**Revisiting Salient Object Detection: Simultaneous Detection, Ranking, and Subitizing of Multiple Sal 检测，排序和子化**

**问题：**  
作者认为，**显著性目标检测领域**迄今为止的工作解决的是一个相当病态的问题。即不同的人对于什么是显著性目标没有一个普遍的一致意见。这意味着一些目标会比另一些目标更加显著，并且不同的显著性目标中存在着一个相对排名。

**本文方法：**  
本文方法解决了考虑了相对排名这个更普遍的问题，并且提出了适合于衡量该问题的数据与度量方法，本文解决方案是基于相对显著性和分段式细化的分层表示的深度网络。该网络也可以解决显著性目标计数问题。

**主要贡献：**  
1.提出一个端到端网络用以解决多个显著对象检测问题，并根据对象的显**著程度**进行排序。

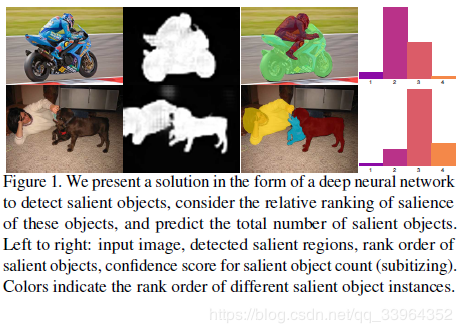
2.提出阶段性细化机制，在该结构上逐步恢复更精细结构的预测。

以下为论文翻译（如有疏漏还望谅解）

**摘要**  
**目标显著性检测**是一个已经详细考虑的问题，并且提出了许多解决方案。在本文中，我们认为迄今为止的工作已经解决了一个相对病态的问题。具体而言，**没有普遍的协议**关于在查询多个观察者时什么构成显著对象。这意味着某些对象比其他对象更容易被判断为显著，并且暗示在显著对象上存在相对等级。本文提出的解决方案解决了考虑相对等级的这个更普遍的问题，并且我们提出了适合于衡量相对对象显著性格局中的成功的数据和度量。基于相对显著性和阶段式细化的分层表示，提出了一种新颖的深度学习解决方案。我们还表明，可以使用相同的网络解决显著对象子化的问题，并且我们的方法超出了所有考虑的所有指标（传统和新提出的）的任何先前工作的性能。

**1.简介**  
显著物体检测中的大多数工作都考虑单个显著物体[37,38,7,8,31,32,9,19,17,24,39,18]或多个显著物体[13,27,36] 但是并不认为突出的东西可能因人而异，某些物体可能会因其重要性而得到更普遍的一致。

缺少数据，包括由多个观察者手工分割的显著对象。 重要的是要注意，由少数观察者（包括一个）提供的任何标签都不允许辨别物体的相对重要性。 基于凝视数据[33]的相对显著性的隐式分配也给出了困难，给出了不同的认知过程而不是涉及手动标记的计算决策[16]。 此外，注视数据对于解释诸如中心偏差，视觉运动约束和其他潜在因素等给定因素是相对具有挑战性的[2,1]。

**图1**.我们以深度神经网络的形式提出一个解决方案来**检测显著对象**，**考虑这些对象显著性的相对排序，并预测显着对象的总数。左向右**：输入图像，检测到的显着区域， 显著对象的等级顺序，显着对象计数的置信度得分（子化）。 颜色表示不同显着对象实例的排名顺序。

因此，在本文中，我们更广泛地考虑了显著对象检测的问题。 这包括**检测图像中的所有显著区域，并通过为不同的显著区域分配置信度来解释观察者间的变异性**。 我们通过进一步处理来**增加PASCAL-S数据集**[23]，以便以相对显著性的形式提groundtruth。 除了传统度量之外，基于显著对象相对于groundtruth排序的等级顺序，针对其他算法测量成功。 最近的工作也考虑了**显著对象的次级化问题**。 我们的观点是，这种确定应该可以**通过提供显著物体检测的模型来实现**（见图1）。 我们还允许我们的网络进行细分。

总的来说，我们的工作概括了显著对象检测的问题，我们提出了一个新的模型，根据这个问题的传统形式，多个显著对象检测和相对排序，以及子化，提供显著对象的预测。 我们的结果显示了所考虑的所有问题的最新性能。

**2.背景**  
**2.1显著性目标检测**  
 **卷积神经网络**（CNN）已经提高了计算机视觉中许多问题的性能标准，包括显著对象检测。 基于CNN的模型能够提取比现代工作中使用的手工制作的功能更具代表性和复杂性的功能[21,34,15]，这些功能促进了广泛采用。

