# 先验知识学习及其在视觉显著性估计中的应用

# 第一章：绪论

## 1.1 显著性研究意义及目的

这些年来，随着科学技术和计算机的快速发展，人们的生活越来越离不开对计算机的依赖，小至拍照购物，大到航天备战，计算机已充分融入到了我们生活中的每一个角落。生活在这样一个成指数化增长的信息年代，计算机在带给我们无数便利的同时也给我们带来了大量的信息干扰，甚至大量的数据让有些人的生活无所适从。每个人都想在海量的数据信息中迅速找到有用的知识，来提升自己的工作效率或者生活节奏。所以，如何从大量数据中排除掉我们不需要的冗余信息，同时得到我们想要的重要信息已经成为计算机领域讨论的热点。

而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力的像素或区域。人类可以毫无压力的从丰富多彩的照片中迅速找到自己感兴趣的物体，是因为人类大脑经过上亿年的自我训练，已近具有超强的视觉注意系统。对于进入人们眼球的一张图像，大脑可以迅速对图像的内容进行判断，分析哪些是目标区域，哪些是背景区域，然后迅速从目标区域选择自己感兴趣的目标，并将自己的注意力定位到该感兴趣区域，从而可以忽略掉图像中大部分不相干的冗余信息，提高处理效率。而随着信息科技的发展与快速交通工具的推广，计算机视觉所接收的信息量呈指数增长，如何从中筛选出人类感兴趣的目标和区域具有重要的意义。视觉领域的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，同时提取图像中感兴趣的区域已经应用到了生活中的各个方面。例如在对图像进行压缩传输的过程中，特别是那些传输带宽不足或者是迫切需要提高传输速度的场合，可以将图像中非显著区域的部分压缩比例变大，对于显著区域的地方则可以正常压缩，这样通过解压的图像会很清楚的得到显著区域的各种细节，而对非显著的地方选择性的忽略了无关紧要的细节；又比如警察通过监控视频追踪嫌疑犯时，通过对视频内容的显著性分析，可以忽略掉视频中大量和车辆无关的背景信息，锁定犯罪分子，这样可以为及时制止犯罪活动提供宝贵的时间；同样显著性检测还可以用到图像的目标检测，目标跟踪及图像分割等多个方面。

本文期望使用现有的深度学习算法，先从大量图像数据库中学习先验知识，这里的先验知识是具有图像高层的语义特征，再结合这些高层语义特征来提升图像显著性估算法的性能。

（1）这里的先验知识具体是指图像的目标性。目标性首先是由Alexe【1】提出，表示一张图像中一个像素点或者一块区域是一个目标组成部分的可能性。通常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。

（2）这里的深度学习指的是自动编码器。随着近年来网络上大量数据库的出现及计算机计算能力的迅速提升，深度学习在各个领域都有着很好的学习能力和检测识别效果，同时具有很强的鲁棒性和适应性。所以这里准备采用深度学习中的自动编码器来实现显著性检测部分，自动编码器是无监督学习算法的一种，采用无监督学习可以减小我们的算法对样本标签的依赖，同时自动编码器由于特殊的自身对称结构，可以实现对图像的自我重构。这里，可以利用图像自我重构的残差作为图像的显著性估计。

（3）这里的高层语义特征是指图像中具有一定语义信息的特征，比如说目标的轮廓，目标的分布情况，图像中是否具有人或者车等。相对于高层特征，图像的底层特征一般是指图像中的颜色、轮廓等不具有语义的一些特征。以高层语义特征为先验知识，相当于在显著性检测时增加了对图像的认知功能，以此来提高图像的显著性分析。

## 1.2 显著性研究现状

近些年来，学者们对于图像显著性的研究已越来越广泛，并已经取得了很好的研究成果。人类大脑具有先进的视觉注意机制，对于生活中的复杂图像，大脑可以迅速反应过来，判别出图像中感兴趣的部分，同时忽略掉图像中不相关的冗余信息。而显著性检测正是为了让计算机能够实现人类大脑的这种视觉注意机制发展起来的。计算机通过显著性算法得到一张与原图同样大小的灰度图，其中灰度图中越白的地方表明原图中该点的显著性越强，反之，灰度图中越暗的地方表明原图中该点的显著性越弱。随着生物学及认知学的研究发展，人们对大脑的视觉注意机制的认知和了解也数来越熟悉。从注意机制的原理上讲，视觉系统识别物体的概念主要是从两个角度来实现的：一是自下而上的由数据驱动的方法，二是自上而下的由高层语义驱动的方法【1-3】。其中，自下而上的研究方法不需要依靠任何的先验知识，主要是要先得到图像的底层特征，如颜色特征，轮廓特征等；而自上而下的研究方法主要是先得到图像的高层语义特征，如图像语义的理解、记忆功能、任务驱动等。而在人类大脑的视觉认知过程中，自下而上与自上而下的两种方式往往是相互协同来实现的。

（1）自下而上的视觉分析计算模型。自下而上的模型主要是得到图像底层的多种特征，然后多种特征之间的关系得到图像的显著性。为了让电脑模拟人类视觉注意力机制，也就是通过算法得到图像或者视频的显著性分析，Itti et al. [1-3】]提出了一个开创性的C-S结构，该结构认为图像中一个像素点的显著性是由该点与其周围环境的区别程度决定的，考虑的主要是图像中局部信息，利用图像的中心区域与外围区域的各种特征比来得到图像的显著性。在过去的十年中，很多学者基于C-S又提出了各种各样的变形结构来估计图像的显著性。Ma 和Zhang [12]提出了一种利用图像的颜色作为感知场，每个像素点作为感知单元，通过区分一个像素点与周围其他单位像素点的区别来得到该像素点的显著性，考虑到该像素点在图像中也存在上下文信息，该方法又利用了图像块之间的特征差异来增强该点的显著性评价【1-16】【1-18】。同样的，Seo和Milanfar [1-32]提出了一种通过计算中心特征矩阵与其周围特征矩阵的矩阵余弦相似度来代表该处的显著性，与其他方法不同的是，其他方法都归一化了周围区域对中心区域的影响。Klein和Frintrop [1-14]使用Kullback-Leibler分歧（KLD）估计特征统计量的C-S差异，并将多个特征的显著图组合成一个显著图。Chen xia【】等人提出了一种利用图像中每个像素点的中心块重构外围块得到的残差来衡量该处的显著性。除了利用图像中的局部信息来计算显著性，也有很多方法是基于图像的全局信息来计算图像显著性的。Achanta等人[2-2]提出了通过计算图像中的一个像素点和整个图像的平均色差的差异来得到该点的显著性，但是这种方法只是简单的考虑了图像中的一阶平均颜色，所以对于自然场景中常见的复杂图片处理效果不太好，同时该方法也没有考虑图像中各个区域的空间位置关系，而空间位置是显著性检测中很重要的一个特征。

（2）自上而下的视觉分析计算模型。自上而下的模型主要是先从图像中提取一些语义信息，如目标位置或者目标轮廓之类的相关特征，然后根据该特征来实现图像的显著性分析。Kanan【2-3】等人提出了一种基于贝叶斯框架显著性模型，该模型包含一个自上而下的组件，可以根据外观引导人们注意场景中可能成为目标的区域，进而得到图像的显著性。Lang等人[2-6]利用无监督学习的方法从没有标记的图像库中得到图像的显著性，他们提出了一种综合多个特征的算法来得到显著性的方法，具体思路是先得到图像的多个特征描述，然后根据从多个特征矩阵的联合分布找到一致的低阶稀疏元素并得到稀疏矩阵，由此来推断图像的显著性。

## 1.3 本文的主要内容

本文的工作主要分为三步：

（1）获得图像的高层语义特征。这里的语义特征主要是指图像的目标性，目标性是指图像中的一个像素点或者一块区域所包含目标的可能性，由目标性特征图，我们可以知道图像中所有目标的大概分布情况，这对图像的显著性检测有着语义上的指导。在求图像目标性的时候我们主要是先将原始图像颜色通道分离，然后每个通道多尺度化，接着对每个尺度的图像进行分割得到多个分割窗口，并对各个窗口进行打分，最后根据窗口的个数及分数情况得到图像的目标性图。

（2）通过自动编码器得到图像的显著性。自动编码器属于深度学习算法中无监督学习的一种，由于自动编码器本身的对称结构，它能够实现对图像的自我重构功能。在搭建自动编码器网络时，前面几层是由RBM堆叠得到的，最后一层是为了适应重构图像的维度而另外添加的；在搭建好了网络结构之后，再利用从图像中随机采样的样本来训练该网络参数；之后是依次遍历图像中的每个像素点，由该像素点的外围块重构中心块得到重构残差，在残差的基础上加上图像的中心先验，作为图像的显著性估计。

（3）图像目标性与显著性的融合。这里主要结合步骤（1）得到的图像目标性及步骤（2）中得到的显著性，来提高整个图像的显著性检测效果，这里给出了两种融合方式。一是将步骤（1）中得到的图像目标性作为先验知识，然后监督步骤（2）中自动编码器的训练过程，以提升自动编码器的优化程度进而提高对应的显著性检测；二是将图像的目标性与由自动编码器重构残差得到的显著性直接融合，来提高该图像的显著性检测。

本文的具体安排如下：

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，并给出了本论文的整体思路。

第二章是本文的相关知识介绍。先介绍了图像的目标性概念，目标性是什么，具有什么意义以及现有的一些实现方法；接着又介绍了图像的显著性概念，显著性的获取方式；最后介绍了显著性实现过程中所用到的一种深度学习网络，即自动编码器，并给出了自动编码器的构建过程及其训练过程。

第三章是本文的主要原理实现部分。先着重讲解了本文中图像目标性的实现过程，并展示了实现效果；然后是自动编码器的构建及训练过程，以及如何利用自动编码器实现图像的显著性检测；最后介绍了怎么利用图像的目标性作为先验知识，来结合由自动编码器得到的显著性共同提高图像的显著性检测效果。

第四章是本文的实验部分。先简单介绍了本论文中实验所用到的数据库，以及实验时所采用的评价指标，然后依次给出了本文的实验过程和实验结果，并给出了实验分析。

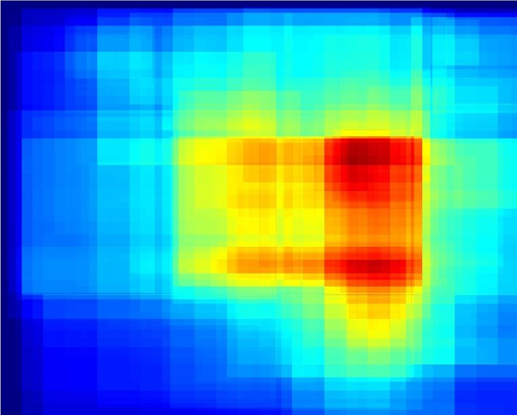
第五章是本文的总结与展望。回顾本文的实现方法及对实验结果的总结，最后给出给出了本文中的不足点以及后续的工作。

# 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

图像的目标性最开始是Alexe【9】等人提出来的，其中的目标是指具有明确的边界，且具有独立的中心的物体，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。同时一般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的；2）图像中的目标区域和周围区域往往有着明显不同的形态，具体形态可能包括颜色、轮廓、复杂度等情况；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。而图像的目标性是图像中的一个像素点或者一个区域所被图像中任意一个目标所包含的可能性。其中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。通常，一张图像的目标性可以由一张与原图对应大小的单通道黑白图像来表示，其中像素点越亮的地方表示对应原图中该处的目标性越大，像素点越暗的地方表示对应原图中该处的目标性越小。具体效果如图所示，这里为了展示效果，我们把黑白图像对应的变成彩色图像，右侧目标性图中越红的地方对应着左边图像目标性越强的地方，我们可以发现左侧图像中的交通标志牌是比较突出的目标，所以该处的目标性比较大。

原图 目标性图

### 2.1.2 目标性的意义

通常，一般人在第一眼看到一张图像时，大脑视觉系统会迅速的给图像中每个区域给定一个语义信息，然后再着重观察图像中比较感兴趣的区域，同时自动忽略掉那些无关紧要的区域。这些语义信息可能是由颜色区分，由图像区域的复杂度来区分，当然更多可能的是由图像中的目标来区分。所以如果我们知道了一张图像对应的目标性，就近似知道了该图像中的目标分布情况。而图像显著性是指我们对一张图像中最感兴趣的地方，一般我们感兴趣的地方也是一个目标或者目标的一部分。所以以图像的目标性为语义指导，我们可以提高图像的显著性检测。一般人分析图像时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。针对现在显著检测算法中一些特征的适应性不足，同时缺少图像的上层语义信息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”两个图像特征描述来指导以前显著性检测的算法框架，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

一般人计算图像的目标性要么是为了以目标性为基础来对图像进行目标检测或者目标识别，要么是为了以目标性为先验知识来提高图像的显著性估计。而对于图像目标性的获取方法，则一般是通过图像底层的基本特征，如图像局部颜色直方图，图像轮廓特征，方向梯度特征等相关特征进行融合推导来得到该图像的目标性特征图。由于现有的对目标性求解的方法比较少，这里只是简单的介绍一种图像目标性的获取方法。

“特征融合与objectness加强的显著目标检测”【11】提出了一种先对图像进行分割然后根据分割图像的窗口来计算图像的目标性的方法。其具体思路是先采用mean-shift分割算法对图像进行分割，得到多个分割窗口，以此来保持每个窗口里面的目标具有一致的目标性。然后再开始计算该图像的目标性，具体过程分为两步：评估图像的像素级目标性和图像的区域级目标性。

计算图像中每个像素点的目标性特征，需要在图像上随机生成N个窗口，然后对每个窗口w计算目标性得分，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显著性得分，以此获取每个像素的目标性特征，公式如下：



