

# 学习准确显著性检测的不确定卷积特征

张萍萍

王东

Huchuan Lu\*Hongyu Wang

宝彩贤大

连理工大学, 中国

jssxzhpp@mail.dlut.edu.cn {wdice, lhchuan, whyu, YBC} @ dlut.edu.cn

## 摘要

深卷积神经网络 (CNN) 在许多计算机视觉任务中提供了卓越的性能。在本文中, 我们提出了一种新的深度全卷积网络模型, 用于精确显著性检测。这项工作的主要贡献是学习深度不确定的卷积特征 (UCF), 它鼓励显著性检测的稳健性和准确性。我们通过在特定的卷积层之后引入重新配置的丢失 (R-dropout) 来构建不确定的内部特征单元集合来实现这一点。另外, 我们提出了一种有效的混合上采样方法来减少解码器网络中去卷积算子的棋盘伪影。所提出的方法也可以应用于其他深度卷积网络。与现有的显著性检测方法相比, 所提出的UCF模型能够结合不确定性以获得更精确的物体边界推断。大量实验证明, 我们提出的显著性模型能够有效地抵御最先进的方法。不确定的特征学习机制以及上采样方法可显著提高其他像素视觉任务的性能。

## 1. 介绍

显著性检测的目标是识别图像中最重要和显眼的物体或区域。作为计算机视觉中的预处理过程, 显著性检测极大地吸引了许多实际应用, 如对象重定位[6, 40], 场景分类[37], 语义分割[34]和视觉跟踪[29, 14]。虽然已经取得重大进展[15, 1, 30, 21, 33, 46], 由于现实世界情景中的复杂因素, 显著性检测仍然非常具有挑战性。在这项工作中, 我们专注于提高显著性检测模型的鲁棒性, 这在文献中已被忽略。

先前的显著性检测方法利用多个手工制作的视觉特征和启发式先验。最近, 深刻



图1. 使用深度CNN的像素级视觉任务中的棋盘工件示例。高分辨率看到更好。

基于学习的方法变得越来越流行, 并为许多数据集设定了基准[27, 49, 3]。它们的优越性能部分归因于对物体外观和各种场景建模的强大代表性。然而, 现有的方法不能提供深度神经网络中“黑盒子”学习的概率解释能力, 并且主要享受模型的卓越性能。合理的概率解释可以提供与预测一致的关系调节, 并使预测系统变得更加稳健[10]。此外, 由于不确定性是任何预测系统的一个自然部分, 建模不确定性至关重要。例如, 物体边界强烈影响显著性模型的预测精度, 因此模型可以提供有意义的不确定性

\*Lu教授是通讯作者。

在不同物体的边界在哪里。据我们所知，没有工作来建模和分析基于深度学习的显着性检测方法的不确定性。

另一个重要问题是像素级视觉任务中的棋盘伪影，其目标是从低分辨率到高分辨率生成图像或特征图。几个典型的例子如图1所示（参考文献[32]）。对于基于CNN的深度方法，奇怪的伪影有时非常致命。例如，当伪影出现在完全卷积网络（FCN）的输出中时，网络训练可能失败并且预测可能完全错误[36]。我们发现这些伪影的真正原因是上采样机制，它通常利用反卷积操作。因此，探索新的上采样方法以更好地减少像素级视觉任务的伪影是非常有意义的。同时，这些伪影也与深度CNN的不确定性学习密切相关。

上面讨论的所有问题都促使我们通过深度网络学习不确定的特征（概率学习），以实现准确的显着性检测。我们的模型具有几个独特的功能，如下所述。

- 与现有的显着性检测方法不同，我们的模型非常简单。它由一个编码器FCN，一个相应的解码器FCN和一个像素级分类层组成。编码器FCN从原始图像分层学习视觉特征，而解码器FCN逐步将编码特征映射上采样到像素级分类的输入大小。
- 我们的模型可以学习深度不确定卷积特征（UCF），以实现更精确的显着性检测。关键因素受到辍学的启发[13]。我们提出一个重新配置的丢包（R-dropout），导致特定卷积层内部特征单元的自适应集合。不需要额外的参数化就可以实现不确定的特征。
- 我们提出了一种新的上采样方法来减少去卷积运算的棋盘伪影。新的上采样方法有两个明显的优点。一方面，它从卷积（提取卷积特征）中分离出上采样（以生成更高分辨率的特征映射），另一方面它与常规反卷积兼容。不确定的特征提取和显着性检测在编码器 - 解码器网络架构中是统一的。所提出的模型参数（即，所有层中的权重和偏差）通过端到端梯度学习被联合训练。
- 我们的方法在显着性检测和其他像素视觉任务方面显示出很好的概括性。没有任何后处理步骤，我们的模型在公共显着性检测，语义分割和眼图数据集方面的性能甚至会更好。

## 2. 相关工作

最近，深度学习在显着性检测方面提供了卓越的性能。例如，王等人。[44]提出了两个深度神经网络来整合局部估计和显着性检测的全局搜索。Li等人[23]训练完全连接的多个CNN层以预测每个超像素的显着程度。为了解决突出物体可能出现在低对比度背景下的问题，Zhao et al. [50]考虑全局和局部背景，并在多环境深层CNN框架中对显着性预测进行建模。这些方法具有优异的性能，然而，它们都包括完全连接的层，这在计算上非常昂贵。更重要的是，完全连接的层会丢弃输入图像的空间信息。为了解决这些问题，李等人。[26]提出了在显着性检测的多任务学习框架下训练的FCN。Wang等人[45]设计一个经常性FCN来利用显着性预测并反驳粗略预测。虽然受到类似精神的激励，但我们的方法在三个方面与[26, 45]显着不同。首先，网络架构是非常不同的。我们使用的FCN采用编码器 - 解码器的形式，这是从主要信息重构的角度来看的。在[26, 45]中，FCN起源于FCN-8s [28]，其设计为用于分割任务的长和短跳过连接。其次，我们的模型不是在[26, 45]中简单地使用FCNs作为预测因子，而是通过使用多个重新配置的丢失来学习不确定的卷积特征，从而提高了显着性检测的稳健性和准确性。第三，我们的模型配备了一种新的上采样方法，自然处理反卷积操作的棋盘伪影。通过训练整个神经网络可以减少棋盘伪影。相反，这些工件由[26, 45]中的手工方法处理。具体而言，[26]使用超像素分割来平滑预测 -

