

基于贝叶斯置信传播的视觉注意层次模型^①

岑君凯^{②***} 霍 宏^{③***} 方 涛^{***}

(* 上海交通大学自动化系 上海 200240)

(** 系统控制与信息处理教育部重点实验室 上海 200240)

摘 要 受人类视觉系统采用层次性信息处理机制预测感兴趣区域和进行快速目标检测的启发,提出了一种新的基于贝叶斯推理的层次性视觉注意模型。该模型利用来自腹侧通路的颜色特征作为自底向上的推理信号,利用来自背侧通路的方向位置信息作为自顶向下的推理信号,并将这些信息在贝叶斯框架下集成,通过推理得到显著图。用自然场景图像集进行了该模型与 IT、GB 等 9 种视觉注意模型的性能对比实验,结果表明该模型的显著图更接近于人类注视图,能够更好地模拟视觉注意。此外,在遥感图像上的实验表明,与性能接近的 GB 模型相比,该模型能更好地实现目标检测。

关键词 视觉注意,显著图,贝叶斯推理,注视图,目标检测,层次性

0 引 言

人类的视觉注意 (visual attention) 机制能帮助人们在有限的脑资源条件下,从复杂的场景中筛选出显著或者感兴趣的区域,而忽略掉那些不关心的物体。近些年来,视觉注意机制及其模型的研究引起了神经心理学、计算机视觉等领域的学者的兴趣,他们在深入理解视觉注意机制的基础上,通过模拟该机制来构建符合生物学机理的视觉注意模型^[1-4],希望这些模型不但能解释神经科学的研究数据和模拟人眼对兴趣点的感知,还能在计算机视觉领域中的视频跟踪、目标检测、图像检索和分割中得到应用。本文在研究人类视觉系统 (human visual system, HVS) 层次性处理视觉信息的机制的基础上,提出了一种基于贝叶斯推理的层次性视觉注意模型,该模型能模拟视觉层信息的层次性处理机制来提取颜色和方向信息,利用置信传播来预测和获取图像的显著区域。

1 相关研究

对大脑的认知实验表明,人类视觉系统 (HVS)

对高对比度的目标更加敏感^[5-6],会更加注意到这些高对比度的显著区域。很多视觉注意模型都能一定程度上模拟这一特性,以此来预测人类注视点^[14,7-12]。这些模型或是基于生物学原理的,或是纯粹利用了一些数学计算方法。在基于生物学原理的模型中,早期 Itti 等^[1]提出的模型(本文称其为 IT 模型)最具代表性。该模型采用颜色、亮度和方向等特征构建高斯金字塔,利用多尺度中央周边算子计算得到各特征图,再把各特征图进行归一并融合成最后的显著图。但是,IT 模型只是对不同特征图进行简单的权值叠加,效果并不佳。Harel 等^[2]则在此基础上提出了基于图论 (Graph-based, GB) 的显著性模型即 GB 模型,此模型通过引入马尔可夫链来计算显著值,尽管计算量较大,但精度有所提高。Bruce 和 Tsotsos^[3]提出了 AIM 模型,它采用独立成分分析从图像块的自信息角度来计算注意显著值,但实验结果显示该方法过多地强调了图像边缘。Murray 等^[4]提出了 ECSF 模型,它在提取颜色、方向特征后引入中央周边机制和对比度敏感函数来分析显著性。在纯粹利用数学计算方法的模型中,Goferman 等^[7]提出的 CA 方法将图像划分成 patch 并通过计算颜色和空间距离来计算不同 patch 之间的差

① 国家自然科学基金青年科学基金 (41101386), 973 计划 (2012CB719903), 国家自然科学基金委创新研究群体 (61221003) 和国家自然科学基金 (41071256) 资助项目。

② 男, 1989 年生, 硕士; 研究方向: 计算机视觉, 模式识别; E-mail: cjkunai327@sjtu.edu.cn

③ 通讯作者, E-mail: huohong@sjtu.edu.cn
(收稿日期: 2014-09-22)

异,从而检测能够代表场景的图像区域。此方法由于采用了多尺度金字塔的方法,很容易受到尺度大小的影响,而且只适用于低分辨率图像。Zhang等^[8]提出的SUN模型对自然图像进行统计,通过贝叶斯推理,认为视觉显著性由图像的自信息和与先验信息相关的互信息这两者决定。此外,还有学者通过数据集的学习来进行视觉注意和预测注视点,如赵志城等^[9]通过最大似然估计法学习眼动数据并作为条件随机场的参数,以更好地将不同特征融合起来。朱辰阳等^[10]采用了多线索集成,通过样本的训练来得到形状线索,并用贝叶斯推理来解决显著问题和目标检测。但是,数据集的训练不仅消耗时间,不能快速地提取显著性,而且往往会由于数据集的不同而产生偏差,受到数据集选择等主观因素的影响。另外,Hou和Zhang^[11]提出的SR模型、Schauerte等^[12]提出的QB模型和Achanta等^[13]提出的FT模型,分别利用傅立叶谱残差、基于离散余弦函数的四元数法和像素点的颜色值与全局平均的差提取低频信息来进行显著性计算和感兴趣区域预测。不过,这几种频域上的方法尽管计算速度较快,但受到图像重建时系数选择的影响,容易受背景噪声所干扰。虽然纯粹利用数学计算方法的注意模型能够在一定程度上模拟视觉注意现象,但这类模型缺乏生物学方面的依据,并不是很具有说服力。