其中w表示W中任意包含像素点x的窗口。其次，计算每个区域i的目标性特征：



计算出所有区域的目标性特征后，将特征值分配到区域所包含的像素上。最后，得到整个图像的目标性特征。

## 2.2 图像显著性介绍

由于在绪论中已经详细的介绍了什么是图像的显著性及显著性的意义，并给出了显著性的研究背景及现状，大家对图像的显著性已经有了比较清楚的了解和认识。所以这里不再对图像显著性的定义及意义给出重复的介绍，该小节会主要介绍一下现有的一些图像显著性检测的方法。

（1）提到显著性图，不得不介绍Itti【1】在显著性领域的一篇开创之作。这篇文章首先提出了一个基于中心外围的模型来估计图像的显著性，其基本思想认为图像中一个像素点的显著性是由该点与其周围环境的区别程度决定的，考虑的主要是图像中局部信息，是利用图像的中心区域与外围区域的各种特征比来得到图像的显著性。如图像中的一个区域中间是蓝色而外围是黄色，或者一个区域中间亮度比较大而外围亮度比较暗，则证明该点处的显著性比较明显。其基本思想及实现如下：

对于一张彩色图像，先对图像进行多尺度化得到9张尺度图，这里的尺度化方法采用的是高斯金字塔。尺度化之后，最大尺度的图像和原图像大小一致，最小尺度化的图像是原图像的1/256。对于尺度化之后的图像，其中最大尺度的图像细节比较多，而尺度稍微小的图像则图像细节比较少，作者这里认为尺度较小的图像在经过了高斯平滑之后，由于图像中的一些细节已近模糊，可以近似的反应图像的背景信息。接着是跨尺度减操作，该操作是指先将尺度较小的图像进行线性插值得到与尺度较大的图像一样的尺寸大小，然后两个图像依次相减。该操作的意义是指将尺度较小的图像看作背景，利用尺度大的图像减去背景，则可以得到图像中每个像素点中心与周围背景像素点的差异，这里称为中央周边差。通过对每对尺度化之后的图像进行跨尺度减操作之后会得到多组中央周边差。多组中央周边差的大小可以反应该图像的中心周边差异程度。在Itti的模型中，将中心周边差又称为特征图。最后将这些特征图输入到一个动态神经网络模型，从而得到原始彩色图像的显著性图。

（2） X.Hou在文献【5】中给出了显著性检测的另外一种模型。该模型和Itti【1】的模型完全是从两个相反的角度出发考虑的。对于一般的显著性检测模型，其主思想是想直接从图像中得到显著的部分，所以往往没有考虑到图像中非显著区域的一些特征。而该模型则采用了一个完全相反的思路，它主要关注图像中非显著区域的特征，然后试图将非显著的区域从原图像中去掉，那么剩下的就是图像中的显著性区域。该模型的一个核心思想是，生活中大部分图像的背景部分满足一定的分布，然后用图像的原始信息减去这个背景的分布就可以得到图像的显著性图。由于该模型思路比较简单，所以没有复杂的公式，只是一般的信号或图像方面的基本变换。接下来简单的介绍该模型的实现过程：

由感知系统及编码知识可以知道，一般的图像信息H(Image)可以由两部分构成：

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

其中，H(Innovation)为图像中显著的前景区域，H（Prior Knowledge）为图像中背景相关的冗余部分。这里只要通过一定的变换将图像中的冗余部分去掉，就能得到我们想要的显著前景区域。

在图像处理领域，自然图像具有一个特别常用的特征，即统计特征具有变换不变性。图像的变换不变性是说图像经过傅里叶变换到频谱空间之后，图像在原有空间的统计特征与在频谱空间的统计特征具有不变性。而这种不变性正是该算法能够有效的保证。对于一张输入图像I(x)，先对其进行傅里叶变换，这样可得到图像在频域中对应的振幅谱和相位谱，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。接着需要对log谱进行平滑，进而得到图像的平均频谱：



其中是一个的矩阵，定义为



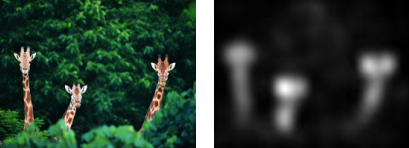
这时在图像的log谱的基础上去掉图像的平均频谱，就可以得到图像的谱残差，该谱残差实际上就是该图像空间域中的显著部分在频谱域中的映射。其计算公式如下：



得到了图像的谱残差之后，最后一步便是将谱残差和相位谱相加，然后经过傅里叶反变换由频域转化到空间域，在加上一个高斯模糊模板就可以得到图像的显著性，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

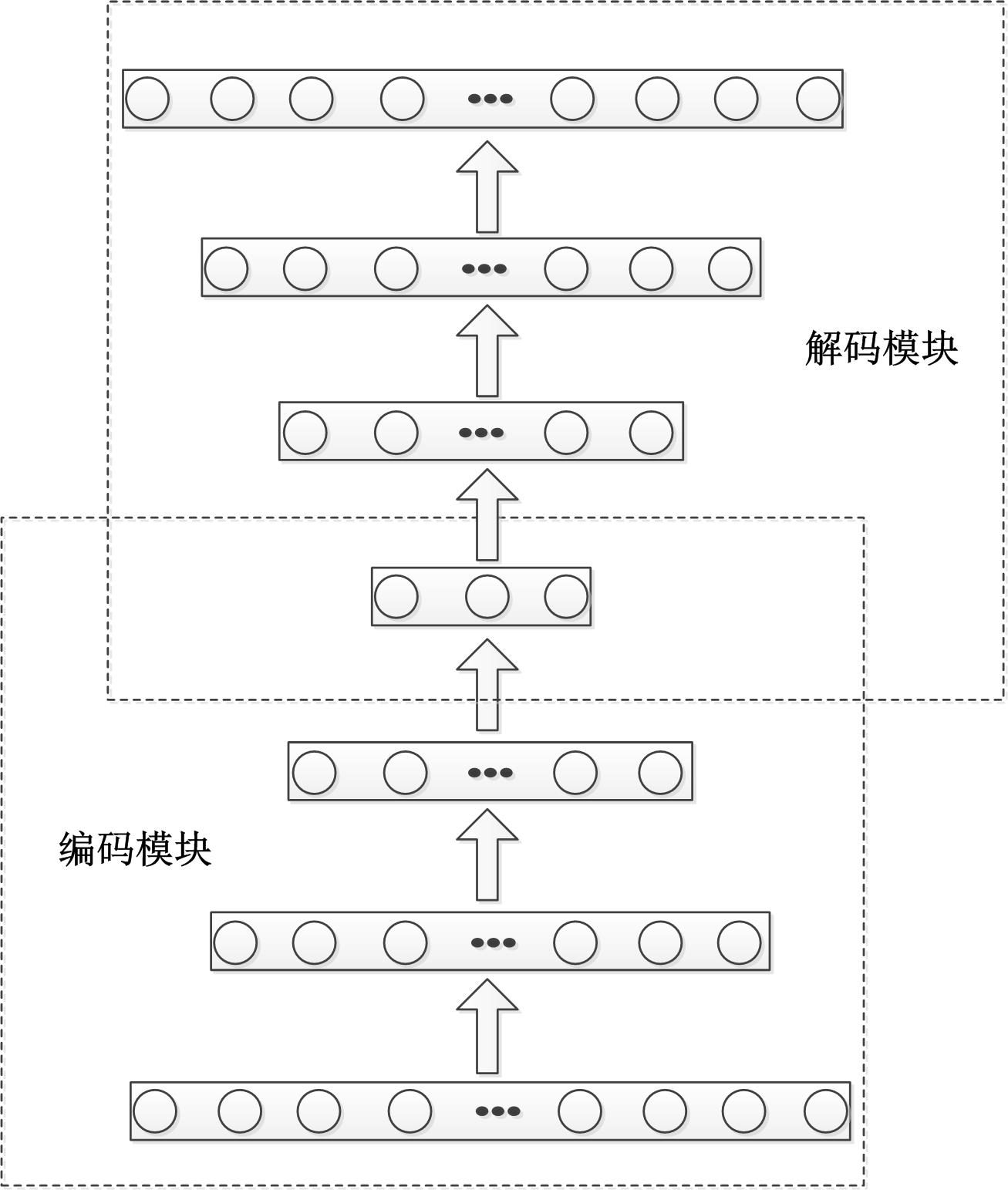
由于本论文中的图像显著性是基于自动编码器实现的，所以这里需要对自动编码器的理论公式及训练过程给出简单介绍。

### 2.3.1 什么是自动编码器

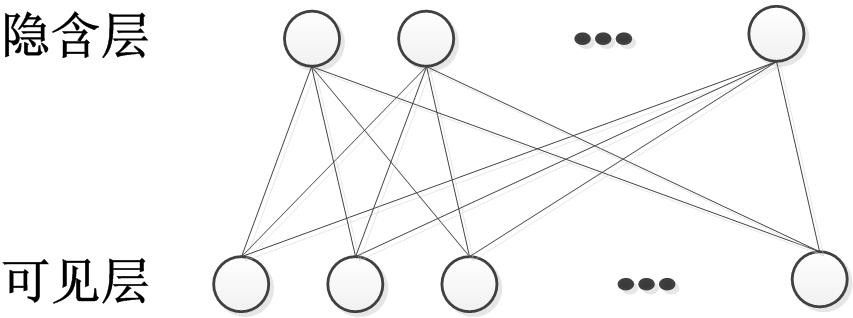
随着计算机科学及数学的多年探讨与研究，以至于无论对于图像领域，语音领域还是文本领域，都能找到很实用的机器学习模型来处理这些领域所面临的问题。机器学习可以对图像进行目标识别，可以将语音和文本互相转换，可以为网络用户个性化推荐等等，而且在解决这些问题的时候往往效果很不错，但有个比较严重的问题是机器学习鲁棒性不够强，一些常见的算法往往需要对特定的问题给出特定的模型。所以深度学习在这种迫切需求下迅速发展起来了。有了网络上足够多的数据样本集和超高计算能力的计算机，深度学习基本可以机器学习领域中的大多数问题。在处理一个问题时，需要解决的一个关键技术就是特征提取，而大多数的机器学习方法也是一个特征提取的过程。而深度学习是一个天然的特征学习工具，它可以通过训练样本和标签，自动的学习到数据集中比较关键的特征。这些年来，深度学习在各个领域扮演着越来越重要的角色，通过广大学者的不懈研究，深度学习已经可以解决很多领域中的各种难题，它可以通过长时间的自我学习来发现大量数据中的复杂结构。特别是在图像、语音及文本处理领域有着非凡的成就。而随着计算机的近一步发展，深度学习必然会有着更广泛的应用，取得更多的成功。为了加速深度学习的工业化，目前很多公司及企业也在急切的研究深度学习相关的加速芯片，这更加保证了深度学习在未来的研究方向不可动摇及替代的一面。

现有的深度学习从训练过程及学习方式来说主要分为三类：监督学习，非监督学习及半监督学习。其中监督学习是指训练样本中必须要有标签，利用样本的标签来指导深度网络的优化及学习过程，这类网络对样本的要求比较高，但是如果有了足够多的样本就能学习的很好；非监督学习是指训练样本并不需要标签，即网络的学习过程对样本的依赖相对低一些，实现起来要求也没有监督学习那么苛刻；而半监督学习是指在训练网络时主要结合少量有标签的数据及多数无标签的数据，总体来说介于监督学习和非监督学习之间。而这里提到的自动编码器就是深度学习中无监督学习的一中方式。

自动编码器在整体上是一个对称结构，第一层和最后一层节点数相同，第二层和倒数第二层节点数相同，其它层同理，最中间的一层为特征层，即特征层是整个网络的对称中心，所以，整个网络的输入和输出是相同的。



在结构上，自动编码器是由多层RBM（限制波尔兹曼机）叠加构成， 一个RBM相当于自动编码器的一层。RBM是由可见层和隐藏层组成的两层网络结构；每层网络中均由多个神经元组成，其中神经元的个数可以人为定义。RBM中的这些神经元一般拥有两种状态，即激活状态和未激活两种状态。同时RBM也是无监督学习方式中的一种，它的学习能力很强，可以从训练集样本中学习到各种复杂的特征。RBM的可见层又叫做网络的输入层，其隐藏层又叫做网络的输出层，输入层和输出层之间的神经元相互连接，而输入层之间的神经元并没有任何联系，即不存在输入层之间的神经元相互连接，同理输出层之间的神经元也没有任何联系。其结构图如下所示：