灰。在[45]中，使用边缘感知侵蚀过程。

我们的工作也与深度学习中的模型不确定性有关。Gal等人[10]在数学上证明，在每个重量层之前应用了失落的多层感知器模型（MLP）相当于对概率深度高斯过程的近似。尽管所提供的理论是可靠的，但是对于深度CNN的全面核实还是被低估了。基于这个事实，我们在这个方向上又迈出了一步，并且表明在卷积层中可以使用重新配置的丢失来学习不确定的特征合奏。另一个代表性的模型不确定性工作是贝叶斯SegNet [20]。贝叶斯Seg-Net能够用模型不确定性的度量来预测以像素为单位的场景分割。他们通过蒙特卡罗采样来实现模型不确定性。在测试时间启动压差以生成像素级标签的后验分布。与[20]不同，我们的模型着重于在训练期间学习不确定卷积特征。



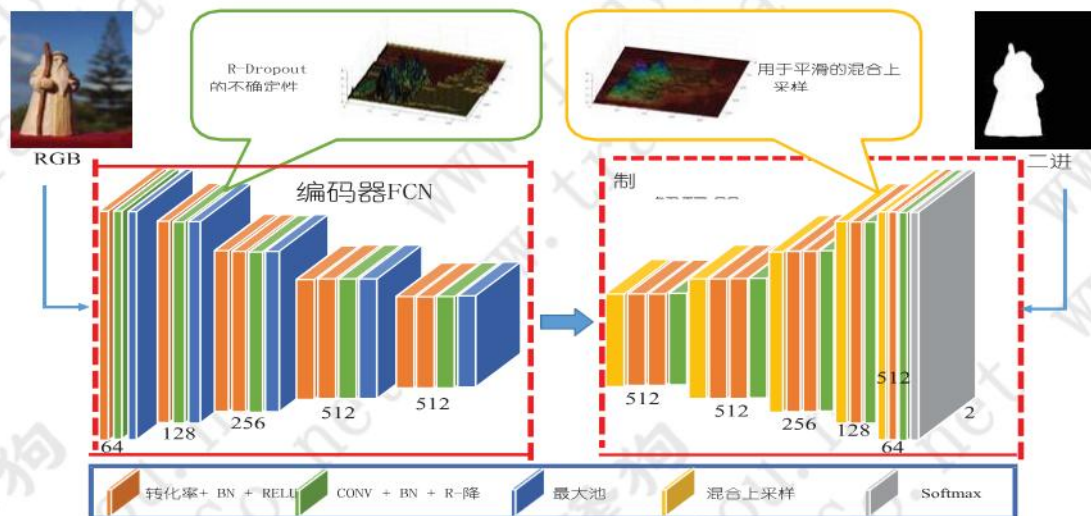


图2. 建议的UCF模型的总体架构。

### 3. 建议的模型

#### 3.1. 网络架构概述

我们的架构部分受到层叠去噪自动编码器的启发[43]。我们将自动编码器推广到深度完全卷积编码器-解码器架构。由此产生的网络形成了一种新型的混合FCN，它由用于高级特征提取的编码器FCN，用于低级信息重构的相应解码器FCN和用于显著性预测的像素级分类器组成。整体架构如图2所示。更具体地说，编码器FCN由多批次归一化卷积层（BN）[16]和整流线性单元（ReLU）组成，其次是非重叠最大池。相应的解码器FCN还引入了上采样操作以构建从低分辨率到高分辨率的特征映射。我们使用softmax classifier进行像素级显著性预测。为了获得学习的卷积特征的不确定性，我们在多个卷积层之后利用重新配置的丢失（称为R-Dropout）。详细的网络配置包含在补充材料中。我们将在以下小节中全面阐述R-Dropout，我们的新的取样方法和培训策略。

#### 3.2. 深度不确定卷积特征集成的R-Dropout

辍学通常被解释为包装大量的个人模型[13, 39]。虽然大量的实验表明完全连接层的丢失提高了深度网络的泛化能力，但是缺乏对其他类型层（如卷积层）使用丢失的研究。在本小节中，我们表明在卷积层之后使用修改的丢失可以被解释为一种概率特征合奏。

鉴于这个事实，我们提供了一个学习不确定性的策略，卷积特征。

**R-D 卷积中的旋转：**假设  $X^{H \times W \times C}$  是3D张量，而  $f(X)$  是CNN中的卷积运算，通过参数  $W$  和  $b$  将  $X$  投影到  $R^{H' \times W'}$ 。  

$$f(X) = WX + b. \quad (1)$$

令  $g(\cdot)$  是一个非线性激活函数。当原始丢失[13]被应用于  $g(f)$  的输出时，我们可以得到其受到干扰的版本  $(f)$

$$g(f) = g(WX + b), \quad (2)$$

独立于  $M_{i,j}$ ，表示训练期间有丢失的激活，公式(2)表示测试时的激活。另外，Srivastava等人。[39]建议在测试时间用  $p$  来缩放激活  $g(f)$  以获得单位激活的近似平均值。许多常用的激活函数，如Tanh, ReLU和LReLU[12]，具有  $g(0) = 0$  的性质。因此，方程(3)可以重写为R-Dropout公式，

$$g(f) = g(M * (WX) + M * b) \quad (4)$$

$$= g((M \otimes W)X + M * b) \quad (5)$$

$$= g(SX + M * b), \quad (6)$$

$$= g(SX + M * b), \quad (7)$$

其中  $\otimes$  表示交叉通道元素方面的产品。从上面的方程中，我们可以推导出当  $S = MW$  仍然是二元时，方程(7)意味着在激活函数的输入端应用了一种随机特性<sup>1</sup>。让

<sup>1</sup>随机性意味着可以在每次训练迭代期间使用特定概率分布来生成可学习张量  $S$ 。  $S$  的更新形成了一个随机过程，而不是一个特定的决定。

