基于贝叶斯框架融合的 RGB-D 图像显著性检测

王松涛 [1, 2, 3] 周 真 [1] 靳 薇 [2, 3] 曲寒冰 [2, 3]

摘 要 为了有效融合 RGB 图像颜色信息和 Depth 图像深度信息, 提出一种基于贝叶斯框架融合的 RGB-D 图像显著性检测方法. 通过分析 3D 显著性在 RGB 图像和 Depth 图像分布的情况, 采用类条件互信息熵 (Class-conditional mutual information, CMI) 度量由深层卷积神经网络提取的颜色特征和深度特征的相关性, 依据贝叶斯定理得到 RGB-D 图像显著性后验概率. 假设颜色特征和深度特征符合高斯分布, 基于 DMNB (Discriminative mixed-membership naive Bayes) 生成模型进行显著性检测建模, 其模型参数由变分最大期望算法进行估计. 在 RGB-D 图像显著性检测公开数据集 NLPR 和 NJU-DS2000 上测试, 实验结果表明提出的方法具有更高的准确率和召回率.

关键词 贝叶斯融合, 深度学习, 生成模型, 显著性检测, RGB-D 图像

引用格式 王松涛, 周真, 靳薇, 曲寒冰. 基于贝叶斯框架融合的 RGB-D 图像显著性检测. 自动化学报, 2020, **46**(4): 695−720

DOI 10.16383/j.aas.2018.c170232

Saliency Detection for RGB-D Images Under Bayesian Framework

WANG Song-Tao[1, 2, 3] ZHOU Zhen[1] JIN Wei[2, 3] QU Han-Bing[2, 3]

Abstract In this paper, we propose a saliency detection model for RGB-D images based on the deep features of RGB images and depth images within a Bayesian framework. By analysis of 3D saliency in the case of RGB images and depth images, class-conditional mutual information (CMI) is computed for measuring the dependence of deep features extracted by CNN, then the posterior probability of the RGB-D saliency is formulated by applying the Bayes′ theorem. By assuming that color- and depth-based deep features are Gaussian distributions, a discriminative mixed-membership naive Bayes (DMNB) model is used to calculate the final saliency map. The Gaussian distribution parameter can be estimated in the DMNB model by using a variational inference-based expectation maximization algorithm. The experimental results on the RGB-D image NLPR and NJU-DS2000 datasets show that the proposed model performs better than other existing models.

Key words Bayesian fusion, deep learning, generative model, saliency detection, RGB-D images

Citation Wang Song-Tao, Zhou Zhen, Jin Wei, Qu Han-Bing. Saliency detection for RGB-D images under Bayesian framework. *Acta Automatica Sinica*, 2020, **46**(4): 695−720

显著性检测是计算机视觉中的一项重要研究内容, 是指模拟人类视觉注意机制实现准确、快速地检测图像中最感兴趣区域的过程. Borji 等给出了显著性 (Saliency) 定义, 直观地描述场景中相对其邻近区域突出的目标或区域, 是对于观察者而言的场景特性[1]. 人类视觉注意机制 (Visual attention mechanism) 通过优先处理少数几个显著区域或目标, 而忽略或舍弃其他的非显著区域或目标, 能够有选择地分配计算资源, 从而极大地提高视觉信息处理的工作效率. 因此, 基于视觉注意机制的显著性计算模型得到广泛的研究. 计算机在处理输入图像或者视频时, 通过检测显著性区域来实现判断其视觉信息的重要性, 应用到目标检测[2−3]、目标识别[4]、图像检索[5−6]、视频质量评估[7]、视频压缩[8]、图像自动裁剪[9−10] 和目标跟踪[11−12] 等领域. 随着 3D 显示技术的发展与设备的成熟, 显著性检测应用到 3D 图像和视频中[13], 例如 3D 视频目标重定位 (Retargeting)[14], 3D 视频质量评估[15−16] 和 3D 超声波图像处理[17−18] 等.

基于视觉注意机制的 RGB 图像显著性检测模

收稿日期 2017-05-02 录用日期 2018-04-16
Manuscript received May 2, 2017; accepted April 16, 2018
国家自然科学基金 (91746207), 北京市西城区优秀人才培养资助项目, 北京市科技计划 (Z161100001116086) 资助
Supported by National Natural Science Foundation of China (91746207), Personnel Training Program of Beijing Xicheng District, Beijing Science and Technology Program (Z161100001116086)
本文责任编委 刘跃虎
Recommended by Associate Editor LIU Yue-Hu
1. 哈尔滨理工大学测控技术与仪器省高校重点实验室 哈尔滨 150080
2. 北京市科学技术研究院人工智能与大数据研究中心 北京 100012
3. 北京市新技术应用研究所大数据研究中心 北京 100094
1. The Higher Educational Key Laboratory for Measuring and Control Technology and Instrumentations of Heilongjiang Province, Harbin University of Science and Technology, Harbin 150080 2. Research Center for Artificial Intelligence & Big Data Analysis, Beijing Academy of Science and Technology, Beijing 100012 3. Key Laboratory of Big Data Analysis, Beijing Institute of New Technology Application, Beijing 100094

型采用低层 (包括亮度、颜色、方向和纹理等[19−20]) 特征对比计算得到显著性, 其中包括全局特征对比计算模型[21]、局部特征对比计算模型[22] 以及结合全局特征和局部特征对比计算模型[23].

为了提高检测的准确率, 结合先验知识作为高层特征的显著性检测模型被提出来[24−25], 其中包括位置先验[26−27]、背景先验[28−29]、颜色先验[30]、形状先验[31] 和边界先验[32] 等.

针对 2D 图像背景复杂的显著性检测情况, 存在显著性区域相对背景区域没有明显的低层特征差异, 或者显著性区域分布情况与先验知识不符等难题. 随着深度学习在目标检测领域成功应用, 将目标类别特征引入显著性检测[33−34], 解决 2D 图像颜色等低层特征不显著等问题. Zhao 等[33] 和 Li 等[34] 引入深度学习方法, 提出基于类别对比差异来进行显著性检测方法, 即采用高层类别特征 (High-level category feature) 计算显著性. Zhao 等利用深层卷积神经网络 (Convolutional neural network, CNN) 提取全局和局部上下文语义特征建模显著性计算[33]. Li 等采用深层卷积神经网络提取多尺度分割得到图像超像素的深层对比特征 (Deep contrast feature), 同时融合低层特征得到 2D 显著性[34].

然而, 大多数基于人类视觉注意机制的 2D 图像显著性检测模型忽略了一个事实, 即人类视觉注意机制是作用于实际 3D 场景. 观察由 3D 图像建立的人眼注视集 (Eye fixation data) 中, 发现其深度信息提供 RGB 图像额外重要的显著性检测信息[35−36]. Desingh 等探讨了基于深度表征 (Appearance)、深度引起的模糊化 (Depth-induced blur) 及深度中心偏置 (Centre-bias) 的 3D 显著性检测方法[36]. Niu 等[37] 和 Ju 等[38] 提出利用深度信息进行 3D 显著性检测. Niu 等基于视差对比及立体摄影专业领域知识进行深度显著性检测[37], Ju 等提出基于 Depth 图像的各向异性中心 − 周围差分法的深度显著性检测模型[38]. 进一步, Lang 等[35] 和 Ren 等[39] 分别提出结合深度先验知识来进行复杂背景下 RGB-D 图像显著性检测, 均说明深度信息在 3D 显著性检测的有效性.

综上所述, RGB-D 图像相比 RGB 图像增加了 Depth 图像, 所以基于 RGB-D 图像的显著性计算模型中必须考虑 Depth 图像对显著性的影响. 因此, 在 RGB-D 图像显著性检测过程中存在两个重要的挑战, 即如何计算 Depth 图像的显著性以及如何与 RGB 图像的显著性融合得到最终的 3D 显著性.

本文提出一种新的 RGB-D 图像显著性检测方法: 1) 利用监督迁移方法将提取 RGB 图像的显著特征的 Clarifai 深层卷积神经网络应用到提取 Depth 图像的显著特征; 2) 假定 RGB 图像的显著特征和 Depth 图像的显著特征在给定显著性类别下条件独立分布, 在贝叶斯框架下进行显著特征融合计算最终的 3D 显著性. 在公开数据集 NLPR[40] 和 NJU-DS2000[38] 与现有基于边界先验的流行排序 2D 显著性检测方法 GMR[32], 基于深度学习的 2D 显著性检测方法[33−34], 基于 Depth 图像深度对比特征的显著性检测方法[38] 以及基于深度显著图融合的 3D 显著性检测方法[39−40] 比较的实验结果显示, 提出的方法具有更好的准确率和召回率.

本文的内容组织如下: 第 1 节主要介绍 RGB-D 图像显著性检测方法及存在的问题; 第 2 节引入了深度学习方法提取 RGB-D 图像显著特征, 主要描述基于 Clarifai 深层卷积神经网络提取 RGB 图像和 Depth 图像显著特征的过程; 第 3 节详细阐述基于贝叶斯框架下融合颜色显著特征和深度显著特征的 RGB-D 图像显著性检测方法; 第 4 节介绍实验设计, 重点分析提出的方法与现有方法的对比实验结果; 第 5 节对文章进行总结, 并提出未来研究方向.

1 问题描述

根据融合深度信息不同的策略, 将 RGB-D 图像显著性检测方法分为如下三类, 如图 1 所示. 这些模型的区别在于是否进行深度显著性的计算以及融合深度信息的方式. 深度特征融合模型不进行深度显著性计算, 直接将 Depth 图像特征与 RGB 图像显著特征进行加权融合; 深度显著图融合模型进行深度显著性计算得到深度显著图, 然后将深度显著图与 2D 显著图进行融合; 基于机器学习框架融合模型进行深度显著性计算, 将深度显著图 (特征) 与 2D 显著图 (特征) 进行训练得到模型.

1) 深度特征融合模型

这类模型将 Depth 图像深度特征和 RGB 图像显著特征进行融合, 用深度特征加权由 RGB 图像得到的 2D 显著特征来计算 RGB-D 图像显著性[41−44]. Fang 等提出一种新的 3D 显著性检测方法, 该框架采用自适应权重系数线性融合颜色特征、亮度特征、纹理特征和深度对比特征得到 3D 显著性[41]. Ciptadi 等提出利用深度测量得到的 3D 布局和形状特征融合的 3D 显著性检测方法[42]. Wu 等计算 RGB-D 图像的颜色对比特征和深度对比特征来构造多特征融合方法, 并通过多尺度增强来提高 3D 显著性检测的准确率[43]. Iatsun 等提出将 2D 显著特征与深度特征融合的 3D 显著性方法, 其中深度特征由二叉分割树 (Binary partition tree) 构造[44]. 深度特征融合模型通过深度特征与 RGB 图像显著特征加权计算 3D 显著性, 该模型把深度特征作为权重系数加入到原有 2D 显著特征. 因为该

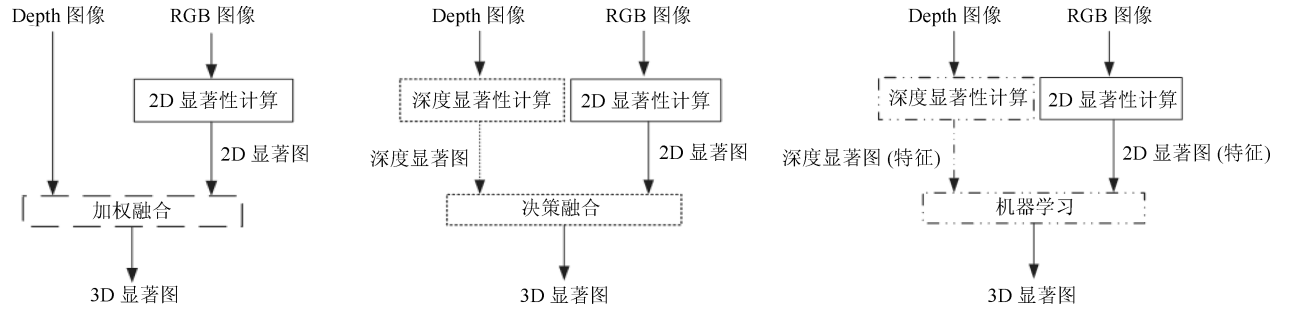


图 1　RGB-D 图像显著性检测方法分类

Fig. 1　Methodologies of the RGB-D saliency detection

模型没有计算深度显著性过程, 具有计算复杂度低、易于与 2D 显著性检测模型融合等优点.

2) 深度显著图 (Depth saliency maps) 融合模型

这类模型通过 Depth 图像显著性计算得到深度显著图, 然后与 RGB 图像显著性计算得到 2D 显著图进行融合得到 3D 显著性[40, 45−47]. Ouerhani 等将场景深度显著性融合到视觉注意模型中, 由深度特征得到显著图线性加权到 2D 显著图中[45]. Xue 等提出通过流行排序算法分别得到 RGB 图像的显著图和 Depth 图像的显著图, 然后进行显著图融合实现 3D 显著目标检测[46]. Peng 等提出一个简单的融合 RGB 图像显著图和 Depth 图像显著图的 RGB-D 图像显著性检测框架, 其中 RGB 图像显著图检测模型采用已有 2D 显著性检测模型, Depth 图像显著图基于多种深度上下文对比特征计算得到[40]. Wang 等提出直接线性加权融合 2D 显著图和深度显著图的 3D 视觉注意模型, 其中深度显著图基于贝叶斯方法由合成激励的跟踪实验人眼移动数据得到[47]. 这类融合模型首先根据 Depth 图像得到深度显著图, 然后与 2D 显著图采用不同的决策策略 (Pooling strategy) 得到 3D 显著图. 因为通过不同方法提取深度特征进行深度显著性计算, 有效利用 3D 场景中深度信息弥补基于 RGB 图像的 2D 显著性检测的局限.

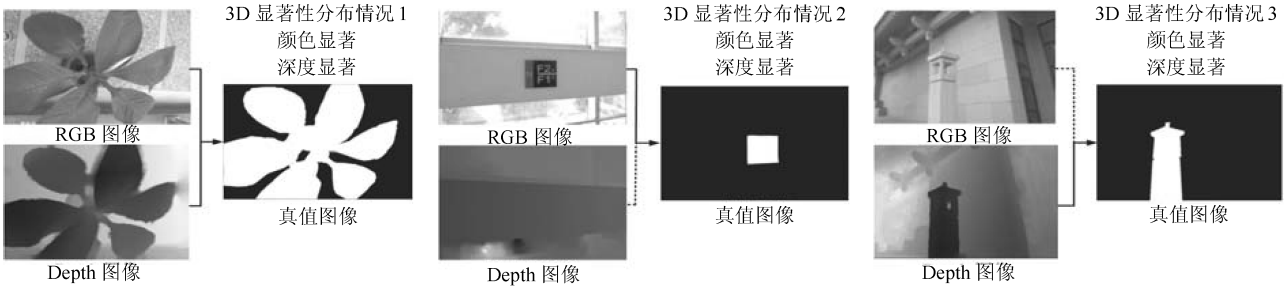
3) 基于机器学习 (Learning-based) 框架融合模型 – 这类模型采用机器学习技术融合 RGB 图像显著特征和 Depth 图像显著特征进行 RGB-D 图像显著性检测[48−52]. Iatsun 等提出基于人工神经网络来训练 RGB-D 图像显著性检测模型, 其中网络参数定义为融合 RGB 图像显著特征和 Depth 图像显著特征的自适应参数[48]. 受到机器学习方法在 2D 显著性检测成功应用的启发, Fang 等提出采用线性支持向量机 (Linear SVM) 融合 RGB 图像和 Depth 图像显著特征的 RGB-D 图像显著性检测方法[49]. Bertasius 等提出 EgoObject 表示法, 对 RGB-D 图像中提取的形状特征、位置特征、尺度特征和深度特征进行编码, 最后训练随机森林回归算子进行 3D 显著性目标检测[50]. 随着深度学习用于 2D 显著性检测, Qu 等提出采用深层卷积神经网络融合低层显著特征得到分层特征 (Hierarchical features) 进行 RGB-D 图像显著性目标检测[52]. 基于机器学习框架融合模型能够自动选取 RGB-D 图像显著特征进行显著性检测, 但是也存在模型训练过程中参数选取、模型收敛等问题.

综上所述, 上述三类 RGB-D 图像显著性检测模型在融合深度特征时, 没有考虑 3D 显著性在 RGB 图像和 Depth 图像分布情况, 而直接将 Depth 图像深度特征与 RGB 图像颜色特征进行融合. 3D 显著性检测模型不同于传统 2D 显著性检测模型的地方是增加评估深度特征对显著性检测的影响, 所以如何融合 Depth 图像的深度特征与 RGB 图像的颜色等特征得到 3D 显著性是研究的重点.