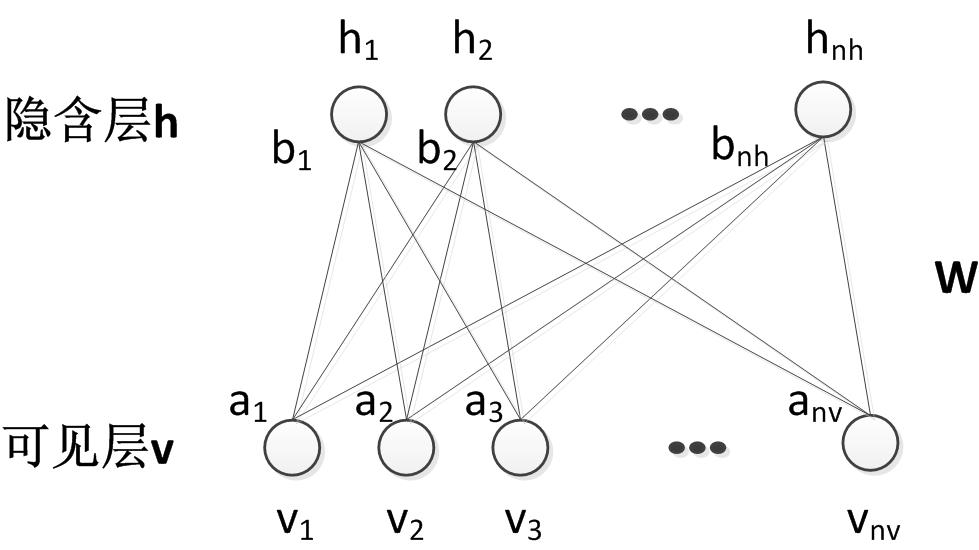
RBM

RBM的主要作用有两种：一种是数据特征提取，第二种是对网络进行参数初始化。对于特征提取是指，由于RBM是具有两层结构的特殊网络，数据通过可见层可以得到网络的隐藏层，同时通过网络的隐藏层又可以得到网路的可见层。这样只要保证数据通过两次来回传输得到的值和最开始送入到网络的值一致，就表明可以将网络的可见层用隐藏层来表示，相当于找到了可见层的特征向量。一般数据可以先经过特征提取，再将提取到的特征交给其他的机器学习方法来处理。对于第二种作用，是对网络参数的初始化。一般比较深层的神经网络如果参数初始化时处于的状态不太好，就大大增加了网络参数优化的难度，有是甚至是无法优化的。而就近一步说就算初始化时处于的状态不是很差，由于大量的参数和太深的层次，网络的参数也很难短时间内优化。这里便可以先采用RBM来实现深层神经网络层与层之间的参数。通过很多实验证明，利用RBM训练得到的权重矩阵和偏移量作为深层神经网络的初始值，会大大减小训练过程中的很多问题，同时缩短训练时间。这里的自动编码器训练时就是利用了RBM的第二个用途，来初始化自动编码器每层之间的参数。

### 2.3.2 自动编码器训练过程

由于自动编码器是由RBM逐层叠加得到的，所以这里先介绍RBM的训练过程，然后再介绍怎么利用RBM来训练自动编码器。

（1）RBM训练过程



如上图所示，RBM包含可见层和隐藏层。V表示可见层神经元的状态，其中，为该层的神经元的总数；h表示的是隐藏层的神经元的状态，其中，为该层的神经元的总数；a表示的是可见层神经元的偏置，其中；b表示的是隐藏层神经的偏置，其中；W指的是隐藏层与可见层之间的连接权重，；这里我们记，表示RBM训练过程中需要调整的参数。

首先，我们需要给定RBM的能量函数，该能量函数定义了可见层和隐藏层神经元状态的之间的函数关系，其计算公式如下：



由可见层神经元和隐藏层神经元的能量函数，可以得到它们之间的状态联合概率分布，其计算公式如下：



其中为归一化因子，其计算公式如下：



当得到了可见层神经元的状态和隐藏层神经元的状态的联合概率分布之后，我们就可以得到边缘概率分布函数，即可见层神经元的状态的边缘概率分布函数和隐藏层神经元的状态的边缘概率分布函数，其计算公式如下：





有了上述的可见层神经元的状态和隐藏层神经元的状态的联合概率分布以及可见层神经元的状态的边缘概率分布函数和隐藏层神经元的状态的边缘概率分布函数。当RBM知道了可见层的输入状态时，此时就可以计算隐藏层上的任何一个神经元被激活的概率，即；同时当RBM知道了隐藏层的输出状态时，同理可以得到可见层上的任何一个神经元被激活的概率，即。

对于RBM模型，其参数主要是可见层和隐藏层之间的权重W，可见层的偏置a以及隐藏层的偏置b，即。在优化该参数时，可以通过一定量的训练样本来反复训练RBM，最终使得训练好的RBM可以很好的与训练样本所吻合，即可以通过训练样本使得RBM到达一个稳定状态。优化参数时可以用最大似然函数来求解，利用梯度上升算法来对参数不断迭代更新，使得RBM的参数不断优化。以前RBM都是采用这种方法来进行训练的，但是这种方法比较大的一个问题是需要很多的样本对网络进行多次的训练才能使得网络参数优化，得到我们想要的参数。后来Hinton教授发明了一种更为简单有效的训练方法，即对比散度（Contrastive Divergence,CD）算法【13】，由于该算法较之前的训练方法有不管是在时间上、复杂度上还是效果上都有了很大的提升，所以现在CD算法已经成为RBM的标准训练算法。

CD算法的训练步骤很简单，思路也很明确，但训练效果却很好，也大大减小了RBM整体参数的训练时间。具体可以描述如下：

对于，首先进行初始化，初始化时可以采用均值为0，方差为1的高斯分布随机初始化，然后对v进行k步Gibbs采样，k是人为设定的一个值，具体根据RBM解决的不同的问题给出不同的值，依据标准是使得合适的k使得RBM的参数能够训练的比较好。然后对k步Gibbs采样中的每一步t(t=1,2,3…,k)，先后利用可见层神经元的边缘概率分布函数采样得到隐藏层神经元的状态，再利用隐藏层神经元的边缘概率分布函数采样得到可见层神经元的状态，其计算公式如下：







其中为参数学习速率。

然后根据得到的可见层及隐藏层的神经元的状态来更新参数，具体更新公式如下：







其中为参数的学习速率。

通过k步的Gibbs采样及训练，每一步都在逐渐的优化RBM的参数同时整个RBM的性能逐步的提升。

1. 自动编码器训练过程

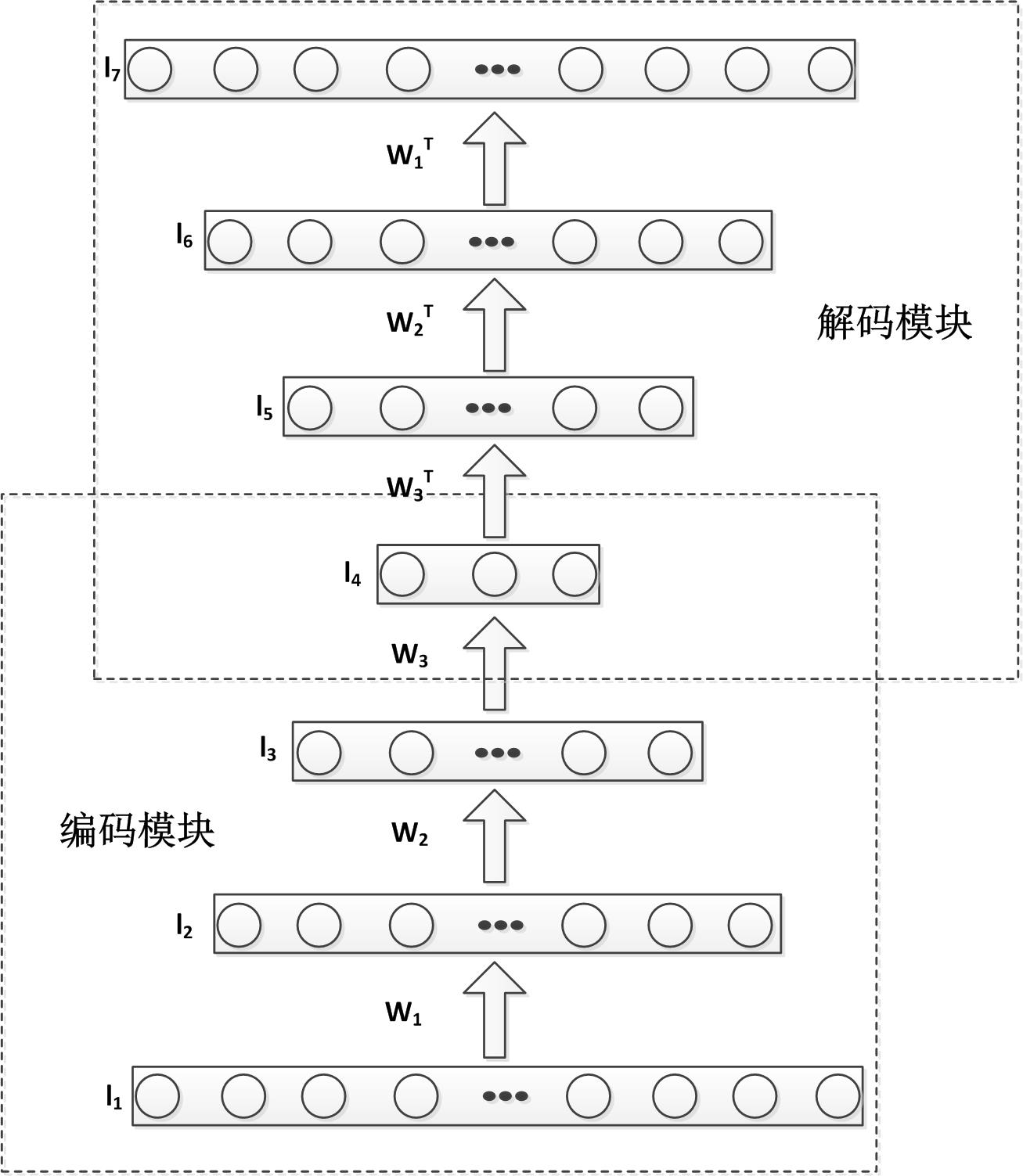
前面提到自动编码器是由多个RBM叠加组合得到的，并且以最中间的特征层为对称中心而两端相互对称的。所以在训练自动编码器时，我们只需要训练整个编码模块，由于编码模块和解码模块相互对称，所以解码模块的参数直接是编码模块的转置得到的，然后将多个RBM的参数组合得到一个自动编码器，最后对组合得到的自动编码器进行微调，即对整个网络利用反向传播算法重新调整一下。所以自动编码器的训练过程主要有三个步骤，具体过程如下：

步骤一，单独训练多个RBM。



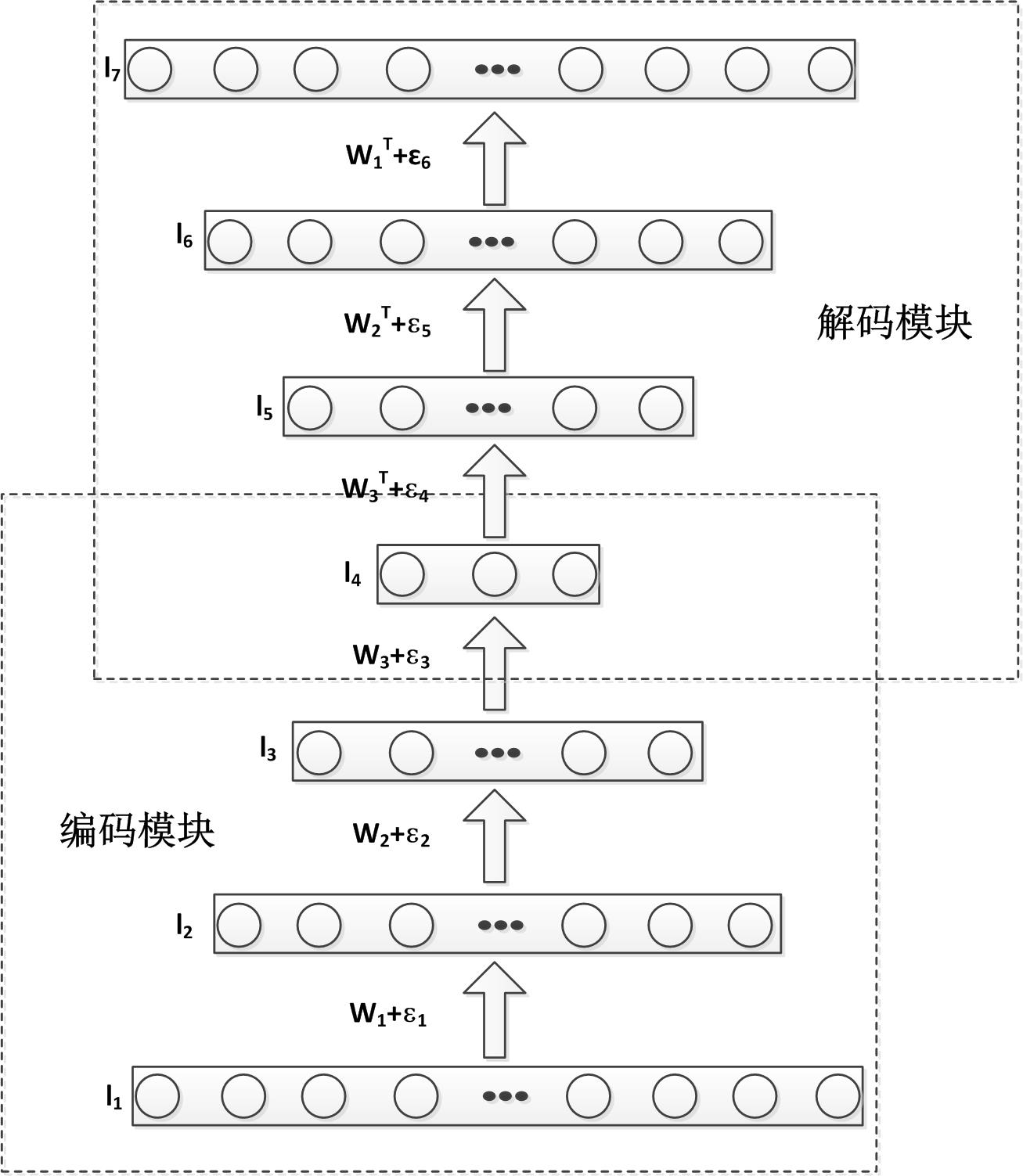
首先，将自动编码器的编码模块分成多个RBM，如上图所示，将编码模块分成了3个RBM，对与每个RBM按照上文中提到的CD算法训练得到RBM的连接权重。

