**先验知识的学习与其在Saliency估计中的应用**

**A Priori Knowledge Learning and Its Applications in Visual Saliency Estimation**

**摘要：**

人类通过视觉系统能够在复杂的场景中快速搜索到自己感兴趣的目标。模拟人类视觉系统得到图像中的显著区域，即显著性区域检测，已经成为计算机视觉领域的热点之一。从本质上说，合理地构建视觉注意模型成为显著性区域检测的关键。以往神经心理学的大量研究表明，视觉注意机制的构建一直以来可以通过两种不同的途径：自下而上的方法以及自上而下的方法。自上而下的方法需要待完成任务的先验知识，而自下而上的方法则是由数据驱动，并且与任务无关。

现在由数据驱动的自下而上的显著性检测方法研究较多，也比较成熟和深入，故本文工作主要关注显著性区域检测中的自上而下的视觉注意模型。现在国际上关于图像检测，识别及分割的人越来越多，且提供的图像数据库也逐渐丰富起来，如ImageNet，iSUN，SALICON，MSCOCO等都是比较大型的图像数据库，这些数据库同时提供了图像的一些特定信息，如MSCOCO等就给出了图片中一些物体的轮廓的等重要信息。本文拟首先从这些大型数据库中提取有用的先验知识，通过深度学习的方式来抽象得到图像的目标概念的描述，之后将这些概念语义加入到显著性估计中去。关于显著性估计，我们可以采用自动编码器等方式来实现。结合概念语义，自动编码器实现的显著性检测结构可以利用到获取的先验知识，提高显著性检测的准确率。

更进一步的，由于图像中有目标轮廓的标记，我们可以进一步标记出目标内部与非目标区域，从而可以学习二类分类器，区分是否有意义的目标。这个可以应用于图像中的任意边线，即给定边缘，可以判断是边缘的哪一侧为目标区域。将这一含有目标信息的先验知识加入到显著性检测中去，进一步提高显著性检测的准确率。

**选题意义：**

近年来，随着现代多媒体技术和互联网的快速发展，越来越多的数字图像被获取、传输、处理和利用，仅仅利用人工的方式对图像进行分析存在很大的局限性，机器智能因此成为数据处理的主要手段。在图像的处理分析工作中，关注的往往是待处理图像中很小的一部分，所以，将计算资源优先分配给图像最显著的区域，将极大地提高图像分析、处理的效率，这就涉及到图像显著性区域检测问题。显著性区域检测的本质是模拟人类视觉系统建立的一种视觉注意模型。该模型利用视觉注意机制得到图像中最容易引起注意的区域，并用响应强度作为显著目标的显著性度量。

人类与生俱来能够在复杂的场景中快速搜索到自己感兴趣的目标，这种选择能力被称为选择注意机制，而机器视觉的目的就是模拟人的视觉神经系统，从而使得机器同样具有智能的视觉注意、理解能力，实现机器智能数据的处理。模拟人类的视觉系统进行显著性检测是目前计算机视觉领域的热点研究方向之一，其在智能机器人、智能工厂、智能武器以及高性能脑模型，图像分割和恢复、图像及视频压缩、目标检测和跟踪，等领域具有广泛的应用价值[[[1]](#endnote-1)]。

**国内外研究现状：**

目前，国内外的许多研究机构对图像显著性区域检测做了广泛的研究，并取得了一定的成果。显著区域检测的本质是一种视觉注意模型。该模型利用视觉注意机制得到图像中最容易引起注意的区域，同时用灰度图像来作为该图像的显著性度量。人类视觉系统能够很快的从复杂的场景中检测到感兴趣的目标，而如何建立机器视觉注意模型来模拟人的视觉系统在图像处理领域得到极大关注。随着认知心理学和神经生物学的发展，人们对视觉注意机制以及视觉神经系统有了更深入的了解和认识，奠定了构建视觉注意机制模型的理论基础。

在复杂的场景中，人类视觉系统釆用串行计算，快速的将注意力转移到自己感兴趣的一个或者几个目标身上，优先处理这些目标区域，这就是视觉注意过程。视觉注意机制在人类的视觉系统中起到重要的心理调节作用，属于视觉感知模型中的一部分，协同记忆等模块完成各项任务，如目标匹配、目标分离以及注意焦点转移。从信息角度来说，视觉神经系统处理信息的资源以及能力都是有限的。所以，人类通过这种选择注意机制对海量的视觉信息进行处理，从而快速地进行分辨、筛选出重要的信息。

模拟人类视觉系统的显著性区域检测计算模型主要从两个角度出发：（1）自下而上（Bottom-up）的视觉分析计算模型；2)自上而下(Top-down)的视觉选择注意模型。其中，前者基于低级视觉特征，即没有任何的先验知识，由数据驱动，自动捕获刺激人眼的区域，而后者则基于高级视觉特征，即先验知识的学习，由知识和任务驱动，以自我意识决定视觉关注区域。研究表明，在人类视觉注意的具体过程中，自下而上和自上而下的选择注意方式通常是共同协作、相互影响的。下面对这两种视觉分析模型进行简要介绍：

