深度无监督显着性检测：多重噪声标签透视

张菁*1, 2，张彤*2, 3，戴宇超†1，梅塔莎哈兰迪2, 3和理查德哈特利2 1西北工业大学，西安，中国
2澳大利亚堪培拉澳大利亚国立大学
3DATA61，CSIRO，堪培拉，澳大利亚

摘要

目前的深度显着性检测方法的成功在很大程度上取决于每像素标记形式的大规模监督的可用性。 这种监督虽然劳动强度大，但并不总是可能，但往往会阻碍学习模式的泛化能力。 相比之下，基于无监督显着性检测方法的传统手工特征尽管已经被深度监督方法所超越，但通常与数据集无关并且可以在野外应用。 这提出了一个自然的问题：“是否可以在不使用标记数据的情况下学习显着性地图，同时提高泛化能力？”。为此，我们提出了一种无监督的新视角[1] 通过学习由“弱”和“嘈杂”无监督手工显着方法产生的多重噪声标签来显着性检测。 我们用于无监督显着性检测的端到端深度学习框架包括潜在显着性预测模块和噪声建模模块，它们协同工作并共同优化。 显式噪声建模使我们能够以概率的方式处理噪声显着图。 对各种基准数据集的广泛的实验结果表明，我们的模型不仅优于所有无监督显着性方法，而且还具有与最近最先进的监督深显着性方法相当的性能。

1. 介绍

显着性检测旨在识别与人类感知一致的图像中视觉上感兴趣的物体，这些物体是各种视觉任务所固有的，例如

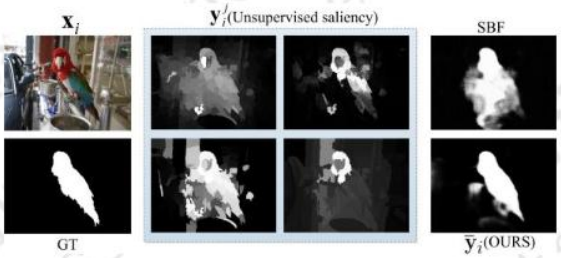


图1. 弱“嘈杂”显着图的无监督显着性学习。 给定输入图像xi及其相应的无监督显着图y^j，我们的框架通过联合优化显着性预测模块和噪声建模模块来学习潜在显着性图yii。与SBF相比[35]也从无监督显着性学习，但采用不同策略，我们的模型取得了更好的性能。

上下文感知图像编辑[36]，图像标题生成[31]。 取决于是否使用了人类注释，显着性检测方法可以大致分为：无监督方法和监督方法。 前者直接基于各种先验计算显着性（例如，中心先验[9]，全球对比之前[6]，之前的背景连通性[43]等），用人类的知识对其进行总结和描述。 后者通过利用大规模人类注释数据库的可用性学习从彩色图像到显着图的直接映射。

基于卷积神经网络（CNN）的强大学习能力，深度监督显着性检测方法[42, 11, 40]实现最先进的表演，大幅超越无监督方法。 这些深度显着性方法的成功在很大程度上取决于具有像素级人类注释的大规模训练数据集的可用性，这不仅是劳动密集型的，而且可能阻碍学习网络模型的泛化能力。 相比之下，无监督显着方法尽管已经被深度监督方法所超越，但通常与数据集无关并且可以在野外应用。

*这些作者在这项工作中做出了同样的贡献。
†Y。 戴（daiyuchao@nwpu.edu.cn）是通讯作者。
[1]在无监督学习中可能存在多种定义，本文中，我们将无监督学习称为学习，而没有任务特定的人类注释，例如我们任务中的密集显着图。

在本文中，我们提出了一种新颖的端对端深度学习框架，用于显着性检测，不含人类注释，因此“无监督”（见图1）1为可视化）。 我们的框架建立在现有高效和有效的无监督显着性方法和深度神经网络的强大能力之上。 无监督显着性方法是用人类知识制定的，不同的无监督显着性方法利用不同的人为设计先验显着性检测。 它们很嘈杂（与地面真实人类注释相比），并且在预测显着性地图时可能具有特定于方法的偏差。通过利用现有的无监督显着图，我们能够消除对劳动密集型人类注释的需求，也可以通过从多个无监督显着性方法联合学习不同先验，我们能够获得这些无监督显着性的补充信息。