步骤二，将训练好的多个RBM连接权重组合得到预训练的自动编码器。



如图所示，步骤一中通过单独训练RBM我们先得到了自动编码器的编码模块的层与层之间的连接权重W1，W2，W3。前面介绍自动编码器时我们也提到自动编码器是由多个RBM对称叠加得到的，所以解码模块的权重直接为对应位置上的编码模块的权重的装置，依次为，，。这一步在自动编码器的训练过程中也称为网络的预训练，因为对于很深的网络，所有参数一起训练时会出现各种问题，如过拟合，训练时间过长，梯度衰减等。利用预训练先得到自动编码器每层之间的参数，可以很好的避免这些问题。

步骤三，对预训练的自动编码器进行微调，得到最终的网络参数。



微调阶段主要是用来对预训练的参数进行整体调整，一般调整的幅度比较少。主要目的是让预训练的参数能够一起协调使用，来适应我们的训练目标。如上图所示，微调之后，在每层的参数后面会多一个调整变量。微调阶段主要是利用SGD（随机梯度下降算法）来实现的。微调结束后，整个自动编码器的训练阶段也就结束了。

### 2.3.3 自动编码器在图像显著性上的应用

自动编码器的主要作用有两点，一是由于自动编码器的输入和输出是相同的，所以可以用来对图像或者其它的信号进行重构；而是由于自动编码器是一个对称结构，整体关于中间特征层而对称，利用这点信息可以用来对数据降维或者升维，当中间特征层的神经元个数少于输入层神经元的个数或者小于输出层神经元的个数，这个时候显现出的作用是对输入数据进行降维，效果和PCA(主成分分析)算法一样。放中间特征层神经元个数多于输入层神经元的个数或者小于输出层神经元的个数，这个时候显现出来的作用是对输入数据进行升维。

在本论文中，我们是需要利用自动编码器来处理输入的彩色图像，得到图像的显著性图。这里是利用的自动编码器的第一个特征，即对图像的重构。具体思路是，对于任意一张图像的任意一个像素点，取其周围的一个大块来重构其中心的小块，然后比较重构得到的中心小块与原始的中心小块的残差。如果这个残差比较小的话，说明该像素点处的显著性比较小，如果这个残差比较大的话，则说明该像素点处的显著性比较明显。具体解释为，一般对于图像中的背景模块，也就是非显著模块，一般像素点的外围块和中心块是很相似的，在颜色，轮廓等其它特征方面也没有特别大的变动，所以，可以用该像素的外围块很好的重构它的中心块，所以重构出来的中心块和原始的中心块的残差会比较小，证明该处的显著性比较小；但对于图像中的前景部分，也就是显著性部分，像素点的外围块和中心块往往具有很大的差异，无论颜色，方向梯度，轮廓等都有突出的差异，这时用该像素点的外围块是不能很好的重构出该处的中心块的，此时重构出来的中心块和原始的块的残差就会比较大，证明该处的显著性比较大。

## 2.4 本章小结

本章先介绍了图像目标性的定义，然后又给出了目标性的意义以及图像目标性的求解过程，其中的求解过程主要是参考以往的算法，让我们清楚的了解了本论文中比较关键的部分；接下来又介绍了图像显著性，主要是给出了一般获得图像显著性的方法，及这些方法在求解思路上的依据；最后介绍了深度学习中的一种特殊网络自动编码器，并对自动编码器的基本组成部分RBM给出了推导过程，同时给出了自动编码器的训练过程，因为本论文中的显著性的求解是基于自动编码器，所有最后给出了我们为什么采用自动编码器的原因。

# 第三章：图像目标性和显著性

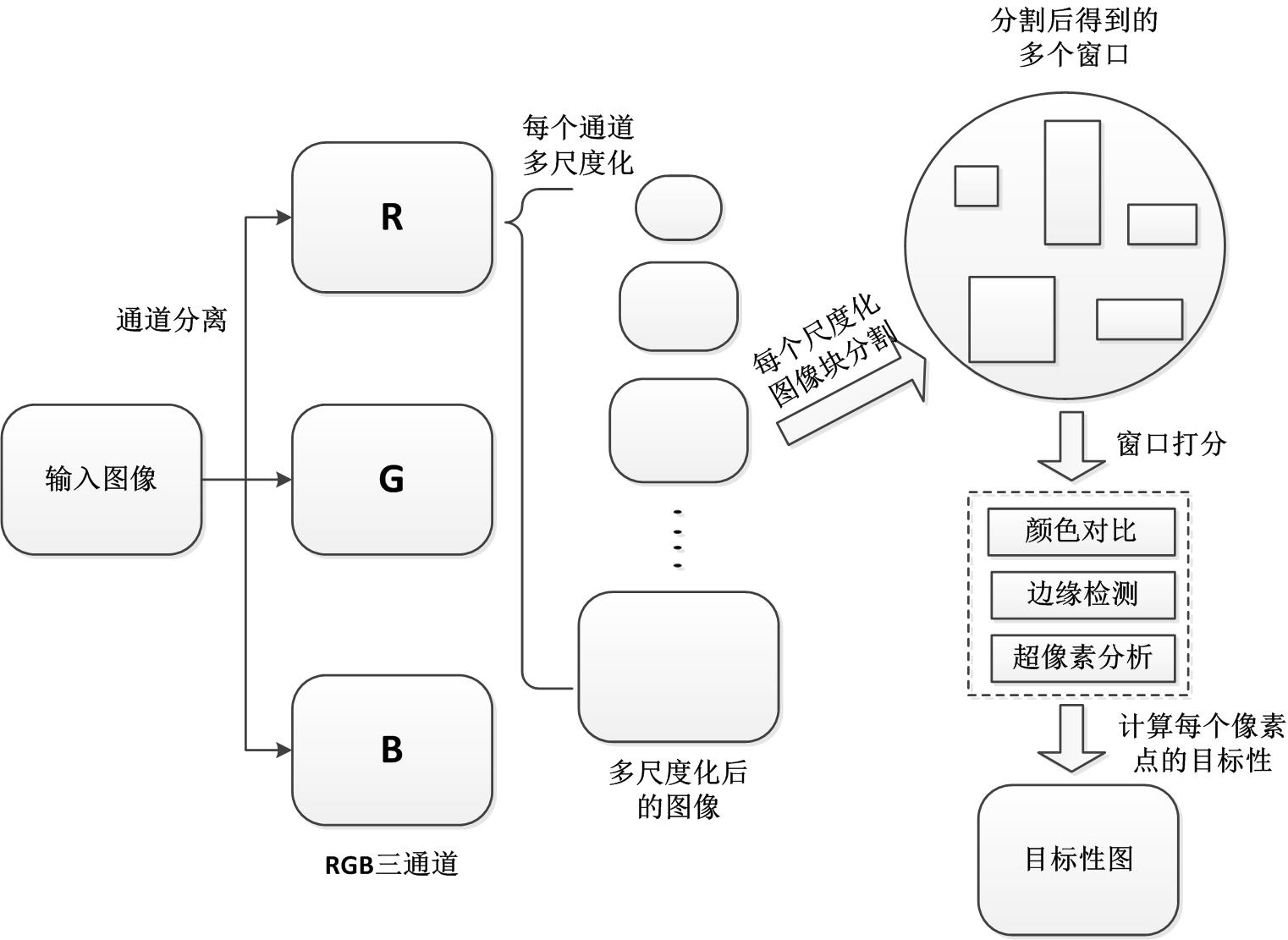
在第二章中我们首先着重介绍了图像的目标性及图像的显著性，再熟悉了目标性和显著性之后，我们又介绍了多个现有方法中图像目标性及图像显著性获取方法，这些方法也都有着不错的检查效果。在本章中，我们会详细介绍本论文完成过程的图像目标性的实现方法及图像显著性的实现方案。

## 3.1 获取图像的目标性

在前面的介绍中，已提到图像的目标性是指图像的一个像素点或者一块区域为一个目标的组成部分的可能性。对于一块区域的目标性，我们很好理解这个定义，即这块区域是一个完整目标的可能性有多大，或者这块区域是一个完整目标的一部分的可能性有多大；而对于一个像素点的目标性，则稍微难理解些，即这个像素点被图像中任意一个目标所包含的可能性有多大。这里反复会提到目标，那什么是目标了？在目标性检测中，目标并不知所有的物体，如大片的天空，如远处的草地。这里的目标本意上是指所有的前景物体，如图中的行人，车辆，交通牌等。一般一个图像中所包含的目标一本含有三个特点：（1）包含一个空间闭合的边界；（2）和自己周围的事物有着不同的形态；（3）有时候是突出的，独特的。

接下来，这里具体介绍一种求解图像目标性的方案。Alexe很早就给出了图像目标性的定义及求解方式，在文献“what is an object”【】中给出了一种思路，在本文的目标性实现过程中，我们参考了Alexe的解决思路。

图像的目标性检测主要由五个步骤组成，具体实现的框架图如下所示：



### 3.1.1 图像颜色三通道分离

现实生活中，不管手机、相机拍摄，或者摄影机录制，得到的图像都是一般都是彩色图像，人眼也如此，物体反射的光通过视网膜进入到大脑中也是彩色的。所以本目标性检测过程中所针对的也是彩色图像。这时，处理方式就有两种，一是直接对彩色图像进行处理，第二种便是先将图像RGB三通道分离，然后对不同的通道分别处理。由于我们是想得到图像中的目标性，所以将颜色通道分开，可以获得更多的轮廓或其它信息。所以，这里的第一步便是图像颜色三通道分离。

该步骤比较简单，就是直接将图像按照RGB三通道分成了三个块，每个块的尺寸和原图像大小一致。

### 3.1.2 多尺度化

现实生活中的物体形态各不一样，大小更是千变万化，当一个物体被相机或者摄影机录制下来的时候，由于拍摄距离不定，拍摄角度的变化，物体在被拍摄或者录制的尺寸大小也就不一样。这种问题在图像处理过程中，是一个经常需要被重视及急待解决的问题。为了解决这个尺度多样化的问题，可以先对原始图像进行金字塔操作，或者对图像进行多尺度放缩，以解决图像目标多尺度化的问题。

对于输入的彩色图像，我们先规定了一系列的尺度放缩比例，,对于原始图像，先将图像按照放缩比例进行缩小，这样就把原始图像扩展到6个尺寸了。可以在一定程度上解决现实生活中图像中目标多尺度的问题。

### 3.1.3 图像分割

在得到了不同尺度的图像之后，我们需要对多尺度图像进行分割以得到不同的图像块。图像分割是指利用图像的底层信息，如颜色信息、轮廓信息或亮度信息来将图像划分为多个区域。这些区域必须满足两个条件：一是同一个区域内具有相同的颜色或亮度信息，且该区域具有一个独立的外围轮廓；而是不同的区域之间则存在很大的差异。现在图像分割的技术已经很成熟，并且分割效果也都相当不错。

这里简单介绍一下现有的常见分割算法。（1）基于图像阈值的图像分割，这种分割方法原理和计算过程都比较简单，主要是先对图像设定好几个阈值，阈值可以由图像的亮度、图像颜色或灰度来设定。让后依次将图像中的所有像素点与这些阈值作比较，把同一个范围内的像素点划分到一个区域，这种方法最后往往需要通过膨胀或腐蚀操作来优化一下；（2）基于边缘的图像分割，这种分割算法中最具有代表性的就是区域生长法。所谓基于边缘是指利用图像中的轮廓信息来区分图像中的像素点，一般处于同一个闭环内的像素点往往具有一定的相似性，而处于轮廓两边的像素点往往有一定的差异。所以可以先根据图像的亮度或颜色信息来得到图像的轮廓信息，然后将同一个轮廓内的像素点划分为一个区域；（3）基于能量泛函的图像分割，这种方法主要是先定义一个能量泛函，其自变量由图像的边缘组成，是在基于边缘的分割算法的基础上发展而来的。通过求能量泛函的最小值来得到图像的边缘，从而来分割图像。

但是在本文中，对这些多尺度的图像块进行分割只是中间一个过程，所以和以往的分割要求就有些不一样。本文中的分割块应该是包含一个目标，或者包含半个目标，或者是包含多个目标的块，在后面的步骤中，我们还需要进一步的对该块进行评比打分。对图像进行分割有很多现成的比较好的方法，如J.R.R. Uijlings2012年在IJCV上发表的了一篇文章“Selective Search for Object Recoginition”【14】中提出了一种选择性搜素算法，其实现过程是先利用图像的基本特征将图像粗糙的划分为很多小的区域集合，然后利用文章中提到的合并策略反复合并相似的区域，并将这些区域添加到区域集合，直到达到文章里的结束条件，这样可以得到一个包含各种大小划分的区域集合，最后对这个区域集合进行打分排名，挑选出其中分数较高的划分区域。Alexe【15】也给出了一种利用图像频谱的方法来对图像分割，具体思路是先利用X. Hou【16】中的方法对图像进行频谱分析，对图像进行二维离散傅里叶变换之后，利用谱残差的思路得到一个特征图，然后根据特征图对原始图像进行分割，得到很多可能包含目标的小窗口。本文中，我们是采用的方法【15】来对图像进行分割的。

### 3.1.4 对分割窗口进行打分

在前一小节中，我们已经对多尺度的图像块进行了分割，进而由每一尺度图像都能得到多个分割窗口，接下来比较关键的就是对这些图像分割块进行打分。打分的目的是来评价该窗口包含目标的可能性，意图是如果一个窗口很好的包含一个目标，那么该窗口的分数会比较高；但是如果一个窗口中没有目标，那么窗口的分数会比较低。打分的依据便是前面所说的一个目标往往是具有一定的特征，如颜色和周围区域不一样，或者目标往往是具有一定闭环的区域等等。本文中主要从三个角度上来对分割后的窗口进行打分，分别是颜色对比，边缘密度，超像素跨越。接下来对这三点给出介绍分析：