$S_{i,j} \in [0, 1]$  和  $\sum_{j=1}^n S_{i,j} = 1$ , 上述方程将严格构造X的内部特征单元的集合。然而, 实际上当然没有证据保持超出限制。 尽管如此, 我们注意到: (1) 随机掩码矩阵S主要取决于掩码生成器<sup>2</sup>; (2) 在训练深度卷积网络时, 卷积层之后的R-Dropout作为卷积特征的不确定集合; (3) 这种特征集合是元素明智概率的, 因此它可以在预测密集标记视觉任务如显著性检测和语义分割方面带来稳健性。

不确定的卷积特征提取: 受上述见解的启发, 我们将R-Dropout用于我们模型的卷积层, 从而生成深度不确定的卷积特征映射。 由于我们的模型由交替的卷积层和合并层组成, 因此我们的模型中存在两个典型案例。为了符号简单, 我们随后放弃批量标准化 (BN)。

1) Conv + R-Dropout + Conv: 如果提议的R-Dropout后面跟着一个卷积层, 则输入的前向传播被表达为

$$f^{l+1} = \text{Conv} (W^{l+1}, (f^l)) \quad (8)$$

$$(f^l) = g(f^l) \quad (9)$$

$$g^{l+1} = g(f^{l+1}) \quad (10)$$

其中l是层数, Conv是卷积运算。 从方程 (9) 可以看出, 干扰激活 $g(f^l)$ 与滤波器 $W^{l+1}$ 卷积以产生卷积特征 $f^{l+1}$ 。 这样, 网络将着重于学习权重和偏置参数, 即W和b, 并且在训练深度网络期间使用R-Dropout的不确定性将被消散。

2) Conv + R-Dropout + Pooling: 在这种情况下, 输入的前向传播变为

$$f^{l+1} = \text{Pool} (\text{Conv} (W^{l+1}, (f^l)) + M * b) \quad (11)$$

$$g_j = \text{汇集} (g(f)_1, \dots, (f)_n), i \in R_{j^0} \quad (12)$$

这里Pooling (•) 表示最大池功能。  $R^l$  是第l层的汇集区域j,  $g(f)_i$  是 $R^l$ 内每个神经元的活动。  $n = |R^l|$  是R中的单位数<sup>3</sup>。 在没有的情况下制定不确定性

我们假设每个汇集区域j中的激活 $(f^l)$ 以非递减顺序排列, 即 $g(f^l)_1 \leq (f^l)_{20} \dots g(f^l)_{n0}$  结果,  $g(f^l)$  将被选为合并激活条件  $\leq$

<sup>2</sup>在R-Dropout中, 生成器可以是任何概率分布。 当发生器是伯努利分布时, 原始丢失是R-Dropout的特例。

(1)  $g(f)_i, g(f)_i, i \in R_{j^0}$  采样根据式(12)

$$P(g^{l+1} = g(f^l)_i) = P_i = pq^{n-i}, p = 1-q.$$

(13) 因此, 在max-pooling操作之前执行R-dropout是从下面的multinomial分布来选择一个指数i, 那么混合激活就是 $g(f^l)_i$ ,

$(P_1, \dots, P_n)$ , (14) 其中 $P_0 (= \text{TF119}) q^n$  是池区内所有单位都退出的特殊事件。

后一种策略通过在卷积层中使用R-Dropout来展现构建不确定性的有效性。 我们采用它来建立我们的网络架构 (参见图2)。 我们将通过实验证明基于R-Dropout的FCN在第4节的显著性检测数据集上产生了惊人的结果。

### 3.3. 混合上采样预测平滑

在本小节中, 我们首先用反卷积算法[42]来解释棋盘伪影的原因。 然后, 我们推导出一种新的上采样方法, 以尽可能减少网络训练和推理中的伪像。

在不失一般性的情况下, 我们将重点放在两个轴上的平方输入 (nn), 方形内核大小 (kk), 相同步幅 (ss) 和相同的零填充 (pp) (如果使用的话)。 由于我们旨在实现上采样, 因此我们设置 $s = 2$ 。 一般来说, 卷积操作C可以用来描述

$$O = C(I, F) = I * F, \quad (15)$$

其中 $I_{(N+F) \times (N+F)}$  是输入,  $F_{k \times k}$  是具有步长s的过滤器, 是离散卷积, O是尺寸为  $(n + 2p - k) / s + 1$  的输出。 卷积C具有由n描述的相关去卷积D,  $k = k, s = 1$  和  $p = k - p - 1$ , 其中n是

通过添加s获得的拉伸输入 每个输入单元之间有1个零点, 并且解卷积的输出尺寸是  $\times$

$O' = s(n' - 1) + k - 2p$ 。 这表明常规反卷积算子等效于对插入了零的新输入 $(n'n')$ 进行卷积。 一个玩具的例子如图3所示。 另外, 当过滤器尺寸k不能被步幅s分开时, 解卷积会引起重叠问题。 如果拉伸的输入是高频或接近周期性的, 即当插入零点时该值非常波动, 解卷积操作的输出结果自然具有像棋盘一样的数值假象。

<sup>3</sup>通过引入, 可以放宽对输入n大小的约束

另一个参数 $t \in 0, \dots, s - 1$ 允许区分, 所有这些导致相同的n。



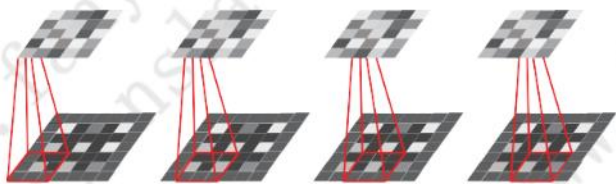


图3. 解卷积的详细解释。使用2 2跨越5 5输入的3 3内核的反卷积填充有零的1 1边界（即， $x_n = 5, k = 3, s_x = 2$ 和 $p = 1$ ）。它相当于使用单位步长（即 $n^i = 3, n^o = 5$ ）在3 3输入（在输入之间插入1零） $k = 3, s \leq 1$ 和 $p = 1$ ）。 $\times$