随着神经生理学、神经心理学等学科的不断发展,人们逐渐揭开了视觉系统和视觉注意的一些秘密,这也促使基于生物学原理的模型能进一步考虑视觉系统对视觉信息的层次性处理机制,并引入高层信息,从而提升对视觉注意的模拟效果。为此,本文从生物视觉角度出发,提出了一种基于贝叶斯推理的层次性视觉注意模型并对其进行了研究,试图通过模拟视觉注意的层次性处理机制^[14],来更好地预测感兴趣区域和进行目标检测。本文模型将来自视觉系统低层V1和V2区的颜色和方向信息,以及来自视觉系统高层的位置信息整合至贝叶斯网络中,通过概率推理计算区域与区域间的对比度,从而得到视觉注意显著图。同其前面提到的9种视觉注意模型的对比实验结果表明,本文提出的模型具有更好的预测能力及检测性能。

2 基于贝叶斯推理的层次性视觉注意模型

2.1 视皮层及视觉注意的层次性处理机制

生物学上的很多研究表明:大脑皮层中视觉信

息处理分为腹侧通路和背侧通路两条并行通路,其中腹侧通路(What通路)主要负责传递颜色、纹理、形状等不同信息,背侧通路(Where通路)主要传递动作信息和空间位置。视皮层中的视觉注意可以认为是一种层次性贝叶斯推理的过程,它将自底向上(bottom-up)的输入信息和自顶向下(top-down)的任务引导结合起来^[15-17]。同时,视觉注意过程涉及视皮层的多个区域,如腹侧通路的V1、V2、V4区,背侧通路的侧顶叶(lateral intraparietal, LIP)等。

视觉注意的层次性处理机制是指在不同的视皮层区域执行不同的信息处理并形成相应的显著性。比如:V1区神经元的简单细胞具有方向选择性和颜色双拮抗特性^[18],能够形成自底向上的显著图^[19],即V1区只涉及bottom-up过程。V2区的神经元不但能接收从V1区传递来的方向和对对比度等信息,而且V2区还能接受来自V4区的top-down显著性调节,通过中央周边感受野和侧抑制形成自己的显著性^[20];同时,它还能够把方向选择性形成的位置信息通过背侧通路传递给LIP等区域,从而进一步引导注意力^[21]。V4区不仅接收来自V2区的信息,还能接收来自前眼区(frontal eye field, FEF)中的记忆内容如颜色等先验知识^[22]、LIP区域传递过来的位置信息^[23]来确定视觉注意中的目标。V4区是腹侧通路中具有强烈注意调节的一个区域,它的显著性是由bottom-up和top-down共同作用产生的^[24],既有基于特征的注意调节,也有基于位置的注意调节。

总之,视觉注意的层次性处理机制表现在:不同的视觉皮层不但利用自己特定的信息处理功能来获得自己的注意显著图,而且在与之关联的皮层之间还存在信息的传递,即将本区的显著性信息传递给其他皮层,供其利用进一步生成其他皮层的注意显著图。显然,视觉注意过程分布在大脑的不同视觉皮层中,并且蕴含着一种层次性。

2.2 模型框架

基于视觉注意的层次性处理机制,本文提出了基于贝叶斯推理的层次性视觉注意模型,其框架如图1所示。主要包括与视觉注意过程相关的V1、V2、V4以及高层LIP和FEF区。首先,输入图像在V1和V2区分别经过空间变换和多通道分离获得自底向上的颜色信息,并进一步传递到V4区,作为V4区进行贝叶斯推理时所需的bottom-up证据。同时,V1区利用Gabor滤波形成的方向信息,传递到V2区形成方向对比图,再通过背侧通路传递到LIP

区域,作为 V4 区进行贝叶斯推理时所需的 top-down 位置先验信息。在 V4 区, bottom-up 证据和 top-down 先验相集成,并通过贝叶斯推理得到位置的后验概率,从而形成显著图。对于多个目标的复杂场景图,在获得显著图以后,可以通过赢者通吃(winner-take-all, WTA)神经网络^[25]进行竞争和返回抑制,实现了注意焦点(focus of attention, FOA)的定位和转移。图中虚线表示 LIP 和 FEF 到 V4 或 V4 到 V2 的 top-down 调节作用。从图 1 中可以看到,基于贝叶斯推理的视觉注意层次模型涉及视觉系统的不同视皮层,模拟了各个视皮层不同的信息处理,并通过证据、先验的传递和集成,模拟了视觉注意信息的层次性处理过程。