为了有效地利用这些有噪声但信息量大的显着图，我们提出了一个新的视角来解决这个问题：用不同的融合策略从无监督显着性方法中去除显着性标记中的噪声[35]，*我们明确地模拟了显着图中的噪声。 如图所示。*2，我们的框架由两个连续的模块构成，即基于当前噪声估计和噪声显着性图形学习从彩色图像到“潜在”显着图的映射的显着性预测模块，以及适合噪声的噪声建模模块噪声显着性映射，并基于更新的显着性预测和噪声显着性图更新不同显着图中的噪声估计。 这样，我们的方法利用了概率方法和确定性方法，其中潜在显着性预测模块以确定性方式工作，而噪声建模模块以概率方式适合噪声分布。 实验表明，我们的策略非常有效，只需要几轮[2]直到收敛。

据我们所知，这是考虑的想法
无监督显着性地图作为从多个噪声标签中学习是全新的并且与现有的无监督深显着性方法不同（例如[35]）。 我们的主要贡献可以概括为：

1) 我们给无监督深显着性检测提供了一种新的视角，并从多种噪声无监督显着性方法学习显着性图。 我们将问题表述为潜在显着性预测模块和噪声建模模块的联合优化。
2) 我们的深层显着模型是以端对端的方式进行培训的，无需使用任何人类注释，从而导致极其便宜的解决方案。
3) 对七个基准数据集进行广泛的绩效评估表明，我们的框架胜过了前 -

[2]在我们的论文中，时代意味着完整的传递所有的训练数据，迭代意味着一个批次的完整传递，而一轮意味着噪声模块的更新。

在无监督的方法上有很大的余地，同时利用最先进的深度监督显着性检测方法获得可比的结果[11, 40].

2. 相关工作

根据是否使用人类注释，显着性检测技术可以大致分为无监督和监督方法。 基于深度学习的方法是后者的特例。 我们还将讨论多个嘈杂标签的学习。

2.1. 无监督显着性检测

在深度学习革命之前，显着性方法主要依赖于不同的先验和手工特征[43, 7, 6, 9]。 我们引用感兴趣的读者[2]和[3]进行调查和基准比较。 之前的颜色对比度已经在超像素级别被利用，6]。 沉和吴
[27]通过利用突出对象之前的稀疏性将显着性检测制定为低秩矩阵分解问题。 突出类似物体区域的物体也被用于[15]来标记作为对象具有更高可能性的区域。朱等人 [43]提出了一个强大的背景测量，即“边界连通性”以及一个优化框架来测量每个超像素的背景。在中心之前，[9]检测表示场景的图像区域，特别是那些靠近图像中心的图像区域。

2.2. 监督显着性检测

传统的监督技术，如[14, 17]，将显着性检测制定为回归问题，并且训练分类器以在像素级或超级像素级分配显着性。 最近，深度神经网络已被成功用于显着性检测[40, 26, 41, 29, 11, 22,
42, 19, 28, 20, 38, 39, 37]。 深度网络可以对高级语义特征进行编码，因此比无监督显着性方法和非深度监督方法更有效地捕捉显着性。 深度显着性检测方法通常训练深度神经网络来为每个像素或超像素指定显着性。李和余[19]使用现有CNN模型的学习功能来替换手工功能。 最近，Cheng等人 [11]提出了一种多分支短连接的深度监督框架，嵌入了高级和低级特征以实现准确的显着性检测。 同样的目的，一个多层次的深度特征聚合框架在[40]。 自顶向下策略和惩罚边缘误差的损失函数在[26].