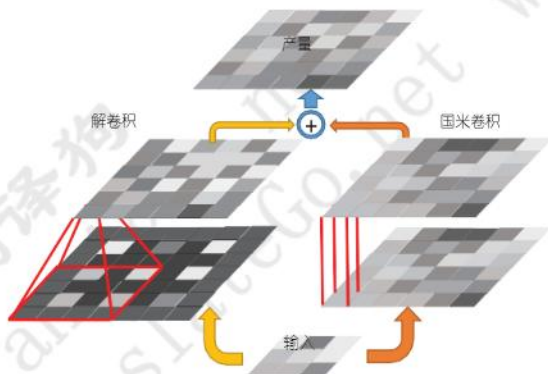


图4. 混合升频采样。共同使用两种策略：  
1) 具有受限滤波器尺寸的解卷积（左分支）；2) 用1x1卷积进行线性插值（右分支）。最终输出

是两种策略的总和结果。

基于上述观察，我们提出了两种策略来避免常规反卷积产生的伪像。第一个是限制过滤器的大小。我们可以简单地确保过滤器尺寸是步幅尺寸的倍数，避免重叠问题，即，

$$k = \lambda s, \quad \lambda \in \mathbb{N}. \quad (1)$$

6) 然后，反卷积将用等效卷积处理零插入输入，从而得到平滑输出。

然而，因为这种方法只关注于改变输出的接收区域，并且不能改变零插入输入的频率分布，所以在很多极端情况下，伪像仍然会泄漏。我们提出了另一种从等值卷积中分离出上采样的替代策略。我们首先通过插值将原始输入调整为所需的大小，然后执行一些等价的卷积。虽然这种策略可能会破坏深CNN中的学习特征，但我们发现通过迭代堆叠这种上采样所构建的高分辨率地图可以令人惊讶地减少伪像。为了充分利用这两种策略的优势，我们通过总结两种策略的输出来引入混合升频方法。图4说明了所提出的上采样方法。在我们提出的模型中，我们使用双线性（或最近

邻居）操作进行插值。这些插值方法是线性运算，可以作为有效的矩阵乘法嵌入到深度CNN中。

### 3.4. 培训整个网络

由于缺乏足够的显着性检测数据来从头开始训练我们的模型，因此我们利用VGG-16模型的前端[38]作为我们的编码器FCN（13个卷积层和5个预先在ILSVRC 2014上进行了训练的汇聚层对于图像分类任务）。我们的解码器FCN是编码器FCN的镜像版本，并具有多个上采样、卷积和修正层系列。批量规范化（BN）被添加到每个卷积层的输出中。如图2所示，我们在特定的卷积层之后以相等的采样率 $p = 0.5$ 添加R-dropout，如图2所示。对于显着性检测，我们随机地初始化解码器FCN的权重并在MSRA10K数据集上调整整个网络[4]，它被广泛用于显着物体检测社区（更多细节将在第4节中描述）。我们将该数据集中每幅图像的地面实况显着图转换为0-1二元图。当我们使用由下式（17）给出的softmax交叉熵损失函数从一般背景中分离显着性前景时，这种变换完全匹配FCN的信道输出。

$$L = - \sum_m \left[ l_m \log(q_m) + (1 - l_m) \log(1 - q_m) \right], \quad (17)$$

升<sub>m</sub>

其中 $l_m$  ( $= 0, 1$ ) 是图像中像素 $m$ 的标签， $q_m$ 是该像素是显着性前景的概率。 $q_m$ 的值是从网络的输出中获得的。在将训练图像放入我们提出的模型中之前，将每个图像与ImageNet平均值[5]相减并重新缩放到格同的大小（448 448）。对于通信，我们还将0-1二进制映射重新调整为相同的大小。该模型使用具有动量，学习速率衰减时间表的小批量随机梯度下降（SGD）端对端训练。参数的详细设置包含在补充材料中。

### 3.5. 显着性推论

因为我们的模型是完全卷积网络，所以在测试时可以将任意大小的图像作为输入。在前馈过程之后，网络的输出由前景激励图（ $M^{FE}$ ）和背景激励图（ $M^{BE}$ ）组成。我们使用 $M^{FE}$ 和 $M^{BE}$ 之间的差异，并剪切负值以获得最终的显着图，即，

$$Sal = \max(M^{FE} - M^{BE}, 0). \quad (18)$$

8) 这种减法策略不仅增加了像素级别的区分度，而且还捕获了上下文对比度信息。可选地，我们可以采用多尺度预测图的集合来进一步提高性能。



## 4. 实验

在本节中，我们从描述显着性检测的实验设置开始。然后，我们彻底评估和分析我们提出的公共显着性检测数据集模型。最后，我们提供了额外的实验来验证我们的方法在其他像素视觉任务上的泛化，即语义分割和眼睛形态。

### 4.1. 实验装置

**显着性数据集：**为了训练提出的网络，我们通过镜像反射和旋转技术（0°，90°，180°，270°）简单地增加MSRA10K数据集[4]

共产生8万张训练图像。

对于检测性能评估，我们采用六个广泛使用的显着性检测数据集如下，

DUT-OMRON [49]。该数据集由5,168张高质量图像组成。此数据集中的图像具有一个或多个显着对象和相对复杂的背景。因此，这个数据集在显着性检测中很困难并且具有挑战性。

ECSSD [48]。该数据集包含1,000个自然图像，包括许多语义上有意义和复杂的图像

在地面真相分割中的结构。

HKU-IS [50]。该数据集包含4,447个具有高质量像素方式注释的图像。这个数据集中的图像被很好地选择为包括多个断开的对象或接触图像边界的对象。

PASCAL-S [27]。该数据集是从PASCAL VOC数据集[8]中精心挑选的，并包含850个图像。

SED [2]。该数据集包含两个不同的子集：SED1和SED2。SED1具有100个图像，每个图像仅包含一个显着对象，而SED2具有100个图像，每个图像包含两个显着对象。