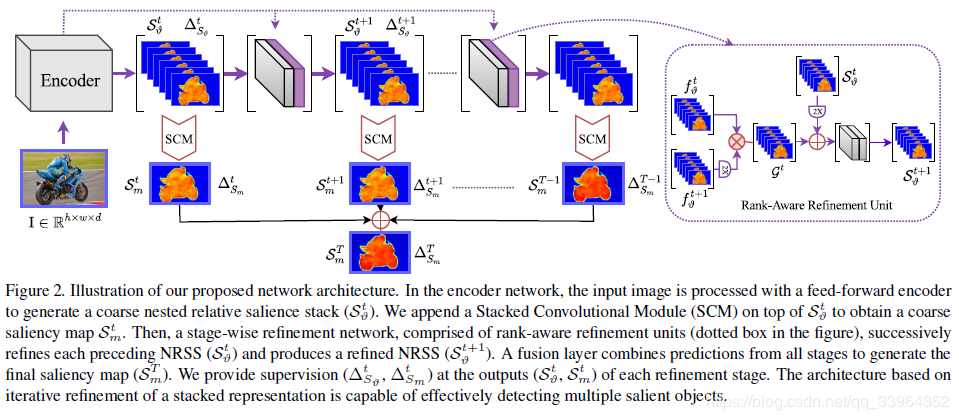
一些基于CNN的方法利用**超像素**和**对象区域提议**来实现准确的显著对象检测 [9,19,17,22,39,18]。 这些方法遵循多分支架构，其中CNN用于跨不同抽象级别提取语义信息以生成初始显著性预测。 随后，添加新分支以获得超像素或对象区域提议，其用于提高预测的精度。

作为超像素和对象区域提议的替代方案，其他方法[26,8,37]通过聚合多级特征来预测每像素的显着性。罗等人。 [26]通过CNN整合本地和全球特征，CNN结构为多分辨率网格。侯等人[8]在浅层和深层特征图之间实现阶段式短连接，以实现更精确的检测，并推断出仅考虑中间层特征的最终显著性图。张等人[37]将多级特征组合为提示，以生成和递归微调多分辨率显着图，这些显著图通过边界保留细化块进行细化，然后融合以产生最终预测。

其他方法[24,31,38]使用端到端**编码器 - 解码器**架构，该架构产生初始粗略显着图，然后逐级细化它以提供显着对象的更好定位。 Liu和Han [24]提出了一种将局部上下文信息逐步与粗略显着图相结合的网络。王等人[31]提出了一种用于显著性检测的循环完全卷积网络，其包括用于校正初始显著性检测错误的先验。张等人[38]在特定卷积层之后引入重新形成的丢失以量化卷积特征中的不确定性，以及用于减少反卷积伪像的新的上采样方法，从而为显著对象检测提供了更好的边界。

与上述方法相反，我们通过应用新颖的机制来控制通过网络的信息流，通过逐步细化来实现空间精确度，同时还重要地包括隐含地携带确定相对显著性所必需的信息的堆叠策略。**2.2显著性目标的细化**  
 最近的工作[35,7]也解决了图像中对显著对象进行细分的问题。此任务涉及计算显著对象的数量，而不管其重要性或语义类别。 [35]中提出的第一个显着对象子网格网络应用前馈CNN将问题视为分类任务。 He等人 [7]通过探索数字和空间表示之间的相互作用，将子化任务与检测结合起来。 我们的建议提供了对显着对象数量的具体确定，识别该数量的可变性，并且还提供输出作为反映这种可变性的分布。

**3.提出的网络结构**  
我们提出了一个新的端到端框架，用于解决检测多个显著对象的问题，并根据对象的显著程度对对象进行排序。我们提出的显著物体检测网络的灵感来自卷积 - 反卷积通道[28,24,12]的成功，其中包括用于初始粗略水平预测的前馈网络。然后，我们提供阶段性的细化机制，在该机制上逐渐恢复更精细结构的预测。图2显示了我们提出的网络的总体架构。编码器阶段用作特征提取器，其将输入图像转换为丰富的特征表示，而细化阶段尝试恢复丢失的上下文信息以产生准确的预测和排名。