1）自下而上（Bottom-up）的视觉分析计算模型

许多研究人员试图通过描绘观察者感兴趣的区域和一些基本特征（如边缘和局部对比性）之间的关系来对显著性区域进行解释。其中，最具代表性的是Itti[[[2]](#endnote-2)]等人在年提出的视觉注意模型。该模型基于特征集成理论[[[3]](#endnote-3)]，采用高斯金字塔并综合考虑亮度、颜色、方向视觉特征，通过算子得到多尺度的显著性度量，最后经过合并、归一化得到最终的显著图像；在Itti工作的基础上，Walther等人[[[4]](#endnote-4)]将其进行扩展，成功应用于目标识别，还建立了显著性检测网站并提供显著性测量工具箱；Harel等人[[[5]](#endnote-5)]提出基于图的视觉检测显著模型来突出显著区域并且可以结合其他图形；Ma和Zhang等人[[[6]](#endnote-6)]提出利用颜色特征对比分析，并提出利用模糊增长算法提取图像的显著性区域。Gao和Mahadevan等人[[[7]](#endnote-7)]提出了基于中心-周围判别理论的显著性检测模型，简称为DICS模型。Chen Xia等人[[[8]](#endnote-8)]提出一种基于非局部重构的显著性区域检测方法，并通过利用非局部其余块对当前中心区域的重构误差来估计显著性。同时Chen xia等人[[[9]](#endnote-9)]在前篇文章的基础上又提出了一种基于深层自动编码器重构的图像视觉显著性区域检测方法，以从全局的角度来挖掘无标记图像数据与显著性值之间的关系，提高图像中显著性区域检测的准确性。

2) 自上而下(Top-down)的视觉分析计算模型

自上而下的视觉分析计算模型的基本思想是从场景中提取出多种基本特征，随后将其集成作为显著图像的表示。受可视化程序的启发，Sprague和Ballard提出了基于“加强学习”的自上而下显著检测模型[[[10]](#endnote-10)]，显示观察者在虚拟环境中的眼球运动；Kanan等人[[[11]](#endnote-11)]提出利用自认数据估计目标在每个位置概率的算法；Liu等人[[[12]](#endnote-12)]提出一种具体的学习机制，其通过条件随机场（自动学习出对各低层特征（包括局部特征、全局特征、区域特征）计算出的显著度权重，从而依据权重计算出图像各个位置总体的显著程度，并且依此进行显著物体的提取。为了更加协调的将多种特征结合到显著性检测算法当中，Lang等人[[[13]](#endnote-13)]将显著性区域检测认为是一个多任务学习的问题，并提出了多任务稀疏追踪算法。

**研究目标：**

期望使用深度学习算法，从有标记图像数据库中学习先验知识，用于提升显著性估计算法的性能。

1. 用已有的深度学习库（cuDNN,caffe）来训练深度网络结构（cnn,auto-encoder）参数。

2，提取现有的大规模图像数据库中图像的objectness特征，objectness是 Alexe 等人提出，表示一个像素或区域包含完整目标的可能性。一个图像中所含目标通常包含下面特点：1)包含一个空间闭合的边界；2)和周围事物有不同的形态；3)有时是独特、突出的。人眼倾向于完整的去识别一个目标是否显著，因此评估每个区域是否属于可识别区域，将这些区域作为先验知识，在显著性检测的时候，可以将这些区域的显著性系数增大，结合深度学习得到的显著图来确定最终的显著图。从而提高显著性检测的准确性。

3，采用自上而下(Top-down) 视觉分析计算模型，来实现显著性的检测。

**研究内容：**

采用自上而下(Top-down) 视觉分析计算模型，基于现有规模图像数据库ImageNet，iSUN，SALICON，MSCOCO等，从现有图像中学习先验知识，进而提高显著性检测的准确性。

目标的抽象（语义）概念对于显著性估计非常重要，由于图像库MSCOCO标记出了目标的轮廓，我们可以先通过深度学习获得目标概念的描述（CALVIN），然后将概念语义加入到显著性估计的过程中，从而提升性能和准确性。

更进一步的，由于图像中有目标轮廓的标记，我们可以进一步标记出目标内部与非目标区域，从而可以学习二类分类器，区分是否有意义的目标。这个可以应用于图像中的任意边线，即给定边缘，可以判断是边缘的哪一侧为目标区域。

**拟解决的关键问题：**

1. 从现有图像库中提取对显著性检测有意义的先验知识；
2. 利用提取的先验知识，增强已有的显著性检测方法，提高检测准确率；
3. 选择好的网络结构来实现显著性检测；

**拟采取的研究方法、技术路线、实验方案及可行性研究：**

研究方法：

本文先通过文献法，调研现有显著性的研究现状，并清楚的了解自上而下和自下而上两种显著性区域检测计算模型，然后确定自己所采用的模型结构；然后通过实验法，根据已定的结构来代码实现，将提取得到的objectness先验知识结合到显著性检测的深度结构中去，实验分析显著效果如何，并根据实验结果来调节网络结构参数；最后是总结法，根据结果分析objectness先验知识对深度结构用来检验显著性性能的提高，给出最终的检测方案。