下面我们将分析 RGB-D 图像显著性区域在 RGB 图像和 Depth 图像分布的情况. RGB-D 图像显著性检测的结果由 RGB 图像和 Depth 图像共同决定, 选取 RGB-D 图像 NLPR[40] 数据集进行分析, 其中 3D 显著性在 RGB 图像和 Depth 图像分布关系如图 2 所示. 在 RGB-D 图像 NLPR 数据集中, Depth 图像由 Microsoft Kinect 设备采集得到, 并且选取 5 名实验人员对实际场景环境下采集的 RGB 图像和 Depth 图像手工标记出显著性区域得到 3D 显著性区域. 3D 显著性区域在 RGB 图像和 Depth 图像分布存在下面三种情况:

情况 1. 颜色 – 深度显著, 即 3D 显著性区域在 RGB 图像和 Depth 图像均是显著的, 定义该数据集为 $D^b = \{I_c^b, I_d^b\}$, 其中 I_c^b 和 I_d^b 分别表示该情况下的 RGB 图像和 Depth 图像.

情况 2. 颜色显著, 即 3D 显著性区域只在 RGB 图像是显著的, 而在 Depth 图像上是非显著的, 定



(a) 颜色−深度显著: RGB 图像和 Depth 图像均是显著的
(a) Color-depth saliency: both RGB images and Depth images are salient

(b) 颜色显著: 只有 RGB 图像是显著的而 Depth 图像是非显著的
(b) Color saliency: only RGB images are salient

(c) 深度显著: 只有 Depth 图像是显著的而 RGB 图像是非显著的
(c) Depth saliency: only Depth images are salient

图 2 3D 显著性在 RGB-D 图像分布情况

Fig. 2 3D saliency situation in RGB-D images

义该数据集为 $D^c = \{I_c^c, I_d^c\}$, 其中 I_c^c 和 I_d^c 分别表示该情况下的 RGB 图像和 Depth 图像.

情况 3. 深度显著, 即 3D 显著性区域只在 Depth 图像是显著的, 而在 RGB 图像上是非显著的, 定义该数据集为 $D^d = \{I_c^d, I_d^d\}$, 其中 I_c^d 和 I_d^d 分别表示该情况下的 RGB 图像和 Depth 图像.

如表 1 所示, 3D 显著性在 RGB 图像和 Depth 图像三种情况图像数量分布比例. 在 RGB-D 图像 NLPR 数据集中, 3D 显著性区域大部分在 RGB 图像和 Depth 图像均是显著的, 即颜色−深度显著情况的图像比例达到 76.7 %, 远远高于颜色显著情况和深度显著情况的图像比例; 而颜色显著情况的图像比例高于深度显著情况的图像比例. 对于 RGB-D 图像 NJU-DS2000 数据集具有相似的分布趋势.

表 1 RGB-D 图像数据集中 3D 显著性分布比例

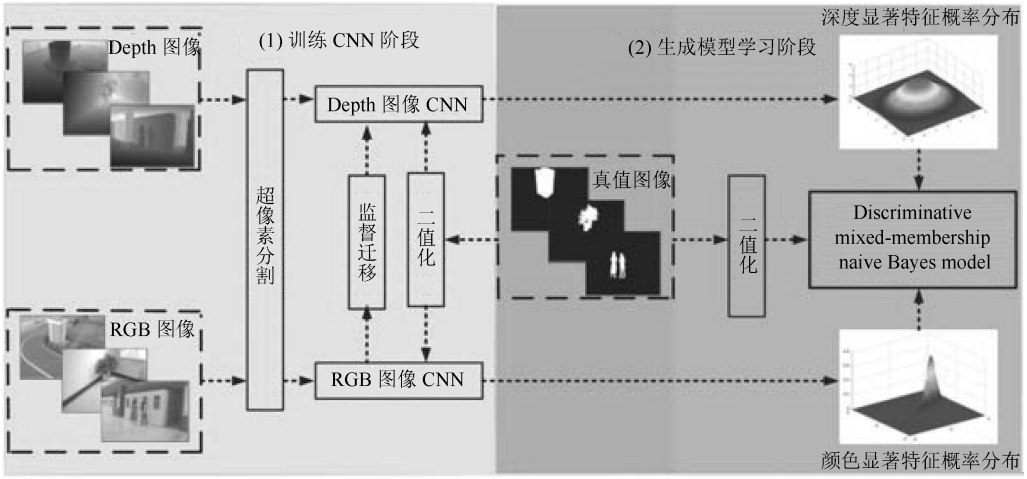
Table 1 3D saliency situation on RGB-D image dataset

数据集	颜色−深度显著	颜色显著	深度显著
NLPR[40]	76.7 %	20.8 %	2.5 %
NJU-DS2000[38]	69.4 %	16.6 %	14.0 %

基于以上分析, 针对 3D 显著性分布的三种情况, 为了有效融合 RGB 图像颜色信息和 Depth 图像深度信息, 保留对 3D 显著性检测有用的颜色信息和深度信息, 抑制无用的颜色信息和深度信息的干扰, 获得比仅基于 RGB 图像颜色特征或者 Depth 图像深度特征更好的显著性检测结果, 提出一种基于贝叶斯框架融合的 RGB-D 图像显著性检测方法, 如图 3 所示. 本文提出的方法属于基于机器学习框架融合模型, 即采用贝叶斯框架融合 RGB 图像的颜色特征和 Depth 图像的深度特征, 基于生成模型进行学习得到 RGB-D 图像显著性分布. 与深度显著图融合模型不同的是, 本文方法建立颜色与深度显著特征之间的分布, 利用了显著性分布关系; 而深度显著图融合模型在显著图进行融合, 只利用显著性值. 与传统基于机器学习框架融合模型不同的是, 本文方法首先通过观察和分析 3D 显著性在 RGB 图像和 Depth 图像分布情况, 建立颜色显著特征和深度显著特征概率分布关系, 利用类条件互信息熵度量其相关性, 利用贝叶斯定理得到显著性的后验概率; 而传统基于机器学习框架融合模型主要采用判别方法进行显著性预测, 没有考虑显著性分布信息. 本文提出的方法首先采用两个深层卷积神经网络分别提取 RGB 图像和 Depth 图像的显著特征, 其中提取 RGB 图像显著特征的深层卷积神经网络采用 Clarifai 网络[33]. 与文献 [33] 不同的是, 本文只提取 RGB 图像的全局上下文特征. 提取 Depth 图像的深层卷积神经网络采用监督迁移 (Supervision transfer) 方法[53] 得到, 通过 Clarifi 网络在 NLPR 数据集的 3D 显著性分布的颜色−深度显著情况的 Depth 图像 $\mathcal{I}_d^b$ 进行微调. 然后计算颜色显著特征和深度显著特征的类条件互信息熵 (Class-conditional mutual information, CMI), 假设当颜色显著特征和深度显著特征在小于给定 CMI 阈值时是条件分布独立的, 采用生成模型 DMNB (Discriminative mixed-membership naive Bayes)[54] 建模 RGB-D 图像显著性分布. 最后将 RGB-D 显著性检测问题视为二分类的贝叶斯推理问题, 采用变分推理[55] 计算后验概率进行预测.

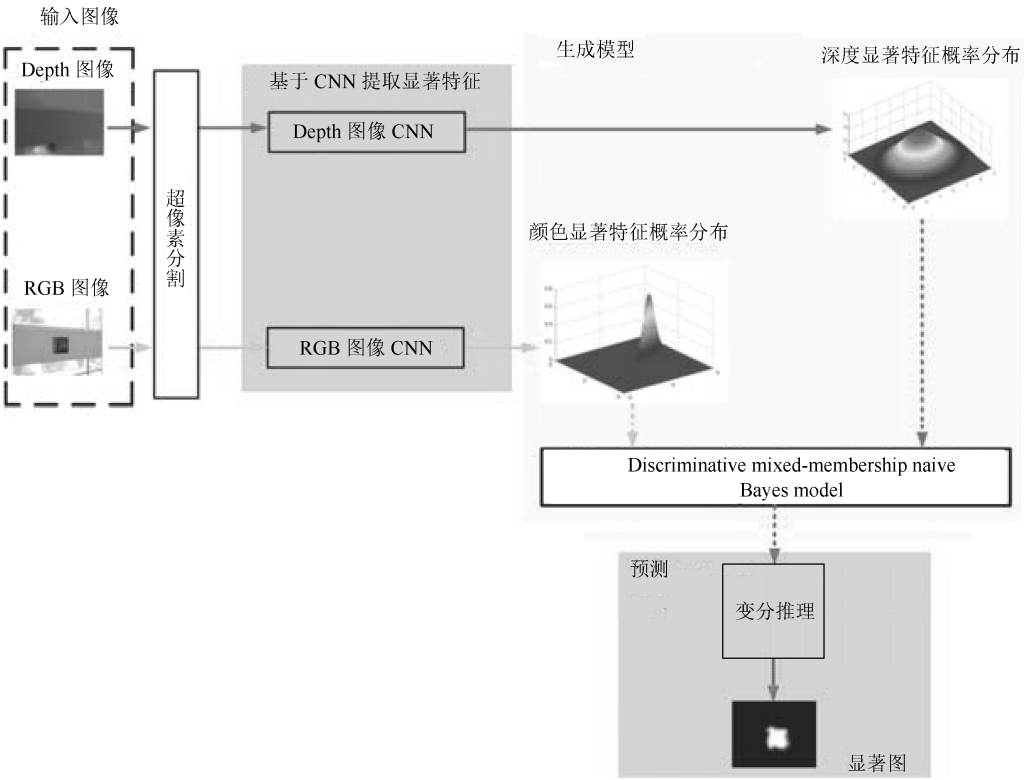
2 基于深层卷积神经网络提取 RGB 图像和 Depth 图像显著特征

近年来, 因为深层卷积神经网络提取图像的高层类别特征有效克服复杂背景中显著性区域低层特征对比度不高的情况, 深层卷积神经网络方法被用于显著性检测[33−34]. 针对 3D 显著性在 RGB 图像和 Depth 图像三种不同分布情况, 采用两个不



(a) 训练阶段, 包括基于监督迁移的 Depth 图像深层卷积神经网络的训练阶段和 DMNB 生成模型的学习阶段

(a) The training stage, including Depth CNN trained based on the supervision transfer and generative process for saliency detection following the DMNB model



(b) 预测阶段, 基于 DMNB 模型的变分推理进行 3D 显著性预测

(b) The testing stage, based on the variational inference to perform saliency prediction

图 3　本文方法框图

Fig. 3　Overview diagram of the proposed model

同的深层卷积神经网络分别单独提取 RGB 图像和 Depth 图像显著特征, 其中提取 RGB 图像的深层卷积神经网络采用 Clarifai 网络[33], 定义为 Υ; 提取 Depth 图像的深层卷积神经网络定义为 Ψ, 采用监督迁移学习方法[53] 得到.

2.1　RGB-D 图像超像素分割

为了提高图像处理效率, 基于全局特征的显著性检测方法从基于像素为处理单位向基于超像素为处理单位转变. 采用 SLIC 超像素分割算法[51] 分别对 RGB 图像和 Depth 图像进行超像素分割时, 由

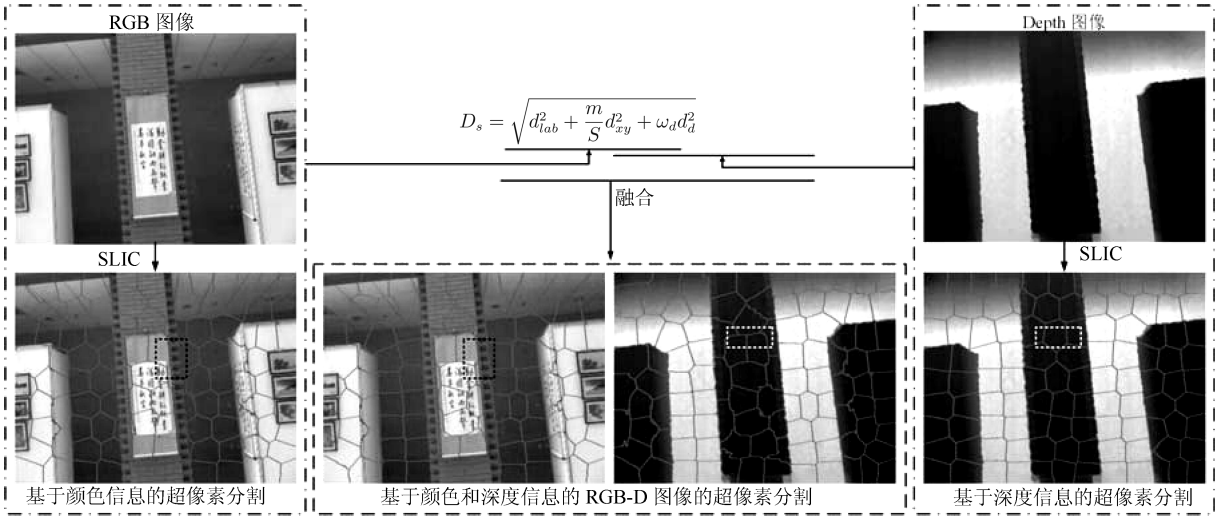


图 4 RGB-D 图像超像素分割 (如 RGB 图像矩形框区域所显示, 兼顾颜色和深度信息超像素分割得到边缘比只考虑颜色信息要准确. 同样情况, Depth 图像矩形框区域显示兼顾颜色和深度信息超像素分割得到边缘比只考虑深度信息要准确)

Fig. 4 Visual samples for superpixel segmentation of RGB-D images (Within the rectangle, the boundary between the foreground and the background segmented by the color and depth cues more accurate than color-based segmentation. Similarly, within the rectangle, the boundary between the foreground and the background segmented by the color and depth cues more accurate than depth-based segmentation)

于没有考虑颜色和深度的相互约束关系, 得到的超像素边界并不准确, 如图 5 所示. 所以, 对于 RGB-D 图像超像素分割需要同时考虑颜色信息和深度信息. 在 SLIC 算法的基础上提出融合颜色信息和深度信息的超像素分割距离度量, 如下式所示:

$$D_s = \sqrt{d_{lab}^2 + \frac{m}{S} d_{xy}^2 + \omega_d d_d^2} \tag{1}$$

其中, $d_d = \sqrt{(d_j - d_i)^2}$ 表示 Depth 图像像素 i 和 j 的距离度量, d_{lab} 和 d_{xy} 分别为原有 SLIC 算法颜色和位置的距离度量. ω_d 和 m/S 分别是深度信息和位置的权重, D_s 是 RGB-D 图像超像素分割的距离度量.

2.2 基于监督迁移学习提取 Depth 图像特征

在显著性检测数据集中, Depth 图像数量远没有 RGB 图像那么多, 所以直接采用深层卷积神经网络对 Depth 图像进行训练提取显著性特征需要克服过拟合问题. 在 RGB-D 图像 NLPR 数据集中, RGB 图像和 Depth 图像是成对的, 并且 RGB 图像和 Depth 图像表示同一场景下两种不同的数据模式, 采用基于监督迁移学习方法利用在 RGB 图像训练得到的 Clarifai 网络在有限的 Depth 图像进行训练, 克服深层卷积神经网络由训练图像数量过少而导致过拟合问题.

假设 RGB 图像显著性数据模式为 M_s, Depth 图像显著性数据模式为 M_d. 存在基于 RGB 图像训练得到的显著性检测 K 层深层卷积神经网络结构为 $\Upsilon = \{\varphi_{M_s}^i, \forall i \in [1, \cdots, K]\}$, 基于 Depth 图像训练得到的显著性检测 L 层深层卷积神经网络结构为 $\Psi = \{\psi_{M_d}^i, \forall i \in [1, \cdots, L]\}$. 在子数据集 D^b 中, 基于 RGB 图像训练得到的深层卷积神经网络监督迁移得到基于 Depth 图像训练得到的深层卷积神经网络过程如下式:

$$\min_{W_d^{[1\cdots L]}} \sum_{(I_d^b, I_c^b) \in D^b} f(\psi_{M_d}^L(I_d^b), \varphi_{M_s}^{i^*}(I_c^b)) \tag{2}$$

其中, $W_d^{[1\cdots L]} = \{w_d^i, \forall i \in [1, \cdots, L]\}$ 是深层卷积神经网络 Ψ 的参数; i^* 是深层卷积神经网络 Ψ 的层数并且满足 $i^* \in [1, \cdots, K]$. $f(\cdot)$ 表示损失函数, 本文采用欧式度量, 并且设定 K 和 L 相等, 如图 5 所示.