我们首先描述如何在3.1节中生成初始粗略显着图。接下来分别对3.2节和3.3节中的

阶段细化网络和多阶段显着图融合进行详细描述。

**3.1粗预测前馈网络**  
最近应用于高级视觉任务的前馈深度学习模型（例如图像分类[6,30]，目标检测[29]）采用由重复卷积阶段和空间汇集组成的级联。通过池化进行下采样允许模型在编码的最深阶段实现具有相对较差的空间分辨率的高度详细的语义特征表示，并且还通过范围大得多的滤波器的空间覆盖来标记。对于识别问题，空间分辨率的损失不成问题;然而，逐像素标记任务（例如，语义分割，显著对象检测）需要像素精确信息以产生准确的预测。因此，我们选择Resnet-101 [6]作为我们的编码器网络（基本构建块），因为它在分类和分段任务方面具有优越的性能。按照像素标记[3,12]的先前工作，我们使用扩张的ResNet-101 [3]来平衡语义上下文和精细细节，从而使输出特征图减少8倍。更具体地说，给定输入图像I\epsilon \mathbb{R}^{h\times w \times d}​，我们的编码器网络产生一个大小为\left \lfloor \frac{h}{8} ,\frac{w}{8}\right \rfloor​的特征图。为了通过自上而下的细化网络扩充编码器网络的主干，我们首先附加一个额外的卷积层和3×3内核和12个通道，以获得嵌套相对显着性堆栈（NRSS） 。然后，我们附加一个Stacked Convolutional Module（SCM）来计算每个像素的粗略水平显著性得分。值得注意的是，我们的编码器网络足够灵活，可以替换为任何其他基线网络，例如： VGG-16 [30]，DenseNet 101 [10]。此外，我们利用金字塔池化[3]来收集更多的全局背景信息。所描述的操作可以表示为：https://img-blog.csdnimg.cn/20181112125402967.png​

I是输入图像，（**W**，\Theta​）表示卷积C的参数.S_{v}^{t}​是阶段t的粗级NRSS，其封装了每个像素的不同显著程度（类似于预测可能同意一个对象的观察者的比例是显着的），S_{v}^{t}​指的是粗级显着著图，并且\xi​是指SCM。Fs（·）表示由编码器网络生成的输出特征映射。SCM由三个卷 积层组成，用于生成所需的显着图。初始卷积层有6个通道，3×3内核，后面是两个卷积层，分别有3个通道，3×3内核和1个通道，1×1内核。 SCM中的每个通道为嵌套的相对显着性堆栈的每个空间位置学习软权重，以便基于它们属于显著对象的置信度来标记像素。

**3.2逐阶段细化网络**  
已经显示出显著对象检测成功的大多数现有作品[24,32,37,8]通常共享阶段式解码的共同结构以恢复每像素分类。尽管编码器的最深阶段具有最丰富的特征表示，但仅依靠**解码阶段**的卷积和解除池化来恢复丢失的信息可能会降低预测的质量[12]。因此，可以从较早的表示逐渐恢复在最深层丢失的空间分辨率。这种概念出现在所提出的基于细化的模型中，其包括编码器和解码器层之间的跳跃连接[25,12,37,8]。但是，如何有效地结合本地和全局背景信息仍然是值得进一步分析的领域。受基于细化的方法[25,11,12]的成功启发，我们提出了一种基于多阶段融合的细化网络，通过将初始粗略表示与早期层表示的更精细特征相结合，在解码阶段恢复丢失的上下文信息。细化网络由等级感知细化单元的连续阶段组成，其尝试在每个细化阶段中恢复丢失的空间细节，并且还保持显著对象的相对等级顺序。每个阶段细化单元将前面的NRSS与更早的更精细尺度表示作为输入，并执行一系列操作以生成精细NRSS，其有助于获得精细显著图。注意，细化分层NRSS意味着细化单元利用不同SCM级别的一致程度来迭代地提高相对等级和总体显著性的置信度。作为最后阶段，**由SCM生成的精确显著性图被融合以获得整体显著性图。**

**3.2.1秩感知细化单元**  
先前的显著性检测网络[32,24]通过直接整合来自早期特征的表示来提出跨不同层次的改进。在[12]之后，我们**将门单元集成在我们的秩感知细化单元中**，该单元控制传递的信息以滤除与图形ground和显著对象相关的模糊性。由前馈编码器生成的初始NRSS（S_{v}^{t}​）为第一细化单元提供输入。注意，**可以将S_{v}^{t}​解释为解码过程中的预测显著性映射**，但是我们的模型强制通道维度与标记显著对象所涉及的参与者数量相同。细化单元采用由以下方法生成的门控特征映射\zeta ^{t}​：门单元[12]作为第二输入。如[12]所示，我们通过组合来自编码器网络的两个连续特征图（f_{v}^{t}​和f_{v}^{t+1}​）来获得（参见图2中的虚线框）。我们首先对前面的S_{v}^{t}​进行上采样，使其尺寸加倍。在上采样的S_{v}^{t}​和\zeta ^{t}​上应用由一系列操作组成的变换函数\tau _{f}​，以获得精细的NRSS（S_{v}^{t+1}​）。然后我们将SCM模块添加到S_{v}^{t+1}​之上以生成**精确显著图S_{m}^{t+1}​**。最后，预测的S_{v}^{t+1}​被馈送到下一级秩感知细化单元。请注意，我们只将NRSS转发到下一阶段，允许网络学习显著对象的不同置信水平之间的对比。与其他方法不同，我们对细化的NRSS和精确的显著性图都进行监督。获得精确NRSS的程序和所有阶段的精确显著性图是相同的。所描述的操作可以总结如下：https://img-blog.csdnimg.cn/20181112125426171.png​