（1）颜色对比

颜色对比（Color Contrast, CC）是说图像中的一个目标往往和周围背景有着不同的外观及颜色分布。根据这一点，如果一个窗口完整包含一个目标，则颜色特征会给出比较高的分数，如果一个串口只包含一半的目标，则分数会有所降低，但一个窗口只包含背景的话，则该窗口的分数会很低。如图所示：在（a）中，若窗口恰好把羊包围住，则该窗口的分数会比较高，当窗口不仅包含了羊，还包含了很多背景草坪时，窗口的分数就会降低。

（a） （b）

在计算窗口w的CC时，我们需要先得到包含该窗口w的一个矩形外环，然后用外环块减去中心块，便得到该窗口w对应的背景区域，这里叫做。如图（a）所示，中心蓝色框为我们将要计算的目标窗口w，黄色框为窗口w对应的矩形外环，则w的背景区域便为黄色框以内，蓝色框以外的区域。具体操作是先将窗口w沿着四个方向各自延伸的尺度，得到对应的矩形外环，这时候会有。计算窗口w和其背景区域的CC特征时是按照颜色直方图的卡方距离来计算的，具体公式如下：

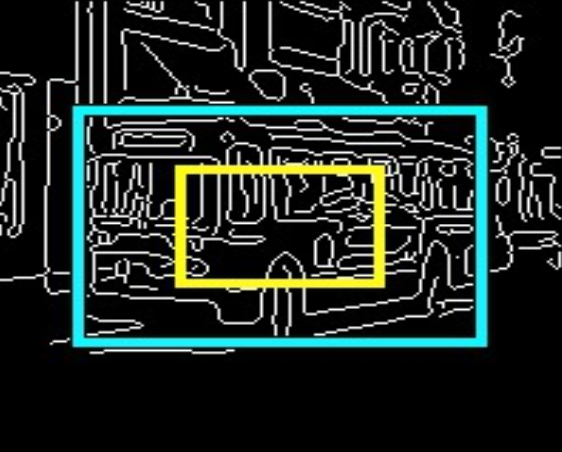


其中，表示延伸尺度，表示卡方距离，表示窗口的LAB（颜色模型）直方图。

CC特征的思想和文献【17】中提到的中心-外围直方图很相似，只是文献【17】中是计算一个像素点的中心区域和外围区域的直方图距离，而本文中的CC是为了得到一个窗口w是否包含目标，是计算本窗口和外围区域的直方图距离。

（2）边缘密度

边缘密度（Edge Density, ED）是指测量窗口w边界的边缘线密度。往往一个目标中边界边缘的数量和目标的边界周长是成正比例关系的，也就是说，随着目标边界周长的增加，其内部边缘的数量也会随着增加。

在计算ED时，我们需要先得到该窗口w的一个矩形内环，这里叫做，具体操作是将窗口w沿着四条边向内收缩的长度，这时候会有，则ED计算公式如下：



其中，表示矩形内环的周长，是一个二值化的值，是统计矩形内环区域内边缘线的个数，其边线先检测时用Canny算法实现的。

1. 超像素跨越

超像素跨越（Superpixels Straddling, SS）是得到目标边界特征的另外一种方法【18】。利用超像素可以根据图像的颜色信息或者纹理信息将图像分割成均匀的小区域，超像素的一个关键特点是可以保留目标的边界，即超像素中的所有像素点都属于同一个目标【19】，所以图像中的一个目标通常可以被分解成几个超像素。本文中根据超像素的这个特点来确定一个窗口是否包含一个目标。对于一个超像素，如果超像素中有一个像素点在该窗口内，且有一个像素在窗口外，就称这个超像素跨越了该窗口。同时一般一个窗口包含一个目标，则该窗口内大部分的面积会被一个超像素所占领。反之，如果一个窗口里面不包含目标，或者包含多个目标的一部分，则该窗口内部会有多个超像素点，同时这些超像素点会跨越该窗口。

在计算SS时，实际上是在计算窗口w内超像素的跨越程度，其计算公式如下：



其中为超像素的集合，该集合是利用文献【18】中的方法得到的，是计算时用到的分割参数；是用来计算该超像素点在窗口w外面的面积，是计算该超像素点在窗口w内部的面积；和中的最小值与窗口w的面积比值来衡量该窗口中超像素跨域的程度。所以越大，说明该窗口w内的超像素越完整，窗口内包含完整目标的可能性越大，反之，说明该窗口w内的超像素跨越情况较多，窗口内包含部分目标或者不包含目标的可能性越大。

### 3.1.5 计算每个像素点的目标性值

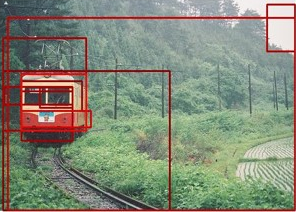
在上一小节中，我们对每一尺度图像中每个窗口从三个角度进行打分，分别是颜色对比，边缘密度，超像素跨越。有了各个窗口的分数之后，就可以计算原始图像中每个像素点的目标性值。该步骤的实现思路比较简单，首先我们需要遍历图像中的每个像素点，对每个像素点按照相同的方式得到目标性值。对于图像中任意一个像素点x，实现公式如下：

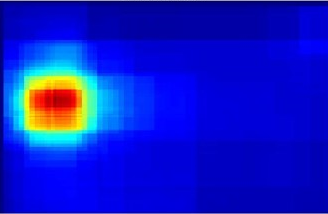
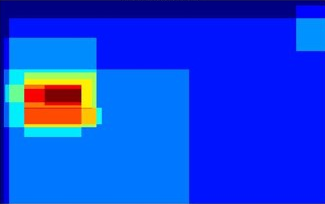
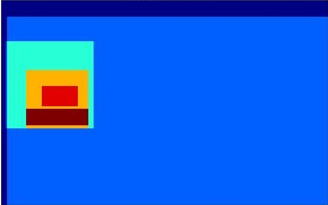
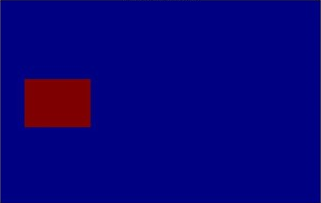


其中是包含像素点x的所有窗口；是通过颜色对比，边缘密度，超像素跨越三个特征来对窗口w给出的分数；是归一化因子。因为一张图像对应的目标性图，其目标性的范围在[0,1]分布区间，值越靠近于1，说明该点的目标性越大，值越靠近于0，说明该点的目标性越小，所以这里会将求得的目标性归一化到[0,1]区间。

通过上面介绍的5个步骤，我们就可以得到输入彩色图像的目标性图。在实验过程中，我们是可以改变算法过程中窗口划分的的总个数，当窗口个数较小时，得到的目标性效果偏差，但算法的计算时间较短；当窗口个数较大时，得到的目标性效果会变好，但算法的计算复杂度会提高，计算时间比较长。这里给出图像目标性的检测效果。







=1 =5 =10 =1000

如上图所示，第一行为原图，第二行对不同下由图像分割得到的分割结果，第三行为分割图对应得到的目标性图。可以发现随着的增加，得到的目标性效果图会越来越好，当=1时，通过分割算法我们只得到了一个窗口，所以整张图像的目标性图中，所有像素点的目标性值只有两个，随着的增加，像素点的目标性分布会越来越大，当=1000时，整个图像的目标性图已经能够很好的反应原始图像中的目标了。在本文后来的实验中所用到的图像的目标性都是在=1000的情况下得到的。

## 3.2 获取图像的显著性估计

在第二章相关背景知识介绍中，我们已经很清楚了介绍了图像显著性的原理和意义，以及常见的显著性检测方法。本节中会详细介绍本文中所用到的显著性检测方法。在介绍自动编码器的时候我们提到自动编码器主要有特征提取，降维及升维的作用，同时，自动编码器还有一个特点就是网络的输出和输入相同，利用这一点可以实现对输入信号的重构，不管输入信号是图像信息，还是语音信号，只要我们设计好对应的网络结构并用足够的样本训练，就能达到我们想要的重构结果。同时在第二章也提到，自动编码器之所以能够利用重构特点来实现对输入图像的显著性分析，主要是因为一般一张图像中背景部分，也就是非显著部分，每个像素点所对应的外围块和该像素点的中心块比较相似，所以通过自动编码器重构出来的中心块和原始的中心块之间的残差会比较小；反之，对于图像中的前景部分，也就是显著部分，每个像素点所对应的外围块和该像素点对应的中心块之间一般会有比较大的差异，这时候利用自动编码器重构出来的中心块和原始像素点的中心块之间的残差就会比较大。这也是利用自动编码器来实现显著性分析的主要原因。

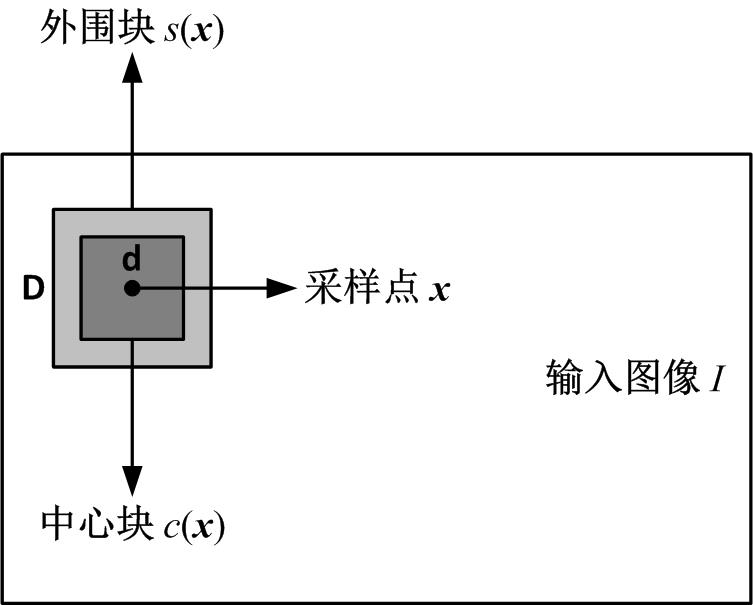
Chen Xia在文献【4】中给出了一种利用变形的自动编码器实现显著性检测的方法，本文中显著性检测部分也是基于该篇文章实现的。其基本思想是先在一张图像中随机采样多个像素点，然后获得每个像素点的外围块作为输入，得到该像素点的中心块作为标签，以此来构建训练样本；接着利用这些训练样本来训练搭建的自动编码器；当网络结构训练结束后，再遍历该图像中的所有像素点，计算由像素点外围块重构的中心块与该像素点实际的中心块之间的残差，来得到该图像的显著性图。基本实现框架图如下所示：



### 3.2.1 全局采样得到训练样本

深度学习总的可以分为三类，分别是监督学习，非监督学习及半监督学习。所谓监督学习是指在训练网络时即需要样本，还需要知道对应样本所属的类别；而非监督学习在训练网络时是不需要知道训练样本的标签的；半监督学习也可以称为强化学习，可以在训练过程中来增强标签。自动编码器则是非监督学习中的一种，因为在训练图像显著性网络结构时，我们是不需要额外提供对应图像的显著性图的。但不管哪种类型的网络，训练样本都是必须的，而且还需要足够的样本。

所以这里的第一步便是得到足够多的训练样本。对于输入图像I，先将图像I从RGB颜色空间装换到LAB颜色空间。转换空间后就可以开始对图像采样获取训练样本了。对于图像I，我们先从图像中随机采样个像素点。这里利用随机采样可以保证训练样本是在图像I中均匀分布的，这样训练好的网络结构对整个图像的各个部分都能重构的比较好。对于中的每个像素点，先得到以像素点为中心，边长为D的外围块，然后再得到以像素点为中心，边长为d的中心块，其中D>d,具体采样方式如图所示：