2.3. 学习嘈杂的标签

尽管深度技术是显着性检测的选择方法，但很少有研究明确地解决了不可靠的显着性学习问题

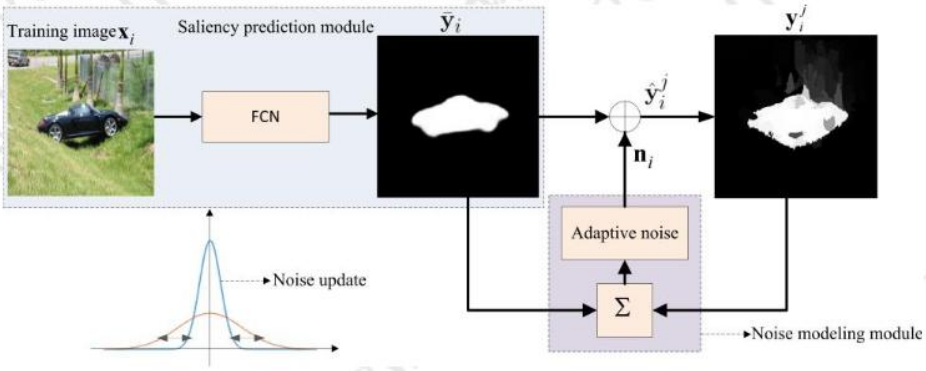


图2. 我们的显着性检测框架的概念图，它由“潜在”显着性预测模块和噪声建模模块组成。 给定输入图像，通过基于手工特征的无监督显着性检测方法生成噪声显着图。 我们的框架在统一损失函数下共同优化了这两个模块。 显着性预测模块针对基于当前噪声估计和噪声显着图的学习潜在显着性图。 噪声建模模块基于更新的显着性预测和噪声显着图更新不同显着图中的噪声估计。 在我们的实验中，整体优化在几轮中收敛。

和嘈杂的标签[35]。 带有噪音标签的学习主要是在存在不准确的班级标签时学习分类模型。 Whitehill等人 [30]解决了基于许多具有不同专业知识的贴标机提供的标签来选择正确标签的问题。 金达尔等人。[16]提出了一个辍学正则化噪声模型，通过增加现有的深层网络和噪声模型来解释标签噪声。 姚等人。[34]提出了一个质量嵌入模型来推断噪声标签的可信度。 与上述带有噪声标签方法的监督学习不同，Lu et al。 [25]提出了一个弱监督的语义分割框架来处理带噪标签。

据我们所知，[35]是第一个也是唯一一个在没有人类注释的情况下学习显着性的深度方法，其中将来自无监督显着性方法的显着性图与人工设计的规则融合为“图像内”融合流和“图像间”融合流以生成学习课程。该方法迭代地用其相应的显着图代替低可靠性的图像间显着图。 它们的递归优化取决于专用设计，并且计算成本很高。 不同于 [35]，我们将无监督显着性学习作为潜在显着性和噪声建模的联合优化。 我们的方法不仅更简单，更容易实施，而且还优于[35]和现有的无监督显着性方法。 此外，与最新的深度监督显着性检测方法相比，我们的方法产生了有竞争力的表现。

3. 我们的框架

针对在没有人类注释的情况下实现深度显着性检测，我们提出了一种端到端噪声模型集成的深层框架，它基于现有的高效和有效的无监督显着性检测方法和深度神经网络的强大能力。

给定一个彩色图像x_i，我们想从它的M个噪声显着性图y^j，j = 1，...，M利用不同的无监督显着性方法学习一个更好的显着图[32,13,21,43]。 一个简单而直接的解决方案是使用嘈杂的显着图作为“代理”人注释，并训练具有这些噪声显着图的深层模型作为监督。 然而，众所周知，网络培训非常容易产生监管信号中的噪音。 由于标签之间强烈的不一致性，多个标签（平均训练，作为多个标签对待）的简单融合也不起作用。 虽然利用噪声显着图可能还有许多其他潜力，但它们都基于人工设计的管线，因此无法有效利用最佳方式。 相反，我们提出了一种有原则的方法来从使用多个噪声标签推断显着性图并同时估计噪声。

3.1. 联合显着性预测和噪声建模

与现有的手动设计程序和基于深度学习的管道相比[35]，我们提出了从无监督显着性学习问题的新视角。如图所示。2，我们的框架由两个连续的模块组成，即显着性预测模块，学习从彩色图像到“潜在”显着图的映射，以及适合噪声的噪声建模模块。 这两个模块协同工作以适应噪声显着图。 通过对噪声进行明确建模，我们能够在没有任何人类注释的情况下训练深度显着性预测模型，从而实现无监督深度显着性检测。

3.2. 损失函数

我们从一组训练图像开始，表示为X = {x_i, i = 1, ... , N}和一组M个不同的显着图这些图像表示为Y = {y_i^j, i = 1, ... , N; j = 1, ... , M}，其中N是训练图像的数量。这些通过应用M个不同的手工"贴标机"进行预先计算。在整个讨论过程中，我编制了训练图像索引，j为手工贴标机编制索引。我们提出了一个具有参数Θ用于显着性检测的神经网络，该神经网络计算每个im-1的显着图y_i= f (x_i, Θ)