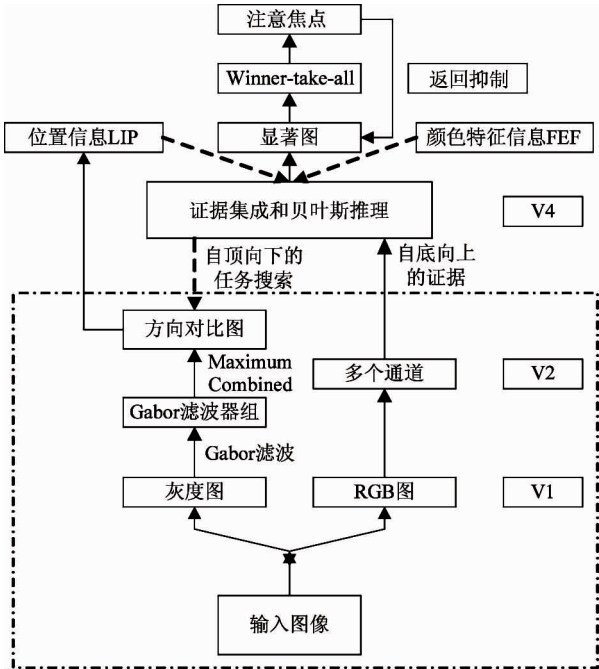


图 1 本文模型框架示意图

2.3 模型的消息传递和推理

各个视皮层完成自身特定的视觉注意加工后,彼此之间还需要进行信息的传递来进一步形成视觉注意显著图。本文采用贝叶斯推理来模拟视觉注意模型中不同皮层之间的信息传递,如图 2 所示。图中 LIP 区的每个神经元都能够对位置信息进行编码,来表示目标可能出现的位置,用 L 表示。FEF 区域中的神经元携带有与颜色相关的先验信息,用 F_c 表示。V4 区的神经元编码某个特征出现在某个指定位置上,用中间变量 C 表示。V2 区接收来自 V1 区处理的信息,并进一步将自底向上的证据输入给 V4 区。不同视皮层间神经元的消息传递在贝叶斯

推理中就是指不同层次内节点之间的信息(也称为证据)传递。图中每一个圆圈表示一个神经元,同一层之间实线和虚线(实线表示神经元与相邻神经元的连接,虚线表示神经元还能与其他多个神经元进行连接)表示该层中神经元的相互连接,带有箭头的实线表示贝叶斯推理中证据的传递方向。

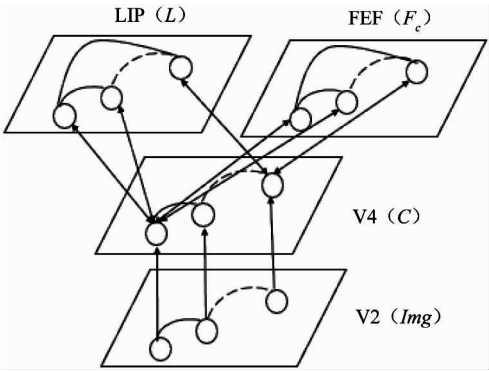


图 2 不同视皮层间消息推理示意图

贝叶斯推理采用置信传播算法^[26]来计算图 2 中各节点的置信度即信息,并实现各节点间的置信度更新和传递(其中 $\pi(\cdot)$ 和 $\lambda(\cdot)$ 分别表示下行消息和上行消息)。节点 F_c 和 L 向节点 C 传递的分别是颜色特征的先验和位置信息,其概率如下:

$$\pi_{F_c \rightarrow C} = P(F_c), \quad \pi_{L \rightarrow C} = P(L) \tag{1}$$

当新证据即图像 Img 输入时,节点 C 接收到新的上行消息为

$$\lambda_{Img \rightarrow C} = P(I = Img | C) \tag{2}$$

节点 C 根据接收的消息更新自己的置信度,然后将消息传递给 F_c 和 L 节点,上行消息 $\lambda_{C \rightarrow F_c}$ 和 $\lambda_{C \rightarrow L}$ 为

$$\begin{aligned} \lambda_{C \rightarrow F_c} &= \sum_L \sum_C P(C | L, F_c) \pi_{L \rightarrow C} \lambda_{Img \rightarrow C} \\ &= \sum_L \sum_C P(C | L, F_c) P(L) P(I = Img | C) \end{aligned} \tag{3}$$

$$\begin{aligned} \lambda_{C \rightarrow L} &= \sum_{F_c} \sum_C P(C | L, F_c) \pi_{F_c \rightarrow C} \lambda_{Img \rightarrow C} \\ &= \sum_{F_c} \sum_C P(C | L, F_c) P(F_c) P(I = Img | C) \end{aligned} \tag{4}$$

节点 F_c 在接收到上行消息 $\lambda_{C \rightarrow F_c}$ 后,更新自己的置信度为

$$\begin{aligned} P(F_c | I = Img) &= \frac{1}{\alpha} \lambda_{C \rightarrow F_c} P(F_c) \\ &= \frac{1}{\alpha} \sum_L \sum_C P(C | L, F_c) P(L) \\ &\quad P(I = Img | C) P(F_c) \end{aligned} \tag{5}$$

节点 L 在接收到上行消息 $\lambda_{c \rightarrow L}$ 后,更新自己的置信度为:

$$\begin{aligned} P(L | I = \text{Img}) &= \frac{1}{\beta} \lambda_{c \rightarrow L} P(L) \\ &= \frac{1}{\beta} \sum_{F_c} \sum_C P(C | L, F_c) P(F_c) \\ &\quad P(I = \text{Img} | C) P(L) \quad (6) \end{aligned}$$

其中 α, β 是保证上述式子概率和为 1 的常数。式(2)中条件概率 $P(\text{Img} | C)$ 正比于输入图像的颜色特征值。对于条件概率 $P(C | L, F_c)$ 的赋值,当颜色特征 F_c 存在并出现在位置 L_0 时,表示目标最有可能出现在位置附近,此时将变量 C 设为以位置为中心的二维正态分布;相反,当颜色特征不存在时,特征图中只有在 $C = 0$ 的地方被激活。具体赋值如表 1 所示 ($\delta, \varepsilon \approx 0.001$, 使得 ($\sum P(C | F_c, L) = 1$))。