2.3 RGB-D 图像显著特征条件独立分析

采用类条件互信息熵来分析深层卷积神经网络提取 RGB 图像和 Depth 图像显著特征的相关性. 定义二值随机变量 z_s 表示显著性类别, RGB 图像显著特征 $\boldsymbol{x}_c$ 和 Depth 图像显著特征 $\boldsymbol{x}_d$ 的类条件互信息熵计算公式定义如下:

$$I(\boldsymbol{x}_c, \boldsymbol{x}_d | z_s) = H(\boldsymbol{x}_c | z_s) + H(\boldsymbol{x}_d | z_s) - H(\boldsymbol{x}_c, \boldsymbol{x}_d | z_s) \tag{3}$$

其中, $H(\boldsymbol{x}_c | z_s)$ 表示 RGB 图像显著特征 $\boldsymbol{x}_c$ 的类条件熵, 定义为

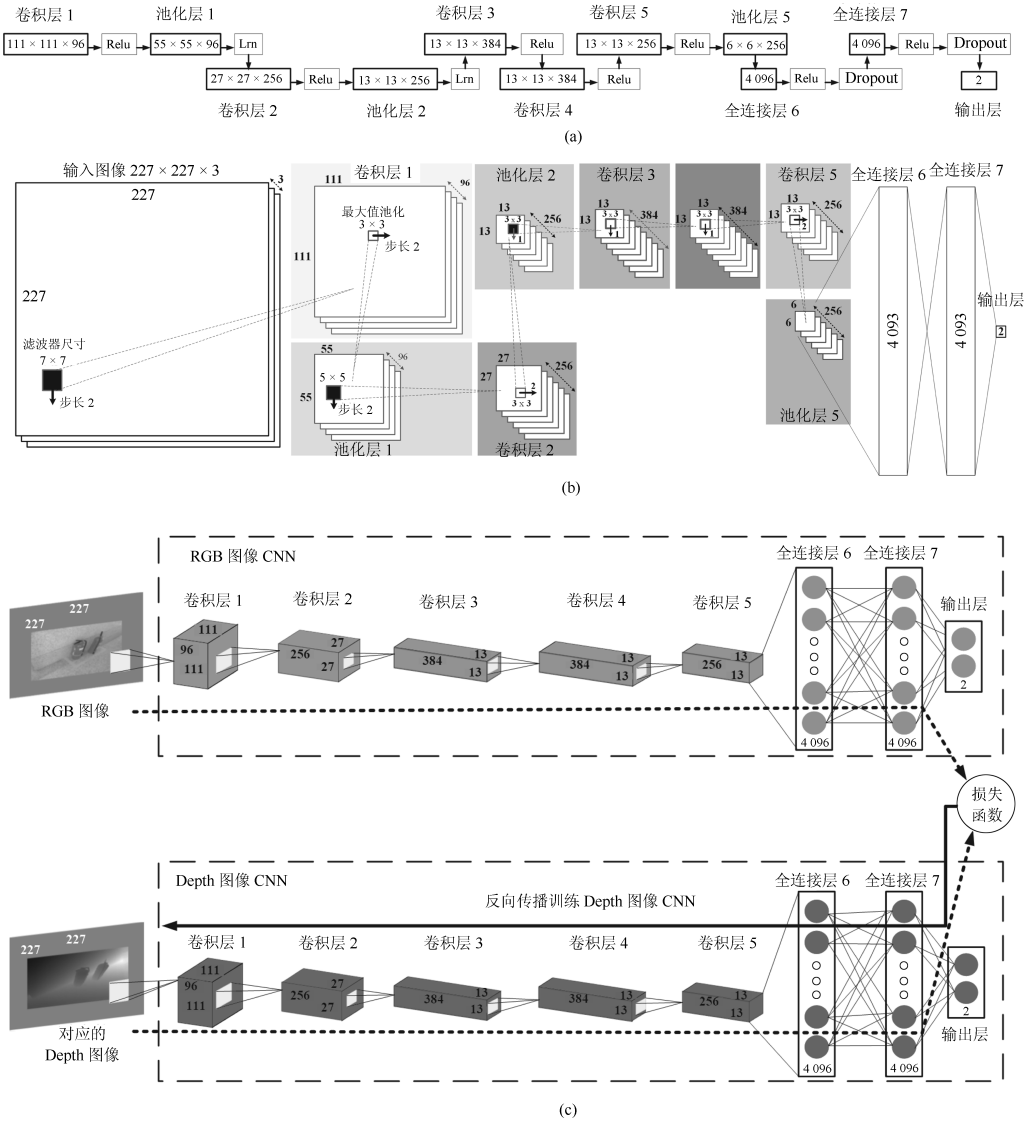


图 5　监督迁移学习过程示意图 ((a) 提取 Depth 图像显著特征的深层卷积神经网络结构图. 其中 Relu 层使用修正线性函数 $\mathrm{Relu}(x) = \max(x, 0)$ 保证输出不为负; Lrn 表示局部响应归一化层; Dropout 表示 Dropout 层, 在训练时以 0.5 比例忽略隐层节点防止过拟合. (b) 基于深层卷积神经网络提取 RGB 图像和 Depth 图像显著特征流程图. 首先图像被裁剪成尺寸为 $227 \times 227 \times 3$ 作为深层卷积神经网络的输入, 在卷积层 1 通过 96 核的尺寸为 7×7 步长为 2 滤波器卷积滤波, 得到卷积图像通过 Relu 函数, 再经过池化层 1 尺寸为 3×3 步长为 2 的最大值池化成 96 个尺寸为 55×55 的特征图, 最后对得到的特征图进行局部响应归一化. 在卷积层 2, 池化层 2, 卷积层 3, 卷积层 4, 卷积层 5 和池化层 5 执行相似的处理. 其池化层 5 输出作为全连接层 6 的输入, 经过全连接层 7 由输出层输出显著类别, 其中输出层采用 softmax 函数. (c) 本文基于监督迁移学习的方法, 在 RGB 图像训练完成的 Clarifai 网络的基础上, 利用与 RGB 图像配对的 Depth 图像重新训练提取 Depth 图像显著特征的深层卷积神经网络)

Fig. 5　Architecture for supervision transfer ((a) The Architecture of Depth CNN, where Relu denotes a rectified linear function $\mathrm{Relu}(x) = \max(x, 0)$, which rectify the feature maps thus ensuring the feature maps are always positive, lrn denotes a local response normalization layer, and Dropout is used in the fully connected layers with a rate of 0.5 to prevent CNN from overfitting. (b) The flowchart of image processed based on Depth CNN. A 227 by 227 crop of an image (with 3 planes) is presented as the input. This is convolved with 96 different 1st layer filters, each of size 7 by 7, using a stride of 2 in both x and y. The resulting feature maps are then: passed through a rectified linear fuction, pooled (max within 3 by 3 regions, using stride 2), and local response normalized across feature maps to give 96 different 55 by 55 element feature maps. Similar operations are repeated in layers 2, 3, 4, 5. The last two layers are fully connected, taking features from the pooling layer 5 as input in vector form. The final layer is a 2-way softmax function, which indicates the image is salient or not. (c) We train a CNN model for depth images by teaching the network to reproduce the mid-level semantic representation learned from RGB images for which there are paired images)

$$-\sum_{z_s} p(z_s) \sum_{\boldsymbol{x}_c} p(\boldsymbol{x}_c|z_s) \log p(\boldsymbol{x}_c|z_s) \tag{4}$$

$H(\boldsymbol{x}_d|z_s)$ 表示 Depth 图像显著特征 $\boldsymbol{x}_d$ 的类条件熵, 定义为

$$-\sum_{z_s} p(z_s) \sum_{\boldsymbol{x}_d} p(\boldsymbol{x}_d|z_s) \log p(\boldsymbol{x}_d|z_s) \tag{5}$$

$H(\boldsymbol{x}_c, \boldsymbol{x}_d|z_s)$ 表示 RGB 图像和 Depth 图像显著特征的联合类条件熵, 定义为

$$-\sum_{z_s} p(z_s) \sum_{\boldsymbol{x}_c, \boldsymbol{x}_d} p(\boldsymbol{x}_c, \boldsymbol{x}_d|z_s) \log p(\boldsymbol{x}_d, \boldsymbol{x}_c|z_s) \tag{6}$$

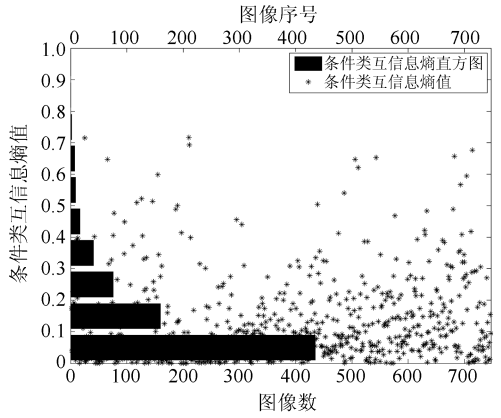
RGB 图像显著特征和 Depth 图像显著特征的类条件互信息熵值分布如图 6 所示, 对于 NLPR 数据集, 3D 显著性分布的颜色−深度显著情况、颜色显著情况和深度显著情况分别有 80.1 %、99.5 % 和 84.0 % 的类条件互信息熵值分布在 0.2 以下; 对于 NJU-DS2000 数据集, 3D 显著性分布的颜色−深度显著情况、颜色显著情况和深度显著情况分别有 84.8 %、96.5 % 和 88.8 % 的类条件互信息熵值分布在 0.2 以下. 对于 3D 显著性分布的颜色显著情况和深度显著情况, 因为 3D 显著性区域只在 RGB 图像或者 Depth 图像是显著的, 对应的类条件互信息熵值较小, 即 RGB 图像显著特征和 Depth 图像显著特征的相关性较低. 而对于 3D 显著性分布的颜色−深度显著情况, 虽然 3D 显著性区域同时存在 RGB 图像和 Depth 图像, 但是大部分情况是 RGB 图像显著区域与 Depth 图像显著区域是互补关系, 即 RGB 图像显著区域与 Depth 图像显著区域部分重叠, 对应的类条件互信息熵值较小; 只有少部分情况是 RGB 图像显著区域与 Depth 图像显著区域完全重叠, 对应的 CMI 值较大.

3 基于贝叶斯框架的 RGB-D 图像显著特征融合

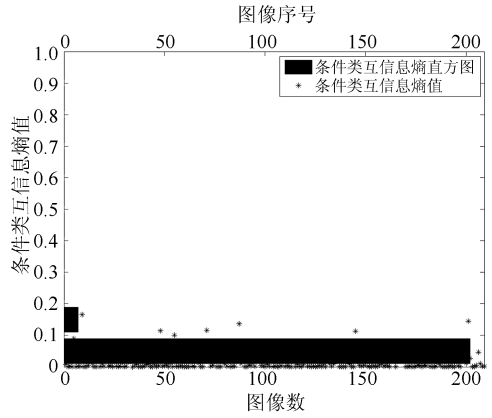
如第 2.2 节定义二值随机变量 z_s 表示 RGB-D 图像超像素是否显著, 给定对应的 RGB 图像的显著特征 $\boldsymbol{x}_c$ 和 Depth 图像的显著特征 $\boldsymbol{x}_d$, 显著性检测可认为估计像素显著性后验概率的贝叶斯推理问题, 定义如下:

$$\mathrm{P}(z_s|\boldsymbol{x}_c, \boldsymbol{x}_d) = \frac{\mathrm{P}(z_s, \boldsymbol{x}_c, \boldsymbol{x}_d)}{\mathrm{P}(\boldsymbol{x}_c, \boldsymbol{x}_d)} \tag{7}$$

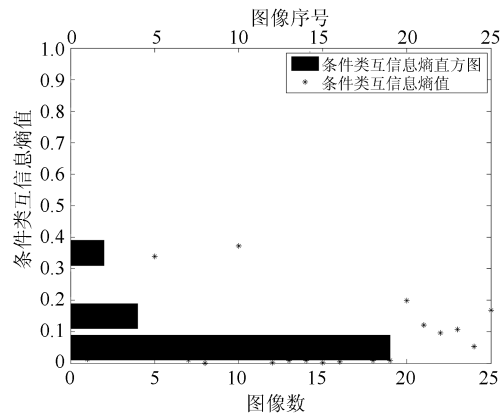
其中, $\mathrm{P}(z_s|\boldsymbol{x}_c, \boldsymbol{x}_d)$ 表示 RGB-D 图像超像素是否显著性的概率, $\mathrm{P}(\boldsymbol{x}_c, \boldsymbol{x}_d)$ 表示观察到的 RGB 图像和 Depth 图像显著特征概率分布, $\mathrm{P}(z_s, \boldsymbol{x}_c, \boldsymbol{x}_d)$ 为隐藏



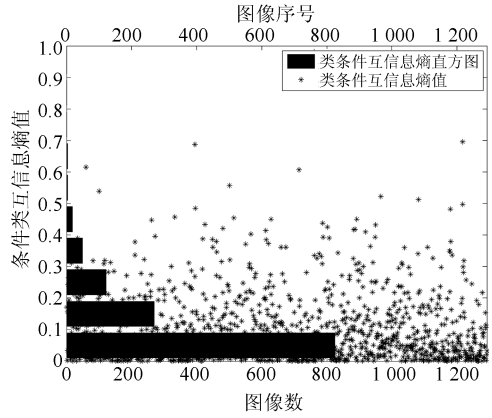
(a)



(b)



(c)



(d)

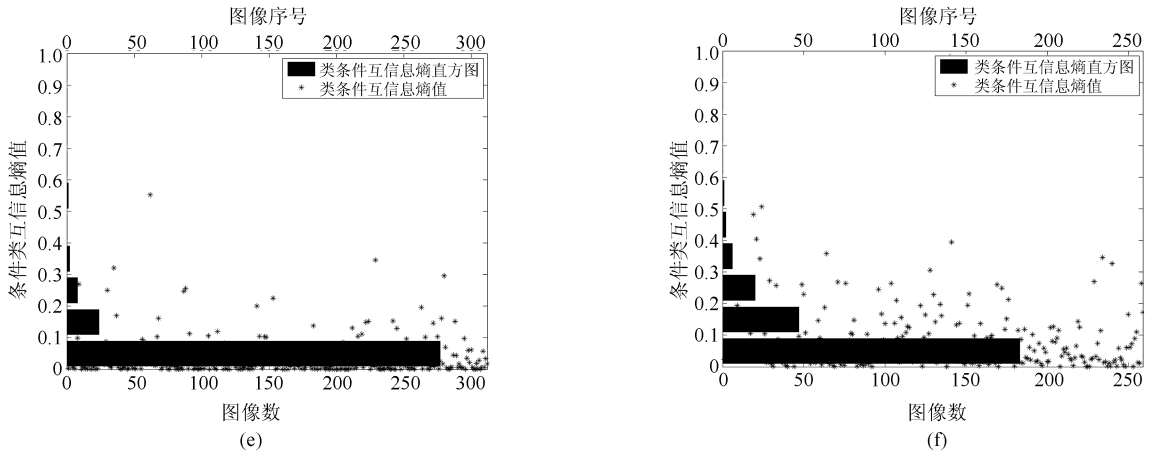


图 6　NLPR 数据集和 NJU-DS2000 数据集 RGB 图像和 Depth 图像显著特征的类条件互信息熵分布图 ((a) NLPR 数据集颜色–深度显著情况; (b) NLPR 数据集颜色显著情况; (c) NLPR 数据集深度显著情况; (d) NJU-DS2000 数据集颜色–深度显著情况; (e) NJU-DS2000 数据集颜色显著情况; (f) NJU-DS2000 数据集深度显著情况)

Fig. 6　Visual result for class-conditional mutual information between color and depth deep features on NLPR and NJU-DS2000 RGB-D image datasets ((a) Color-depth saliency situation in terms of the NLPR dataset, (b) Color saliency situation in terms of the NLPR dataset, (c) Depth saliency situation in terms of the NLPR dataset, (d) Color-depth saliency situation in terms of the NJU-DS2000 dataset, (e) Color saliency situation in terms of the NJU-DS2000 dataset, (f) Depth saliency situation in terms of the NJU-DS2000 dataset.)

类别变量 z_s 和显著特征 $\boldsymbol{x}_c, \boldsymbol{x}_d$ 的联合概率分布, 定义为 $\mathrm{P}(z_s, \boldsymbol{x}_c, \boldsymbol{x}_d) = \mathrm{P}(z_s)\mathrm{P}(\boldsymbol{x}_c, \boldsymbol{x}_d|z_s)$.

计算 RGB 图像显著特征 $\boldsymbol{x}_c$ 和 Depth 图像显著特征 $\boldsymbol{x}_d$ 的类条件互信息熵值, 当给定显著类别 z_s 下 $\boldsymbol{x}_c$ 和 $\boldsymbol{x}_d$ 相互独立分布时类条件互信息熵等于 0; 而当 $\boldsymbol{x}_c$ 和 $\boldsymbol{x}_d$ 是确定函数关系时类条件互信息熵值达到最大值 1[56]. 采用类条件互信息熵阈值 τ 来量化显著性特征的相关性, 即当 $\boldsymbol{x}_c$ 和 $\boldsymbol{x}_d$ 的类条件互信息熵值小于阈值 τ 时, 认为给定显著类别变量 z_s 下的 $\boldsymbol{x}_c$ 和 $\boldsymbol{x}_d$ 是条件独立分布的. 在显著特征条件独立分布的情况下, $\mathrm{P}(\boldsymbol{x}_c, \boldsymbol{x}_d|z_s) = \mathrm{P}(\boldsymbol{x}_c|z_s)\mathrm{P}(\boldsymbol{x}_d|z_s)$, RGB-D 图像显著性计算如下式:

$$\mathrm{P}(z_s|\boldsymbol{x}_c, \boldsymbol{x}_d) \propto \mathrm{P}(z_s)\mathrm{P}(\boldsymbol{x}_c|z_s)\mathrm{P}(\boldsymbol{x}_d|z_s) \qquad (8)$$

3.1　基于 DMNB 模型建模显著性检测

假设 RGB 图像显著特征和 Depth 图像显著特征在给定显著类别下是条件独立分布的, 采用贝叶斯框架融合颜色显著特征和深度显著特征, 基于 DMNB 模型计算 RGB-D 图像显著性. DMNB 模型是朴素贝叶斯 (Navie Bayes) 模型在以下两方面的扩展:

1) 在朴素贝叶斯模型中, 所有特征共享一个分量 (Component); 而在 DMNB 模型中, 每个特征有独立的分量, 并且这些分量分布服从狄利克雷 (Dirichlet)–多项式先验分布.