年龄。我们的想法是模拟每个手工贴标签作为y_i加噪声之和：y_i^j= y_i+ n_i^j，其中n_i^j是

样本是从一些概率（"噪声"）分布q_i中选择的，它将被估计。为了简化这项工作，假定分布q取决于x_i，而不是在标签j上[3]。我们假设噪声分布q_i的简单模型，即它是零均值高斯，对于每个图像x_i的每个像素都是独立的。因此，总分布q = q_1, q_2, ... , q_N被假定为独立的对于所有的i和像素（m, n），并且通过参数 - eter setΣ={$\sigma_{i,mn}$}，其中i表示训练图像，（m, n）表示像素坐标。有时，分布q将被表示为q（Σ）以强调参数Σ的作用。通过这个简单的参数化，很容易为任何i和j生成噪声样本n_i^j。

给定Θ，Σ和输入图像x_i，根据下式生成显着图y_i^j：

$$(1)$$

其中每个n_i^j是从分布q_i（Σ）中抽取的样本。在训练过程，网络参数Θ和Σ的噪声模型被更新以最小化适当的损失函数。损失函数有两部分：

$$L(\Theta,\ \Sigma) = L_{预计值}(\Theta,\ \Sigma) + \lambda L_{噪声}(\Theta,\ \Sigma), \quad (2)$$

λ是调节器来平衡这两个项。在我们的优化框架下，增加噪声模型的方差将会使预测损失$_{预计值}$变大并降低$L_{噪声}$。同时，保持方差较低会降低交叉熵损失$L_{预计值}$，但增加$L_{噪声}$。因此我们的模型在这些之间取得平衡两个损失并且趋于最小化总体状态失利。这两项损失如下所述：

显着性预测：对于潜在显着性预测模块，由于其在特征学习和特征表示方面的优越能力，我们使用完全卷积神经网络（FCN）。我们使用传统的交叉熵损失并且在整个训练图像中

预测性损失$L_{预计值}$旨在衡量预测标签y_i^j与手工制作的一致性标签y_i^j。交叉熵损失用于此目的，以及模型值y和的交叉熵损失

"地面实况"价值y（噪音标签）由下式给出：

$$L_{CE} = -\ (y\ \log\ (y) + (1-y)\ \log\ (1-y)\ )\ 。\quad (3)$$

这适用于所有像素（m, n），所有标记器j和所有测试图像x_i以给出总预测损失。

$$L_{预计值}(\Theta,\ \Sigma) = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{m,n} L_{CE}(Y_{我加}^j,\ y_{我加}^j), \quad (4)$$

其中y_i^j是我们在像素（m, n）处的噪声显着图预测，其可以通过1）元素方式，并且y_i^j被截断以位于[0, 1]的范围内。

噪声建模为了有效处理来自不同无监督显着图贴标机的噪声显着图，我们建立了一个近似噪声的概率模型，并将其与我们的确定性部分（潜在显着性预测模型，如图1所示）2）。通过这种方式，我们的整个模型可以以端对端的方式进行训练，以最大限度地减少整体损失函数Eq。（2).

噪声损失$_{噪声}$测量（针对每个训练图像**xi）噪声分布q_i（Σ）与测量y的经验方差关于**网络的输出y_i= f (x_i; Θ)。更确切地说，给定输入x_i，定义n_i^j= y_i^j - y_i，经验误差每个y_i^j相对于网络预测。对于每个像素位置（m, n），这提供来自零点的M个样本，平均高斯概率分布p_i，其每个像素的方差可写为$\sigma_{我加入}$。p_i的完整参数集表示为$\hat{\Sigma}$={$\sigma_{我加入}$}。

由于估计n_i^j的真实后验分布是棘手的，因此我们建议通过顺序优化先验参数来逼近它。我们假设噪声是由一些随机过程产生的，涉及一个不可观测的连续随机变量集Σ。从编码器的角度来看，未观测到的变量n可以在 - 诠释为一种潜在的表现形式。在这里，我们将y_i^j模型化为一个概率编码器，因为给出了图像x参数$_{我和网络}$