表 1 $P(C | F_c, L)$ 的条件概率赋值

	$F_c = 1, L = L_0$	$F_c = 0, L = L_0$
$C = 0$	$P(C F_c, L) = \delta$	$P(C F_c, L) = 1 - \varepsilon$
$C \neq 0$	$P(C F_c, L) \sim N(L_0, \sigma)$	$P(C F_c, L) = \varepsilon$

2.4 模型的证据与先验

这里主要介绍基于贝叶斯推理的层次性视觉注意模型的自底向上的证据与先验概率的获取。

2.4.1 来自 What 通路的颜色特征

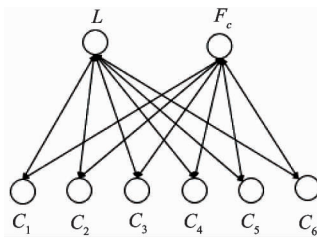
要获得来自 What 通路的颜色特征,需要先将 RGB 颜色空间转换成符合人类视觉系统的 Lab 空间,其中 L 的值域为 $[0, 100]$, a 和 b 的值域均为 $[-128, 127]$ 。先把 L, a, b 归一化,即

$$L_n = L/100, a_n = a/128, b_n = b/128 \quad (7)$$

在此基础上,结合颜色双拮抗理论(double-opponent theory)^[18],将颜色近似划分成 6 个通道,即

$$\begin{aligned} \text{Black} &= 1/2 - L_n, \text{White} = L_n - 1/2 \\ \text{Green} &= 0 - a_n, \text{Red} = a_n - 0 \\ \text{Blue} &= 0 - b_n, \text{Yellow} = b_n - 0 \end{aligned} \quad (8)$$

由于颜色间的互斥性,当一个位置上的目标被认为是红色时,它成为其他颜色的概率值就会相对降低。因此,6 个颜色通道可以认为是条件独立的,并把它作为贝叶斯网络中的 6 个节点,先验概率设为等概率,再将其与位置信息相关联,构成如图 3 所示的贝叶斯网络用于生成显著图。



C_1 到 C_6 分别代表 V4 区编码颜色和位置信息的中间变量

图 3 用于生成显著图的贝叶斯网络

2.4.2 来自 Where 通路的方向证据及位置偏置

在初级视觉皮层中,一般用 Gabor 滤波器来描述 V1 区简单细胞的方向选择性:

$$\begin{aligned} G_{\lambda, \theta, \psi, \gamma}(x, y) &= \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right), \\ &\text{s. t.} \end{aligned} \quad (9)$$

$$x' = x \cos \theta + y \sin \theta, y' = -x \sin \theta + y \cos \theta \quad (10)$$

其中, (x, y) 表示坐标, λ 是波长, σ 是标准差, γ 是空间纵横比, θ 表示方向, ψ 表示相位偏移。本文模型主要选取了 8 个方向 $\theta = \{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2},$

$\frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}\}$, 这样 V1 区中每个神经元就有 8 个方向,将其中最大的值作为该神经元的最优方向。归一化后作为 V2 区的方向对比显著图,传至 LIP 区后形成视皮层中 Where 通路的位置信息。

最初每个位置的先验概率 $P(L)$ 都相等。在方向信息传至 LIP 之后,位置 L 的先验概率更新为

$$P(L) = \frac{P(L) + P(F_o)}{\sum (P(L) + P(F_o))} = \frac{P(L) + P(F_o)}{1 + \sum P(F_o)} \quad (11)$$

$$P(F_o) = \frac{L(x, y)}{\sum L(x, y)} \quad (12)$$

其中, $L(x, y)$ 为该位置的方向强度值, $P(F_o)$ 为该位置的方向强度概率,正比于该位置的方向显著性。

在进行感兴趣区域预测时,方向证据直接作为 LIP 区的位置偏置,并不需要加入自顶向下的方向调节。当对遥感影像进行目标检测时,由于飞机等刚性物体方向强度高,即方向强度较大的地方就是疑似目标的位置。本文把这一过程当作 V2 区受到 V4 的方向注意引导^[24],即

$$L(x, y) = \begin{cases} L(x, y), & L(x, y) > T \times L_{\text{mean}} \\ 0, & \text{其他} \end{cases} \quad (13)$$

其中, L_{mean} 表示平均方向强度, T 为阈值参数。

上述的证据和先验在贝叶斯框架中集成,共同用于计算关于位置的后验概率,进而生成视觉注意的显著图。

2.5 模型的显著图

获取模型的显著图 $\hat{S}(x)$ 的过程就是计算关于 LIP 区位置的后验概率,可以由式(6)得到。在贝叶斯推理中,图像被划分成互不重叠的图像块,且每个图像块都关联上颜色特征和位置信息。将显著图 $\hat{S}(x)$ 经高斯滤波器 $G_\sigma(x)$ 平滑得到最后的显著图:

$$S(x) = G_\sigma(x) \times \hat{S}(x) \tag{14}$$

当某区域的颜色、纹理等对比度较高时,该区域就会吸引人眼的注意。本文通过颜色和方向的层次性处理,利用不同颜色间的竞争,并把方向作为轮廓和疑似目标的感知,用贝叶斯概率推理实现了区域与区域间的对比。当某一区域目标位置的后验概率较大时,就表示该区域对比度较高,其他区域的对比度就会相对降低,那么该区域就更为显著,更加会吸引人眼的注意。