2) 在朴素贝叶斯模型中, 共享分量作为类别标签 (Indicator); 而在 DMNB 模型中, 类别标签由混合隶属度 (Mixed membership) 逻辑回归产生.

基于 DMNB 模型的显著性检测图模型如图 7 所示, 其中假设 RGB 图像显著特征和 Depth 图像显著特征 $\boldsymbol{X} = (\boldsymbol{x}_c, \boldsymbol{x}_d)$ 服从高斯分布以及标签 $\boldsymbol{Y}$ 符合伯努利 (Bernoulli) 分布. 在 NLPR 数据集中选取 M 个 RGB-D 图像超像素特征 $\{(x_{ij})|i = 1, \cdots, M, j = 1, \cdots, N\}$ 和标签 $\{y_i|i = 1, \cdots, M\}$ 进行生成模型训练, 通过生成 $\{\boldsymbol{X}, \boldsymbol{Y}\}$ 的概率最大来估计 DMNB 模型参数, 其中 N 为显著特征维数.

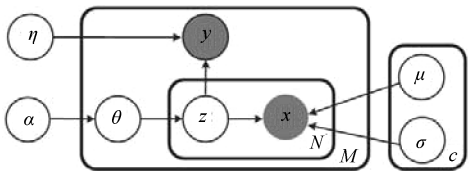


图 7　基于 DMNB 模型显著性检测的图模型 (y 和 $\boldsymbol{x}$ 为可观测变量, $\boldsymbol{z}$ 为隐藏变量. 其中 $\boldsymbol{x}_{1:N} = (\boldsymbol{x}_c, \boldsymbol{x}_d)$ 表示 RGB-D 图像显著特征, 特征 $\boldsymbol{x}_j$ 服从 C 个均值为 $\{\mu_{jk}|j = 1, \cdots, N\}$ 和方差为 $\{\sigma^2_{jk}|j = 1, \cdots, N\}$ 高斯分布. y 是标识超像素是否显著的标签, 取值 1 或者 0, 其中 1 表示显著, 0 表示非显著)

Fig. 7　Graphical models of DMNB for saliency estimation (y and $\boldsymbol{x}$ are the corresponding observed states, and $\boldsymbol{z}$ is the hidden variable, where each feature $\boldsymbol{x}_j$ is assumed to have been generated from one of C Gaussian distribution with a mean of $\{\mu_{jk}|j = 1, \cdots, N\}$ and a variance of $\{\sigma^2_{jk}|j = 1, \cdots, N\}$, y is either 1 or 0 that indicates whether the pixel is salient or not.)

基于 DMNB 模型显著性检测的生成过程如算法 1 所示, 其中 $p(\cdot|\alpha)$ 表示参数为 α 的狄利克雷分布, $p(\cdot|\theta)$ 表示参数为 θ 的多项式分布, $p(\boldsymbol{x}_j|\boldsymbol{z}_j,\Omega_j)$ 表示给定隐藏类别 $\boldsymbol{z}_j$ 参数为 Ω_j 的高斯分布, $p(y|\boldsymbol{z}_j,\eta)$ 为给定隐藏类别 $\boldsymbol{z}_j$ 参数为 η 的伯努利分布. $(\boldsymbol{x}_{1:N},y)$ 的边缘分布如下:

$$p(\boldsymbol{x}_{1:N},y|\alpha,\Omega,\eta)= \int p(\theta|\alpha)\prod_{j=1}^{N}\sum_{\boldsymbol{z}_j}P(\boldsymbol{z}_j|\theta)p(\boldsymbol{x}_j|\boldsymbol{z}_j,\Omega_j)p(y|\boldsymbol{z}_j,\eta)\mathrm{d}\theta \tag{9}$$

其中, θ 是 C 分量的先验分布, $\Omega=\{(\mu_{jk},\sigma_{jk}^2)|j=1,\cdots,N,k=1,\cdots,C\}$, $p(\boldsymbol{x}_j|\boldsymbol{z}_j,\Omega_j):=\mathcal{N}(\boldsymbol{x}_j|\mu_{jk},\sigma_{jk}^2)$, $p(y|\boldsymbol{z}_j,\eta):=Bern(y|\eta)$. 由于式 (9) 中存在隐藏变量 $\boldsymbol{z}$ 而无法直接计算其概率, 采用变分推理的方法进行求解.

算法 1. 基于 DMNB 模型的显著性检测生成过程

输入: C.
初始化: α,η.
抽取分量比例: $\theta\sim p(\theta|\alpha)$
对于特征:
选择一个分量 $\boldsymbol{z}_j\sim p(\boldsymbol{z}_j|\theta)$;
选择一个特征值 $\boldsymbol{x}_j\sim p(\boldsymbol{x}_j|\boldsymbol{z}_j,\Omega_j)$.
抽取标签: $y\sim p(y|\boldsymbol{z}_j,\eta)$;
输出: 联合概率分布 $p(\boldsymbol{x}_{1:N},y|\alpha,\Omega,\eta)$.

3.2 变分求解

为了求解 $\log p(y,\boldsymbol{x}_{1:N}|\alpha,\Omega,\eta)$ 的下界, 引进变分分布 $q(\boldsymbol{z}_{1:N},\theta|\gamma,\phi)$ 来近似含有隐藏变量的真值后验分布 $p(\boldsymbol{z}_{1:N},\theta|\alpha,\Omega,\eta)$. 直接应用 Jensen's 不等式[57], $\log p(y,\boldsymbol{x}_{1:N}|\alpha,\Omega,\eta)$ 的下界:

$$\log p(y,\boldsymbol{x}_{1:N}|\alpha,\Omega,\eta)\geq \boldsymbol{H}(q(\boldsymbol{z}_{1:N},\theta|\gamma,\phi))+ \mathrm{E}_q(\log p(y,\boldsymbol{x}_{1:N},\boldsymbol{z}_{1:N}|\alpha,\Omega,\eta)) \tag{10}$$

注意到 $\boldsymbol{x}_{1:N}$ 和 y 在给定 $\boldsymbol{z}_{1:N}$ 是条件独立的, 得到变分分布如下:

$$q(\boldsymbol{z}_{1:N},\theta|\gamma,\phi)=q(\theta|\gamma)\prod_{j=1}^{N}q(\boldsymbol{z}_j|\phi) \tag{11}$$

其中, $q(\theta|\gamma)$ 是 C 维 θ 的狄利克雷分布, $q(\boldsymbol{z}_j|\phi)$ 是 $\boldsymbol{z}_j$ 的多项式分布. 采用 $\mathcal{L}$ 表示下界函数, 如式 (12) 所示:

$$\begin{aligned}\mathcal{L}=&\mathrm{E}_q[\log p(\theta|\alpha)]+\mathrm{E}_q[\log p(\boldsymbol{z}_{1:N}|\theta)]+\\ &\mathrm{E}_q[\log p(\boldsymbol{x}_{1:N}|\boldsymbol{z}_{1:N},\gamma)]-\mathrm{E}_q[\log q(\theta)]-\\ &\mathrm{E}_q[\log q(\boldsymbol{z}_{1:N})]+\mathrm{E}_q[\log p(y|\boldsymbol{z}_{1:N},\eta)]\end{aligned} \tag{12}$$

其中, $\xi>0$ 是新引入的变分参数, 如式 (13) 所示:

$$\mathrm{E}_q[\log p(y|\boldsymbol{z}_{1:N},\eta)]\geq\sum_{k=1}^{C}\phi_k(\eta_k y-\frac{e^{\eta_k}}{\xi})-\frac{1}{\xi}-\log\xi \tag{13}$$

根据变分参数 γ_k, ϕ_k 和 ξ 最大化下界函数 $\mathcal{L}(\gamma_k,\phi_k,\xi;\alpha,\Omega,\eta)$ 得到参数更新如式 (14), 式 (15) 和式 (16) 所示:

$$\phi_k\propto \mathrm{e}^{\Psi(\gamma_k)-\Psi\left(\sum\limits_{l=1}^{C}\gamma_l\right)+\frac{\eta_k y_i-\frac{e^{\eta_k}}{\xi}-\sum\limits_{j=1}^{N}\frac{(\boldsymbol{x}_{ij}-\mu_{jk})^2}{2\sigma_{jk}^2}}{N}} \tag{14}$$

$$\gamma_k=\alpha+N\phi_k \tag{15}$$

$$\xi=1+\sum_{k=1}^{C}\phi_k\mathrm{e}^{\eta_k} \tag{16}$$

在变分推理优化对数似然函数的下界过程中, 使累积下界 $\sum_{i=1}^{M}\mathcal{L}(\gamma^*,\phi^*,\xi^*,\alpha,\Omega,\eta)$ 在所有训练数据 $(\boldsymbol{x}_i,y_i)$ 最大, 由变分参数 (γ^*,ϕ^*,ξ^*) 来产生估计参数 μ, σ 和 η, 如式 (17), 式 (18) 和式 (19) 所示:

$$\mu_{jk}=\frac{\sum\limits_{i=1}^{M}\phi_{ik}x_{ij}}{\sum\limits_{i=1}^{M}\phi_{ik}} \tag{17}$$

$$\sigma_{jk}=\frac{\sum\limits_{i=1}^{M}\phi_{ik}(x_{ij}-\mu_{jk})^2}{\sum\limits_{i=1}^{M}\phi_{ik}} \tag{18}$$

$$\eta_k=\log\left(\frac{\sum\limits_{i=1}^{M}\phi_{ik}y_i}{\sum\limits_{i=1}^{M}\frac{\phi_{ik}}{\xi_i}}\right) \tag{19}$$

$$(\gamma_i^m,\phi_i^m,\xi_i^m)=\arg\max\mathcal{L}(\gamma_i,\phi_i,\xi_i;\alpha^{m-1},\Omega^{m-1},\eta^{m-1}) \tag{20}$$

$$(\alpha^m,\Omega^m,\eta^m)=\arg\max_{(\alpha,\Omega,\eta)}\sum_{i=1}^{N}\mathcal{L}(\gamma_i^m,\phi_i^m,\xi_i^m;\alpha,\Omega,\eta) \tag{21}$$

基于变分推理和参数估计更新公式, 直接构造变分期望最大算法估计参数 α, Ω 和 η, 如算法 2 所示.

算法 2. 基于 DMNB 模型显著性建模的变分期望最大求解算法

输入: 收敛阈值 $\varepsilon_{\mathcal{L}}$.
初始化: α^0, Ω^0, η^0.

计算期望: 在训练数据集基于参数 $(\alpha^{m-1},\Omega^{m-1},\eta^{m-1})$ 下根据式 (20) 计算最优变分参数, 并计算 $\log p(y_i,\boldsymbol{x}_{1:N}|\alpha,\Omega,\eta)$ 的下限 $\mathcal{L}(\gamma_i^m,\phi_i^m,\xi_i^m;\alpha,\Omega,\eta)$.

最大化过程: 根据式 (21) 更新 DMNB 模型参数 (α,Ω,η), 计算逼近下限

$$\mathcal{L}^m=\sum_{i=1}^{N}\mathcal{L}(\gamma_i^m,\phi_i^m,\xi_i^m;\alpha^m,\Omega^m,\eta^m),$$

$$\mathcal{L}^{m+1}=\sum_{i=1}^{N}\mathcal{L}(\gamma_i^{m+1},\phi_i^{m+1},\xi_i^{m+1};\alpha^{m+1},\Omega^{m+1},\eta^{m+1}),$$

UNTIL $\mathcal{L}^m-\mathcal{L}^{m+1}\le\varepsilon_{\mathcal{L}}$.

输出: α, Ω, η.

通过变分期望最大算法得到 DMNB 模型参数后, 基于参数 η 给定特征 $\boldsymbol{x}_{1:N}$ 进行显著性预测, 如式 (22) 所示:

$$\mathrm{E}[\log p(y|\boldsymbol{x}_{1:N},\alpha,\Omega,\eta)]=\begin{cases}\eta^{\mathrm{T}}\mathrm{E}[\bar{\boldsymbol{z}}]-\mathrm{E}[\log(1+\mathrm{e}^{\eta^{\mathrm{T}}\bar{\boldsymbol{z}}})], & y=1\\ 0-\mathrm{E}[\log(1+\mathrm{e}^{\eta^{\mathrm{T}}\bar{\boldsymbol{z}}})], & y=0\end{cases}\tag{22}$$

其中, $\bar{\boldsymbol{z}}$ 为所有观察到特征值的 $\boldsymbol{z}_{1:N}$ 的均值. 由于无法直接计算 $\mathrm{E}[\bar{\boldsymbol{z}}]$ 值, 引入分布 $q(\boldsymbol{z}_{1:N},\theta)$ 并且计算 $\mathrm{E}_q[\bar{\boldsymbol{z}}]$ 作为 $\mathrm{E}[\bar{\boldsymbol{z}}]$ 的近似. 在实际求解中, $\mathrm{E}_q[\bar{\boldsymbol{z}}]=\phi$, 因此式 (22) 求解只需比较 $\eta^{\mathrm{T}}\phi$ 和 0.

3.3 参数分析和设置

本文涉及的参数如表 2 所示, 本节主要讨论如何确定类条件互信息熵阈值参数 τ 和 DMNB 模型混合分量参数 C.

表 2　参数表

Table 2　Summary of parameters

变量名	取值范围	参数描述
τ	(0, 1)	类条件互信息熵阈值
α	(0, 20)	狄利克雷分布参数
θ	(0, 1)	多项式分布参数
η	(−10.0, 3.0)	伯努利分布参数
Ω	((0, 1), (0, 0.2))	高斯分布参数
N	>2	特征维度
C	>2	DMNB 模型分量参数
$\varepsilon_{\mathcal{L}}$	(0, 1)	EM 收敛阈值

在训练 DMNB 模型时, 为了满足条件独立的假设, 我们选取数据集中 RGB 图像显著特征和 Depth 图像显著特征的类条件互信息熵值小于 τ 作为训练样本, 而对于类条件互信息熵阈值参数选取决定训练样本的数量. 通过第 2.3 节分析, 对于 NLPR 数据集有 84.3 % 的 RGB 图像和 Depth 图像的类条件互信息熵值小于 0.2, 对于 NJU-DS2000 数据集有 87.3 % 的 RGB 图像和 Depth 图像的类条件互信息熵值小于 0.2. 选择 $\tau=0.2$, 选取数据集中样本的 80 % 且类条件互信息熵值小于 τ 用于训练, 剩余 20 % 的样本用于测试.

对于算法 1 中 DMNB 模型分量参数 C 的选择, 可通过狄利克雷过程混合模型 (Dirichlet process mixture model, DPMM)[55] 基于训练样本找到合适的数值. DPMM 通过混合分量数值随着训练集增加而增长, 提供一个非参数的混合模型参数的先验值, 如图 8 所示. 通过观察图 8 (b) 和图 8 (d), 通过狄利克雷过程混合模型算法得到参数 C 分别为 24 和 28, 表示 NJU-DS2000 数据集比 NLPR 数据集更复杂.

使用交叉认证方法来验证参数 C, 给定一个由狄利克雷过程混合模型得到的参数 C 的取值范围, 选取 NLPR 数据集训练样本的 90 % 作为训练集, 选取训练样本的 10 % 作为验证集, 结果如图 9 所示. 复杂度 (Perplexity) 定义为 Perplexity $=\exp(-\sum_{i=1}^{n}\frac{\log p(\boldsymbol{x}_i)}{n})$, 其值越小表示 DMNB 生成模型描述数据能力越好, 其中 n 为训练集和测试集中选取特征 $\boldsymbol{x}_i$ 的数量. 对于生成模型 DMNB, 较大的参数 C 更易在训练集得到较低的 Perplexity 值, 因为较大的参数 C 增加了模型的复杂度来拟合训练集. 然而, 模型的复杂度会降低泛化的能力而在测试集得到较高的 Perplexity 值. 在下面对比实验中, 选取 $C=24$, 对于 NLPR 数据集和 NJU-DS2000 数据集 DMNB 模型分别进行学习和测试.