明智地计算损失函数元素。

[3]假设分布q_i也依赖于标签者j被观察到不会改善结果

Θ它产生代码n的可能值的分布（例如高斯）。 该过程包含两个步骤：（1）从先前的某个生成噪声图n_i分布$q(\Sigma^*)$； （2）产生噪声图n_i'并估计相应的参数σ_i

相应的噪声损失被定义为分布p_i和q_i之间的KL散度。

$$L_{噪声}(\Theta, \Sigma) = \sum_{i}^{N} KL(q(\Sigma_i) p(\Sigma\hat{}i))。 \quad (5)$$

由于我们采用高斯分布作为噪声模型的先验分布，因此KL散度具有封闭形式的解决方案：

$$KL(q(\sigma)\,p(\sigma)) = \log(\sigma/\sigma) + \frac{\sigma^2 + (\mu - \hat{\mu})^2}{2\hat{\sigma}^2} - \frac{1}{2}, \quad (6)$$

基于这个方程，我们可以将每个坐标$_i$（m，n）的σ^2更新为

$$t)\,{}_i^2)^2 = (\sigma^t{}_i)^2 + \alpha\ (\ (_i\sigma^t)^2\ _i \quad (7)$$

通过区分方程（6）关于σ^2，其中α是步长，本文设α = 0.01。

对于不同的图像，我们有相应的噪声图，它遵循具有不同方差的iid高斯分布。 因此，如果同时优化FCN参数Θ和噪声参数Σ，则很难收敛。 为了平滑训练整个网络，在预测损失收敛后更新噪声模块的参数。 给定图像的噪声图在一轮中从相同的分布中采样，但是它们在每一轮中都被更新。 在第一轮中，我们将噪声方差初始化为零，并且训练FCN直到它收敛。 根据显着性预测和噪声标签的方差，我们更新每幅图像的噪声方差并重新训练网络。 通过最小化损失函数Eq。（2）用这个程序，我们可以训练网络并估计相应的噪声图。

3.3。 基于深度噪声模型的显着性检测器

网络体系结构我们在DeepLab网络上构建我们的潜在显着性预测模块[4]，其中深CNN（ResNet-101 [10]特别是）最初设计用于图像分类的目的是通过1）将所有完全连接的层转化为卷积层和2）通过扩张卷积来增加特征分辨率[4]。 数字2显示了我们框架的整体结构。具体来说，我们的模型需要一个重新缩放的图像x_i 425 ×425作为输入。 对于训练，噪声模型用于迭代更新显着性预测y^j，并且在测试阶段将其排除，其中潜在显着性预测输出y_i2是我们预测的显着图。

实现细节：我们使用Caffe [12]，最大时期为20.我们通过使用经过训练的图像分类的深度残差模型初始化我们的模型[10]。 我们用动量0.9和学习速率降低的随机梯度下降法

培训损失没有减少时为90%。 基础学习 -

使用“多聚”衰减策略将初始化速率初始化为1e-3 [12]。为了进行验证，我们将“test iter”设置为500（测试批量大小1）以覆盖全部500个验证图像。 在一台装有NVIDIA Quadro M4000 GPU的个人电脑上，培训需要4个小时，培训批量为1，“iter size”为20。

4. 实验结果

在本节中，我们报告了各种显着性检测基准数据集的实验结果。

4.1. 建立

数据集：我们评估了我们提出的模型在7个显着性基准数据集上的表现。 来自MSRA-B数据集的3,000幅图像[24]用于获取有噪声的标签（其中2,500个用于训练的图像和500个用于验证的图像），剩余的2,000个图像用于测试。 MSRA-B数据集中的大多数图像只有一个显着的对象。 ECSSD数据集[32]包含1,000个具有语义意义但结构复杂的图像。 DUT数据集[33]包含5,168个图像。 SOD显着性数据集[14]包含300个图像，其中许多图像包含多个低对比度的显着物体。 SED2 [1]数据集包含100个图像，每个图像包含两个显着对象。 PASCAL-S [23]数据集是从PASCAL VOC [8]数据集并包含850个图像。 THUR数据集[5]包含五个班的6,232幅图像，即“蝴蝶”，“咖啡杯”，“狗跳”，“长颈鹿”和“飞机”。

无监督显着性方法：在本文中，我们从现有的无监督显着性检测方法中学习无监督显着性。 在我们的实验中，我们选择RBD [43]，DSR [21]，MC [13]和HS [32]由于其有效性和效率，如[3].