3 实验分析和对比

为了验证本文模型的有效性,主要进行了两个实验。第一个实验采用了公开的 Bruce 数据集^[3],对感兴趣区域进行预测;第二个实验采用了遥感影像数据集,用来评价本文模型的显著性性能及在目标检测时的性能。两个实验均与其他注意模型进行了对比和分析。

3.1 实验参数设置及环境

实验中,先将原图 ($w \times h$) 进行双线性插值至

$(\max(w, h) \times \max(w, h))$,图像块的大小划为 $K = 3\% \times \max(w, h)$; Gabor 滤波器的各个参数: $\lambda = 6, \sigma = 1, \gamma = 0.5, \psi = -2/\pi$ 。高斯滤波器的宽度为 $\sigma = K/4$, 条件概率赋值时以位置为中心的正态分布的标准差为 $\sigma = 1$; 阈值 $T = 4$ 。实验硬件环境: CPU: Intel Xeon X5690 3.47GHz; 内存: 32GB; 操作系统: 64 位 Windows Server2008; 开发平台 MATLAB R2013a。

3.2 自然场景的显著图对比实验

显著图对比实验采用 Bruce 数据集^[3], 该数据集包含 120 张大小为 681×511 的自然场景图, 其中每一张图都有相应的眼动跟踪数据 (eye tracking data), 分别由 20 个实验者在无任务情况下自由浏览得到。这些数据较客观地反映了人类对各种场景中兴趣点的感知与定位。本文将这些数据形成的注视图 (fixation map) 作为评估不同视觉注意模型的标准。图 4 给出了 Bruce 数据集的部分图像及其 fixation map。

这里主要选取了 9 种模型同本文提出的基于贝叶斯推理的层次性视觉注意模型进行对比, 包括 IT^[1]、GB^[2]、AIM^[3]、ECSF^[4]、CA^[7]、SUN^[8]、SR^[11]、QB^[12] 和 FT^[13]。图 5 显示了不同模型的显著图对比情况。在得到各自的显著图后, 绘制受试者操作特性 (ROC) 曲线并计算 ROC 曲线之下面积 (AUC) 值, 以更直观地显示 10 种模型在该数据集上的性能, 如图 6 和图 7 所示。从定性和定量的实验结果可以看出, 本文模型在预测感兴趣区域时对背景纹理不太敏感, 跟人类注视图最为接近, AUC 值也比其他模型高, 具有最佳的视觉注意模拟效果, 而与本文模型最接近的是 GB 模型。

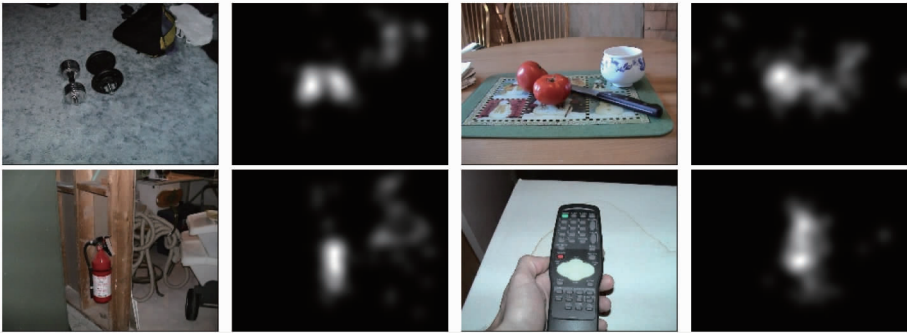
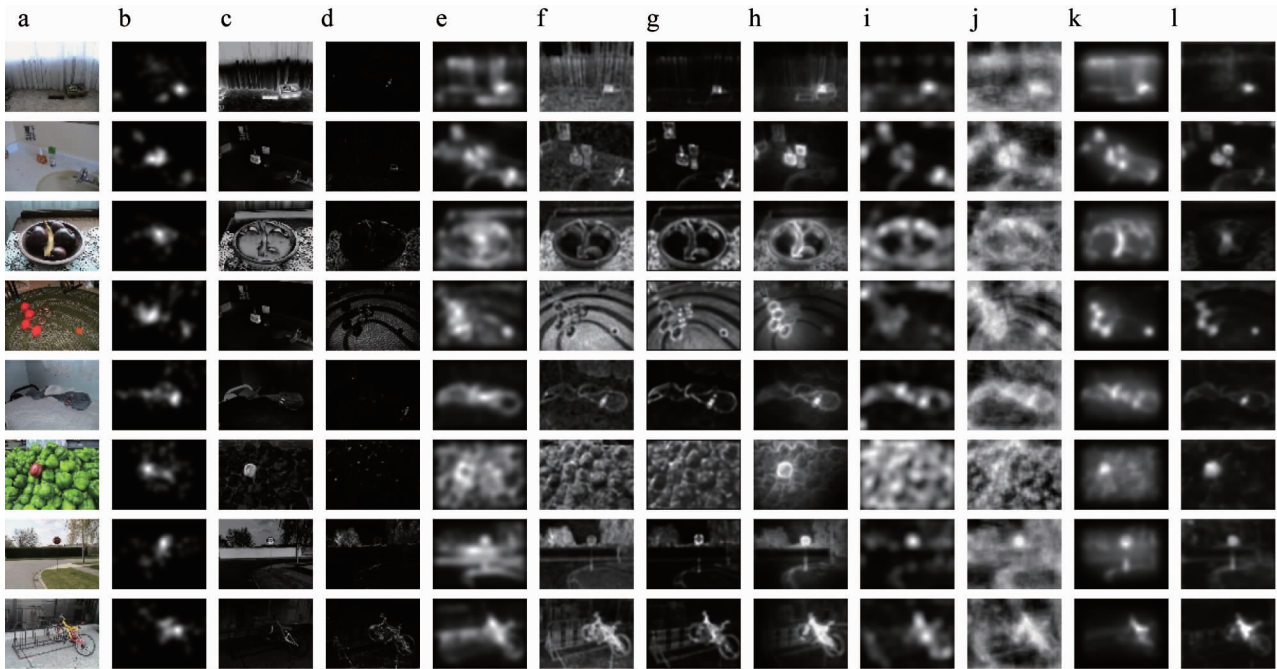