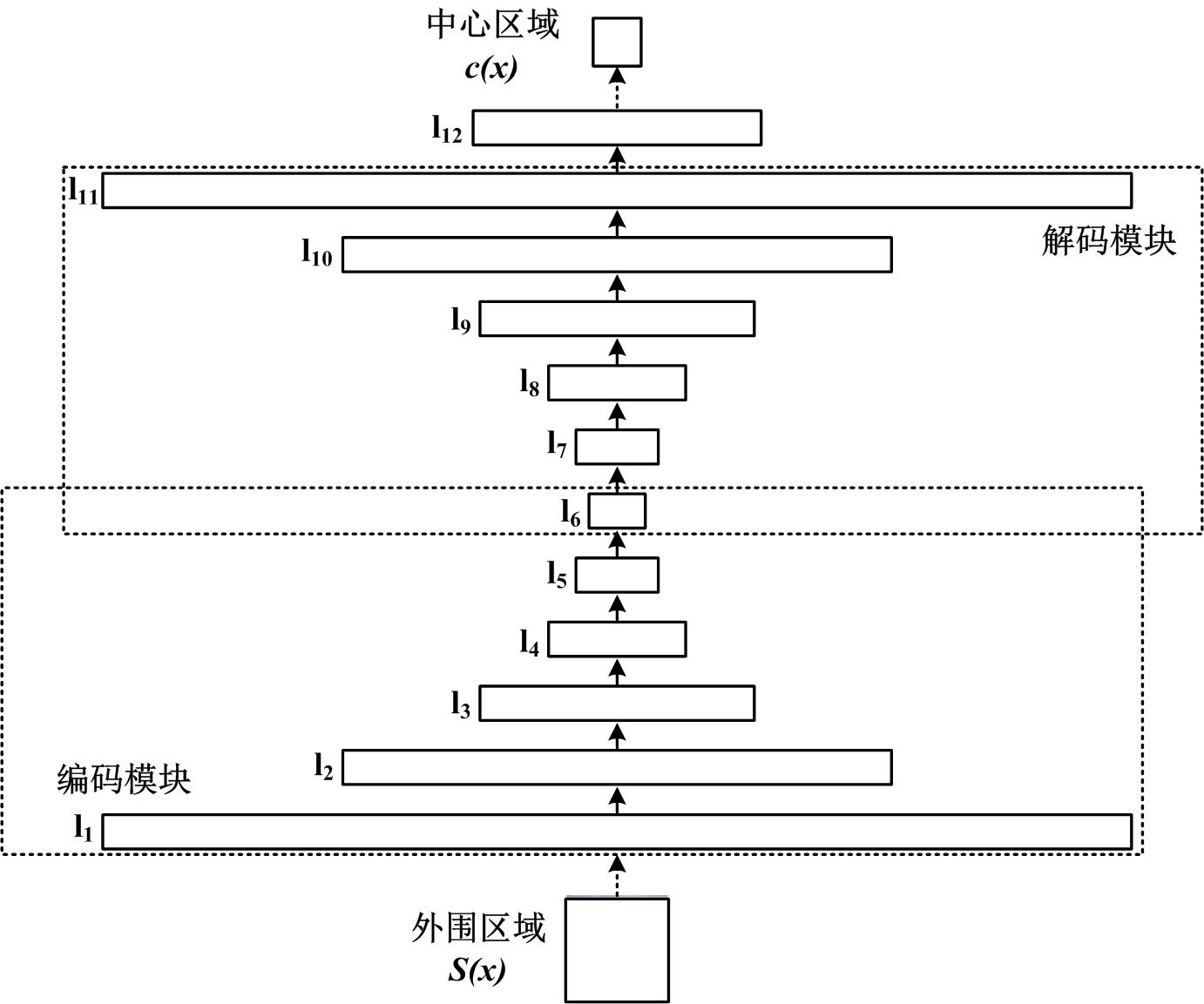


在具体的实验过程中，像素点x的外围块边长D的取值为15；，像素点x的中心块边长d的取值为7，图像I中总的采样个数的取值为8000；采样完成之后，还需要对这些样本进行归一化处理。归一化处理是深度学习中对样本预处理时采用的最常见的方法之一。

### 3.2.2 构建并训练自动编码器网络

一般的自动编码器是由对称的RBM叠加组成，但是在本文中，我们需要在自动编码器的网络结构上利用像素点x的外围块重构该像素点对应的中心块，这就意味着这里的自动编码器输入维度和输出维度不同，和传统的自动编码有一定的区别和变形。为了解决这个问题，就需要在传统意义上的自动编码器后面添加一个输出层，这样导致该自动编码器的训练过程也有稍微的不同。

1. 构建自动编码器



如图所示，整个自动编码器是由12层网络构成的，其中前11层是一个典型的自动编码器结构，其中到为自动编码器的编码模块，是由5个RBM叠加构成；到为自动编码器的解码模块，由编码模块的5个RBM对称叠加构成，其中编码模块与解码模块关于中心层成对称结构。最后一层是为了让整个网络能够重构出像素点的中心块而在自动编码器的后面自主添加的一层。该网络的输入是像素点x的外围块，输出是像素点x的中心块；每一层网络的神经元个数依次为，由于输入的外围块是边长为D的正方形彩色图像块，其中D为15，图像块为3通道，所以第一层的神经元个数为675；最后一层为重构的中心块，其边长d为7，图像块同样为3通道，所有最后一层的神经元个数为147；第二层到第六层的神经元个数依次为256,128，64,32,8；由于第六层到第十一层与前面层相对称，所以神经元个数依次为8,32,64,128,256。

1. 网络参数初始化

在第二章中我们已经介绍了自动编码器网络参数的初始化过程，即利用RBM来对自动编码器的每一层进行初始化。在本自动编码器中，前十一层网络以第六层成对称结构，所以只需要利用RBM结构来得到前六层的初始化参数，其后面对称的六层只需要将其对称层的参数直接转置就可以得到。这里网络的前6层是由5个RBM叠加得到的，采用CD算法便可以得到前6层的初始化参数，这样，后面对称层的初始化参数也就随之得到了。对于最后一层，由于是为了使得网络的输出维度和像素点的中心层一致才加上去的，所以在初始化的时候也比较特殊，这里无法使用RBM来初始化，采用的是一般深度学习中的参数初始化方法，利用均值为0，方差为1的高斯分布来随机初始的。

1. 微调网络参数

该网络经过步骤（2）的初始化之后，已经有了一些拟合能力，即已经有了重构的一些可能，但是与我们期望中的重构能力还是有差距的。所以接下来需要对整个网络的参数进行微调，所谓微调，就是在参数初始化的基础上进行轻微调动，其目的一是为了让整个网络的参数联合起来，更加适应训练样本；二是增强网络的重构能力。整个网络的微调过程是基于梯度下降算法实现的。

具体操作是，首先将训练样本分成多批，然后每一次拿一批样本来训练网络，将像素点的外围块输入到网络，然后依次沿着网络前向传播，直到最后一层输出为，其中为该网络根据外围块重构得到的中心块，其维度和原本的中心块一致，但是由于该网络还处于训练阶段，所以和必然存在很大的交叉熵误差。网络微调的过程本质上就是减少这个误差的过程，其中误差计算方法如下



其中i表示最后一层网络中的第i个节点；为log函数。

接下来需要采用梯度下降算法，将交叉熵误差沿着网络依次反向传播，同时计算误差在每一层参数上的梯度，这里一般是链式求导过程。每一层参数有了对应的梯度误差之后，就可以根据梯度误差来更新了。反复前面的前向传播、误差计算、反向传播、梯度计算及参数更新，直到整个网络最后的交叉熵误差很小时，说明该网络的参数已经训练好，网络已具备很好的重构能力。这时候整个自动编码器网络训练结束。

### 3.2.3 对图像进行显著新估计

当自动编码器网络的参数学好之后，该网络对图像中的背景部分已经有了很好的重构能力，但对于前景模块，重构误差会比教大。这是因为整个网络在训练过程中，训练样本是随机采样的，而对于一整张图像而言，图中大部分的区域都是非显著区域，只有很少部分才是显著区域，所以整个网络的参数基本是基于非显著区域的样本学到的，因此对图像非显著区域就有很好的重构能力，重构误差小；对于图像中显著区域的重构能力差，重构误差大。

（1）估计图像的重构误差

在对图像I进行显著性估计的时候，需要遍历图像中的每一个像素点，对于每一个像素点x，依旧如训练网络样本采样一样的方式，分别得到该像素点x对应的外围块及该像素点x对应的中心块；然后将外围块输入到训练好的自动编码器中，沿着网络方向前向传播直到最后一层，最后得到由该网络重构的中心块；这时候就可以计算重构中心块与原始中心块之间的残差，其残差计算方式如下：



其中，为当前像素点对应的重构误差，为向量的范数。

（2）计算图像中心先验

通常步骤（1）中的重构误差可以直接作为图像的显著性估计，但是为了使得显著性估计能够更加符合人类的视觉注意力机制，这里又计算了图像的中心先验。所谓中心先验是指与原图像同样大小的概率图模型，其作用是越靠近图像的中心区域的像素点，该点处对应的中心先验越大，越靠近图像边缘的像素点，该点出对应的中心先验越小。中心先验其作用机制是，一般对于一张图像，人们往往会潜意识的先关注图像的中心部分，然后才会向周围延伸。所以这里在求图像显著性的时候人为加入了中心先验。一般一张图像的中心先验可以通过一个高斯模板来实现，计算公式如下：



其中，为图像中心的像素点，为图像I中任意像素点到中心像素点距离的平方，为设定好的方差，为指数函数。

同时中心先验也可以采用距离优先函数来实现，其计算公式如下：



其中，为图像中心的像素点，为图像I中任意像素点到中心像素点的距离；本算法中采用的该方法来得到的图像中心先验。

（3）由重构残差和中心先验计算图像显著性

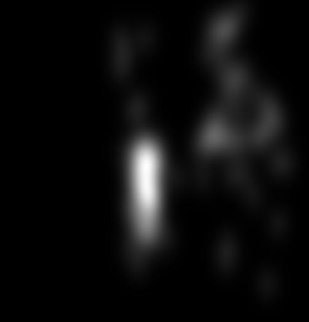
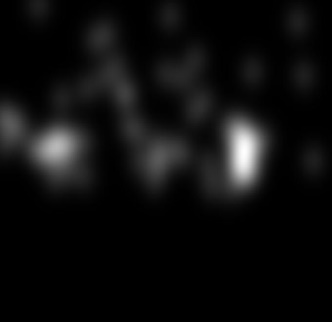
得到了图像的重构残差和中心先验之后，其显著性值就很好计算了，这里是直接将两个模板对应的值相乘就可以了，其计算公式如下：

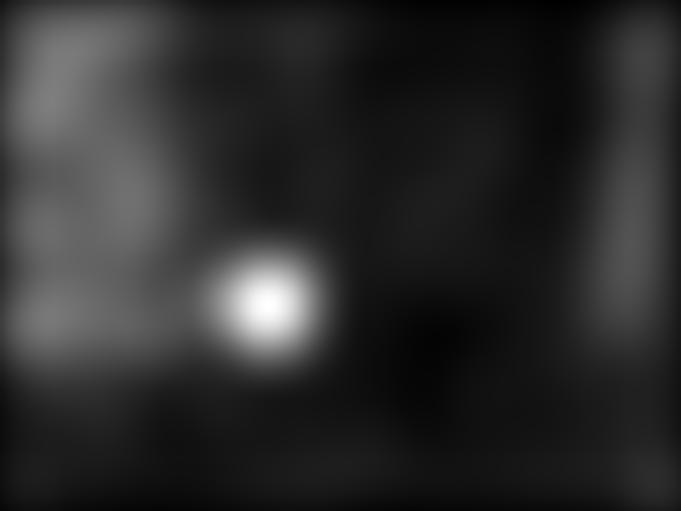


其中表示图像I的显著性图，为一张和输入图像I同样大小的灰度图，取值范围为[0,1]，其值越靠近1的话，表明该点处的显著性值越大，越靠近0的话，表明该点处的显著性越小。

由该方法得到的显著图如下所示：

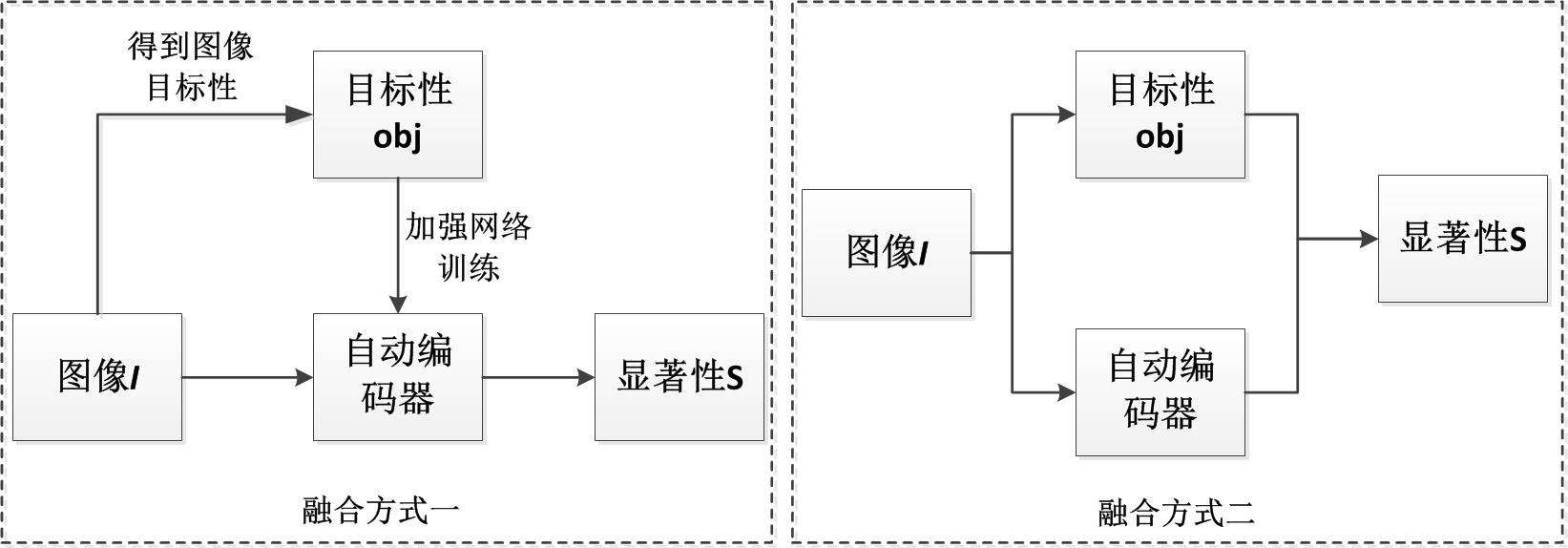
其中第一行为原图，第二行为原图对应的真值图，第三行为由自动编码得到的显著性图。

## 3.3 目标性与显著性估计的融合

在前面两小节中，我们分别单独给出了图像目标性及图像显著性估计的检测方法，其中显著性估计时采用的是利用深度学习中的自动编码器实现的【4】，是一种无监督的检测方法，该方法在现有非监督检测方法中的效果已经不错，比普通的非深度学习的方法结果已经要好。同时前面在显著性检测方法的介绍中提到，一般显著性检测可以分为两大类：一类是自下而上的以数据为驱动的显著性检测方法，该方法主要是结合图像底层特征，如方向梯度、局部颜色对比、颜色强度对比，物体轮廓及超像素来综合得到图像的显著性；另一类是自上而下的以任务目标为驱动的显著性检测方法，该方法主要是结合人类的视觉认知功能，如首先给出图片中含有人、汽车、动物等高层语义信息，再结合高层语义信息来得到图像的显著性。而现有显著性检测的方法中，一般是采用自下而上的以数据为驱动的检测思想，很少有用到结合语义信息的自上而下的显著性检测方法。而在本文中，最终目标便是探讨如何利用上层语义信息来提高图像的显著性检测方法，即采用自上而下的显著性检测思想。