竞争方法：我们将我们的方法与10种最先进的深显着性检测方法（带有干净标签）进行了比较：DSS [11]，NLDF [26]，护身符[40]，UCF [41]，SRM [35]，DMT [22]，RFCN [28]，DeepMC [42]，MDF [19]和DC [20]，5种传统的基于手工特征的显着性检测方法：DRFI [14]，RBD [43]，DSR [21]，MC [13]和HS [32]，这被证明在[3]作为深度学习革命之前的最新技术方法，以及最近的无监督深显着性检测方法SBF [35].

评估指标：我们使用3个评估指标，包括平均绝对误差（MAE），F-measure以及Precision-Recall（PR）曲线。 MAE可以更好地估计估计的和地面真实显着图之间的差异性。 它是地面实况与估计显着图之间的平均每像素差异，归一化为[0,1]，定义为：

$$MAE = \frac{1}{W_x H_x} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x, y) - GT(x, y)|, \quad (8)$$

其中W和H是相应显着图S的宽度和高度，GT是地面真实显着图。

F-measure（F_β）被定义为加权谐波

表1. 在7个基准数据集上，包括我们在内的不同方法的平均F-度量（F_β）和MAE的性能。

	MSRA-B		ECSSD		DUT		SED2		帕斯卡		THUR		草皮	
方法	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE
BL1	.7905	.0936	.7205	.1444	.5825	.1369	.7773	.1112	.6714	.2206	.5953	.1339	.6306	.1870
BL2	.6909	.1710	.6542	.2170	.4552	.2951	.7232	.1406	.6776	.2409	.5119	.2545	.5928	.2566
BL3	.8879	.0587	.8717	.0772	.7253	.0772	.8520	.0819	.8264	.1525	.7368	.0749	.7922	.1231
OURS	.8770	.0560	.8783	.0704	.7156	.0860	.8380	.0881	.8422	.1391	.7322	.0811	.7976	.1182

表2. 对于包括我们在内的七种基准数据集中的不同方法（F_β）和MAE的性能（最佳粗体）。 从DSS到DC是基于深度学习的监督方法，从DRFI到HS是基于手工特征的无监督方法，SBF和OURS是基于深度学习的无监督显着性检测方法。

	MSRA-B		ECSSD		DUT		SED2		帕斯卡		THUR		草皮	
方法	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE	F_β	MAE
DSS [11]	.8941	.0474	.8796	.0699	.7290	**.0760**	.8236	.1014	.8243	.1546	.7081	.1142	.8048	.1118
NLDF [26]	**.8970**	.0478	**.8908**	.0655	**.7360**	.0796	-	-	.8391	.1454	-	-	**.8235**	**.1030**
护身符[40]	-	-	.8825	**.0607**	.6932	.0976	**.8745**	**.0629**	.8371	**.1292**	.7115	.0937	.7729	.1248
UCF [41]	-	-	.8521	.0797	.6595	.1321	.8444	.0742	.8060	.1492	.6920	.1119	.7429	.1527
SRM [29]	.8506	.0665	.8260	.0922	.6722	.0846	.7447	.1164	.7766	.1696	.6894	.0871	.7246	.1369
DMT [22]	-	-	.7589	.1601	.6045	.0758	.7778	.1074	.6657	.2103	.6254	.0854	.6978	.1503
RFCN [28]	-	-	.8426	.0973	.6918	.0945	.7616	.1140	.8064	.1662	.7062	.1003	.7531	.1394
DeepMC [42]	.8966	.0491	.8061	.1019	.6715	.0885	.7660	.1162	.7327	.1928	.6549	.1025	.6862	.1557
MDF [19]	.7780	.1040	.8097	.1081	.6768	.0916	.7658	.1171	.7425	.2069	.6670	.1029	.6377	.1669
DC [20]	.8973	**.0467**	.8315	.0906	.6902	.0971	.7840	.1014	.7861	.1614	.6940	.0959	.7603	.1208
DRFI [14]	.7282	.1229	.6440	.1719	.5525	.1496	.7252	.1373	.5745	.2556	.5613	.1471	.5440	.2046
RBD [43]	.7508	.1171	.6518	.1832	.5100	.2011	.7939	.1096	.6581	.2418	.5221	.1936	.5927	.2181
DSR [21]	.7227	.1207	.6387	.1742	.5583	.1374	.7053	.1452	.5785	.2600	.5498	.1408	.5500	.2133
MC [13]	.7165	.1441	.6114	.2037	.5289	.1863	.6619	.1848	.5742	.2719	.5149	.1838	.5332	.2435
HS [44]	.7129	.1609	.6234	.2283	.5205	.2274	.7168	.1869	.5948	.2860	.5157	.2178	.5383	.2729
SBF [35]	-	-	.7870	.0850	.5830	.1350	-	-	.7780	.1669	-	-	.6760	.1400
OURS	.8770	.0560	.8783	.0704	.7156	.0860	.8380	.0881	**.8422**	.1391	**.7322**	**.0811**	.7976	.1182