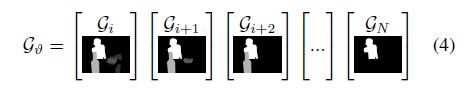
其中u代表上采样操作; w^{b}​和w_{s}^{b}​分别表示变换函数\tau _{f}​和SCM（等式中的\xi​）的参数。 注意，t指的是细化过程的特定阶段。

**3.3多级显著性图融合**  
在细化单元的不同阶段的预测显著性图能够找到具有越来越尖锐边界的显著区域的位置。 由于所有秩感知细化单元彼此堆叠在一起，因此网络允许每个阶段学习在细化过程中有价值的特定特征。 这些现象促使我们将不同层次的SCM结合起来预测，因为它们之间的内部连接没有明确地存在于网络结构中。 为了促进交互，我们在**网络的末端添加融合层，该融合层连接不同阶段的预测显著性映射，从而产生融合特征映射**S_{m}^{\hat{f}}​**。 那么，我们应用1×1卷积层**\gamma​**生成我们网络的最终预测显着性图**S_{m}^{T}**​**。 请注意，我们的网络具有T个预测，包括一个融合预测和T-1个阶段预测。 我们可以编写如下操作：

https://img-blog.csdnimg.cn/20181112125827730.png​

其中ð表示跨渠道连接; wf是用于获得最终预测的结果参数。

**3.4 Ground-truth的堆叠表示**  
显著对象检测或分割的**ground-truth**包含一组数字，用于定义每个像素的显著程度。生成二元掩模的传统方法是通过阈值处理，这意味着没有相对显著性的概念。由于我们的目标是明确地模拟观察者协议，因此使用传统的二元groug-truth掩模是不合适的。为了解决这个问题，我们建议生成一组堆叠的图，这些图对应于不同的显著性水平（由观察者间协议定义）。给定ground-truth显著图\zeta _{m}​，我们得到N个ground-truth的堆叠\zeta_{v}​（\zeta _{i}​，\zeta_{i+1}​，.....，\zeta_{N}​） 其中**每个映射\zeta _{i}​包括二进制指示，该二进制指示至少i个观察者判断对象是显著的（以每像素级别表示）**。 N是标记显著对象的不同参与者的数量。**叠加的groundtruth显著性映射\zeta_{v}​为多个显著对象提供了更好的分离**（参见方程式（4）进行说明），**并且自然地充当相对等级顺序，允许网络学习专注于显著程度**。重要的是要注意叠加的ground-truth的嵌套性质，其中\zeta_{i+1}​\subseteq​\zeta _{i}​。这在概念上是重要的，其中\zeta _{i}​= 1\Leftrightarrow​恰好与i个观察者一致，导致ground-truth堆叠中的零层，以及基于一致程度的微小差异对ground-truth的大的改变。

​

**3.5显著性目标子化网络**  
以前的工作[35,7]将次级化视为一项简单的分类任务。与我们的多个显著对象检测网络类似，子网络化网络也基于ResNet-101[6]，除了我们删除最后一个块。我们在末尾附加一个完全连接层，为输入图像中存在的0,1,2,3和4+显著对象中的每一个生成置信度分数，然后是另一个完全连接的层，从而为每个类别生成最终置信度分数。这背后的原因是，**单层允许积累与显著性相关的置信度，而两层允许推理相对显著性。 我们使用我们预先训练的检测模型来训练子化网络。作为分类器，子化网络减少了ground-truth中的显著对象数n和总预测目标数的两个交叉熵损失**l_{sub}^{1}​**（c，n）和**l_{sub}^{f}​**（cf，n）。**

**显著对象次级化的新数据集**：由于显著对象次级化不是一个广泛解决的问题，因此创建了有限数量的数据集[35]。为了便于在更复杂的场景中研究这个问题，我们为Pascal-S数据集[23]创建了子化ground-truth，它提供了实例计数作为标签。 Pascal-S数据集中关于不同类别的图像分布如表1所示。从表中可以明显看出，有相当数量的图像具有两个以上的显著对象，但只有少数图像具有超过七个。 我们最初在标签过程中包括所有显著对象的实例。 为了减少不同类别之间的不平衡，我们创建了另一个ground-truth集，我们只将图像分类为1,2,3和4+显著对象。

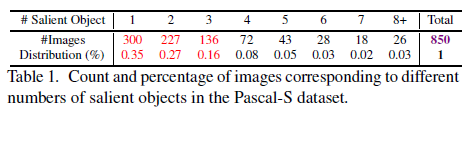
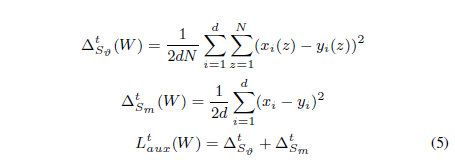
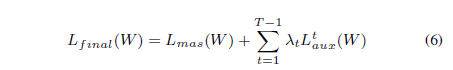
​