DMNB 模型由 M 个 RGB-D 图像超像素进行训练[58], 其超像素分割参数为 $S=40$, $m=20$ 和 $\omega_d=1.0$. 其超像素显著特征由 RGB 图像深层卷积神经网络和 Depth 图像深层卷积神经网络的倒数第二层输出, 超像素显著特征维数 $N=8\,192$. 对于算法 1 中参数 Ω 由训练数据的均值和方差初始化, 并且初始化参数 $\alpha=M_c/M$, 其中 M_c 为训练集中分量 c 的特征数量. 最后, 设置算法 2 中参数 $\varepsilon_{\mathcal{L}}=0.001$ 进行 DMNB 模型参数求解.

4 实验结果和分析

本文提出的方法均以 BFSD (Saliency detection based on Bayesian fusion) 简称, 采用 Matlab 7.12 实现算法, 并在 Intel Core (TM) i5-6400 CPU, 8 GB 内存 PC 机上完成所有实验. 将所提出的方法与 6 种 sate-of-the-art 显著性检测方法进行比较, 其中包括基于图像边界先验知识的流形排序 2D 显著性检测方法 GMR[32], 基于深层卷积神经网络提取全局和局部特征的 2D 显著性检测方法 MC[33], 基于深层卷积神经网络提取多尺度局部特征的 2D 显著性检测方法 MDF[34], 基于 Depth 图像深度对比特征的显著性检测方法 ACSD[38], 基于 2D

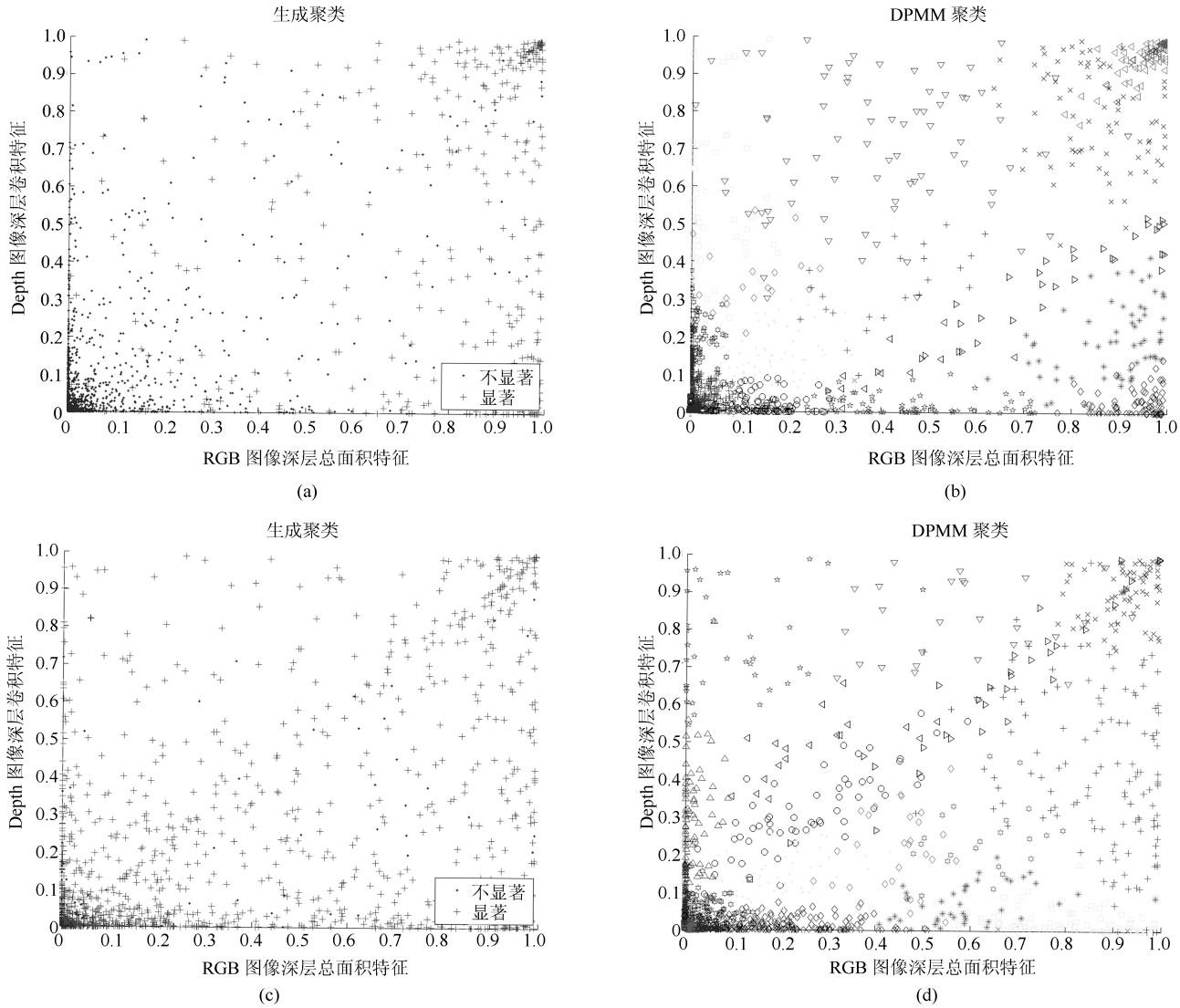


图 8 对比基于生成聚类和狄利克雷过程聚类方法确定 DMNB 模型混合分量参数 C ((a) 针对 NLPR 数据集显著特征生成聚类图. (b) 针对 NLPR 数据集基于狄利克雷过程的显著特征聚类图, 其中不同图例的数目代表 DMNB 模型混合分量参数 C. 对于 NLPR 数据集, 得到 $C = 24$. (c) 针对 NJU-DS2000 数据集显著性特征生成聚类图. (d) 针对 NJU-DS2000 数据集基于狄利克雷过程的显著特征聚类图, 其中不同图例的数目代表 DMNB 模型混合分量参数 C. 对于 NJU-DS2000 数据集, 得到 $C = 28$)

Fig. 8 Visual result for the number of components C in the DMNB model: generative clusters vs DPMM clustering ((a) Generative clusters for NLPR RGB-D image dataset. (b) DPMM clustering for NLPR RGB-D image dataset, where the number of colors and shapes of the points denote the number of components C. We find $C = 24$ using DPMM on the NLPR dataset. (c) Generative clusters for NJU-DS2000 RGB-D image dataset. (d) DPMM clustering for NJU-DS2000 RGB-D image dataset, where the number of colors and shapes of the points denote the number of components C. We find $C = 28$ using DPMM on the NJU-DS2000 dataset.)

显著图与深度显著图融合的 3D 显著性检测方法 LMH[40] 以及融合深度先验的 3D 显著性检测方法 GP[39]. 这些方法的显著性结果均通过运行原作者公开的源代码得到.

4.1 实验数据和评价标准

在本节中, 我们选取 NLPR 数据集和 NJU-DS2000 数据集进行对比实验, 如表 3 所示. NLPR 数据集通过 Kinect 设备采集不同真实场景下的 1 000 张 3D 图片, 分别存储成 RGB 图像、Depth 图像和真值图像 (Ground truth, GT), 其中真值图像由 5 位参与者选出引人注意的显著物体区域得到, 即手工标记出显著的区域为 255, 非显著区域为 0. 该数据集由 Peng 等[40] 提供, 以便客观地评估 3D

显著性检测模型的性能[1]. NJU-DS2000 数据集采用双目视觉方法得到不同场景下的 2 000 张 3D 图片, 其中真值图像由 4 名自愿者手工标定. 与 NLPR 数据集不同的是, 其 Depth 图像由两个视角的 RGB 图像通过立体匹配得到[59]. 该数据集由 Ju 等[38] 提供, 在 NJU-DS400 数据集基础上进行扩充[2].

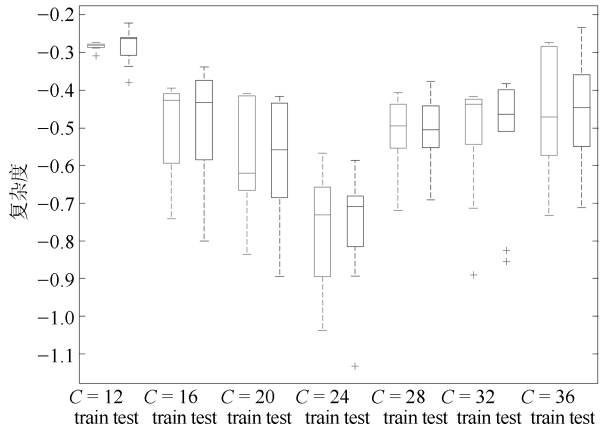


图 9　对于 NLPR 数据集交叉验证 DMNB 模型混合分量参数 C, 给定一个由 DPMM 模型得到的参数 C 的取值范围, 采用 10-fold 进行交叉验证

Fig. 9　Cross validation for the parameter C in the DMNB model in terms of NLPR dataset, we use 10-fold cross-validation with the parameter C for DMNB models, the C found using DPMM was adjusted over a wide range in a 10-fold cross-validation

表 3　NLPR 数据集和 NJU-DS2000 数据集分布情况

Table 3　The benchmark and existing 3D saliency detection dataset

数据集	图片数	显著目标数	场景种类	中央偏置
NLPR	1 000	(大多数) 一个	11	是
NJU-DS2000	2 000	(大多数) 一个	> 20	是

目前为止, 还没有特定的标准来评价 3D 显著性检测方法的性能指标. 然而, 对于 2D 情况下有不同指标应用于评价显著性检测方法的性能. 引入两种评价标准来评估 3D 显著性检测方法的性能: 第一种是 F 测度 (F-measure), 第二种是 ROC (Receiver operating characteristic) 曲线.

F 测度是一种通过加权准确率 (Precision) 和召回率 (Recall) 的整体性能评价标准, 定义如下:

$$F_\beta = \frac{(1+\beta^2)\text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}} \tag{23}$$

准确率是检测到且属于真值图像的部分所占检测到所有部分的比例, 召回率是检测到且属于真值图像的部分占真值图像的比例. 为了强调在显著性检测中准确率相对召回率的重要性, 将 β^2 设置为 0.3, 同文献 [32] 一致.

计算 F 测度时, 选取显著图平均值的 2 倍作为动态阈值, 定义如下:

$$T = \frac{2}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} S(i,j) \tag{24}$$

其中, W 和 H 分别表示图像的宽度和高度, $S(i,j)$ 是 (i,j) 位置处像素点的显著性值.

第二种评价标准是 ROC (Receiver operating characteristic) 曲线和 AUC (The area under the ROC curve) 值. ROC 曲线是一种分析决策性能的定量方法, 用来评价显著性检测结果是否与真实值图像一致性. 将算法得到显著性图归一化为 $[0, 255]$ 的范围, 然后选取阈值从 0 到 255 依次变化分割显著图得到二值图像. 同时将真值图像二值化, 其中显著区域像素值为 255, 非显著区域像素值为 0. 最后参照真值图像计算真正率 (True positive rate) 和假正率 (False positive rate) 绘制 ROC 曲线. AUC 值为 ROC 曲线下的面积, 其 AUC 值越大表示算法性能越好.

4.2　NLPR 数据集实验比较

4.2.1　运行时间比较

表 4 给出本文提出的方法 BFSD 与 GMR、MC、MDF、ACSD、LMH 和 GP 6 种方法在 NLPR 数据集上处理一幅 RGB-D 图像的平均运行时间的对比, 其中 2D 显著性检测方法 GMR、MC 和 MDF 只处理 RGB 图像得到显著图, ACSD 显著性检测方法只处理 Depth 图像得到显著图, BFSD 和 3D 显著性检测方法 LMH、GP 同时处理 RGB 图像和 Depth 图像得到显著图. 由于实验 PC 机没有 GPU 卡进行加速, 基于深层卷积神经网络提取显著特征的方法 MC、MDF 和 BFSD 运行时间均较长, 其中 MDF 运行时间最长是因为采用多尺度超像素分割进行显著性检测, 以不同尺度下得到的超像素为中心生成 3 个不同尺寸的矩形框区域由深层卷积神经网络提取特征, 最后将不同尺度下得到的显著图经过条件随机场 (Conditional random field) 框架进行显著图融合. 而 ACSD 运行时间最短是因为该算法只处理 Depth 图像, 通过 SLIC 超像素分割算法[51] 得到超像素区域, 计算超像素的各向异性–周边差异进行显著性检测. 运行时间最长的 MDF 方法与运行时间最短的 ACSD 方法处理时间的差异在于三个方面的原因: 第一个原因是 MDF

[1] http://sites.google.com/site/rgbdsaliency
[2] http://mcg.nju.edu.cn/en/resource.html

方法是基于多尺度超像素分割, 而 ACSD 是单尺度超像素分割; 第二个原因是 MDF 方法基于深层卷积神经网络提取显著特征, 即以不同尺度下得到的超像素为中心生成 3 个尺寸的矩形区域采用深层卷积神经网络提取特征, 而 ACSD 方法采用超像素的各向异性–周边差异进行显著特征提取; 第三个原因是 MDF 方法采用条件随机场对不同尺度得到的显著图进行融合, 而 ACSD 方法不需要融合. 其中最耗费时的是采用深层卷积神经网络提取特征的部分, 因为实验机器没有 GPU 卡而通过 CPU 模式运行深层卷积神经网络. 未来采用 GPU 卡运行深层卷积神经网络提取特征, MDF 方法运行时间会大幅降低, 也包括本文提出的方法 BFSD.

表 4 NLPR 数据集处理一幅 RGB-D 图像平均时间比较

Table 4 Comparison of the average running time for per RGB-D image on the NLPR dataset

数据集	GMR	MC	MDF	ACSD	LMH	GP	BFSD
NLPR	2.9 s	72.7 s	942.3 s	0.2 s	2.8 s	38.9 s	80.1 s

4.2.2 颜色 – 深度显著情况实验比较

对于 3D 显著性分布的颜色–深度显著情况, 各显著性检测方法得到的显著图如图 10 所示. 对于颜色–深度显著的情况, 显著性区域在 RGB 图像和 Depth 图像均有分布, 采用融合 RGB 图像显著特征和 Depth 图像显著特征的方法 LMH、GP 利用 Depth 图像深度特征克服 RGB 图像背景的干扰, 如图 10 第 3 行所示的黑色木雕. 然而, RGB 图像中存在显著区域与周围区域颜色等低层特征对比度不高的情况, 基于深层卷积神经网络的显著性检测方法 MC、MDF 提取的高层类别特征克服 RGB 图像显著区域颜色等低层特征对比度不高的情况, 如图 10 第 1 行所示的白色 “请勿泊车” 的牌子与周围白色水泥路面. 采用深层卷积神经网络提取 Depth 图像的高层类别特征, 在贝叶斯框架下与深层卷积神经网络提取 RGB 图像的高层类别特征融合, 能够克服 3D 显著性区域在 RGB 图像和 Depth 图像与周围区域低层特征对比度不高的情况, 如图 10 第 10 行所示的花叶.

如图 11 (a) 和图 12 (a) 所示, 本文提出的方法 BFSD 与 GMR、MC、MDF、ACSD、LMH 和 GP 6 种方法在 NLPR 数据集颜色–深度显著情况的 ROC 曲线、准确率、召回率和 F 测度对比. 与图 10 观察一致, 基于深层卷积神经网络的显著性检测方法 MC、MDF 优于基于低层特征的显著性检测方法 GMR, 因为深层卷积神经网络提取 RGB 图像的高层特征含有的类别信息能够克服显著区域与周围背景区域低层特征对比度不高的情况. 由于 Depth 图像深度低层特征的局限, 对于仅仅提取深度对比特征的显著性检测方法 ACSD 较其他显著性检测方法明显处于劣势. 如图 12 (a) 所示, 融合深度先验的 3D 显著性检测方法 GP 优于基于边界先验显著性检测方法 GMR, 而简单融合 RGB 图像显著图与 Depth 图像显著图的 3D 显著性检测方法 LMH 没有比 GMR 方法具有明显优势. 采用深层卷积神经网络提取 Depth 图像的高层类别特征, 在贝叶斯框架下与深层卷积神经网络提取 RGB 图像的高层类别特征融合, 在检测准确率、召回率和 F 测度均优于其他 6 种方法. 如图 11 (a) 所示, 由于颜色–深度显著情况下颜色显著特征与深度显著特征不是严格条件独立, 所以提出的方法与同样基于深层卷积神经网络提取显著特征的 MC、MDF 方法 ROC 曲线相近, 对应的 AUC 值比较如表 5 第 1 行所示.

4.2.3 颜色显著情况实验对比

对于 3D 显著性分布的颜色显著情况, 上述显著性检测方法得到的显著图如图 13 所示. 由于 3D 显著性区域只在 RGB 图像是显著的, 而在 Depth 图像是非显著的, 基于 Depth 图像深度特征的显著性检测方法 ACSD[38] 很难检测到正确的显著性区域, 如图 13 第 5 行中的雕像; 基于简单融合 RGB 图像显著图和 Depth 图像显著图的 3D 显著性检测方法 LMH 受到深度显著图的干扰, 也很难检测到正确的显著性区域, 如图 13 第 2 行中的球柱建筑.