SOD [48]。该数据集有300个图像，最初是为图像分割而设计的。[18]生成了明显物体的像素智能注释。

**实现细节：**我们使用改进的Caffe工具箱[20]实现了基于MATLAB R2014b平台的方法。我们在配备i7-4790 CPU（16G内存）和一个NVIDIA Titan X GPU（12G内存）的四核PC机上运行我们的方法。我们的模型的培训过程需要近23小时，并在最小批量SGD的200k次迭代后收敛。所提出的显着性检测算法以大约7fps的速度运行448 × 448分辨率（23 fps，224 × 224决议）。源代码可以在这里找到

<http://ice.dlut.edu.cn/lu/>。

**显着性评估指标：**我们采用三种广泛使用的指标来衡量所有算法的性能，即精确回忆（PR）曲线，F-measure和平均绝对误差（MAE）[3]。通过对预测的显着性图进行阈值处理并将二进制图与实际情况进行比较来计算精确度和召回率。数据集的PR曲线表示平均精度和召回率

显着地图在不同的阈值。F-measure是平均精确度和平均召回的均衡平均值，可以通过计算

$$F\beta = \frac{(1 + \beta^2) \times \text{精确} \times \text{召回}}{\beta^2 \times \text{精度} \times \text{召回}}. \quad (19)$$

根据现有的工作[48] [44] [3] [49]，我们设定 $\beta^2$ 为0.3 重量精确度超过召回率。我们报告了每个显着图以图像相关阈值进行自适应二值化时的性能。阈值被确定为图像的平均显着性的两倍：

$$T = \frac{2}{W \times H} \sum_{x=1}^W \sum_{y=1}^H S(x, y), \quad (20)$$

其中W和H是图像的宽度和高度， $S(x, y)$ 是 $(x, y)$ 处像素的显着性值。

我们还计算了公平比较的平均绝对误差（MAE），如[3]所示。MAE评估显着性检测的准确性

$$MAE = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - G(x, y)|, \quad (21)$$

其中G是二元地面真实面具。

### 4.2. 与最先进的性能比较

我们将提出的UCF算法与其他10种最先进的算法进行比较，其中包括6种基于深度学习的算法（DCL [24]，DS [26]，ELD [22]，LEGS [44]）。

MDF [50]，RFCN [45]）和4个常规对应物

（BL [41]，BSCA [33]，DRFI [18]，DSR [25]）。来源

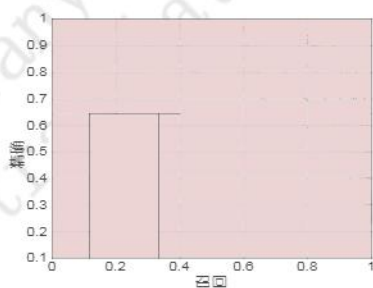
采用推荐参数的代码或竞争方法的显着图进行公平比较。如图5和Tab所示。如图1所示，我们提出的UCF模型在所有评估指标方面可以在几乎所有数据集上始终超越现有方法，这有力地表明了所提出方法的有效性。

有关更多结果，请参阅补充材料DUT-OMRON，HKU-IS，PASCAL-S和SOD数据集。

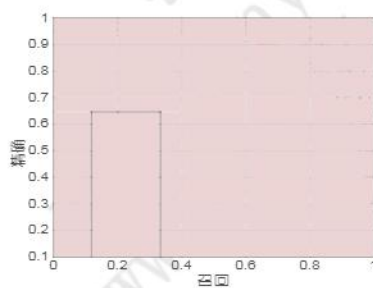
从这些结果中，我们有几个基本的观察结果：（1）我们的UCF模型在F-measure和MAE方面优于ECSSD和SED数据集上的其他算法。更具体地说，我们的模型分别将ECSSD和SED数据集上性能最好的现有算法所实现的F-measure分别提高了3.9%和6.15%。MAE得到持续改进。

（2）尽管我们提出的UCF并不是最好的HKU-IS和PASCAL-S数据集，但它仍然非常有竞争力（我们的模型在这些数据集中排名第二）。有必要指出的是，只有增强的MSRA10K数据集用于训练我们的模型。RFCN，DS和DCL方法

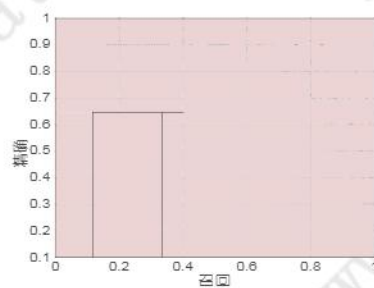




(a) ECSSD



(b) SED1



(c)

SED2图5. 与其他最先进的方法相比, 该算法的性能。

	DUT-OMRON		ECSSD		HKU-IS		PASCAL-S		SED1		SED2	
方法	$F_\beta$	MAE	$F_\beta$	MAE	$F_\beta$	MAE	$F_\beta$	MAE	$F_\beta$	MAE	$F_\beta$	MAE
UCF	0.6283	0.1203	0.8517	0.0689	0.8232	0.0620	0.7413	0.1160	0.8647	0.0631	0.8102	0.0680
V-E	0.6135	0.1224	0.7857	0.0795	0.7716	0.0785	0.6303	0.1284	0.8128	0.0732	0.7576	0.0851
V-D	0.5072	0.1345	0.6942	0.1195	0.6851	0.0967	0.5695	0.1624	0.7754	0.0844	0.6930	0.0954
V-C	0.6165	0.1210	0.8426	0.0711	0.8156	0.0670	0.7201	0.1203	0.8665	0.0653	0.8014	0.0795
V-B	0.6168	0.1305	0.8356	0.0781	0.8060	0.0651	0.6845	0.1254	0.8547	0.0685	0.7905	0.0709
V-A	0.6128	0.1409	0.8166	0.0811	0.7346	0.0988	0.6172	0.1367	0.7641	0.1023	0.6536	0.1044
DCL [24]	0.6842	0.1573	0.8293	0.1495	0.8533	0.1359	0.7141	0.1807	0.8546	0.1513	0.7946	0.1565
DS [26]	0.6028	0.1204	0.8255	0.1216	0.7851	0.0780	0.6590	0.1760	0.8445	0.0931	0.7541	0.1233
ELD [22]	0.6109	0.0924	0.8102	0.0796	0.7694	0.0741	0.7180	0.1232	0.8715	0.0670	0.7591	0.1028
LEGS [44]	0.5915	0.1334	0.7853	0.1180	0.7228	0.1193	-	-	0.8542	0.1034	0.7358	0.1236
密度板 [50]	0.6442	0.0916	0.8070	0.1049	0.8006	0.0957	0.7087	0.1458	0.8419	0.0989	0.8003	0.1014
RFCN [45]	0.6265	0.1105	0.8340	0.1069	0.8349	0.0889	0.7512	0.1324	0.8502	0.1166	0.7667	0.1131
BL [41]	0.4988	0.2388	0.6841	0.2159	0.6597	0.2071	0.5742	0.2487	0.7675	0.1849	0.7047	0.1856
BSCA [33]	0.5091	0.1902	0.7048	0.1821	0.6544	0.1748	0.6006	0.2229	0.8048	0.1535	0.7062	0.1578
DRFI [18]	0.5504	0.1378	0.7331	0.1642	0.7218	0.1445	0.6182	0.2065	0.8068	0.1480	0.7341	0.1334
DSR [25]	0.5242	0.1389	0.6621	0.1784	0.6772	0.1422	0.5575	0.2149	0.7909	0.1579	0.7116	0.1406