表1.与Pascal-S数据集中不同数量的**显著对象**  相对应的图像的计数和百分比。

**3.6训练网络**  
我们提出的网络在每个细化阶段产生一系列嵌套相对显著性堆栈（NRSS） 和显著性映射;但是，我们主要对最终的融合显著性图感兴趣。**鼓励网络的每个阶段通过利用先前的NRSS表示来重复产生具有越来越精细细节的NRSS和显著性图。我们在每个细化阶段的输出处应用辅助损耗以及在网络末端的总体主要损失。**这两种损失都有助于优化过程。更具体地说，令I\epsilon\mathbb{R}^{h\times w\times 3}​是具有ground-truth显著图\zeta _{m}\in\mathbb{R} ^{h\times w}​的训练图像。如3.4节所述，我们生成一堆groundtruth显著图\zeta _{v}\in \mathbb{R}^{h\times w\times 12}​。为了对NRSS（S_{v}^{t}​）和显著图S_{m}^{T}​进行监督，我们首先将\zeta_{v}​和\zeta _{m}​下采样到每个阶段生成的S_{v}^{t}​的大小，在每个阶段得到\zeta _{v}^{t}​和\zeta_{v}^{m}​。然后，**在每个细化阶段，我们定义逐像素欧几里德损失**\bigtriangleup _{S_{v}}^{t}​**和**\Delta _{S_{m}}^{t}​**分别测量（**S_{v}^{t}​**，**\zeta _{v}^{t}​**）和（**S_{m}^{t}​**，**\zeta _{m}^{t}​**）之间的差异。**我们可以将这些操作概括为：

​

其中x\in \mathbb{I}\mathbb{R}^{d}​和y\in \mathbb{I}\mathbb{R}^{d}​（d表示空间分辨率）是矢量化的ground-truth和预测的显著性图。 xi和yi分别表示S_{v}^{t}​和\zeta _{v}^{t}​的特定像素。**W**表示整个网络的参数，**N**表示ground-truth片的总数（在我们的例子中N = 12）。 结合主和辅助损耗的网络的最终损失函数可写为：

​

其中Lmas（W）是指在最终预测显着图S_{m}^{T}​上计算的欧几里德损失函数。我们将所有阶段的​ \lambda _{t}设置为1以平衡损失，这仍然是连续可微的。每个预测阶段都包含与两个预测相关的信息，允许我们的网络从深层传播监督信息。 这也开始于将权重与初始粗略表示对齐，从而导致粗略到精细的学习过程。 融合预测通常看起来比其他阶段性预测好得多，因为它包含来自所有细化阶段的聚合信息。 对于显著性推断，我们可以简单地将任意大小的图像提供给网络，并使用融合预测作为我们的最终显著性图。