这里所用到的上层语义信息便是指的图像的目标性，因为目标性本意上是指一张图片中，一个区域或者像素点被一个目标包含的可能性大小。而图像显著性是指一个图片中，最能引起人们视觉注意力机制的目标，那么既然显著性大的地方必然是一个物体或者目标，那么以图像的目标性概念为依托来求图像的显著性，也就相当于利用了图像中上层的语义信息。前面我们也得到了整个图像中每个像素点的目标性，接下来就是看怎么将目标性这种概念融合到图像显著性检测中去。

本文中主要采用了两种融合方式：第一种是利用图像的目标性来增强自动编码器的学习过程，即在训练自动编码器时，将图像目标性加入进去，来引导自动编码器参数的学习；第二种方式是将图像的目标性和利用自动编码器得到的显著性相互融合，以得到更好的显著性。简单的框架图如下所示：



### 3.3.1 目标性增强自动编码器的学习

前面在介绍自动编码器学习的过程中，我们首先需要在输入图像中随机采样8000个点，对每个像素点x，都需要得到它对应的外围块和中心块来构成训练样本。由这些点训练好的自动编码器对图像中非显著区域的像素点重构残差很小，而对于图像中显著区域的像素点重构残差比较大。之所以会这样，是因为虽然这8000个样本是在图像中随机采样的，也就是各个点的采样概率都一样，都可能成为训练样本，但是由于图像中非显著部分的区域远远大于显著区域部分，而非显著性区域部分中，像素点的外围块和中心块是很相似的，所以由这些样本学习到的网络只会对非显著部分重构的很好。其实在样本采样的过程中，对于显著区域的像素点采样得到的样本对于自动编码器的学习是不利的。因为这些像素点外围块和中心块本来差异就比较大，当训练网络时强制自动编码器通过外围块重构的中心块和原始中心块残差较小，反而会让自动编码器朝一个不好的方向学习。幸好所有采样样本中，处于显著区域的像素点比较少。

前面在训练网络时之所以会选择随机采样，是因为采样时我们不知道图像中哪些区域属于显著性区域，所以只能选择随机采样。但当我们得到了输入图像对应的目标性图，也就是清楚了图像中哪一部分属于目标性的可能性比较大，换而言之，就是知道了图像的上层语义信息。那么就可以利用该语义信息来指导我们的采样过程，也就可以间接提高自动编码器网络的训练过程。

具体实现方式是，首先对于输入图像I，我们先利用前面介绍的方法得到该图像对应的目标性，其中目标性上每个像素对应的值均在[0,1]之间。接下我们仍然从图像I中随机采样8000个点，但是对于每一个点，我们会随机生成一个[0,1]之间的随机数，如果大于，则保留这个采样点；反之，如果小于，则舍弃这个采样点。这样则使得得到的8000个样本基本是从图像I中非目标区域得到的。然后利用该样本来按照前面的方式来训练自动编码器。

### 3.3.2 目标性和显著性相互融合

第一种方式是在自动编码器训练之前就加入了图像的高层信息，即图像的目标性。第二种方式则是将通过自动编码器得到的残差图和该图像的目标性图直接融合来得到图像的显著性图。相比于第一种利用目标性来指导自动编码器的训练过程，这种方法将高层语义信息用的更为直接。因为在这种融合方式下，图像的目标性直接参与了图像的显著性评估。从原理上来讲，一张图像中越显著的地方，代表该处是目标的可能性越大，因为显著性是根据人类的视觉注意机制得到的，而人眼一定时会关注图像比较感兴趣的目标，而不会首先关注非目标区域。这就说明如果我们知道了该图像中的所有目标信息，目标分布或者目标位置，该处的视觉显著性也越可能比较明显。所以在第二种融合方式中，我们直接对图像的目标性值和图像的残差图进行比较融合。

具体实现方式是对于输入图像I，先求得该图像的目标性图，同时得到该图像经过自动编码器之后的残差图；然后在这个两个特征图上做显著性的探讨，具体探讨过程我们会在第四章实验部分给出详细过程及结果。

## 3.3 本章小结

该章是本论文中算法的原理部分，着重讲解了图像的目标性及图像的显著性获得过程，然后又引出了图像目标性及显著性的融合方式。获得图像的目标性的时候，我们将图像三通道分离，然后对每个通道分别多尺度化以适应图像目中目标的多尺度问题，接着再对每个尺度进行图像分割并对分割后的图像块窗口根据颜色、轮廓、超像素等底层特征来打分，最后根据分割后的窗口及窗口的分数来求得整个图像的目标性图；获得图像的显著性的时候，我们先在原始图像中进行随机采样以得到足够多的训练样本，接着由多层RBM构建一个自动编码器，并在自动编码器后面添加了一层重构层以适应网络输入和输出维度不同的问题，然后利用样本来反复训练以优化该自动编码器网络参数，最后利用训练好的网络得到输入图像对应的显著性图；本章节的最后简单的引出了如何利用图像的高层语义信息，即图像的目标性来提高图像的显著性估计，并给出了图像目标性及显著性的两种融合方式，具体实验过程将会在下一章中详细讲解。

# 第四章：基于目标性的显著性实验与分析

第三章中已经详细给出了本论文中算法的原理部分，在本章中，我们会着重分析在该算法下的实验探讨过程及实验结果。在此之前，我们会先引出本文中实验所用到的数据库及显著性检测的评价指标。

## 4.1 实验数据库介绍

由于显著性这个课题已经研究了比较长的时间，所以现在已有很多的相关数据库在网络上公开，以便大家在研究这个课题的时候有数据支撑，同时可以在同一个数据集上同其他人的算法做分析比较。一般公认的显著性图像数据库会包含各种场景下的图像以检验算法的鲁棒性及通用性，其显著性真值的计算一般是利用眼动仪来追踪观察者的眼球移动轨迹来得到的。具体过程是，先找到多个实验者，因为每个人的视觉注意机制都会不一样，对图像中的关注点也就不一样，所以这里需要多个实验者，然后对多个实验者的数据求平均以得到图像的显著性真值；对于每个实验者，让其距离图像一定的距离来观察图像一定的时间，同时观察不同的图像时还需要保持一定的间隔。这一步中影响显著性真值的因素主要有三个，分别是观察者距离图像的距离，每张图像的观察时间及两张图像切换时的时间间隔；在实验者观察图像的过程中，会用眼动仪来实时跟踪实验者的眼球在图像中的移动轨迹及停留时间，有了这些数据之后，我们就可以根据停留时间及轨迹图得到观察者在图像中的扫视点，最后将该扫视点进行高斯模糊之后便得到了图像的显著性真值图。

本文中的所有实验都是基于两个显著性检测图像数据库来进行的，这里分别记做DS1【20】，DS2【21】，数据库的部分信息如下表所示：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 数据库 | 数量 | 观察者 | 图像大小 | 观察距离（cm） | 观察时间/间隔（s） | 备注 |
| DS1 | 120 | 20 | 681×511 | 75 | 4/2 | 用的最多的数据库，包括室内及室外多种场景图 |
| DS2 | 1003 | 15 | 多种尺寸 | 60 | 3/1 | 最新的数据库，包含779幅景观图及228幅肖像图 |

其中，DS1是由Bruce and Tsotsos [20]提出来的，该图像库中总共有120幅图像，是显著性算法研究过程用的最多的一个库，其中的图像主要包含室内与室外多种场景。图像库中所有的图像均为长为681，宽为511的固定尺寸，图像的真值是由20个实验者实验得到的。获取真值时，观察者距离图像的距离为75厘米，每张图像观察4秒钟，图像切换时，需要停留2秒钟。DS2是由MIT大学的Judd et al【21】提出来的，该图像库中包含1003幅图像，是显著性检测领域中比较新的一个数据库。其中的图像是包含779幅各种场景图及228幅各种肖像图，图像库中所有的图像尺寸并没有统一大小，最大的维度为1024个像素点，其它维度分布在405到1024像素点之间。图像的真值是由15个观察者得到的。获取真值时，观察者距离图像的距离设定为60厘米，每张图像观察3秒钟，图像切换时，需要停留1秒钟。

## 4.2 实验评价方式介绍

每个实验过程中都对结果与真值给出评价指标，来衡量实验结果与真值之间的差异，以及每个实验算法的好坏。常用简单的评价指标有准确率，召回率，F-Score,复杂一点的评价指标有ROC，AUC等等。在本实验中所使用的评价指标主要是两种AUC，分别记做AUC\_Judd和AUC\_Borji，接下来简单介绍一下AUC的评价方式。

ROC（Receiver Operating Characteristic）中文名为受试者工作特征曲线，主要用来评价一个二分类问题。

|  |  |  |  |
| --- | --- | --- | --- |
|  | | 预测值 | |
| 正样本 | 负样本 |
| 真实值 | 正样本 | 真正类（TP） | 假负类（FN） |
| 负样本 | 假正类（FP） | 真负类（TN） |

如表所示，对于一个二分类问题，主要会出现下面四种情况：

真正类（TP）：即一个正样本被算法预测为正样本；

假正类（FP）：即一个负样本被算法预测为正样本；

真负类（TN）：即一个负样本被算法预测为负样本；

假负类（FN）：即一个正样本被算法预测为负样本；

从上述四种情况中又可以得到两个新概念：真阳率（TPR）和假阳率（FPR），真阳率表示被算法预测为正样本占所有正样本的比例，假阳率表示算法错认为正样本的负样本占所有负样本的比例：





这里当二分类的阈值连续变动时，ROC正是一条随着阈值变动的曲线，其中横轴为假阳率，纵轴为真阳率。

AUC（Area Under Curve）是指ROC曲线下的面积。在实验过程中，我们总是期望真阳率越高越好，假阳率越低越好，但是随着二分类阈值的连续变动时，真阳率和假阳率总会同时的提高或降低，不会两个指标都变好，所以AUC便是兼顾真阳率和假阳率的一种中和方式。

## 4.3 实验过程及分析

在清楚了本论文的数据集及评价指标之后，接下来就是实验部分。该部分也是本论文的主要实践模块，主要由四个实验构成。实验一，利用第三章中所提到的由自动编码器对输入图像重构得到的残差图加上中心先验模板来进行显著性估计；实验二，利用以图像目标性为先验知识来监督自动编码器的训练样本的采样，然后利用采样样本训练自动编码器，并得到图像的显著性来进行估计；实验三，先获取图像的目标性及图像经过自动编码器得到残差图，然后对目标性及残差图直接融合来进行显著性估计。接下来是对实验的结果及详细分析模块。

### 4.3.1 实验一：重构残差进行显著性估计

实验一是利用自动编码器对图像的重构残差来进行图像显著性估计。

体实验流程是对于DS1及DS2图像库中的每一张图像，按照3.2小节中给出的显著性求解流程，先在图像中随机采取8000个样本，利用该样本集训练构建好的自动编码器网络，然后再依次遍历图像中的每一个像素点x，将像素点的外围块输入到网络重构出该像素点的中心块，并得到重构残差，同时求出该图像的中心先验模板，最后结合重构残差和中心先验模板得到该图像的显著性。得到了DS1及DS2图像库中每一张图像的显著性之后，接着是将得到的显著性与图像库给出的真值进行比较，这里是采用的是AUC\_Judd和AUC\_Borji两个评价指标，具体评价方式是先对数据库中的每一张图像给出评价分数，然后再对所有图像的评级分数求平均。下面给出由两个评价指标给出该实验在两个数据库上的显著性检测结果，同时文献【4】中给出了一些其他人在显著性领域的算法及其算法在DS1数据集上的根据AUC\_Judd的评价结果，这里为了作为对比，也给出了他们的结果,如表所示：

表一

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 算法 | GBVS[73] | Wu[74] | CovSal[75] | Ren[76] | Shen[77] | NCSR[78] |
| AUC\_Judd | 0.819 | 0.782 | 0.821 | 0.805 | 0.777 | 0.818 |

表二

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DS1（120幅图像） | | DS2（1003幅图像） | |
| AUC\_Judd | AUC\_Borji | AUC\_Judd | AUC\_Borji |
| 实验一 | 0.796 | 0.786 | 0.776 | 0.770 |

其中，表一中的GBVS【73】、Wu【74】、CovSal[75]、Ren[76]、Shen[77]、NCSR[78]的检测结果是由文献【4】提出的，且该文中只给出了这些算法在DS1图像库上利用AUC\_Judd评价指标给出的评价结果，所以这里无法给出这些算法在DS2上的评价结果及其在DS1上利用AUC\_Borji得到的评价结果。

实验结果分析：（1）由表一及表二中的第二列可以看出，利用自动编码器对图像的重构残差来进行图像显著性估计的思路是比较好的，虽然实验结果比该列中大多的算法要差，但只是差了一点点。这也证明了图像的中心块和周围块的差异是可以表明该处的显著性的。（2）由表二可以看出，不同的评价指标给出的评价结果有差异。（3）同时由表二可以看出，同一个算法在不同的数据库上检测得到的显著性也有差异，说明不同场景的图像显著性是有区别的。