表1. 常用数据集上不同显著性检测方法的F-measure和MAE。最好的三个结果分别以红色, 绿色和蓝色显示。所提出的方法在这些数据集上排名第一和第二。

在PASCAL VOC分割数据集[9]上预先培训, 该数据集与PASCAL-S和HKU-IS数据集重叠。这个事实可能解释他们在两个数据集上的成功。但是, 他们在其他数据集上的表现明显较差。(3) 与其他相比

方法, 我们提出的UCF在大多数数据集上实现较低的MAE。这意味着我们的模型更加确信了不确定特征学习的预测区域。

典型图像上不同方法的视觉比较如图6所示。我们的显著图可以可靠地突出显示各种具有挑战性的场景中的显著对象, 例如, 对象和背景之间的低对比度(前两行), 多个断开的显著对象(3-4行)和图像边界附近的物体(5-6行)。另外, 我们的显著图提供了更精确的显著物体边界(1, 3, 6-8行)。

消融研究: 为了验证每个组件的贡献, 我们还用不同的设置评估了所提出的UCF模型的几个变体, 如表1所示。2. 相应的性能在Tab。1. VA模型是DeconvNet [31]的近似值。VA和VB的比较表明我们的

设置	V-A	V-B	V-C	V-D	V-E	UCF
+降	✓	✓	✓			✓
+ R-降						
+休息解除	✓	✓	✓	✓	✓	✓
+国米						✓

表2. 我们的UCF模型的变体。请注意, Rest Deconv和Inter指示混合升频方法。



图7. 不同上采样算法的比较。(a) 输入图像; (b) 解卷积; (c) 插值; (d) 我们的方法;

(e) 事实真相。补充材料中的更多例子。不确定的学习机制确实可以帮助学习更强大的特征来进行准确的显著性推断。VB和VC之间的比较显示了两种上采样策略的效果。结果意味着插值策略在显著性检测中表现得更好。VB, VC和UCF的联合比较表明, 我们的混合上采样方法能够更好地控制输出显著图。一个关于视觉效果例子



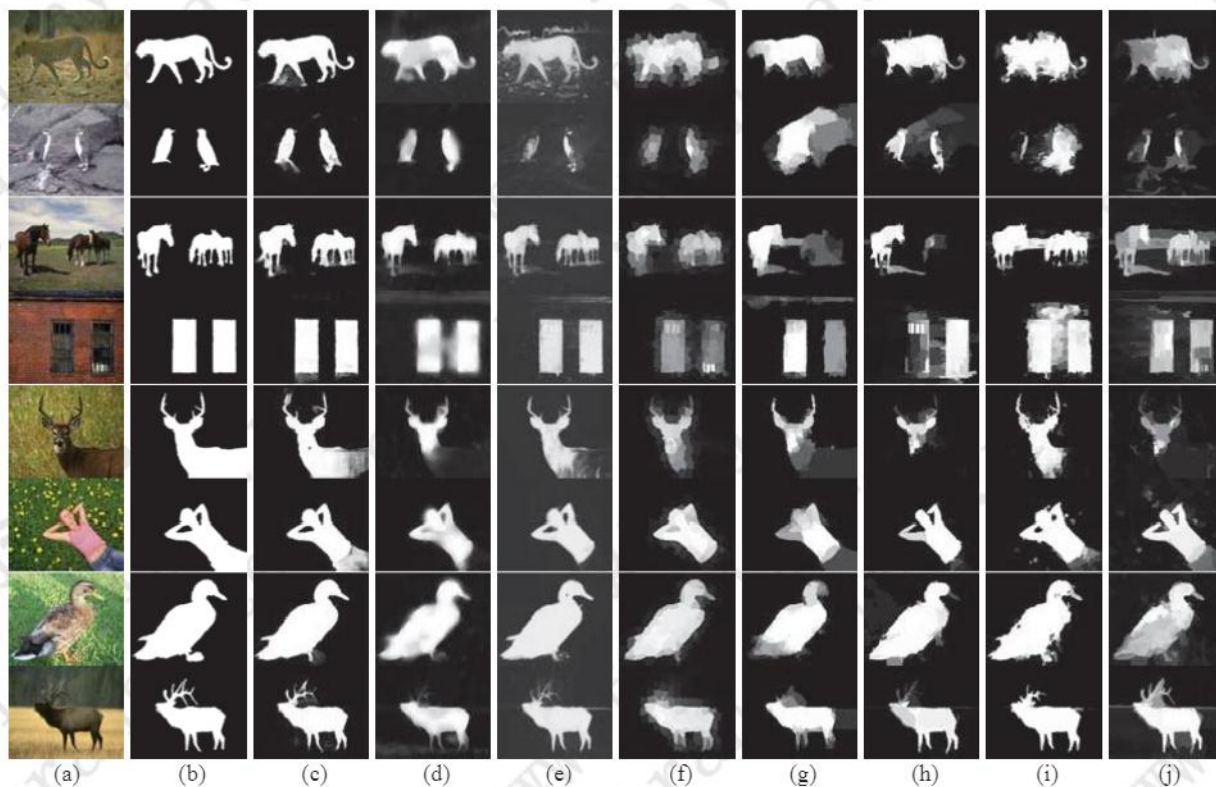


图6. 显著图的比较。(a) 输入图像; (b) 事实真相; (c) 我们的方法; (d) RFCN; (e) DCL; (f) DS; (g) LEGS; (h) 中密度纤维板; (i) ELD; (j) DRFI。每个数据集中的更多示例可以在补充材料中找到。