如图 11 (b) 所示, 对于 3D 显著性分布的颜色显著情况, 显著性区域只在 RGB 图像有分布, 而在 Depth 图像上是不显著的, 对于直接融合深度特征的 3D 显著性方法 GP、LMH 明显受到 Depth 图像显著性特征的干扰, 而基于 RGB 图像的显著性检测方法 GMR、MC、MDF 则没有什么影响, 其 AUC 值对比如表 5 第 2 行所示. 如图 11 (b) 所示,GP 和 LMH 方法对于颜色显著情况检测的准确率、召回率和 F 测度指标都劣于颜色–深度显著的情况. 分析深层卷积神经网络提取的 RGB 图像和 Depth 图像的显著特征的相关性, 采用基于贝叶斯框架进行颜色特征和深度特征的融合, 优于直接融合深度特征的策略.

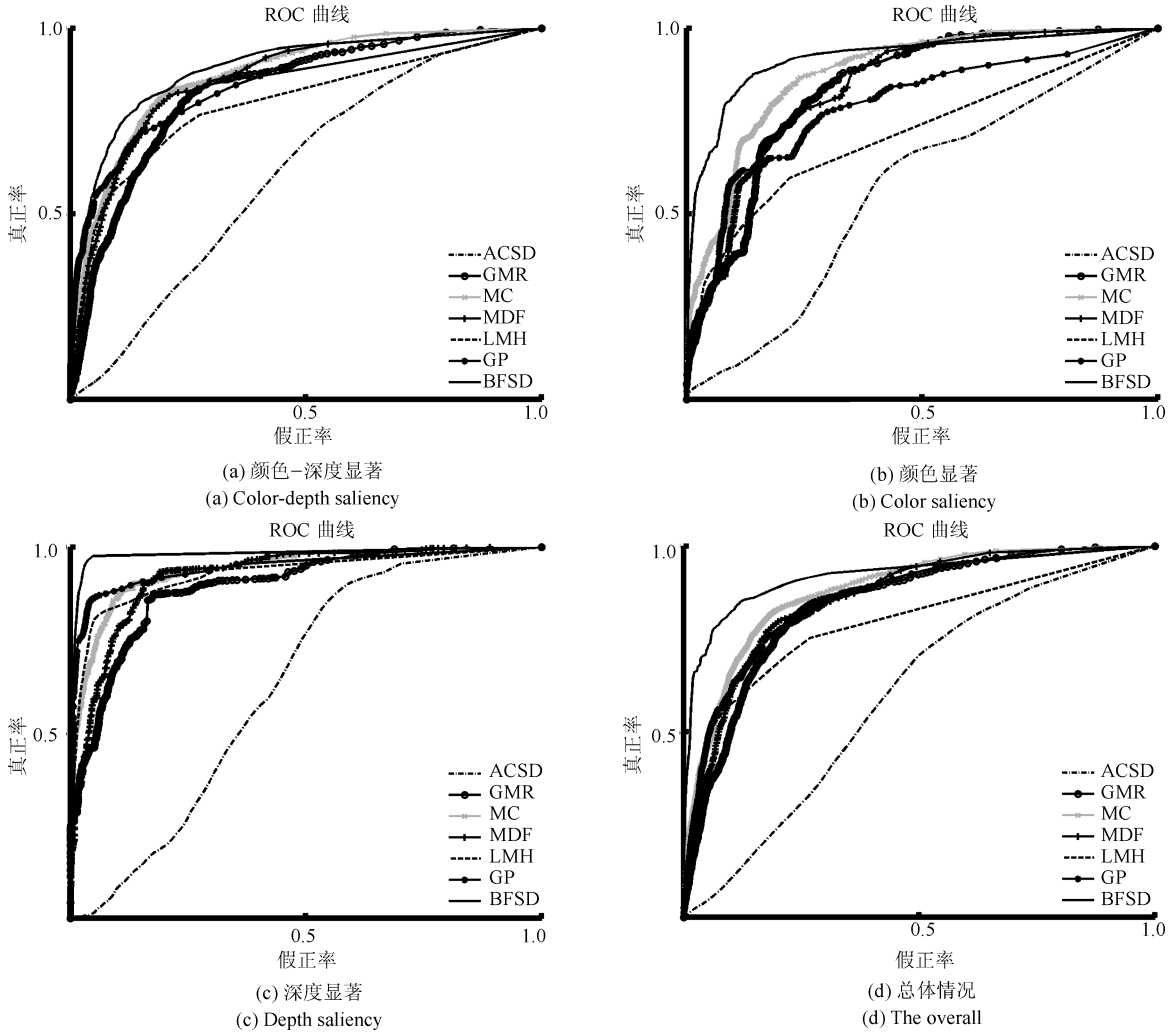
4.2.4 深度显著情况实验对比

对于 3D 显著性分布的深度显著情况, 上述显著性检测方法得到的显著图如图 14 所示. 由于 3D 显著性区域在 RGB 图像是不显著的, 基于 RGB 图像 2D 显著性检测方法 GMR、MC、MDF 很难得到正确的显著性区域, 如图 14 第 2 行所示黑色的鼎. 基于边界先验的显著性检测方法 GMR 方法假设显著性区域绝大数分布在图像的中心位置, 而背景分布在图像的边缘, 所以很难处理显著性区域分布在图

像边界的情况, 如图 10 中第 8 行所示的白色石柱、如图 13 中第 5 行所示的塑像、如图 14 中第 3 行所示的石灯. 由于 3D 显著性区域只在 Depth 图像是显著的, 直接融合 RGB 图像显著图和 Depth 图像显著图的 3D 显著性检测方法 LMH 和利用深度先

图 10　NLPR 数据集颜色–深度显著情况显著图对比. ((a) RGB 图像; (b) Depth 图像; (c) 真值图; (d) ACSD 方法; (e) GMR 方法; (f) MC 方法; (g) MDF 方法; (h) LMH 方法; (i) GP 方法; (j) 本文方法)

Fig. 10　Visual comparison of the saliency detection in the color-depth saliency situation in terms of NLPR dataset ((a) RGB image, (b) Depth image, (c) Ground truth, (d) ACSD, (e) GMR, (f) MC, (g) MDF, (h) LMH, (i) GP, (j) BFSD)



(a) 颜色−深度显著
(a) Color-depth saliency

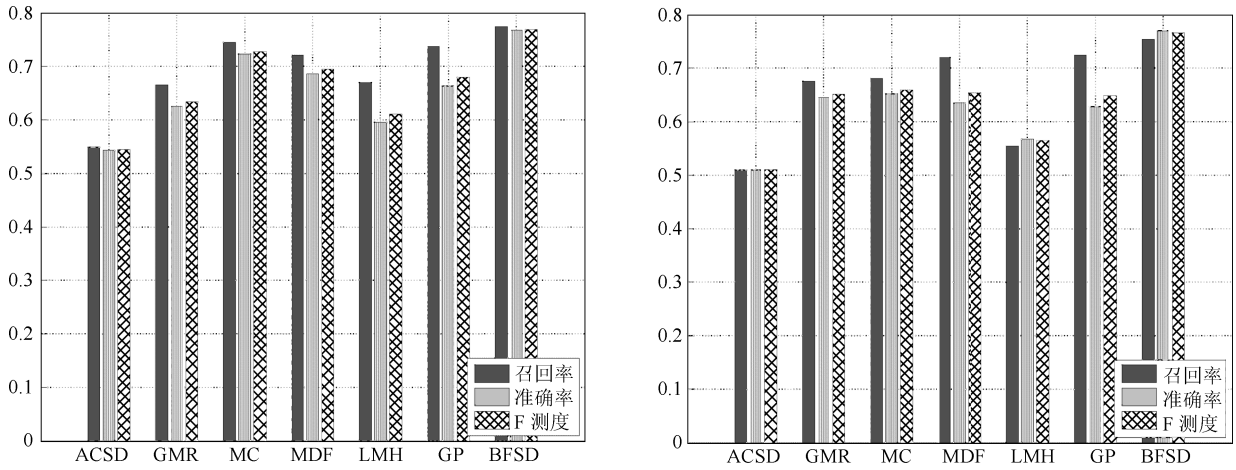
(b) 颜色显著
(b) Color saliency

(c) 深度显著
(c) Depth saliency

(d) 总体情况
(d) The overall

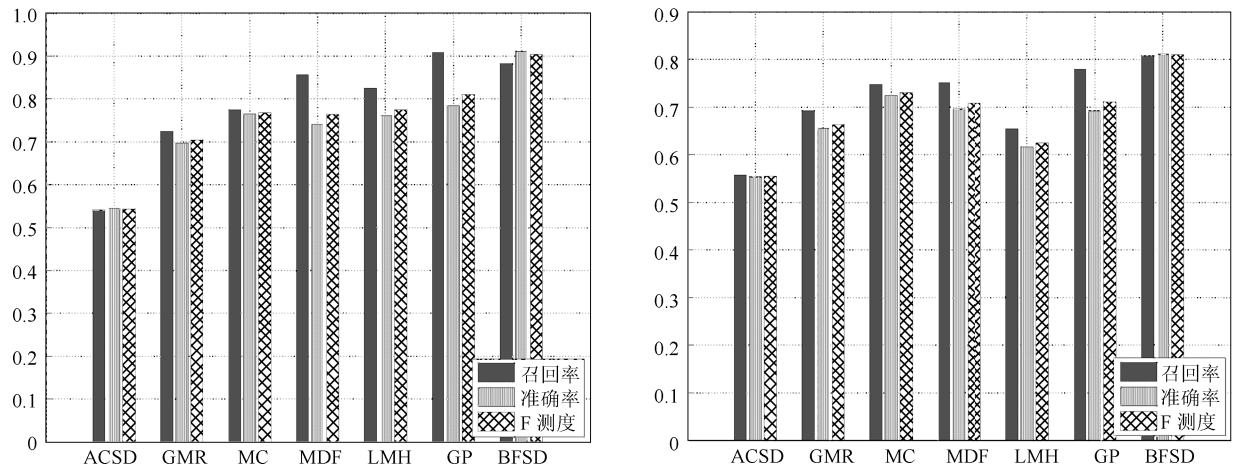
图 11 NLPR 数据集 ROC 曲线对比图

Fig. 11 The ROC curves of different saliency detection models in terms of the NLPR dataset



(a) 颜色−深度显著
(a) Color-depth saliency

(b) 颜色显著
(b) Color saliency

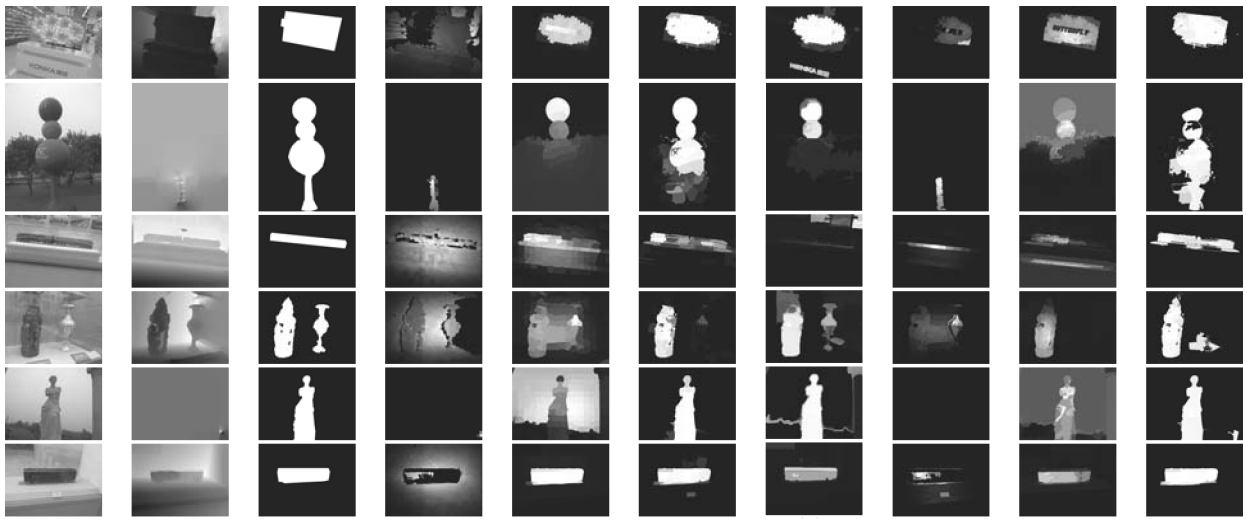


(c) 深度显著
(c) Depth saliency

(d) 总体情况
(d) The overall

图 12　NLPR 数据集 F 测度结果对比图

Fig. 12　The F-measures of different saliency detection models when used on the NLPR dataset



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

图 13　NLPR 数据集颜色显著情况显著图对比 ((a) RGB 图像; (b) Depth 图像; (c) 真值图; (d) ACSD 方法; (e) GMR 方法; (f) MC 方法; (g) MDF 方法; (h) LMH 方法; (i) GP 方法; (j) 本文方法)

Fig. 13　Visual comparison of the saliency detection in the color saliency situation in terms of NLPR dataset ((a) RGB image, (b) Depth image, (c) Ground truth, (d) ACSD, (e) GMR, (f) MC, (g) MDF, (h) LMH, (i) GP, (j) BFSD)

验的 3D 显著性检测方法 GP 通过利用 Depth 图像深度特征能够有效减少 RGB 图像对 3D 显著性区域检测的干扰, 如图 14 第 4 行的灯光和第 5 行的柜子. 本文提出的方法 BFSD 考虑深层卷积神经网络提取的 RGB 图像和 Depth 图像显著性特征是条件独立分布的, 基于贝叶斯框架进行融合显著特征, 由生成模型建模 3D 显著性检测得到显著性区域, 如图 14 第 3 行中的石灯.

如图 11 (c) 所示, 对于 3D 显著性分布的深度显著情况, 显著性区域只在 Depth 图像是显著的, 而在 RGB 图像上是不显著的, 而融合深度特征的 3D 显著性检测方法 GP 和 LMH 利用 Depth 图像深度特征克服 RGB 图像不显著, 优于基于 RGB 图像的显著性检测方法 GMR、MC 和 MDF, 其 AUC 值对比如表 5 第 3 行所示. 如图 12 (c) 所示, GMR、MC 和 MDF 方法在深度显著情况的准确率、召回率和 F 测度指标劣于颜色–深度显著的情况, 而优于颜色显著的情况. 与颜色显著情况相似, 在深度显著情况下, 深层卷积神经网络提取的 RGB 图像和 Depth 图像的显著特征是条件独立分布的, 基于贝叶斯框架进行颜色特征和深度特征的融合, 优于直接融合深度特征的策略.

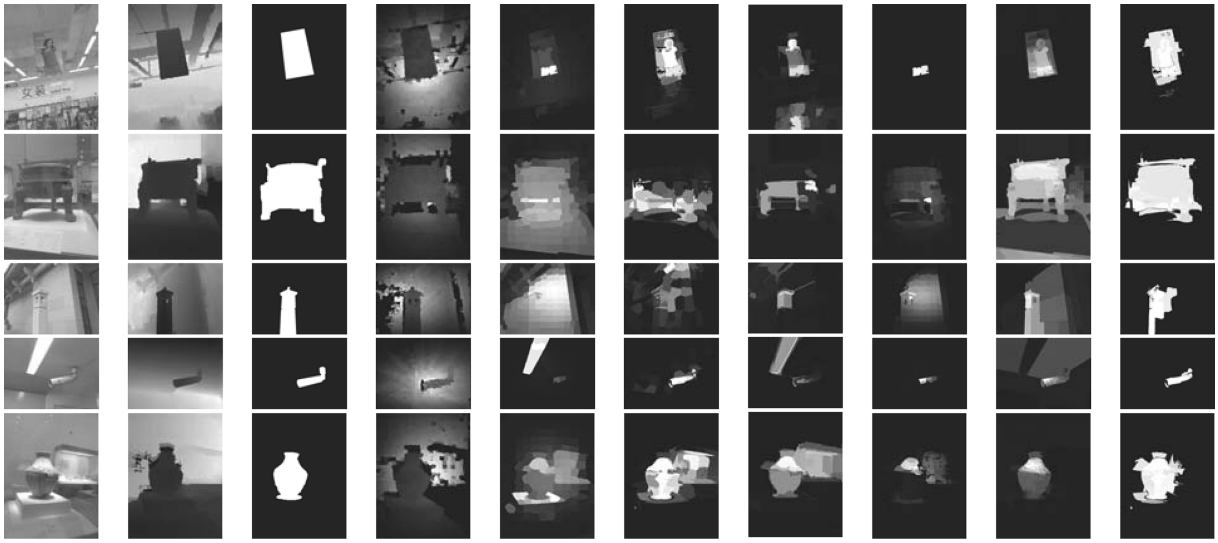


图 14 NLPR 数据集深度显著情况显著图对比 ((a) RGB 图像; (b) Depth 图像; (c) 真值图; (d) ACSD 方法; (e) GMR 方法; (f) MC 方法; (g) MDF 方法; (h) LMH 方法; (i) GP 方法; (j) 本文方法)

Fig. 14 Visual comparison of the saliency detection in the depth saliency situation in terms of NLPR dataset ((a) RGB image, (b) Depth image, (c) Ground truth, (d) ACSD, (e) GMR, (f) MC, (g) MDF, (h)LMH, (i) GP, (j) BFSD)

表 5 给出本文提出的方法 BFSD 与 GMR、MC、MDF、ACSD、LMH 和 GP 6 种方法在 NLPR 数据集上的 AUC 值对比. 基于深度特征的显著性检测方法 ACSD 在 3D 显著性分布深度显著情况得到比颜色–深度显著情况和颜色显著情况较高的 AUC 值, 但是总体比其他方法的 AUC 值较低, 仅仅基于 Depth 图像的深度特征的 3D 显著性检测无法得到较高的性能; 在 3D 显著性分布的深度显著情况, 融合深度特征的 3D 显著性检测方法 LMH、GP 和 BFSD 均比基于 RGB 图像显著特征的方法 GMR、MC 和 MDF 要好, 说明深度特征能够在 RGB 图像不显著时有助于 3D 显著性检测; 在 3D 显著性分布的颜色显著情况, 融合深度特征的 3D 显著性检测方法 LMH、GP 却比基于 RGB 图像显著性特征的方法 GMR、MC 和 MDF 要差, 说明在融合深度特征的时候也要考虑颜色特征和深度特征的分布关系. 采用类条件互信息熵分析深层卷积神经网络提取的 RGB 图像和 Depth 图像显著特征的相关性, 假设条件独立下基于贝叶斯框架下进行特征融合, 在 3D 显著性分布的三种情况均得到较高的 AUC 值.