**4.实验**  
我们模型的核心遵循基于ResNet-101 [6]的结构，其具有预先训练的权重以初始化编码器部分。 提出了基本体系结构的一些变体，并且我们报告了以下变体的数字： **RSDNet**：该网络包括扩展的ResNet-101 [3]+ NRSS + SCM。 **RSDNet-A**：该网络与RSDNet相同，除了ground-truth按比例缩放1000倍，鼓励网络明确地学习更深层次的对比。 **RSDNet-B**：结构遵循RSDNet，除了添加了一个巨大的金字塔池模块。 **RSDNet-C**：RSDNet-B +ground-truth扩展。 RSDNet-R：RSDNet，具有逐级的秩感知细化单元+多级显著图融合。 **4.1数据集和评估指标**  
**数据集**：Pascal-S数据集包括850个自然图像，其中多个复杂对象来自PASCAL VOC 2012验证集[4]。我们将Pascal-S数据集随机分成两个子集（425用于训练，425用于测试）。在该数据集中，显著对象标签基于使用12个参与者来标记显著对象的实验。实际上，所有用于显著对象分割或检测的现有方法将ground-truth显著性阈值化以获得二元显著性图。这个操作似乎有点武断，因为阈值可能需要k个观察者之间达成共识，并且k的值在不同研究之间变化。这是最常用的显著对象分割数据集之一，但在具有由合理样本大小的观察者提供的多个明确标记的显著区域方面是独特的。由于这项工作的一个关键目标是对图像中的显著对象进行排序，因此我们使用原始的ground-truth图（每个像素的值对应于认为它是显著对象的观察者的数量），而不是试图预测基于可论证的有争议的阈值处理过程的二进制输出。 **评估指标：**对于多重显著对象检测任务，我们使用四种不同的标准指标来衡量性能，包括**精确召回（PR）曲线**，F-测量（沿曲线最大），**ROC曲线下面积**（AUC）和平均绝对值误差（MAE）。由于其中一些依赖于二元决策，我们根据认为对象显著的参与者数量来确定ground-truth显著性图，从而产生12个二元ground-truth图。对于每个二元ground-truth图，预测显著性图的多个阈值允许计算真阳性率（TPR），假阳性率（FPR），精确度和召回率以及相应的ROC和PR曲线。鉴于在此工作之前的方法是基于变化的阈值进行训练并考虑二元ground-truth图，基于产生最佳AUC或F-测量分数的二元ground-truth图来报告分数（并且显示相应的曲线）。还报告了最大F-测量，平均F-测量和中值F-测量，以提供性能如何随所选阈值而变化的感觉。我们还报告了MAE分数，即预测显著性图和产生最小分数的二元ground-truth图之间的平均像素差异。为了评估显著对象的等级顺序，我们引入了显著对象排名（SOR）度量，其被定义为Spearman在地面真实等级顺序和显著对象的预测等级顺序之间的秩次顺序相关性。为便于解释，SOR分数归一化为[0 1]。根据考虑整个数据集的每种方法的平均SOR分数报告分数。 **4.2性能比较**  
 评估显著检测模型的问题本身具有挑战性，这导致所使用的基准之间的差异。鉴于这些考虑因素，我们应用于所有方法的具体评估旨在消除一种算法相对于另一种算法的任何优点。我们将我们提出的方法与最新的最新方法进行比较，包括Amulet [37]，UCF [38]，DSS [8]，NLDF [26]，DHSNet [24]，MDF [18]，ELD [17] ]，MTDS [22]，MC [39]，HS [34]，HDCT [15]，DSR [21]和DRFI [14]。为了公平比较，我们根据[20]中提供的公开代码构建评估代码，我们使用模型作者提供的显着性图，或者通过运行带有推荐参数设置的预训练模型。  
**定量评估：**表2显示了我们模型的所有变体的性能得分，以及其他最近的显著对象检测方法。很明显，RSDNet-R优于其他所有评估指标的近期方法，这确定了我们提出的有效性。从结果我们得到的基本观察结果很少： （1）我们的网络在Pascal-S数据集上以相当大的差距改进了最大F-度量，这表明我们的模型足够通用，可以实现更高的精度。更高的召回率（见图3）。 （2）我们的模型降低了Pascal-S数据集的整体MAE，并且在ROC曲线（AUC）得分下获得了比图3中所示的基线更高的面积。 （3）尽管我们的模型仅在一个子集上训练Pascal-S，它明显优于同样利用大规模显着性数据集的其他算法。

**总体而言**，该分析暗示了所提出的分层堆叠细化策略的优势，以提供更准确的显着性图。此外，值得一提的是，RDSNet-R优于所有最近用于显着对象检测/分割的基于深度学习的方法，而没有任何后处理技术，例如通常用于提升分数的CRF。

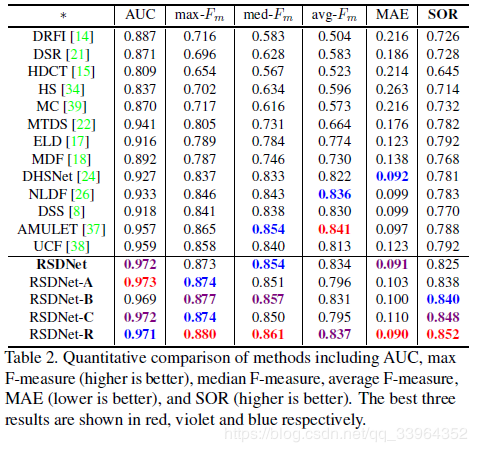
​

表2.方法的定量比较，包括AUC，最大F-测量（越高越好），中位数F-测量，平均F-测量，MAE（越低越好）和SOR（越高越好）。最好的三个结果是 分别以红色，紫色和蓝色显示。

