# 第一章：绪论

## 1.1 研究背景、目的及意义

### 1.1.1 研究背景

近年来，随着信息技术和网络的大力发展，人们对计算机的应用越来越广泛，上至太空旅行，下至买菜做饭，计算机已经渗入到了我们生活的方方面面，然而在这样的信息爆炸的时代，人们都想用最短的时间做尽可能多的事情，计算机给我们带来生活便利的同时也让人们在无尽的信息中苦苦挣扎，因此如何用更少的数据表达更丰富的意义对人们至关重要。在海量数据中提取重要信息，忽略其它辅助性的次要信息成为很多计算机相关领域研究的热点问题。

人们能够在大量不相关的复杂信息的干扰下快速地找到显著性的物体或者感兴趣的目标，是由于人类和其他灵长目动物具有视觉注意机制、选择性注意机制，利用图像的局部特征，首先选定环境景象中某个特定的区域，然后通过眼睛运动扫描，将该区域置于具有更高分辨率的视网膜的中央四区，人们此时将注意力转移到该区域，对该区域的目标或者物体进行更深层次的分析和处理。

经学者们研究表明，人类视觉的选择性注意机制大致分为两个阶段：（自底向上的基于视觉场景中显著度计算的、快速的预注意阶段，这是一种低级的认知过程。（自顶向下的基于调整选择准则来适应外界命令要求的、慢速的注意阶段，这是一种高级的认知过程。人们希望用计算机来模拟这种类似用人眼观察周围场景的活动，因而产生了计算机视觉这门学科。计算机视觉在各个领域都给人们带来了很大的方便的同时也取得了很大的发展，但对于模拟人类视觉机制上仍然存在很大的障碍，其主要原因在于人类本身对于自己在生理和心理上感知周围世界的物体的机制仍然没有一个清晰的轮廓，这也是制约着该领域发展的最大障碍。

在计算机软硬件突飞猛进的今天，提取图像中感兴趣的区域已经应用到了

生活中的方方面面，例如在公安机关追踪肇事车辆逃逸时，能够根据目标车辆的特征对视频中的目标进行快速的蹄选，帮助公安机关在破案时节省了宝贵的时间。

显著性检测是计算机视觉领域中非常具有代表性的问题，它的目的是定位出那些最吸引人视觉注意的像素或区域。随着信息科技的发展与快速交通工具的推广，人类所接收的信息量呈指数级增长，如何筛选出人类感兴趣的目标和区域具有重要的意义。显著性区域与人视觉感知关系极为紧密， 并具有一定的主观性， 开展显著性检测的研究非常有利于图像处理基本任务的完成。例如，能够大致定位出兴趣目标，减小图像分割与检测的搜索范围和计算量; 精确的显著性检测使目标跟踪与识别更加容易

### 1.1.2 研究目的

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。

4，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。

### 1.1.3 研究意义

借鉴人类视觉注意机制，计算机视觉领域的研究人员展开了对适用于计算机模拟的视觉注意模型的研究。他们通过将视觉注意模型作为机器视觉系统中进行信息筛选和计算资源优先分配的关键组件，用于提升计算机视觉系统处理海量数字媒体的能力，提高数字媒体资源的利用率。构建有效的视觉注意模型首先需要解决的关键性问题就是视觉显著性检测。然而目前由计算机模拟的视觉显著性检测算法的检测结果准确度还不高，这将会限制计算机视觉系统的应用效果和应用范围。因而构建检测性能更好的视觉显著性检测算法是目前计算机视觉领域亟待解决的关键性问题。

## 1.2 显著性研究现状

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等人在年提出的视觉注意模型。该模型基于特征集成理论，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[1]；在Itti工作的基础上，Walther等人将其进行扩展，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[2]；Ma和Zhang等人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[3]。同时Chen xia等人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[4]。

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，显示观察者在虚拟环境中的眼球运动；Kanan等人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[5]。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[6]。

## 1.3 显著性的相关应用

图像的显著性检测能够降低对于图像内容的理解与图像的低层特征之间存在的隔阂，这为其他层次的图像处理提供了极大的便利，例如图像分割【8】，目标检测，图像和视频压缩，目标重定位、图像编辑

[11-14]