### 4.3.2 实验二：目标性监督网络重构进行显著性估计

实验二是用以图像目标性为先验知识来监督自动编码器训练样本的采样，然后利用采样样本训练自动编码器，并得到图像的显著性来进行估计。

具体实验流程是对于DS1及DS2图像库中的每一张图像，先按照3.1小节给出的目标性的求解方式，得到图像的目标性，然后依据图像的目标性从图像中采样得到8000个样本，具体方式是对于每一个点，我们会随机生成一个[0,1]之间的随机数，如果大于，则保留这个采样点；反之，如果小于，则舍弃这个采样点。有了训练样本之后再按照实验一的方式得到两个图像库中的每个图像的显著性。最后利用AUC\_Judd和AUC\_Borji来得到每个图像库中的显著性的得分。这里给出实验结果：

表三

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DS1（120幅图像） | | DS2（1003幅图像） | |
| AUC\_Judd | AUC\_Borji | AUC\_Judd | AUC\_Borji |
| 实验一 | 0.796 | 0.785 | 0.776 | 0.770 |
| 实验二 | 0.798 | 0.788 | 0.778 | 0.772 |

实验结果分析：由表三可知，对比实验一，实验二在数据集DS1和DS2上的结果会比实验一的结果稍微的好一些，说明利用图像的目标性来指导自动编码器训练样本的采样，使得采样得到的样本基本为图像的非目标区域，由这些样本训练得到的自动编码器对图像的背景部分重构的残差更小，而对图像的前景部分重构的残差会更大些，以此来提高了显著性检测的效果。

### 4.3.1 实验三：目标性和重构残差线性融合进行显著性估计

实验三是先获取图像的目标性及图像经过自动编码器得到的残差图，然后对目标性及残差图直接融合来进行显著性估计。

具体实验流程是对于DS1及DS2图像库中的每一张图像，先按照3.1小节给出的目标性的求解方式，得到图像的目标性；然后再按照3.2小节给出的显著性的求解方式，得到图像的残差图；接着对图像的目标性和残差图进行线性相加来综合得到图像的显著性；具体相加公式如下所示：



其中，a和b分别为图像目标性及图像残差前面的系数，且a和b满足以下条件：



这里实验时采用遍历的方法，对a依次从0.1开始递增，每次增加0.1直到0.9，b则根据限制条件随着a更新。得到了DS1和DS2图像库中所有图像的显著性后，再利用AUC\_Judd和AUC\_Borji来计算每个图像库的显著性得分。这里给出实验结果：

表四

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | DS1（120幅图像） | | DS2（1003幅图像） | |
| a | b | AUC\_Judd | AUC\_Borji | AUC\_Judd | AUC\_Borji |
| 0.1 | 0.9 | 0.807 | 0.797 | 0.780 | 0.774 |
| 0.2 | 0.8 | 0.815 | 0.806 | 0.790 | 0.784 |
| 0.3 | 0.7 | 0.821 | 0.812 | 0.798 | 0.792 |
| 0.4 | 0.6 | 0.825 | 0.816 | 0.804 | 0.798 |
| **0.5** | **0.5** | **0.827** | **0.819** | 0.807 | 0.801 |
| **0.6** | **0.4** | 0.827 | 0.818 | **0.809** | **0.803** |
| 0.7 | 0.3 | 0.824 | 0.815 | 0.808 | 0.802 |
| 0.8 | 0.2 | 0.820 | 0.812 | 0.806 | 0.799 |
| 0.9 | 0.1 | 0.816 | 0.807 | 0.802 | 0.795 |

表五

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DS1（120幅图像） | | DS2（1003幅图像） | |
| AUC\_Judd | AUC\_Borji | AUC\_Judd | AUC\_Borji |
| 实验一 | 0.796 | 0.785 | 0.776 | 0.770 |
| 实验二 | 0.798 | 0.788 | 0.778 | 0.772 |
| 实验三 | **0.827** | **0.819** | **0.809** | **0.803** |

实验结果分析：（1）由表四可知，由图像的目标性和重构残差进行线性融合而得到的图像的显著性在评价指标上有了很好的表现，不管是在DS1还是DS2上， AUC\_Judd和AUC\_Borji两种评价指标下实验三中其最好结果都比实验一高出了3%，这在AUC评价指标中已经算提升了很多。说明利用图像的高层语义信息，确实能够提高视觉图像的显著性估计。（2）由表四可知，随着a或者b的升高，线性叠加得到的显著性会先变好再变坏，一般是a和b接近时得到的显著性效果最好。（3）由表五可知，实验三要比实验二的结果好很多，说明图像的目标性已经有了很多图像上层语义相关的信息，将目标性用在显著性检测算法过程中的后面部分会比用在前面部分效果要好。（4）由表五中的实验三和表一可知，实验三中的显著性通过AUC\_Judd指标得到的评价结果要比其他人的方法要好，再次证明了利用图像的目标性能够提高图像的显著性检测。

通过前面三个实验的对比及分析，总的说明了在对图像进行显著性分析时，加入图像的高层语义信息作为先验知识，是可以提升图像的显著性检测效果的。其次，本文中所用到的先验知识是指图像的目标性，目标性是图像高层信息中语义性很强的一种特征，它能给出图像中目标大概的分布情况，这对显著性检测具有很强的指导意义。因为图像中显著性很强的区域一般也是一个目标或者目标的一部分，如果有了图像中目标的分布情况，那图像中的显著性区域的分布也有了相应的参考，所以利用目标性来提高图像的显著性检测是可行的。

## 4.4 本章小结

本章是该论文的实验部分，首先介绍了实验过程所用到的数据库，并对数据库中的图像场景及显著性真值的获取情况给出了简单介绍；接下来介绍了实验结果的评价方式，有了评价方式之后，实验结果才能得以量化，同时能够与别人的方法进行比较与分析；最后是实验展示与分析模块，该部分是本章的重点，主要从三个实验来分析图像显著性检测效果，即利用自动编码得到的图像残差、将目标性作为先验知识来训练自动编码器以得到的显著性及将目标性与自动编码器得到的残差线性融合得到的图像显著性。实验结果也验证了利用图像高层语义信息能够提高图像的显著性检测。

# 第五章：总结与展望

## 5.1 工作总结

本文主要采样自上而下的显著性检测方法，探讨以图像的高层语义信息作为先验知识，来分析图像的显著性检测。一般的显著性检测是基于自下而上的方法，即以图像中底层的纹理特征、颜色特征、超像素特征为基础，进行多种特征的融合及判断来得到该图像的显著性。而本文首先是求得图像的目标性，一种可以表明图像中目标所在位置的高层语义特征，又利用自动编码器对图像进行自我重构得到重构残差；然后简单的检验了一下图像的目标性及重构残差对图像显著性的贡献；再利用图像目标性这个高层语义特征来监督自动编码器的训练过程，利用训练好的自动编码器重构原始图像，根据重构残差来得到对应图像的显著性；接着又分析了利用图像的目标性与自动编码器得到的重构残差之间的线性叠加来得到显著性。通过多个实验分析，我们验证了利用图像的高层语义特征能够提高图像的显著新分析。

本文实验中的高层语义特征是指图像的目标性，所谓目标性是指图像中的一个像素点或一块区域是目标的可能性的几率。具体获取方式是先将原始图像三通道分离；然后对分离后的每个通道进行多尺度化以适应图像中目标尺度多变的情况；之后对于每个尺度的图像进行分割得到多个图像块窗口；有了窗口之后，我们需要对这些窗口进行打分，来评价该窗口是否包含一个目标或者半个目标或者不包含目标，具体评价指标主要是依赖图像的底层特征，这里分别是颜色对比、边缘密度、超像素跨越三个特征；每个窗口有了分数之后，最后需要遍历图像中的每一个像素点，利用包含该像素点的所有窗口及窗口打分来得到每个像素点的目标性值。

本文实验中的利用自动编码器网络根据图像的外围块重构图像的中心块得到的残差来作为显著性，是指对于每一张图像，先在图像中随机采样8000训练样本，然后利用该样本集训练自动编码器，等网络参数优化以后，再对原始图像逐点扫描，将每一个点的外围块输入到网络并计算网络输出的中心块和理论中心块之间的误差；将误差与图像的中心模板相乘作为该图像的显著性。

本文中实验中利用图像的高层语义特征来提高图像的显著性检测，主要采取了两种方式：一是利用图像的目标性来加强自动编码器训练时的样本采样过程，在采样时，如果该采样点的目标性值比较大，则跳过该采样点，如果该采样点的目标性值比较小，则保留该采样点，这样则能保证所有的训练样本都处于图像中非目标区域，进而可以使得由该样本训练得到的自动编码器对图像非显著区域重构误差小，而对于图像显著性区域重构误差较大。二是直接将图像对应的目标性值及经过自动编码器得到的残差加上中心先验之后进行线下叠加，来作为该图像的显著性。这两种方法都可以改善图像的显著性检测，而第二种方式能够很好的提升显著性检测效果。通过实验证明，利用图像的高层语义信息是可以提高图像的显著性检测。

## 5.2 工作展望

显著性检测是图像处理领域非常有用的一门技术，好的显著性检测算法可以在视频监控或者图像处理上大大缩短处理时间，同时提高处理的效率。但是现实生活中，一张图片中可能包含各种尺寸的目标及复杂的背景，这大大增加了显著性检测的难度。以往的显著性检测算法大多从图像的底层特征，如颜色、方向梯度、轮廓、超像素等方向来实现的，而本文则利用了图像的高层语义特征，即图像的目标性来加强图像的显著性分析，实验结果证明图像的高层语义特征确实是能够提高图像的显著性检测。我们知道，一般人在观察图像时，首先一瞬间会在脑海中自动对图像中的每一部分给出认知的定义，然后才会聚焦自己感兴趣的部分。对于图像的高层语义特征，本文中只用到了图像的目标性，还没有尝试其他的语义特征，所以在后续的工作中，可以尝试别的高层语义特征，比如先给出图像中人或者车的概念或者轮廓信息或者位置信息，然后再以该语义特征为导向，来提高图像的显著性检测。

另外本文中在进行目标性检测或者由自动编码器得到图像的显著性过程中，都是采用的非监督的方法，非监督方法的好处是不需要带有标签的训练样本，所以在进行实验时对数据的依赖程度比较低。但是一般而言，监督学习的的方法都会比非监督学习的方法效果要好，所以后续可以找到合适的数据库，采用监督学习的方法来对图像进行显著性检测，这样势必会得到更好的检测效果。

# 致谢

# 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

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

【13】Training products of Experts by Minimizing Contrastive Divergence

【14】Selective Search for Object Recoginition

【15】What is an object ?

【16】X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In CVPR, 2007

【17】T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum. Learning to detect

a salient object. In CVPR, 2007.

【18】P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. IJCV, 59(2):167–181, Sep 2004.

【19】B. C. Russell, A. A. Efros, J. Sivic, W. T. Freeman, and A. Zisserman.

Using multiple segmentations to discover objects and their extent in

image collections. In CVPR, 2006

【20】N. D. B. Bruce and J. K. Tsotsos, “Saliency based on information

maximization,” in Proc. Annu. Conf. Neural Inf. Process. Syst., Montreal,QC, Canada, Dec. 2005, pp. 155–162

【21】T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to predict where humans look,” in Proc. 12th IEEE Int. Conf. Comput. Vis., Kyoto, Japan, Sep. 2009, pp. 2106–2113

【1-3】L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual

attention for rapid scene analysis,” IEEE Trans. Pattern Anal. Mach.

Intell., vol. 20, no. 11, pp. 1254–1259, Nov. 1998

【1-16】J. Wu, F. Qi, G. Shi, and Y. Lu, “Non-local spatial redundancy reduction for bottom-up saliency estimation,” J. Vis. Commun. Image Represent., vol. 23, no. 7, pp. 1158–1166, Oct. 2012

【1-18】A. Borji and L. Itti, “Exploiting local and global patch rarities for saliency detection,” in Proc. 25th IEEE Conf. Comput. Vis. Pattern Recognit., Providence, RI, USA, Jun. 2012, pp. 478–485.

【1-14】D. A. Klein and S. Frintrop, “Center-surround divergence of feature statistics for salient object detection,” in Proc. 13th IEEE Int. Conf. Comput. Vis., Barcelona, Spain, Nov. 2011, pp. 2214–2219

【1-32】H. J. Seo and P. Milanfar, “Static and space-time visual saliency detection by self-resemblance,” J. Vis., vol. 9, no. 12, pp. 1–27,2009

【2-2】R.Achana, frequency-tuned salient region detection

[2-6] 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

【2-3】SUN: Top-down saliency using natural statistics

【73】J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in Proc. Adv. Neural Inf. Process. Syst., Vancouver, BC, Canada, Dec. 2006, pp. 545–552.

【74】J. Wu, F. Qi, G. Shi, and Y. Lu, “Non-local spatial redundancy reduction for bottom-up saliency estimation,” J. Vis. Commun. Image Represent.,vol. 23, no. 7, pp. 1158–1166, Oct. 2012

【75】E. Erdem and A. Erdem, “Visual saliency estimation by nonlinearly integrating features using region covariances,” J. Vis., vol. 13, no. 4, pp. 1–20, Mar. 2013.

【76】Z. Ren, S. Gao, L.-T. Chia, and D. Rajan, “Regularized feature reconstruction for spatio-temporal saliency detection,” IEEE Trans. Image Process., vol. 22, no. 8, pp. 3120–3132, Aug. 2013

【77】C. Shen and Q. Zhao, “Learning to predict eye fixations for semantic contents using multi-layer sparse network,” Neurocomputing, vol. 138, pp. 61–68, Aug. 2014.

【78】C. Xia, P. Wang, F. Qi, and G. Shi, “Nonlocal center-surround reconstruction-based bottom-up saliency estimation,” in Proc. 20th IEEE Int. Conf. Image Process., Melbourne, VIC, Australia, Sep. 2013,pp. 206–210.