如图 12 (d) 所示, 本位提出的方法 BFSD 与 GMR、MC、MDF、ACSD、LMH 和 GP 6 种方法在整个 NLPR 数据集上的准确率、召回率和 F 测度的对比. 对比不同的融合策略来融合 RGB 图像的颜色特征和 Depth 图像的深度特征发现, 采用相乘融合方法的 3D 显著性检测方法 LMH 与基于低层颜色特征 2D 显著性检测方法 GMR 有相近的 F 测度; 而采用融合深度先验的 3D 显著性检测方法 GP 得到比基于低层颜色特征 2D 显著性检测方法 GMR 更高的 F 测度. 基于深层卷积神经网络的 MC 和 MDF 方法采用高层类别特征较基于融合低层颜色特征和深度特征的 3D 显著性检测方法 LMH 和 GP 有较高的准确率、召回率, 说明高层特征比低层特征有助于显著性检测; 采用深层卷积神经网络提取 RGB 图像和 Depth 图像高层显著特征, 在贝叶斯框架下进行融合, 在 3D 显著性分布的三种情况均得到较好的准确率、召回率和 F 测度.

表 5 AUC 值比较

Table 5 Comparison of the AUC on the NLPR dataset

显著分布情况	ACSD	GMR	MC	MDF	LMH	GP	BFSD
颜色–深度显著	0.61	0.73	0.81	0.82	0.70	0.79	0.83
颜色显著	0.56	0.74	0.84	0.83	0.61	0.65	0.84
深度显著	0.63	0.71	0.76	0.74	0.75	0.81	0.90
总体	0.60	0.73	0.81	0.80	0.69	0.78	0.85

4.3 NJU-DS2000 数据集实验比较

NJU-DS2000 数据集深度信息获取方式与 NLPR 数据集深度信息获取方式不同, 其中 NLPR 数据集深度信息由三维点云表示的绝对深度值, 而 NJU-DS2000 数据集深度信息是由视差图表示的相对深度值. 由于 LMH 和 GP 方法针对 NLPR 数据集设计, 融合的深度显著性需要绝对深度值计算得到, 所以在 NJU-DS2000 数据集对比实验缺少 LMH 和 GP 方法的实验结果.

本文提出的方法 BFSD 与 GMR、MC、MDF 和 ACSD 4 种方法在 NJU-DS2000 数据集的实验结果如图 15 ～ 图 19 所示. 图 15 表示 NJU-DS2000 数据集颜色 – 深度显著情况下显著图对比, 在该情

图 15　NJU-DS2000 数据集颜色 – 深度显著情况显著图对比 ((a) RGB 图像; (b) Depth 图像; (c) 真值图; (d) ACSD 方法; (e) GMR 方法; (f) MC 方法; (g) MDF 方法; (h) 本文方法)

Fig. 15　Visual comparison of the saliency detection in the color-depth saliency situation in terms of NJU-DS2000 dataset ((a) RGB image, (b) Depth image, (c) Ground truth, (d) ACSD, (e) GMR, (f) MC, (g) MDF, (h) BFSD)

图 16 NJU-DS2000 数据集颜色显著情况显著图对比. ((a) RGB 图像; (b) Depth 图像; (c) 真值图; (d) ACSD 方法; (e) GMR 方法; (f) MC 方法; (g) MDF 方法; (h) 本文方法)

Fig. 16 Visual comparison of the saliency detection in the color saliency situation in terms of NJU-DS2000 dataset ((a) RGB image, (b) Depth image, (c) Ground truth, (d) ACSD, (e) GMR, (f) MC, (g) MDF, (h) BFSD)

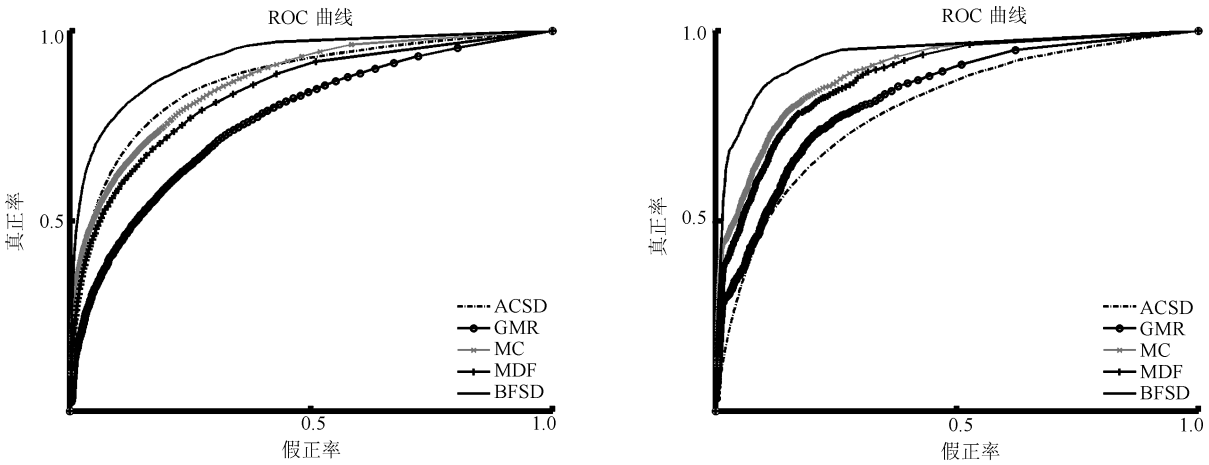
况下显著区域的颜色信息和深度信息存在互补的关系, 提出的方法 BFSD 通过融合颜色和深度信息相比只考虑颜色信息的 GMR、MC 和 MDF 方法和只考虑深度信息的 ACSD 方法得到更完整的显著图. 如图 18 (a) 和图 19 (a) 所示, BFSD 得到的 ROC 曲线以及准确率、召回率和 F 测度均优于上述 4 种方法.

如图 16 所示, NJU-DS2000 数据集颜色显著情况下显著图检测结果对比. 由于 3D 显著区域只在 RGB 图像是显著的, 而在 Depth 图像是不显著的, 只考虑深度信息进行显著性检测的 ACSD 方法较难得到准确的显著区域, 如图 18 (b) 所示, 基于颜色信息进行显著性检测的方法比 ACSD 方法均获得较好的 ROC 曲线. 如图 19 (b) 所示, 基于颜色信息进行显著性检测方法中, 由于 MC 和 MDF 方法采用深层卷积神经网络提取高层类别特征得到比基于低层特征的 GMR 方法得到更高的召回率.

如图 17 所示, NJU-DS2000 数据集深度显著情况下显著图检测结果对比. 在 NJU-DS2000 数据集深度显著情况下, 由于 3D 显著区域只在 Depth 图像是显著的, 而在 RGB 图像是不显著的, 只考虑颜色信息进行显著性检测的 GMR、MC 和 MDF 方法较难得到准确的显著区域. 相比颜色显著情况, 基于低层特征的 GMR 方法检测准确率迅速降低, 如

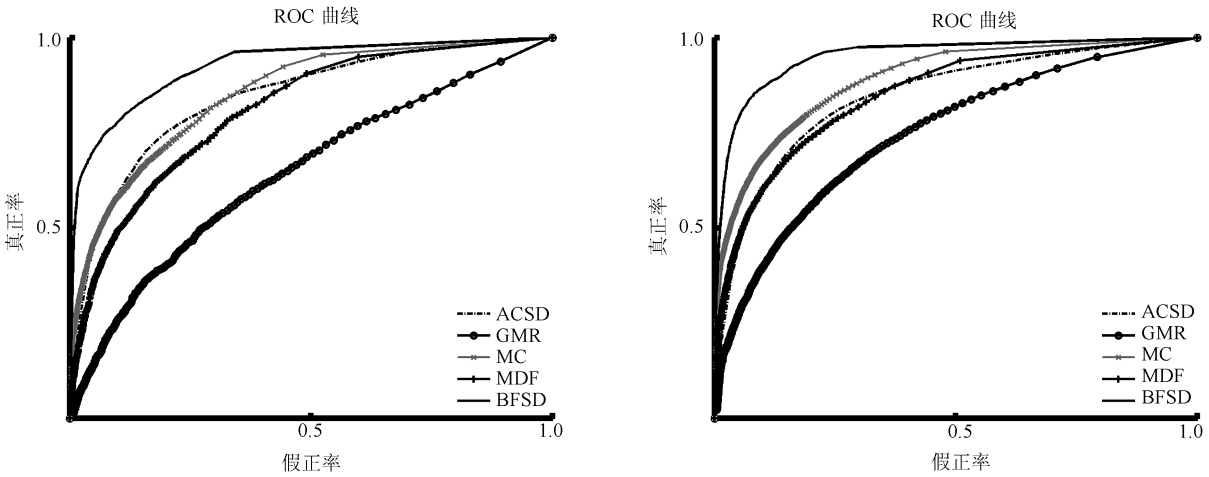
图 17　NJU-DS2000 数据集深度显著情况显著图对比 ((a) RGB 图像; (b) Depth 图像; (c) 真值图; (d) ACSD 方法; (e) GMR 方法; (f) MC 方法; (g) MDF 方法; (h) 本文方法)

Fig. 17　Visual comparison of the saliency detection in the depth saliency situation in terms of NJU-DS2000 dataset ((a) RGB image; (b) Depth image; (c) Ground truth; (d) ACSD; (e) GMR; (f) MC; (g) MDF; (h) BFSD)



(a) 颜色−深度显著
(a) Color-depth saliency

(b) 颜色显著
(b) Color saliency

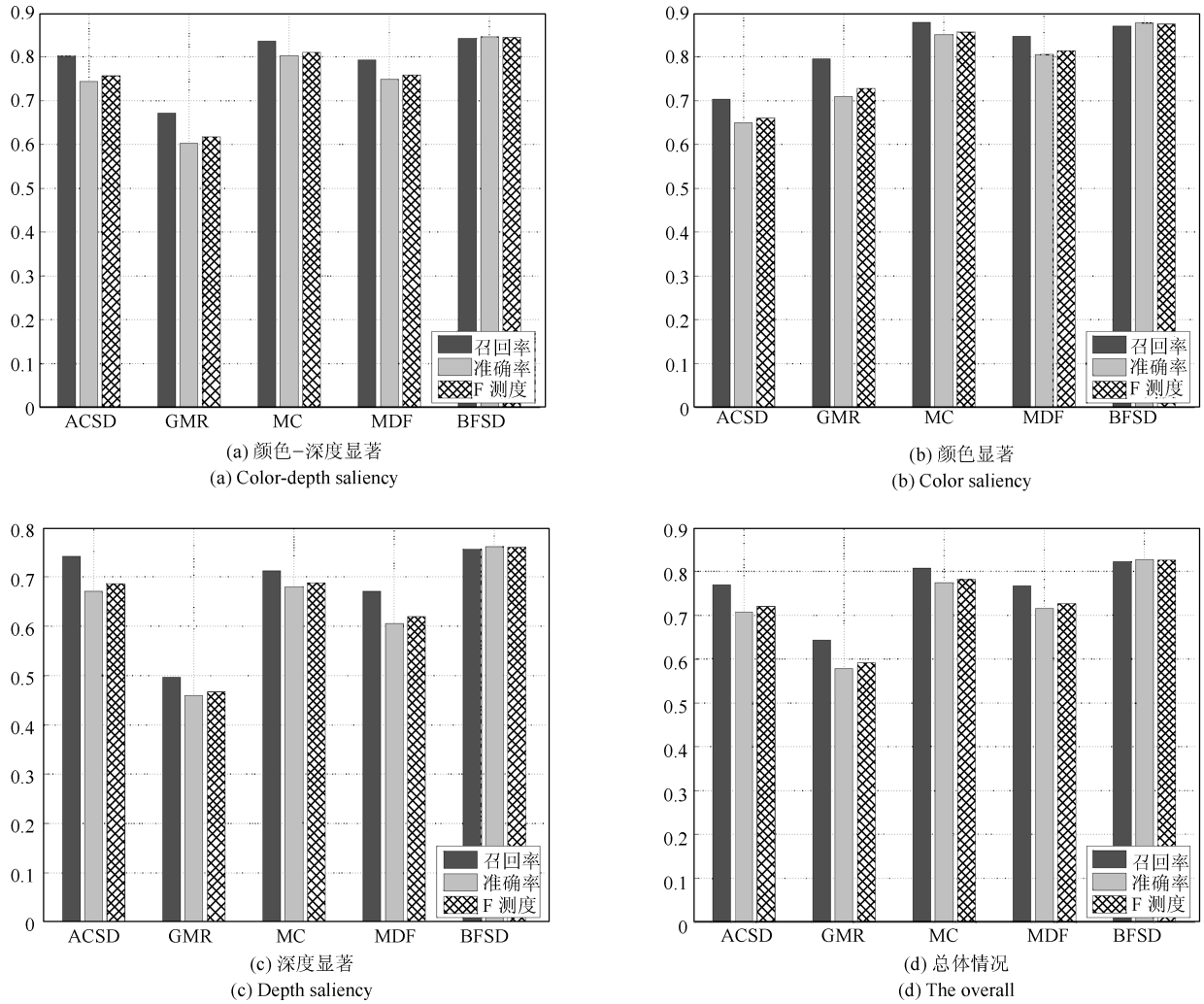


(c) 深度显著
(c) Depth saliency

(d) 总体情况
(d) The overall

图 18 NJU-DS2000 数据集 ROC 对比图

Fig. 18 The ROC curves of different saliency detection models in terms of the NJU-DS2000 dataset



(a) 颜色−深度显著
(a) Color-depth saliency

(b) 颜色显著
(b) Color saliency

(c) 深度显著
(c) Depth saliency

(d) 总体情况
(d) The overall

图 19 NJU-DS2000 数据集 F 测度结果对比图

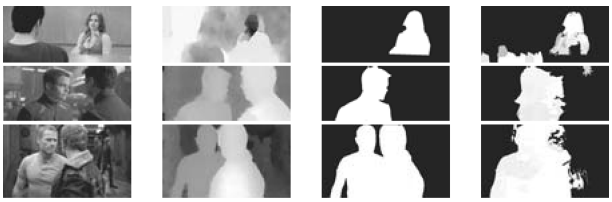
Fig. 19 The F-measures of different saliency detection models when used on the NJU-DS2000 dataset

图 19 (b) 和图 19 (c) 所示. 而基于深度信息进行显著性检测的 ACSD 方法比基于颜色信息进行显著性检测的方法得到较高的召回率, 如图 19 (c) 所示.

本文提出的方法 BFSD 利用显著区域的颜色和深度特征分布关系基于贝叶斯框架下进行融合, 在颜色显著情况和深度显著情况下均得到较好的 ROC 曲线, 如图 18 (b) 和图 18 (b) 所示. 最后, 如图 19 (d) 所示, 对于 NJU-DS2000 数据集总体情况本文提出的方法 BFSD 在检测准确率、召回率和 F 测度均优于 GMR、MC、MDF 和 ACSD 4 种方法.