等等。除此之外，还可以应用到

车牌和路标的检测、高性能脑模型、高智能武器系统和医疗设备等方面。下面列举

出了两个图像显著性区域在其他图像处理过程中的应用。

显著性区域检测技术有着广泛的应用，在目标自动检测⑴、图像检索、物体识别图像分割视频的快速浏览和汇总、图像和视频压缩、图像自动化修剪和内容感知的图像编辑】】】』〃】等领域都可以应用。目前显著性检测技术的应用还无法令人满意，除了显著图本身的质量不够高以外，显著性信息应用方式和方法还不够成熟，需要找到更加令人满意的实现方法。

## 1.4 本文的主要内容

本文的工作主要分为三步：

1. 提取先验知识。这里的先验知识主要是指图像的目标性。图像目标性表示图像中的一个像素或者一快区域包含完整目标的可能性。一张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。

（2）构建深度学习网络。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。

（3）先验知识与深度学习网络结构的融合。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）中的图像目标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）中得到的目标性特征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，最后一起来训练深度网络。

本文的主要内容安排如下：

第一章是绪论，主要是提出本论文的研究背景及意义，分析了研究现状，并给出了整个论文的整体思路。

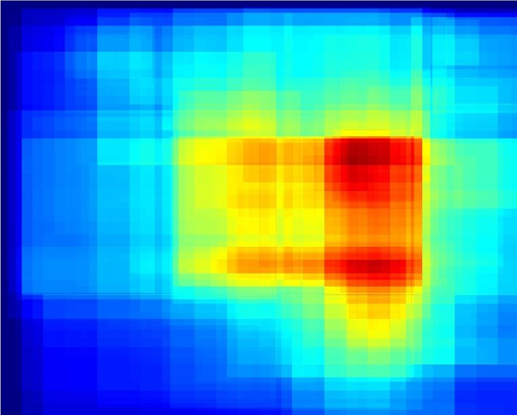
第二章是论文相关的知识介绍

# 第二章：相关知识介绍

## 2.1 图像目标性介绍

### 2.1.1 什么是图像目标性

图像目标性最开始是Bogdan Alexe【9】等人提出来的，其中目标是指具有明确定义的边界，且具有独立的东西和中心，如汽车、电话和人等，而不是没有形态的背景环境，如天空、草坪和道路等。图像目标性是指对于图像中的一个区域，该区域是目标的可能性有多大，具体到图像中的每一个像素点，是指该像素点被图像中某目标包含的可能性大小。一张图像的目标性表示方式为一张对应大小的单通道黑白图像，其中图像中像素点越亮的地方表示对应原图中目标性越大，像素点越暗的地方表示对应原图中的目标性越小。

原图 目标性图

### 2.1.2 目标性的意义

通常，一般人会主动去判断一张图像中的目标是否显著，以此来确定该目标区域是否具有大量信息，是否是可识别区域，然后再对该区域做细致分析，得到我们需要的图像信息。由Bogdan Alexe提出的目标性我们可以知道，一张图像中的目标往往都有一定的明显特征：1）目标基本都是被一个封闭的区间所包含；2）图像中目标区域和周围区域有着明显不同的形态；3)目标都是突出显著的，有时候是独一无二的【10】。许多目标对象都同时具有上述特征中的几个或者全部。针对显著检测中特征的适应性不足以及当前一些算法出现多检与漏检的问题，提出从“目标在哪儿”与“背景在哪儿”两个角度描述显著性的框架，进行特征融合来提高显著目标检测的准确率。

### 2.1.3 目标性的获得方法

对于图像目标性的先验知识，现在已有很多的研究和方法实现，一般是通过图像底层的基本特征，如图像局部颜色，轮廓，方向梯度来获得相关特征，得到该图像的目标性特征图。

（1）“特征融合与 objectness 加强的显著目标检测”【11】提出了一种先利用mean-shift分割算法分割图像区域，来保持每个目标的 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。

计算每个像素的 objectness 特征，需要在图像上随机分布 N 个窗口，每个窗口w用文献[3]的方法计算 objectness 得分，并记为 P(w)。随后对所有窗口W，统计包含每个像素点的窗口的 objectness得分，以此获取每个像素的 objectness 特征，公式如下：