图 4 Bruce 数据集的部分图像及其 Fixation map^[3]



a: 原图; b: Fixation map; c-k: FT, SR, IT, SUN, AIM, CA, QB, ECSF, GB 模型; l: 本文模型

图 5 10 种模型在 Bruce 数据集上得到的显著图对比结果

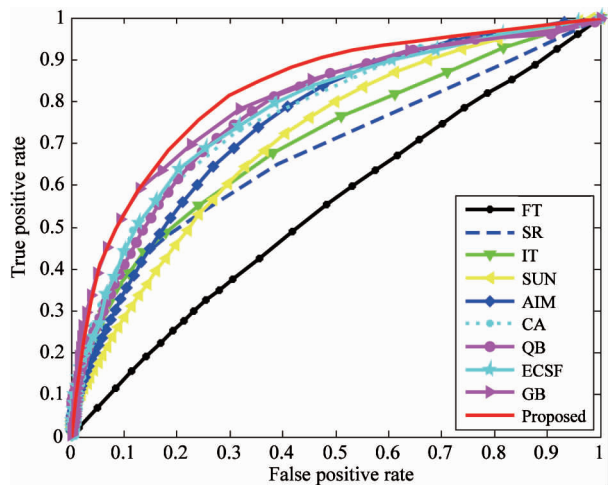


图 6 10 种模型的 ROC 曲线

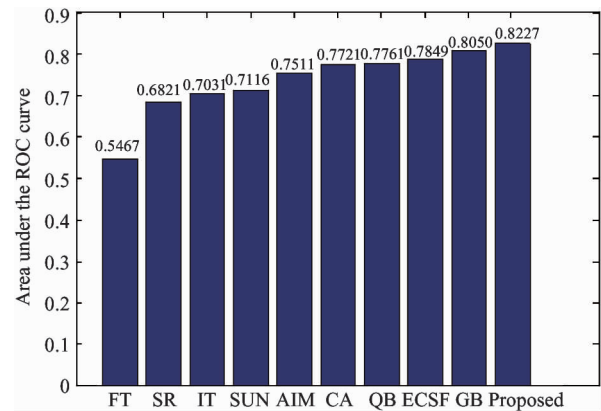


图 7 10 种模型的 AUC 值

3.3 遥感影像的显著图评价及目标检测

实验采用的遥感影像集是从 Google Earth 上手工获取的,一共有 100 幅影像,人工标注了每一幅影像的飞机目标所在矩形区域,作为 ground-truth 图。其中这些影像中目标飞机的方向、形状、大小及其出现在影像上的位置都不尽相同,有的影像背景不复杂,仅仅包含跑道上的飞机。有的影像复杂些,除了飞机外,还包含机场建筑物和自然植被等。

3.3.1 遥感影像集上的显著图评价

遥感影像显著图评价实验主要对比了本文模型和性能与本文接近的 GB 模型^[2]。根据文献[27]提出的评价指标,计算了本文模型和 GB 模型与 ground-truth 图的相似度 (similarity),线性相关系数 (linear correlation coefficient, LCC) 和 KL 距离等定量参数,如表 2 所示。其中 AUC、Similarity 和 LCC 的值越大表示显著图与 ground-truth 图越相近,KL 的值越小表示显著图与 ground-truth 图的概率分布

表 2 本文模型与 GB 模型在几种评价指标上的性能比较

评价指标	本文模型	GB 模型
AUC	0.9488	0.8965
Similarity	0.2029	0.1751
LCC	0.5304	0.4715
KL	1.9473	2.1180

距离越小,即越相近。从这几个指标可以看出,本文模型在遥感影像集上也优于 GB 模型。

3.3.2 遥感影像集的目标检测

由于遥感影像自身的多目标和复杂性,显著图并不能很直观地衡量目标间与背景间的显著关系。在这里,利用(赢者通吃 WTA)神经网络^[25]在显著图提取注意焦点(FOA)作为目标检测的结果,进一步探讨本文模型在目标检测方面的应用。由于篇幅所限,图 8 显示了本文模型与 GB 模型的部分实验结果图及对比情况,其中圆圈的中心代表 FOA,直线表示焦点的转移。影像(a)包含了 5 个 FOA,由于背景较为平滑和简单,两个模型都能够正确检测到飞机。影像(b)包含 4 个 FOA,影像(c)包含 5 个 FOA,影像(d)和(e)包含了 7 个 FOA,对于这 4 组