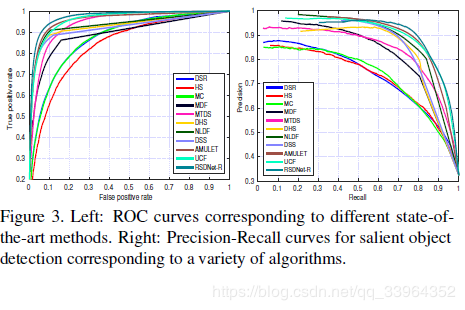
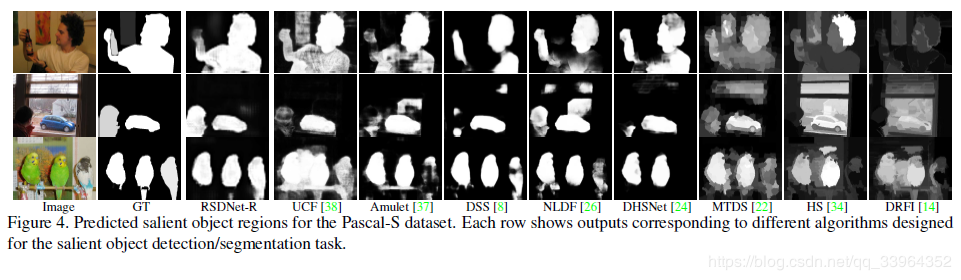
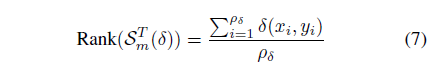
​

图3**.左：**ROC曲线对应于不同的最新方法。

**右：**对应于各种算法的显着物体检测的精确调用曲线。

**定性评估**：图4描绘了RSDNet-R与其他最先进方法的视觉比较。我们可以看到，我们的方法可以准确地预测显着区域，并在各种挑战中产生更接近地面真实图的输出 例如，触摸图像边界的实例（第1和第2行），同一对象的多个实例（第3行）。 每个阶段的嵌套相对显着性堆栈提供了不同的表示，以区分多个显着对象，并允许推断它们的相对显着性

**4.2.1检测排序**  
  由于显著的实例排名是一个全新的问题，因此没有现有的基准。为了促进研究这个问题的这个方向，我们有兴趣从预测的显著性图中找到显著对象的排名。通过平均该实例掩码内的显著程度来获得显著实例的排序。

 ​其中x代表预测显著性图（S_{m}^{T}​）的特定实例，\delta​包含的像素总数\rho _{\delta }​，并且\delta​（xi，yi）指像素（xi，yi）的显著性得分。 虽然可能存在用于定义排名顺序的替代方案，但这是分配该分数的直观方式。话虽如此，我们希望这是进一步探索问题的另一个有趣的细微差别; 特别是显著性与规模，以及部分整体关系。 请注意，**我们无需更改网络体系结构即可获得所需的排名。 相反，我们使用提供的实例分割和显著性映射来计算每个图像的排名。** 为了证明我们的方法的有效性，我们将总体排名得分与最近的最新状态方法进行了比较。值得注意的是，没有先前的方法报告显著实例排名的结果。我们应用建议的SOR评估指标来报告不同模型如何衡量相对显著性。表2中的最后一列显示了我们的方法的SOR分数以及与其他最先进方法的比较。对于我们模型的最佳变体，我们获得85.2％的相关性得分。所提出的方法在对多个显著对象进行排序方面明显优于其他方法，并且我们的分析表明学习显着对象检测在某种程度上隐含地学习排名，但是明确学习排名也可以改善显著对象检测，而不管如何定义基础事实。图5显示了为显著对象检测设计的现有技术方法的定性比较。请注意，排名对三个以上对象的作用尤其明显。

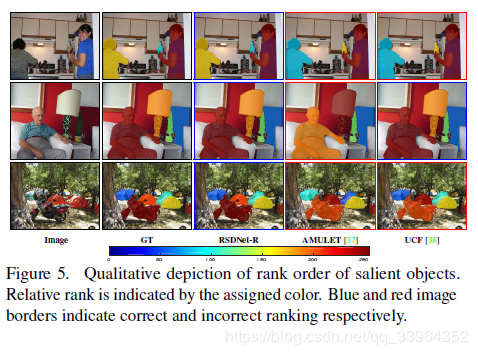
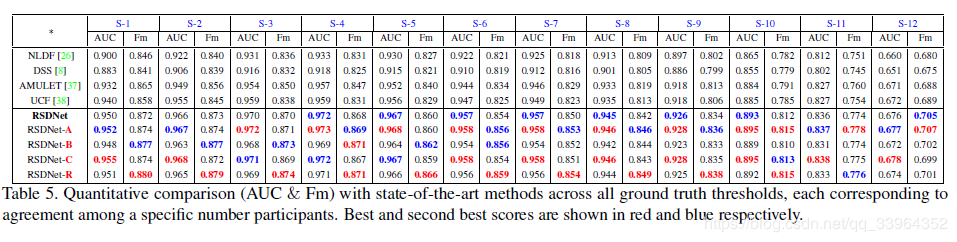
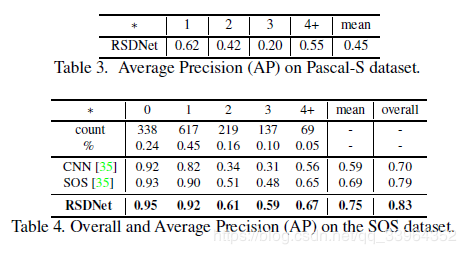
​

