Wang Xing-wei, Hou Mei-jia, Huang Min, et al. A Game Theory Based Fuzzy Unicast QoS Routing Scheme [J]. *Chinese Journal of Computers*, 2007, 30(1): 10-17(in Chinese)
王兴伟, 侯美佳, 黄敏, 等. 一种基于博弈论方法的模糊 QoS 单播路由机制[J]. *计算机学报*, 2007, 30(1): 10-17
- [10] Lippman S B, Lajoie J, Moo B E. C++ Primer Plus[M]. Li Shixian, Jiang Ai-jun, Mei Xiao-yong, et al., eds. Beijing: Posts & Telecom Press, 2012: 122(in Chinese)
Lippman S B, Lajoie J, Moo B E. C++ Primer Plus[M]. 李师贤, 蒋爱军, 梅晓勇, 等译. 北京: 人民邮电出版社, 2012: 122
- [11] Liu Wei. Proficient in MATLAB and C/C++ mixed programming design [M]. Beijing: Beihang University Press, 2008 (in Chinese)
刘维. 精通 MATLAB 与 C/C++ 混合程序设计[M]. 北京: 北京航空航天大学出版社, 2008
- [12] Koufos K, Ruttik K, Jantti R. Distributed Sensing in Multiband Cognitive Networks[J]. *IEEE Transactions on Wireless Communications*, 2011, 10(5): 1667-1677
-
- (上接第 88 页)
- [3] Harel J, Koch C, Perona P. Graph-based visual saliency[C]//NIPS. 2007: 545-552
- [4] Cheng M M, Zhang G X, Mitra N J, et al. Global contrast based salient region detection[C]//IEEE CVPR. 2011: 409-416
- [5] Zhang J M, Sclaroff S. Saliency Detection: A Boolean Map Approach[C]//ICCV. 2013: 153-160
- [6] Judd T, Ehinger K, Durand F, et al. Learning to predict where humans look[C]//ICCV. 2009: 2106-2113
- [7] Borji A. Boosting bottom-up and top-down visual features for saliency estimation[C]//CVPR. 2012: 438-445
- [8] Zhao Q, Koch C. Learning a saliency map using fixated locations in natural scenes[J]. *Journal of Vision*, 2011, 11(3): 1-15
- [9] NUSEF[OL]. <http://mmas.comp.nus.edu.sg/NUSEF.html>
- [10] MSRA[OL]. http://research.microsoft.com/en-us/um/people/jiansun/SalientObject/salient_object.htm
- [11] MIT[OL]. <http://people.csail.mit.edu/tjudd/WherePeopleLook/index.html>
- [12] PASCAL VOC 2007[OL]. <http://pascallin.ecs.soton.ac.uk/challenges/VOC/>
- [13] Torralba A. Modeling global scene factors in attention[J]. *Journal of the Optical Society of America*, 2003, 20(7): 1407-1418
- [14] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [15] Erdem E, Erdem A. Visual saliency estimation by nonlinearly integrating features using region covariances[J]. *Journal of Vision*, 2013, 13(4): 1-20
- [16] Margolin R, Zelnik-Manor L, Tal A. Saliency for image manipulation[J]. *The Visual Computer*, 2013, 29(5): 381-392
- [17] Hou X, Harel J, Koch C. Image Signature: Highlighting sparse salient regions[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(1): 194-201
- [18] Zhang L, Tong M, Marks T, et al. SUN: A bayesian framework for saliency using natural statistics[J]. *Journal of Vision*, 2008, 8(7): 32:1-20
- [19] Judd T, Durand F, Torralba A. A benchmark of computational models of saliency to predict human fixations[R]. MIT, 2012
- [20] Itti L, Koch C. Computational modeling of visual attention[J]. *Nat Rev Neurosci*, 2001, 2(3): 194-203
- [21] Hou X, Zhang L. Saliency detection: A spectral residual approach[C]//CVPR. 2007: 1-8
- [22] Judd T, Durand F, Torralba A. MIT saliency benchmark[OL]. <http://people.csail.mit.edu/tjudd/SaliencyBenchmark/>
- [23] Bruce N, Tsotsos J. Saliency based on information maximization [C]//NIPS 18. 2006: 155-162
- [24] Cheng M-M, Zhang Z, Lin W-Y, et al. BING: Binarized normed gradients for objectness estimation at 300fps[C]//IEEE CVPR. 2014: 2386-2393
- [25] Borji A, Itti L. State-of-the-art in visual attention modeling[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 185-207
- [26] Vig E, Dorr M, Cox D. Large-Scale Optimization of Hierarchical Features for Saliency Prediction in Natural Images[C]//IEEE CVPR. 2014: 2798-2805
- [27] Borji A, Sihite D N, Itti L. Salient object detection: A benchmark [C]//ECCV. 2012: 414-429
- [28] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection[J]. *PAMI*, 2012, 34(10): 1915-1926
- [29] Kienzle W, Wichmann F A, Scholkopf B, et al. A Nonparametric Approach to Bottom-Up Visual Saliency[C]//NIPS. 2007: 689-696