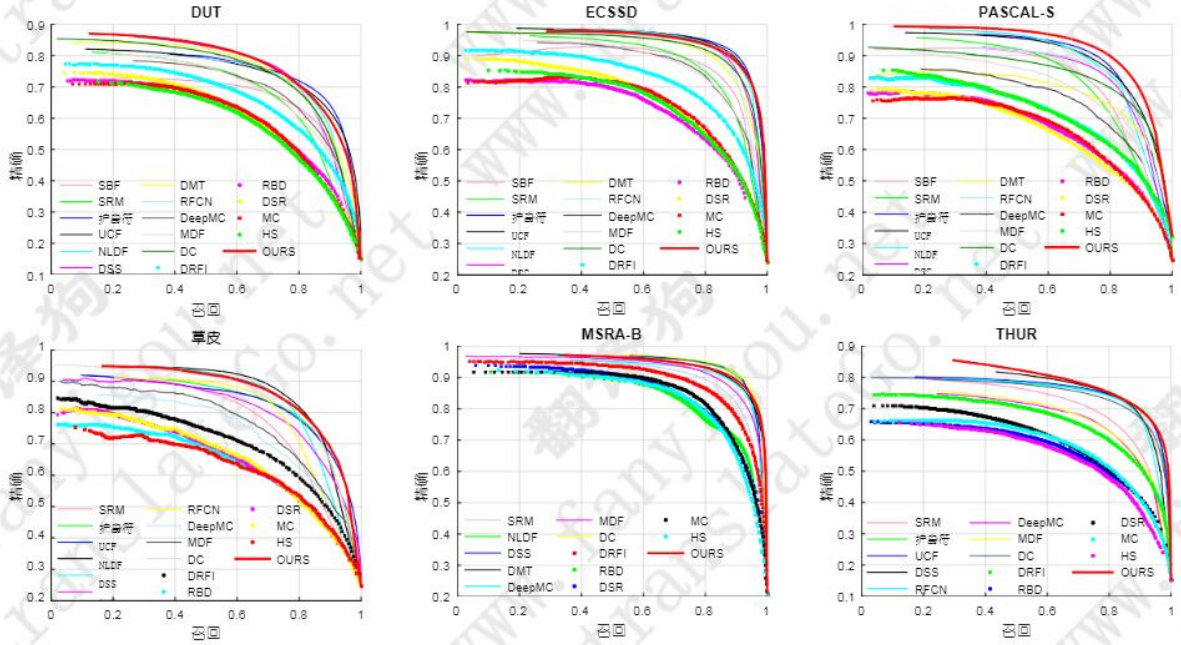


图3. 六个基准数据集（DUT，ECSSD，PASCAL-S，SOD，MSRA-B，THUR）上的PR曲线。 在屏幕上最好看。

精确度和召回率的意思是：

$$F_\beta = \frac{(1+\beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall}, \quad (9)$$

其中$\beta^2 = 0.3$，Precision对应于正确检测的显着像素的百分比，Recall是检测到的显着像素相对于地面实况的比例

显着像素的数量。 PR曲线是通过在[0,255]范围内对显着图进行阈值处理得到的。

4.2. 基线实验

由于可能有不同的方法来利用多重噪声显着图，并且为了公平地比较我们的任务的简单解决方案，我们运行以下三个基线方法，结果在表格中报告1.