技术路线：

本文通过对显著性检测方法的调研及分析，对比比较了自上而下和自下而上两种显著性区域检测计算模型，提出了一种结合先验知识与显著性检测相结合的方案来提高显著性检测的准确性。首先，我们采用深度学习的方法从大型的图像数据库中获取objectness特征，objectness表示一个像素或区域包含完整目标的可能性。一个图像中所含目标通常包含下面特点：1)包含一个空间闭合的边界；2)和周围事物有不同的形态；3)有时是独特、突出的。人眼倾向于完整的去识别一个目标是否显著，因此评估每个区域是否属于可识别区域，将这些区域作为先验知识；然后，我们设计一个深度网络的显著性检测结构，该结构可以采用卷积神经网络或者自动编码器等效果比较好的结构，利用该结构训练得到一个初步的显著图；最后，我们将结合先验知识得到的objectness特征，更新该显著图，更新方式可以为设置不同的权重，或者其他的方式。或者我们可以把objectness特征等先验知识直接加入到深度网络结构的训练中去，即训练网络的数据不是直接用数据库中的数据，而是先把先验知识加入到训练数据中，再用带有先验知识的数据来训练显著性深度网络结构。

实验方案：

1. 提取先验知识

在现有的大型图像数据库中提取objectness特征，图像库有ImageNet，iSUN，SALICON，MSCOCO，目前拟采用MSCOCO数据库。objectness表示一个像素或区域包含完整目标的可能性。一个图像中所含目标通常包含下面特点：1)包含一个空间闭合的边界；2)和周围事物有不同的形态；3)有时是独特、突出的，这些特点和显著的概念大体上一致。考虑到要保持目标的区域完整性，这里可以使用 mean-shift 分割算法分割图像区域保持每个目标的 objectness 特征具有一致性。其次计算 objectness，过程分为两步：评估像素级 objectness 和区域级 objectness。这样之后得到整个图像的objectness图。上述方式只是提取objectness的一种方式，我们会根据结果的好坏适当调整该方法。

1. 训练深度网络

由于目前卷积神经网络在图像处理方面的效果特别好，同时自动编码器也有很好的表现，拟从这两种网络结构中选取一种比较好的结构，在图像数据库上训练，得到一种能够很好表示图像显著性的结构及参数。

1. 先验知识与深度网络结构的融合

结合步骤（1）及步骤（2），来提高整个图像的显著性效果，最简单的结合方式是将步骤（1）中的objectness性按照一定的权重加入到步骤（2）的显著图中。另外也可以将步骤（1）中得到的objectness性直接加入到训练图像中去，将步骤（2）的训练作为预训练，根据加上先验知识的数据，最后再整个训练网络。

可行性研究：

目前直接利用自动编码器等深度结构网络来做显著性已取得了很多好的结果，同时图像的objectness性也有很多相关的论文及方法，这个也是可以得到了。那么本文中将objectness性与深度结构网络相结合，提高图像的显著性检测，只需要提出一种融合的结合方式，这个在我们实验及尝试的基础上，是可以实现的。

**研究计划及预期取得的研究成果：**

研究计划：

1. 调研：调研现有的显著性检测的方案，包括自上而下及自下而上两种模型的各种方案，其中主要放在自上而下的模型调研上；
2. 模型比较：根据调研的结果，比较效果突出的显著性检测模型，与我们文章中的研究思路做对比，优化自己提出的模型；
3. 先验知识获得：根据已定的结构模型，从现有的大型有标记数据库中提出想要的先验知识；
4. 训练模型：将已获得的先验知识结合已有的显著性结构，训练新的检测结构；
5. 结构优化：根据检测结果与已有的结构检测结果做对比，适当调整自己的结构，让其检测准确性更高。

研究结果：

期望能够发一篇文章

**已具备的实验条件和研究工作积累：**

1. 具有深度学习的基础，用深度学习方法（卷积神经网络，自动编码器）做过图像及声音方面的运算及研究。
2. 协助师姐写过显著性检测方面的专利，对于显著性检测的目的，一些基本方式及显著性结果表示都有一定的了解。
3. 师姐及导师发过显著性检测方面的文章，可以在研究期间给予指导和帮助。

**已取得的研究成果：**

1. [↑](#endnote-ref-1)
2. [↑](#endnote-ref-2)
3. [↑](#endnote-ref-3)
4. [↑](#endnote-ref-4)
5. [↑](#endnote-ref-5)
6. [↑](#endnote-ref-6)
7. [↑](#endnote-ref-7)
8. [↑](#endnote-ref-8)
9. [↑](#endnote-ref-9)
10. [↑](#endnote-ref-10)
11. [↑](#endnote-ref-11)
12. [↑](#endnote-ref-12)
13. [↑](#endnote-ref-13)