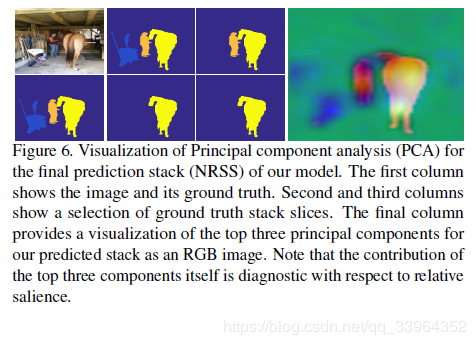
图5.显著对象的排名顺序的定性描述。 相对等级由指定的颜色指示。

蓝色和红色图像边框分别表示正确和不正确的排名。

**4.2.2应用：显著对象子化**  
如前所述，显着对象检测，排序和子化是相互关联的。因此，自然要考虑显着区域预测和排名是否为子化提供指导。进一步训练检测网络的副本以在Pascal-S上执行子化。为简单起见（并与先前的工作[35,7]一致），我们训练我们的系统仅用于预测1,2,3或4+的对象，并报告表3中的平均精度（AP）[5]。 由于这是在Pascal-S数据集上执行子图化的第一项工作，因此我们没有任何基线可供比较。为了使比较成为可能，我们在SOS数据集[35]上微调和评估我们的模型，并报告 表4中的AP和加权AP（总体）得分。与基线相比，我们提出的模型在该数据集上实现了最先进的结果。

​

**表5.**在所有ground-truth阈值中使用最先进方法的定量比较（AUC和Fm），每个方法对应于特定数量参与者之间的一致性。 最佳和第二好成绩分别以红色和蓝色显示。

​  
**图6.**我们模型的最终预测堆栈（NRSS）的主成分分析（PCA）的可视化。

第一列显示图像及其groud-truth。 第二列和第三列显示了一组groud-truth堆栈切片。最后一列提供了我们预测堆栈的前三个主要组件的可视化，作为RGB图像。

请注意，前三个组件本身的贡献在相对显著性方面是诊断性的。

**4.3检查嵌套相对显著性堆栈**  
嵌套相对显著性堆叠的切片的比较可能是具有挑战性的，因为一些层对之间的差异可能是微妙的，并且对比度可以跨层不同。因此，我们通过主成分分析（PCA）检查NRSS层之间的可变性，以确定存在最大可变性（和信号）的区域。图6示出了作为RGB图像的前三个主要分量，其中第一主成分（其捕获跨层的最大方差）被映射到R通道，第二主成分被映射到G通道，依此类推。groud-truth中的显著区域在各层的可变性中被捕获，表明我们的显著性排序机制的价值。此外，几乎可以直接从该可视化读取相对等级，其中前2个特征向量的高值导致黄色，第一个仅为红色等。  
 我们还报告了表5中每个切片的AUC分数和最大F-测量值（表示为S）。与基线相比，我们提出的方法在所有基本真值阈值上获得更好的分数，这对应于表示同意的不同参与者数量。一个对象是显著的。这进一步显示了堆叠机制的有效性和预测相对显著性，无论如何确定groud-truth（如果被视为二进制数量），这导致改进。

**4.4失败案例**  
尽管大多数案件表现良好; 有些实例更具挑战性（见图7）。 有时，groud-truth有多个具有相同程度显著性的对象（参与者同意的关系）（参见图7中的第1行）。 当关于图像中显著的一致性（如第二行中所示）或者在具有相对接近的显著程度的两个对象之间存在遮挡（如最后一行中所示）时，发生其他的排序失败。

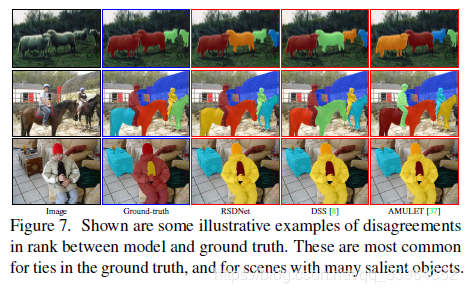
​

图7.显示了模型和ground-truth之间等级上的分歧的一些说明性示例。

这些对于ground-truth中的关系以及具有许多显著对象的场景是最常见的。

**5.总结**  
在本文中，我们提出了一个神经网络框架，用于**检测，排序和子化多个显著对象，引入堆栈细化机制以实现更好的性能**。 这种方法成功的**关键在于如何在groud-truth和网络中以产生稳定性能的方式表示相对显著性**。 我们强调的是，到目前为止，显著对象检测已经假设了一个相对有限的，有时是不一致的问题定义。 综合实验表明，所提出的架构在广泛的指标范围内优于最先进的方法。