如图7所示。另外，VD模型和VE模型分别验证去卷积和插值上采样的有用性。VB和VC模型实现了比其他显著性更好的结果

方法。这进一步说明了我们方法的优势。

### 4.3. 泛化评估

为了验证我们方法的普遍性，我们在其他像素视觉任务上执行了其他实验。

**语义分割：**根据现有的作品[31, 20]，我们简单地将分类器更改为21个类，并在PASCAL VOC 2012数据集上执行语义分割任务[9]。我们的UCF模型使用Berekeley的扩展注释[11]，通过PASCAL VOC 2011培训和验证数据进行培训。我们实现了富有表现力的结果（平均IOU：68.25，平均像素准确度：92.19，像素准确度：77.28），这与其他最先进的分割方法非常相似。此外，尽管分割性能差距并不像显著性检测那么大，但我们的新上采样方法确实比常规解卷积性能更好（平均IOU：67.45对65.173，平均像素精度：91.21对90.84，pix.accuracy：76.18 vs 75.73）。

**眼睛固定：**眼睛预测的任务与我们的分类任务本质上不同。我们使用欧几里得损失来进行注视预测。我们提交结果

	AUC-J	sAUC	CC	NSS	IG
MIT300 [19]	0.8584	0.7109	0.7423	2.14	-
iSUN [47]	0.8615	0.5498	0.8142	-	0.1725
SALICON [17]	0.7621	0.6326	0.8453	-	0.3167

表3. 眼部数据集的结果。第一行的度量可以在[19, 17]中找到。

到MIT300 [19]，iSUN [47]和SALICON [17]标准设置的服务器。我们的模型也获得了Tab. 3. 以上关于语义分割和眼睛任务任务的结果表明，我们的模型在其他像素任务中具有很强的泛化能力。

### 5. 结论

在本文中，我们提出了一种新颖的用于显著性检测的完全卷积网络。利用重新建立的退出来促进概率训练和推理。这种不确定的学习机制使我们的方法能够学习不确定的卷积特征并产生更准确的显著性预测。还提出了一种新的上采样方法来减少去卷积操作的伪像，并且明确地强制网络为显著性检测学习精确的边界。广泛的评估表明，我们的方法可显著提高显著性检测的性能，并在其他像素视觉任务上显示出良好的泛化。确认。我们感谢Alex Kendall分享Seg-网络代码。本文得到中国自然科学基金会#61472060，#61502070，#61528101和#61632006的支持。