其中 w 表示 W 中任意包含像素 x 的窗口。实验中 N 设置为 10000。其次，计算每个区域i 的objectness 特征：



计算出所有区域的objectness 特征后，将特征值分配到区域所包含的像素上。最后，得到整个图像的objectness 特征，标记记为。

其实现效果如图所示

原图 objectness图

（2）

## 2.2 图像显著性介绍

### 2.2.1 什么是图像显著性

目前图像显著性的求解方法有很多种，从原理角度上大致分为三类，分别是仿生物学的方法、纯计算的方法、仿生物学与纯计算相结合的方法；同时从计算角度又分为自上而下、自下而上两中方法。

### 2.2.2 图像显著性的获取方法

（1）“[A Model of saliency Based Visual Attention for Rapid Scene Analysis](http://blog.csdn.net/chenjiazhou12/article/details/39456589)”【12】这篇文章是图像显著性领域最具代表性的文章，是在1998年Itti等人提出来的，到目前为止引用的次数超过了5000，在它的基础上发展起来的有关图像显著性论文更是数不胜数，论文的提出主要是受到灵长类动物早期视觉系统的神经结构和行为所启发而产生了视觉注意系统。灵长类动物具有很强的实时处理复杂场景的能力，视觉信息进行深入的处理之前，对所收集到的感觉信息进行选择，这些选择可能减少场景理解的复杂性，这个选择过程在一个空间有限的视野区域即所谓的注意焦点（focus of attention， FOA）中完成的，它搜索场景的方式可以是快速、自下而上（bottom-up，BU）、显著驱动和任务独立的方式，也可以是慢速、自上而下（top-down，TD）、意志控制和任务相关的方式。注意模型包括“动态路由”模型，在此模型中，通过皮层视觉继承，从一个小的视野区域中得到的信息可以不断前行。通过皮层连接的动态修正或在TD和BU的控制下对活跃区建立特定的瞬时模式，来选择注意区域。

这篇文章所使用的模型建立在由Koch和Ullman所提出的生物模型和其他几个模型。它和所谓的特征整合模型相关，解释了人类视觉搜索策略。视觉输入首先被分为一系列特征地形图。然后在每个图中，不同的空间位置通过竞争获取显著性，只有从周围脱颖而出的位置才能保留。所有的特征图以纯粹的BU方式输入到高级的显著性图，它对整个视觉场景的局部醒目度进行地形编码。在灵长类动物中，人们相信这种图存在于后顶叶皮层（posterior parietal cortex）中，同时也存在于枕核丘脑（pulvinar nuclei ofthalamus）中。模型的显著性图被认为是产生注意转移的内部动力。因此这个模型表示了BU显著性可以引导注意转移，不需要TD。这个模型在可以进行并行处理，提高运算速度，而且可以根据特征的重要性，为特征加上权值，特征越重要，权值越大。

（2）“Saliency Detection: A Spectral Residual Approach”【5】提出了一个图像视觉显著性的简单计算模型，这个模型和Irri提出的模型是两个截然不同的模型，Irri模型对于图像视觉显著性主要关注整幅图片突出的部分，通过各种特征的融合提取显著性图，而Hou的这个模型直接关注的点就不在一张图片里突出的地方，而是背景，观察是否大部分图片的背景在某个空间上都满足什么变化，最后剔除背景，自然就只剩下图片突出的部分了，没有很复杂的公式变换，都是图像处理基本的变换公式。

Barlow提出的有效编码假说指出，感知系统的信息处理过程与外界信号的统计特性密切相关。初级感知系统的作用就是去除输入信号的统计冗余。也就是说，大脑之所以能够对外界环境自适应，是因为复杂的外界刺激存在冗余，而大脑的神经元能够有效地去除这些冗余，从而可以利用较少的资源尽可能有效地表达更多的信息。

从信息论的角度来看，有效编码假说将图像信息H(Image)分为两部分：

H(Image)=H(Innovation)+H(Prior Knowledge)

H(Innovation)表示突出的部分，H（Prior Knowledge）则表示冗余的信息。通过去除图像冗余信息，就可以获得图像与众不同的部分，即显著目标。

自然图像的统计特性具有变换不变性：即将图像从原来的空间坐标变换到频率坐标系中，图像在空间中具有的统计特性在频域中仍然保留，这种不变性恰好保证了采用能量谱来刻画自然图像空间相关性的可靠性。