影像,两个模型都能检测到最显著的飞机目标区域,但经过 WTA 网络的反馈抑制后,次显著的区域出现了差异。本文模型仍然能够将其他飞机检测出来,实现注意的转移,而 GB 模型却出现了检测误差,把一些建筑物和草地等背景当作了显著的目标。两个模型检测的具体情况如表 3 所示,其中“/”前为模型检测到的目标个数,“/”后为影像中实际包含的目标个数。对 100 幅遥感影像集的检测结果如下:本文模型从 319 架形状、大小各异的飞机中,利用 WTA 选取 FOA 的方法,一共检测到了 296 架,检测率为 92.79%;而 GB 模型检测到 268 架,检测率为 84.01%。这说明本文模型加入自顶向下的注意引导后,能够在较为复杂的遥感影像中取得更好的目标检测性能。

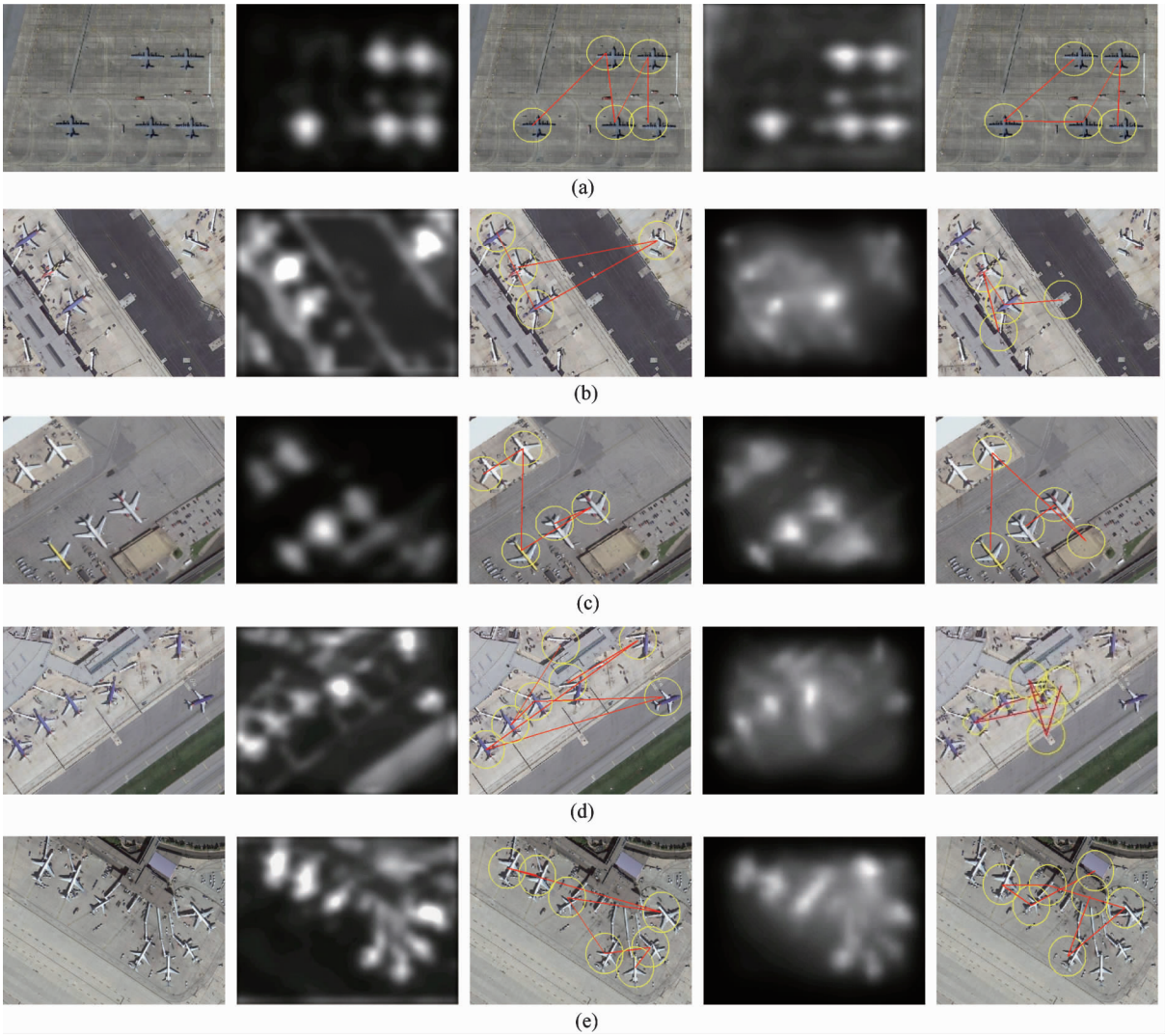


图 8 从左至右分别是原始遥感影像,本文模型的显著图,本文模型注视焦点图,GB 模型的显著图及其注视焦点图。

表 3 本文模型与 GB 模型的飞机检测数对比

图像	本文模型	GB 模型
(a)	5/5	5/5
(b)	4/4	2/4
(c)	5/5	3/5
(d)	7/7	2/7
(e)	7/7	4/7

4 结 论

本文从神经生物学角度进行了视觉注意研究,结合人类视觉系统采用的层次性的信息处理机制,提出了一种基于贝叶斯推理的层次性视觉注意模型。该模型通过模拟视皮层信息的层次性处理机制来提取颜色和方向信息,并构建贝叶斯网络,利用置信传播来预测和获取图像的显著区域。在自然图像集上与其他 9 种具有代表性的模型进行的对比实验表明,本文模型的显著图更接近于人类注视图,能更好地模拟视觉注意。在遥感影像数据集的实验也表明本文模型有较好的检测结果。为了更好地进行目标检测,本文模型有待集成更多的视觉特征如纹理等。将来的工作还包括研究大脑皮层中尚未了解的注意机制来更好地模拟视觉选择性注意,以及将所提出的方法用于目标检索和识别。