4.4　失败情况

通过观察 NLPR 数据集提出 3D 显著性区域在 RGB 图像和 Depth 图像分布的三种情况, 但是对于 NJU-DS2000 数据集存在上述假设之外的情况. 如图 20 所示, 第一行和第二行的背向观察者的男子在真值图像中标记为不显著的, 而在第三行中背向观察者的男子在真值图像中标记为显著的. 从 RGB 图像中观察, 背向观察者的男子没有面向观察者 (第一行和第三行) 的男子或者 (第二行的) 女子显著, 但是从 Depth 图像观察, 背向观察者的男子由于距离观察者相对较近, 则会吸引较多的注意. 由于提出的方法采用了贝叶斯框架下融合颜色显著特征和深度显著特征进行显著性检测, 通过有监督的生成模型进行学习, 依赖有标签的 RGB 图像和 Depth 图像数据进行训练. 当样本中存在 3D 显著性分布不一致的情形, 尤其 RGB 图像和 Depth 图像存在 "冲突的" 的区域, 如图 20 所示的背向观察者的男子, 本文方法无法准确地估计其显著性.



(a) RGB 图像 (a) RGB image　(b) Depth 图像 (b) Depth image　(c) 真值图 (c) Ground truth　(d) 本文方法 (d) BFSD

图 20　失败情况

Fig. 20　Some failure cases

5　结论

提出一种基于贝叶斯框架融合 RGB 图像颜色信息和 Depth 图像深度信息的 RGB-D 图像显著性检测方法. 首先分析 3D 显著性在 RGB 图像和 Depth 图像分布的情况, 采用类条件互信息熵分析由卷积神经网络提取的颜色特征和深度特征的条件独立性, 根据贝叶斯定理得到 RGB-D 图像显著性的后验概率. 假设颜色特征和深度特征符合高斯分布, 采用 DMNB 生成模型进行显著性检测建模, 其模型参数由变分最大期望算法进行估计. 在 RGB-D 图像显著性目标公开数据集 NLPR 和 NJU-DS2000 上测试, 对于 3D 显著性分布的三种情况, 提出的方法 BFSD 与 6 种 state-of-the-art 显著性检测方法比较, 其中包括 2D 显著性方法 ACSD[38]、GMR[32]、MC[33]、MDF[34] 和 3D 显著性检测方法 LMH[40]、GP[39], 实验结果表明提出的方法均获得较高的准确率和召回率.

对于基于贝叶斯框架的 RGB-D 图像显著性检测问题, 未来仍有有几个关键问题需要进行深入研究, 主要包括先验参数的在线估计以及变分推理算法的优化问题.

References

1 Borji A, Itti L. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(1): 185−207

2 Wang Wen-Guan, Shen Jian-Bing, Shao Ling, Porikli Fatih. Correspondence driven saliency transfer. *IEEE Transaction on Image Processing*, 2016, **25**(11): 5025−5034

3 Ding Zheng-Hu, Yu Ying, Wang Bin, Zhang Li-Ming. Visual attention-based ship detection in multispectral imagery. *Journal of Computer-Aided Design & Computer Graphics*, 2011, **23**(3): 419−425
(丁正虎, 余映, 王斌, 张立明. 选择性视觉注意机制下的多光谱图像舰船检测. 计算机辅助设计与图形学学报, 2011, **23**(3): 419−425)

4 Gao D S, Han S Y, Vasconcelos N. Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(6): 989−1005

5 Jian M W, Dong J Y, Ma J. Image retrieval using wavelet-based salient regions. *The Imaging Science Journal*, 2011, **59**(4): 219−231

6 Wang Xiang-Yang, Yang Hong-Ying, Zheng Hong-Liang, Wu Jun-Feng. A color block-histogram image retrieval based on visual weight. *Acta Automatica Sinica*, 2010, **36**(10): 1489−1492
(王向阳, 杨红颖, 郑宏亮, 吴俊峰. 基于视觉权值的分块颜色直方图图像检索算法. 自动化学报, 2010, **36**(10): 1489−1492)

7 Feng Xin, Yang Dan, Zhang Ling. Saliency variation based quality assessment for packet-loss-impaired videos. *Acta Automatica Sinica*, 2011, **37**(11): 1322−1331
(冯欣, 杨丹, 张凌. 基于视觉注意力变化的网络丢包视频质量评估. 自动化学报, 2011, **37**(11): 1322−1331)

8 Gupta R, Chaudhury S. A scheme for attentional video compression. In: Proceedings of the 4th International Conference on Pattern Recognition and Machine Intelligence. Moscow, Russia: IEEE, 2011. 458−465

9 Guo C L, Zhang L M. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Transactions on Image Processing*, 2010, **19**(1): 185−198

10 Kim W, Kim C. A novel image importance model for content-aware image resizing. In: Proceedings of the 18th IEEE International Conference on Image Processing. Brussels, Belgium: IEEE, 2011. 2469−2472

11 Jiang Xiao-Lian, Li Cui-Hua, Li Xiong-Zong. Saliency based tracking method for abrupt motions via two-stage sampling. *Acta Automatica Sinica*, 2014, **40**(6): 1098−1107
(江晓莲, 李翠华, 李雄宗. 基于视觉显著性的两阶段采样突变目标跟踪算法. 自动化学报, 2014, **40**(6): 1098−1107)

12 Li Wan-Yi, Wang Peng, Qiao Hong. A survey of visual attention based methods for object tracking. *Acta Automatica Sinica*, 2014, **40**(4): 561−576
(黎万义, 王鹏, 乔红. 引入视觉注意机制的目标跟踪方法综述. 自动化学报, 2014, **40**(4): 561−576)

13 Le Callet P, Niebur E. Visual attention and applications in multimedia technologies. *Proceedings of the IEEE*, 2013, **101**(9): 2058−2067

14 Wang J L, Fang Y M, Narwaria M, Lin W S, Le Callet P. Stereoscopic image retargeting based on 3D saliency detection, In: Proceedings of 2014 International Conference on Acoustics, Speech, and Signal Processing. Florence, Italy: IEEE, 2014. 669−673

15 Kim H, Lee S, Bovik A C. Saliency prediction on stereoscopic videos. *IEEE Transactions on Image Processing*, 2014, **23**(4): 1476−1490

16 Zhang Y, Jiang G Y, Yu M, Chen K. Stereoscopic visual attention model for 3D video. In: Proceedings of the 16th International Conference on Multimedia Modeling. Chongqing, China: Springer, 2010. 314−324

17 Uherčík M, Kybic J, Zhao Y, Cachard C, Liebgott H. Line filtering for surgical tool localization in 3D ultrasound images. *Computers in Biology and Medicine*, 2013, **43**(12): 2036−2045

18 Zhao Y, Cachard C, Liebgott H. Automatic needle detection and tracking in 3D ultrasound using an ROI-based RANSAC and Kalman method. *Ultrasonic Imaging*, 2013, **35**(4): 283−306

19 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, **20**(11): 1254−1259

20 Hu Zheng-Ping, Meng Peng-Quan. Graph presentation random walk salient object detection algorithm based on global isolation and local homogeneity. *Acta Automatica Sinica*, 2011, **37**(10): 1279−1284
(胡正平, 孟鹏权. 全局孤立性和局部同质性图表示的随机游走显著目标检测算法. 自动化学报, 2011, **37**(10): 1279−1284)

21 Cheng M M, Mitra N J, Huang X L, Torr P H S, Hu S M. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(3): 569−582

22 Tang Yong, Yang Lin, Duan Liang-Liang. Image cell based saliency detection via color contrast and distribution. *Acta Automatica Sinica*, 2013, **39**(10): 1632−1641
(唐勇, 杨林, 段亮亮. 基于图像单元对比度与统计特性的显著性检测. 自动化学报, 2013, **39**(10): 1632−1641)

23 Guo Ying-Chun, Yuan Hao-Jie, Wu Peng. Image saliency detection based on local and regional features. *Acta Automatica Sinica*, 2013, **39**(8): 1214−1224
(郭迎春, 袁浩杰, 吴鹏. 基于 Local 特征和 Regional 特征的图像显著性检测. 自动化学报, 2013, **39**(8): 1214−1224)

24 Xu Wei, Tang Zhen-Min. Exploiting hierarchical prior estimation for salient object detection. *Acta Automatica Sinica*, 2015, **41**(4): 799−812
(徐威, 唐振民. 利用层次先验估计的显著性目标检测. 自动化学报, 2015, **41**(4): 799−812)

25 Shi K Y, Wang K Z, Lu J, B Lin L. PISA: pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors. In: Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013. 2115−2122

26 Judd T, Ehinger K, Durand F, Torralba A. Learning to predict where humans look. In: Proceedings of the 12th International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 2106−2113

27 Liu T, Yuan Z J, Sun J, Wang J D, Zheng N N, Tang X O, et al. Learning to detect a salient object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(2): 353−367

28 Wei Y C, Wen F, Zhu W J, Sun J. Geodesic saliency using background priors. In: Proceedings of the 12th European Conference on Computer Vision. Firenze, Italy: Springer, 2012. 29−42

29 Qian Sheng, Chen Zong-Hai, Lin Ming-Qiang, Zhang Chen-Bin. Saliency detection based on conditional random field and image segmentation. *Acta Automatica Sinica*, 2015, **41**(4): 711−724
(钱生, 陈宗海, 林名强, 张陈斌. 基于条件随机场和图像分割的显著性检测. 自动化学报, 2015, **41**(4): 711−724)

30 Shen X H, Wu Y. A unified approach to salient object detection via low rank matrix recovery. In: Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012. 853−860

31 Jiang H Z, Wang J D, Yuan Z J, Liu T, Zheng N N, Li S P. Automatic salient object segmentation based on context and shape prior. In: Proceedings of 2011 British Machine Vision Conference. Dundee, UK: BMVA Press, 2011. 110.1−110.12

32 Yang C, Zhang L H, Lu H C, Ruan X, Yang M H. Saliency detection via graph-based manifold ranking. In: Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013. 3166−3173

33 Zhao R, Ouyang W L, Li H S, Wang X G. Saliency detection by multi-context deep learning. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 1265−1274

34 Li G B, Yu Y Z. Visual saliency detection based on multiscale deep CNN features. *IEEE Transactions on Image Processing*, 2016, **25**(11): 5012−5024

35 Lang C Y, Nguyen T V, Katti H, Yadati K, Kankanhalli M, Yan S C. Depth matters: influence of depth cues on visual saliency. In: Proceedings of 12th European Conference on Computer Vision. Firenze, Italy: Springer, 2012. 101−115

36 Desingh K, Krishna K M, Rajan D, Jawahar C V. Depth really matters: improving visual salient region detection with depth. In: Proceedings of 2013 British Machine Vision Conference. Bristol, England: BMVA Press, 2013. 98.1−98.11

37 Niu Y Z, Geng Y J, Li X Q, Liu F. Leveraging stereopsis for saliency analysis. In: Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE, 2012. 454−461

38 Ju R, Ge L, Geng W J, Ren T W, Wu G S. Depth saliency based on anisotropic center-surround difference. In: Proceedings of 2014 IEEE International Conference on Image Processing. Pairs, France: IEEE, 2014. 1115−1119

39 Ren J Q, Gong X J, Yu L, Zhou W H, Yang M Y. Exploiting global priors for RGB-D saliency detection. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Boston, MA, USA: IEEE, 2015. 25−32

40 Peng H W, Li B, Xiong W H, Hu W M, Ji R R. RGBD salient object detection: a benchmark and algorithms. In: Proceedings of 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 92−109

41 Fang Y M, Wang J L, Narwaria M, Le Callet P, Lin W S. Saliency detection for stereoscopic images. *IEEE Transactions on Image Processing*, 2014, **23**(6): 2625−2636

42 Ciptadi A, Hermans T, Rehg J. An in depth view of saliency. In: Proceedings of 2013 British Machine Vision Conference. Bristol, United Kingdom: BMVA Press, 2013. 112.1−112.11

43 Wu P L, Duan L L, Kong L F. RGB-D salient object detection via feature fusion and multi-scale enhancement. In: Proceedings of 2015 CCF Chinese Conference on Computer Vision. Xi′an, China: Springer, 2015. 359−368

44 Iatsun I, Larabi M C, Fernandez-Maloigne C. Using monocular depth cues for modeling stereoscopic 3D saliency. In: Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing. Florence, Italy: IEEE, 2014. 589−593

45 Ouerhani N, Hugli H. Computing visual attention from scene depth. In: Proceedings of the 15th International Conference on Pattern Recognition. Barcelona, Spain: IEEE, 2000. 375−378

46 Xue H Y, Gu Y, Li Y J, Yang J. RGB-D saliency detection via mutual guided manifold ranking. In: Proceedings of 2015 IEEE International Conference on Image Processing. Quebec City, QC, Canada: IEEE, 2015. 666−670

47 Wang J L, Da Silva M P, Le Callet P, Ricordel V. Computational model of stereoscopic 3D visual saliency. *IEEE Transactions on Image Processing*, 2013, **22**(6): 2151−2165

48 Iatsun I, Larabi M C, Fernandez-Maloigne C. Visual attention modeling for 3D video using neural networks. In: Proceedings of 2014 International Conference on 3D Imaging. Liege, Belglum: IEEE, 2014. 1−8

49 Fang Y M, Lin W S, Fang Z J, Lei J J, Le Callet P, Yuan F N. Learning visual saliency for stereoscopic images. In: Proceedings of 2014 IEEE International Conference on Multimedia and Expo Workshops. Chengdu, China: IEEE, 2014. 1−6

50 Bertasius G, Park H S, Shi J B. Exploiting egocentric object prior for 3D saliency detection. arXiv:1511.02682, 2015.

51 Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(11): 2274−2282

52 Qu L Q, He S F, Zhang J W, Tian J D, Tang Y D, Yang Q X. RGBD salient object detection via deep fusion. *IEEE Transactions on Image Processing*, 2017, **26**(5): 2274−2285

53 Gupta S, Hoffman J, Malik J. Cross modal distillation for supervision transfer. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 2827−2836

54 Shan H H, Banerjee A, Oza N C. Discriminative mixed-membership models. In: Proceedings of the 9th IEEE International Conference on Data Mining. Miami, Florida, USA: IEEE, 2009. 466−475

55 Wang S T, Zhou Z, Qu H B, Li B. Visual saliency detection for RGB-D images with generative model. In: Proceedings of the 13th Asian Conference on Computer Vision. Taipei, China: Springer, 2016. 20−35

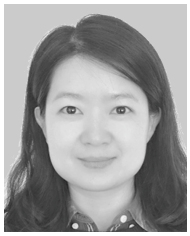
56 Rish I. An empirical study of the naive Bayes classifier. *Journal of Universal Computer Science*, 2001, **3**(22): 41−46

57 Blei D M, Jordan M I. Variational inference for dirichlet process mixtures. *Bayesian Analysis*, 2006, **1**(1): 121−143

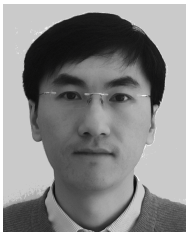
58 Sun D Q, Roth S, Black M J. Secrets of optical flow estimation and their principles. In: Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 2432−2439

王松涛 北京市科学技术研究院北京市新技术应用研究所助理研究员. 哈尔滨理工大学测控技术与仪器省高校重点实验室博士研究生. 主要研究方向为计算机视觉, 模式识别, 深度学习.
E-mail: wangsongtao1983@163.com
(**WANG Song-Tao** Assistant professor at Beijing Institute of New Technology Applications, Beijing Academy of Science and Technology. Ph. D. candidate at the Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentations of Heilongjiang Province, Harbin University of Science and Technology. His research interest covers computer vision, pattern recognition, and deep learning.)

周 真 哈尔滨理工大学测控技术与通信工程学院教授. 主要研究方向为可靠性工程技术, 生物信息检测. 本文通信作者. E-mail: zhzh49@126.com
(**ZHOU Zhen** Professor at the School of Measurement-Control Technology and Communications Engineering, Harbin University of Science and Technology. His research interest covers reliability engineering technology and biological information detection. Corresponding author of this paper.)



靳 薇 博士, 北京市科学技术研究院北京市新技术应用研究所副研究员. 全国公共安全基础标准技术委员会委员. 主要研究方向为机器学习, 计算机视觉, 模式识别, 生物特征识别.
E-mail: jinwei201002@163.com
(**JIN Wei** Associate professor at Beijing Institute of New Technology Applications, Beijing Academy of Science and Technology. She is also a committee member of National Technical Committee for Basic Standards of Public Safety. Her research interest covers machine learning, computer vision, pattern recognition, and biometrics.)



曲寒冰 博士, 北京市科学技术研究院北京市新技术应用研究所副研究员. 中国自动化学会智能自动化专业委员会委员. 主要研究方向为机器学习, 计算机视觉, 模式识别, 生物特征识别.
E-mail: quhanbing@gmail.com
(**QU Han-Bing** Associate professor at Beijing Institute of New Technology Applications, Beijing Academy of Science and Technology. He is also a committee member of Intelligent Automation Committee of Chinese Association of Automation (IACAA). His research interest covers machine learning, computer vision, pattern recognition, and biometrics.)