在自然图像的统计特性中尺度不变性是最经典也是研究最广泛的特性，这种特性也被称为1/f法则，即自然图像集合的平均傅里叶谱的幅值A(f)服从下式的分布：



对于一个系统模型而言，考虑输入刺激信号的统计相似性可以极大地减少多余的视觉信息，这是因为相似性意味着冗余。不同的图像数据，其log谱却有着相似的分布趋势，而且曲线满足局部线性条件。因此，在不同的log谱中我们只需要关注其差异部分忽略相似部分。

给定一幅图像I(x)首先计算2维离散傅里叶变换，将其从空间域转到频域，并得到振幅谱和相位谱，对幅值取对数后得到log谱，变换公式如下：







其中表示傅里叶变换，表示求傅里叶变换后的幅值。

由于log曲线满足局部线性条件，所以用局部平均滤波器h(f)对其进行平滑，获得平均频谱：



其中是一个的矩阵，定义为



因此谱残差就是log谱和其进行均值滤波后的差，可按下面的式子计算：

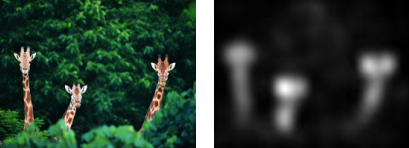


最后将谱残差和相位谱相加，并求出自然指数，然后对其进行傅里叶反变换，最后经过一个高斯模糊滤波就得到了所谓的显著性区域值，具体公式如下：



其中，g(x)为高斯模糊模板，为傅里叶反变换。

得到的图像显著图如图所示：



## 2.3 自动编码器介绍

由于本论文中的图像显著性是基于自动编码器实现的，所以这里对自动编码器的理论公式及训练过程给出简单介绍。

### 2.3.1 什么是自动编码器

### 2.3.2 自动编码器训练过程

### 2.3.3 自动编码器在图像显著性上的应用

## 2.4本章小结

## 参考文献

[1] Itti, L.; Koch, C.; Niebur, E., "A model of saliency-based visual attention for rapid scene analysis," in Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.20, no.11, pp.1254-1259, Nov 1998

[2] Scholkopf;lkopf, B.; Platt, J.; Hofmann, T., "Graph-Based Visual Saliency," in Advances in Neural Information Processing Systems 19:Proceedings of the 2006 Conference , 1, MIT Press, 2007, pp.545-552

[3] Chen Xia; Pengjin Wang; Fei Qi; Guangming Shi, "Nonlocal center-surround reconstruction-based bottom-up saliency estimation," in Image Processing (ICIP), 2013 20th IEEE International Conference on , vol., no., pp.206-210, 15-18 Sept. 2013

[4] Chen Xia; Fei Qi; Guangming Shi, "Bottom-up Visual Saliency Estimation with Deep Autoencoder-based Sparse Reconstruction," IEEE Trans. Neural Netw. Learn. Syst., accepted (December 20, 2015)

[5] Tie Liu; Jian Sun; Nan-Ning Zheng; Xiaoou Tang; Heung-Yeung Shum, "Learning to Detect A Salient Object," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on , vol., no., pp.1-8, 17-22 June 2007

[6] Congyan Lang; Guangcan Liu; Jian Yu; Shuicheng Yan, "Saliency Detection by Multitask Sparsity Pursuit," in Image Processing, IEEE Transactions on , vol.21, no.3, pp.1327-1338, March 2012

[8] Han J, Ngan K, Li M, et al. Unsupervised extraction of visual attention objects in color

images[J]. Circuits and Systems for Video Technology, 2006, 16(1) :141　145.

[6] Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification

[C]//Proceedings of the 9th European Conference on Computer Vision, Graz: Springer Verlag,

2006:490-503.

【9】B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. PAMI, 2012

【10】特征融合与 objectness 加强的显著目标检测

[11] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. PAMI, 2012

【12】[A Model of saliency Based Visual Attention for Rapid Scene Analysis](http://blog.csdn.net/chenjiazhou12/article/details/39456589)