## 参考

- [1] B. Alexe, T. Deselaers和V. Ferrari. 什么是对象? 在 CVPR, 第73-80页, 2010年
- [2] A. Borji. 什么是显著对象? 数据集和显著物体检测的基准模型. IEEE TIP, 24 (2) : 742-756, 2015
- [3] A. Borji, M.-M. Cheng, H. Jiang和J. Li. 突出物体检测: 基准. IEEE TIP, 24 (12) : 5706-5722, 2015. 1, 6
- [4] M.-M. Cheng, NJMitra, X. Huang, PHS Torr和S.-胡先生. 基于全局对比度的显著区域检测. IEEE TPAMI, 37 (3) : 569-582, 2015. 5, 6
- [5] J. Deng, W. Dong, R. Socher, L.-J. 李, 李, 和费菲. Imagenet: 一个大规模的分层图像数据库. 在 CVPR, 2009年5
- [6] Y. Ding, J. Xiao和J. Yu. 图像重定向的重要性过滤. 在 CVPR, 第89-96页, 2011年
- [7] A. Dosovitskiy和T. Brox. 用卷积网络反转视觉表示. arXiv: 1506.02753, 2015
- [8] M. Everingham, LV Gool, CKI Williams, JM Winn和A. Zisserman. pascal可视对象类 (voc) 挑战. IJCV, 88: 303-338, 2010
- [9] M. Everingham, L. Van Gool, CKI Williams, J. Winn和A. Zisserman. PASCAL Visual Object Classes Challenge 2012 (VOC2012) 结果. 7, 8
- [10] Y. Gal和Z. Ghahramani. 辍学作为贝叶斯近似: 见解和应用. 在深度学习研讨会上, ICML, 2015. 1, 2
- [11] B. Hariharan, P. Arbelaez, L. Bourdev, S. Maji和J. Malik. 反向检测器的语义轮廓. 在ICCV, 2011. 8
- [12] K. He, X. Zhang, S. Ren和J. Sun. 深入研究直肠系统: 在imagenet分类上超越人类级别的性能. 在CVPR, 第1026-1034页, 2015年. 3
- [13] GE Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever和RR Salakhutdinov. 通过防止特征检测器的共同适应来改进神经网络. arXiv: 1207.0580, 2012. 2, 3
- [14] S. Hong, T. You, S. Kwak和B. Han. 用卷积神经网络学习判别显著图进行在线跟踪. arXiv: 1502.06796, 2015
- [15] X. Hou和L. Zhang. 显著性检测: 光谱残差方法. 在 CVPR中, 1-8页, 2007年
- [16] S. Ioffe和C. Szegedy. 批量标准化: 通过减少内部协变量来加速深度网络培训. arXiv: 1502.03167, 2015. 3
- [17] M. Jiang, S. Huang, J. Duan和Q. Zhao. Salicon: 在上下文中的显著性. 在CVPR, 2015年. 8
- [18] P. Jiang, H. Ling, J. Yu和J. Peng. 通过飞碟的显著区域检测: 唯一性, 对焦性和对象性. 在ICCV, 第1976-1983页, 2013年6, 7页
- [19] T. 贾德, 杜兰德和A. 托拉尔巴. 预测人类xation的显著性计算模型的基准. MIT技术报告, 2012年. 8
- [20] A. 肯德尔, V. 巴德里纳拉亚南和R. 西波拉. 贝叶斯分段: 用于场景理解的深度卷积编码器 - 解码器体系结构中的模型不确定性. arXiv: 1511.02680, 2015. 2, 6, 8
- [21] DA Klein和S. Frintrop. 显著物体检测的特征统计的中心 - 环绕发散. 在ICCV, 第2214-2219页, 2011年
- [22] G. Lee, Y.-W. Tai和J. Kim. 编码的低级别距离图和高级特征的深度显著性. 在CVPR中, 2016年6月6, 7日
- [23] G. Li和Y. Yu. 基于多尺度深度特征的视觉显著性. 在CVPR, 第5455-5463页, 2015. 2
- [24] G. Li和Y. Yu. 用于显著物体检测的深度对比学习. 在CVPR, 第478-487页, 2016. 6, 7
- [25] X. Li, H. Lu, L. Zhang, X. Ruan和M.-H. 杨. 通过密集和稀疏重建进行显著性检测. 在ICCV, 第2976-2983, 2013. 6, 7
- [26] X. Li, L. Zhao, L. Wei, M.-H. Yang, F. Wu, Y. Zhuang, H. Ling和J. Wang. Deep saliency: 用于显著物体检测的多任务深度神经网络模型. IEEE TIP, 25 (8) : 3919-3930, 2016. 2, 6, 7
- [27] Y. Li, X. Hou, C. Koch, J. Rehg和A. Yuille. 显著物体分割的秘密. 在CVPR, 第280-287页, 2014. 1, 6
- [28] J. Long, E. Shelhamer和T. Darrell. 用于语义分割的完全卷积网络. 在CVPR, 第3431-3440页, 2015年. 2
- [29] V. Mahadevan和N. Vasconcelos. 基于显著性的判别式追踪. 在CVPR中, 第1007-1013页, 2009年
- [30] L. Marchesotti, C. Cifarelli和G. Csurka. 用于图像缩略图的应用程序的视觉显著性检测框架. 在ICCV, 第2232-2239页, 2009年
- [31] H. Noh, S. Hong和B. Han. 学习反卷积网络进行语义分割. 在ICCV, 第1520-1528页, 2015年. 1, 7, 8
- [32] A. Odena, V. Dumoulin和C. Olah. 解卷积和棋盘文物. [http://distill.pub/2016/deconv- 棋盘/](http://distill.pub/2016/deconv-棋盘/), 2016. 2
- [33] Y. Qin, H. Lu, Y. Xu, and H. Wang. 通过元胞自动机进行显著性检测. 在CVPR, 第110-119页, 2015年1, 6, 7
- [34] C. Rother, V. Kolmogorov和A. Blake. Grabcut: 使用迭代图切割的交互式前景提取. 在ACM TOG, 第23卷, 第309-314页, 2004年
- [35] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford和X. Chen. 改进的gans技术. arXiv: 1606.03498, 2016
- [36] W. Shi, J. Caballero, F. Huszar, J. Totz, AP Aitken, R. Bishop, D. Rueckert和Z. Wang. 使用有效的亚像素卷积神经网络的实时单个图像和视频超分辨率. 在CVPR, 2016. 2
- [37] C. Siagian和L. Itti. 使用与视觉关注共享的特征进行快速生物启发的场景分类. IEEE TPAMI, 29 (2) : 300-312, 2007
- [38] K. Simonyan和A. Zisserman. 用于大规模图像识别的非常深的卷积网络. arXiv: 1409.1556, 2014. 5
- [39] N. Srivastava, GE Hinton, A. Krizhevsky, I. Sutskever和R. Salakhutdinov. 辍学: 一种防止神经网络过拟合的简单方法. JMLR, 15 (1) : 1929-1958, 2014. 3
- [40] J. Sun和H. Ling. 缩放和对象感知图像重定向用于缩略图浏览. 在ICCV, 第1511-1518页, 2011年

- [41] N. Tong, H. Lu, X. Ruan和M.-H. 杨。通过自举学习的显著物体检测。在CVPR, 第1884-1892页, 2015年6月7日
- [42] D. 文森特和弗朗切斯科。深度学习的卷积算法指南。arXiv: 1603.07285, 2016
- [43] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio和P.-A. Manzagol。叠加去噪自动编码器: 在具有局部去噪标准的深度网络中学习有用的表示。JMLR, 11 (Dec) : 3371-3408, 2010. 3
- [44] L. Wang, H. Lu, X. Ruan和M.-H. 杨。深度网络通过局部估计和全局搜索进行显著性检测。在CVPR, 第3183-3192页, 2015. 2, 6, 7
- [45] L. Wang, L. Wang, H. Lu, P. Zhang和X. Ruan。循环完全卷积网络的显著性检测。在ECCV, 第825-841页, 2016年. 1, 2, 6, 7
- [46] T. Wang, L. Zhang, H. Lu, C. Sun和J. Qi。用于显著性检测的核化子空间排序。在ECCV, 第450-466页, 2016年
- [47] P. Xu, KA Ehinger, Y. Zhang, A. Finkelstein, SR Kulkarni和J. Xiao。Turkergaze: 采用基于网络摄像头的眼动追踪的众包显著性。arXiv预印本arXiv: 1504.06755, 2015. 8
- [48] Q. Yan, L. Xu, J. Shi和J. Jia。分层显著性检测。在CVPR, 第1155-1162页, 2013年. 6
- [49] C. Yang, L. Zhang, H. Lu, X. Ruan和M.-H. 杨。通过基于图形的多方排名进行显著性检测。在CVPR, 第3166-3173页, 2013. 1, 6
- [50] R. Zhao, W. Ouyang, H. Li, and X. Wang。多情境深度学习的显著性检测。在CVPR中, 第1265-1274页, 2015. 2, 6, 7