参考文献

[1] 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

[2] Harel J, Koch C, Perona P. Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 2007, 19:545-552

[3] Bruce N D B, Tsotsos J K. Saliency based on information maximization. *Advances in Neural Information Processing Systems*, 2006, 18: 155-162

[4] Murray N, Vanrell M, Otazu X, et al. Saliency estimation using a non-parametric low-level vision model. In: *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition*, Providence, USA, 2011. 433-440

[5] Einhauser W, Konig P. Does luminance-contrast contribute to saliency map for overt visual attention? *Journal of Neuroscience*, 2003,17(5):1089-1097

[6] Grossberg S, Raizada R D. Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits

of primary visual cortex. *Vision research*, 2000,40 (10-12):1413-1432

[7] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(10): 1915-1926

[8] Zhang L, Tong M H, Marks T K, et al. SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision*, 2008, 8(7):1-20

[9] 赵志城,白雅娟,周仁来等. 基于 Tobii 眼动和多尺度条件随机场的视觉显著性模型. *高技术通讯*,2013,23 (3):263-269

[10] 朱辰阳,霍宏,方涛. 基于贝叶斯推理的多线索视觉注意模型. *计算机工程与设计*,2013,34(7):2470-2475

[11] Hou X, Zhang L. Saliency detection: a spectral residual approach. In: *Proceedings of the 20th Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, 2007. 1-8

[12] Schauerte B, Stiefelhagen R. Predicting human gaze using quaternion dct image signature saliency and face detection. In: *Proceedings of the 12th IEEE Workshop on the Applications of Computer Vision*. Breckenridge, 2012, 9-11

[13] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, 2009. 1597-1604

[14] Kastner S, Pinsk M A. Visual attention as a multilevel selection process. *Cognitive, Affective, & Behavioral Neuroscience*, 2004, 4(4): 483-500

[15] Lee T S, Mumford D. Hierarchical Bayesian inference in the visual cortex. *Journal of Optical Society of America*, 2003, 20(7): 1434-1448

[16] Chikkerur S, Serre T, Tan C, et al. What and where: A Bayesian inference theory of attention. *Vision research*, 2010, 50(22): 2233-2247

[17] Rao R P N. Bayesian inference and attentional modulation in the visual cortex. *Neuroreport*, 2005, 16(16): 1843-1848

[18] Chatterjee S, Callaway E M. Parallel colour-opponent pathways to primary visual cortex. *Nature*, 2003, 426 (6967): 668-671

[19] Koene A R, Li Z P. Feature-specific interactions in saliency from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, 2007, 7 (7):1-14

[20] VanRullen R. Visual saliency and spike timing in the ventral visual pathway. *Journal of Physiology*, 2003, 97 (2): 365-377

- [21] Deco G, Lee T S. A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing*, 2002, 44: 775-781
- [22] Zhou H, Desimone R. Feature-Based Attention in the Frontal Eye Field and AreaV4 during Visual Search. *Neuron*, 2011, 70 (6): 1205-1217
- [23] McAdams C J, Maunsell J H R. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*. 1999, 19(1): 431-441
- [24] Melloni L, Leeuwen S, Alink A, et al. Interaction between bottom-up saliency and top-down control: how saliency maps are created in the human brain. *Cerebral cortex*, 2012, 22(12): 2943-2952
- [25] Walther D B. Saliency Toolbox. <http://www.saliency-toolbox.net/index.html>; 2013
- [26] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers. 1988. 275-279
- [27] Riche N, Duvinage M, Mancas, et al. Saliency and Human Fixations: State-of-the-Art and Study of Comparison Metrics. In: Proceedings of International Conference on Computer Vision, Sydney, Australia, 2013. 1153-1160

A hierarchical visual attention model based on Bayesian inference

Cen Junkai^{***}, Huo Hong^{***}, Fang Tao^{***}

(^{*} Department of Automation, Shanghai Jiao Tong University, Shanghai 200240)

(^{**} Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai 200240)

Abstract

Inspired by the human visual system' hierarchical processing mechanism for prediction of interesting regions and quick detection of targets, a novel hierarchical visual attention model based on Bayesian inference is proposed. The model utilizes color feature information from the ventral pathway as the bottom-up inference cue, and uses orientation location information from the dorsal pathway as the top-down cues. All the cues are integrated in a Bayesian framework, and saliency maps are obtained by inference. The natural image experiments were conducted to compare the performance of the proposed model with that of the other nine state-of-the-art visual attention models IT, GB, AIM, ECSF, CA, SUN, SR, QB and FT, and the results show that the saliency maps created by the proposed model are much closer to the eye fixation data, and it can better simulate visual attention. Moreover, the results of the experiments on remote sensing dataset demonstrate that the proposed model can better detect targets compared to the GB model with the similar performance.

Key words: visual attention, saliency map, Bayesian inference, fixation maps, object detection, hierarchical