基线1使用噪声无监督显着伪基准真实：对于给定的输入图像x_i和其基于M的手工特征基于显着性图y^j，j = 1，...，M，我们得到具有带噪声标签的M个图像对x_i，y^j，j = 1，...，M。然后我们训练一个深层模型[10]基于这些噪声标签，结果在表格中显示为“BL1”1.

基线2：使用平均无监督显着性作为伪基础事实：我们不使用所有四个无监督显着性作为基础事实，而是使用那些无监督显着性的平均显着性图作为伪基础事实，并在表中训练另一个基线模型“BL2”1.

基线3：带有地面实况监督的监督式学习：我们提出的框架由显着性预测模块和噪声建模模块组成，以有效利用噪声显着图。 为了说明我们的模型可以实现的最佳性能以及为我们的框架提供基线比较，我们直接用干净的标签来训练我们的潜在显着性模块，这自然给出了显着性检测性能的上限。 表格中报告了结果“BL3”1.

分析：在表中1，我们将我们的无监督显着性方法与上述基准配置进行比较。 我们的方法在很大程度上明显优于BL1和BL2，显示了我们的端到端学习框架的优越性。 如表中所示1，BL1的性能优于BL2的性能。这是因为：1）对于BL1，我们有12000个训练图像对（4个无监督显着方法），而对于BL2，我们有3000个平均噪声标签； 2）由于那些无监督显着性方法往往倾向于不同的先验显着性检测，并且它们的显着性图可以在某种程度上是互补或有争议的。 简单地平均这些显着性图导致更糟糕的代理显着性图监督。 与BL3相比，我们的无监督方法获得了高度可比的结果，而BL3是使用地面真实清洁标签进行训练并且没有噪音。这表明，通过共同学习潜在显着性图并在统一框架中对噪声进行建模，即使没有任何人类注释，我们也可以学习所需的可靠显着图。

4.3. 与最先进的技术进行比较

定量比较我们比较了我们的方法与十一种最重要的显着性方法和五种常规方法。 结果列于表中2和图。3，其中“OURS”代表我们模型的结果。 表2 表明在这7个基准数据集上，深度监督方法明显优于传统方法，MAE降低2%-12%，这进一步证明了深度显着性检测的优越性。

MSRA-B是一个相对简单的数据集，大多数显着对象在整个图像中占主导地位。 最近的深度监督显着性方法[11] [26] [40]可以达到0.8970的最高平均F-measure，而我们的无监督方法无人注释可以达到0.8770的平均F-measure，这只是稍微更糟。 DUT数据集具有超过25%的显着性占用小于4%的图像。 小型突出物体检测非常具有挑战性，这增加了该数据集的难度。 与所有竞争方法相比，我们实现了第三高的平均F-measure。THUR数据集是我们在本文中使用的最大的数据集，并且大多数图像具有复杂的背景。 现有竞争方法的平均F-measure / MAE达到0.7115 / 0.0854，而我们的方法达到最佳平均F-measure和MAE为0.7322 / 0.0811。SBF [35]使用图像间和图像内置信度图作为伪地面真值来训练基于无监督显着性的无监督深度模型，这与我们从无噪声显着性预测显着性的方法（如从有噪标签学习）的预测完全不同。 表2 表明我们的框架导致更好的表现，平均F值提高10%，MAE平均降低3%。图。3 显示了我们的方法的PR曲线与四个基准数据集上的竞争方法之间的比较。 对于PASCAL-S和THUR数据集，我们的方法几乎排在第1位，对于其他三个数据集，与竞争性深度监督方法相比，我们的方法实现了有竞争力的表现。 这些实验共同证明了我们提出的无监督显着性检测框架的有效性。

定性比较图4 演示了几种视觉比较，其中我们的方法始终优于竞争方法，尤其是我们用来训练模型的那四种无监督显着性。 第一个图像是一个简单的场景，大多数竞争方法可以获得好的结果，而我们的方法在大部分背景区域被抑制的情况下达到最佳效果。 第三幅图像的背景非常复杂，所有竞争方法都无法检测到明显的对象。 使用适当的噪声标签，与无监督显着性方法和深度显着性方法相比，我们获得最佳结果。 第四个图像的对比度非常低，大多数竞争方法未能捕捉到最后一只企鹅误检的整个显着物体，特别是那些无监督显着性方法。 我们的方法正确捕捉所有三只企鹅。最后一行中的显着对象非常小，并且竞争方法没有成功

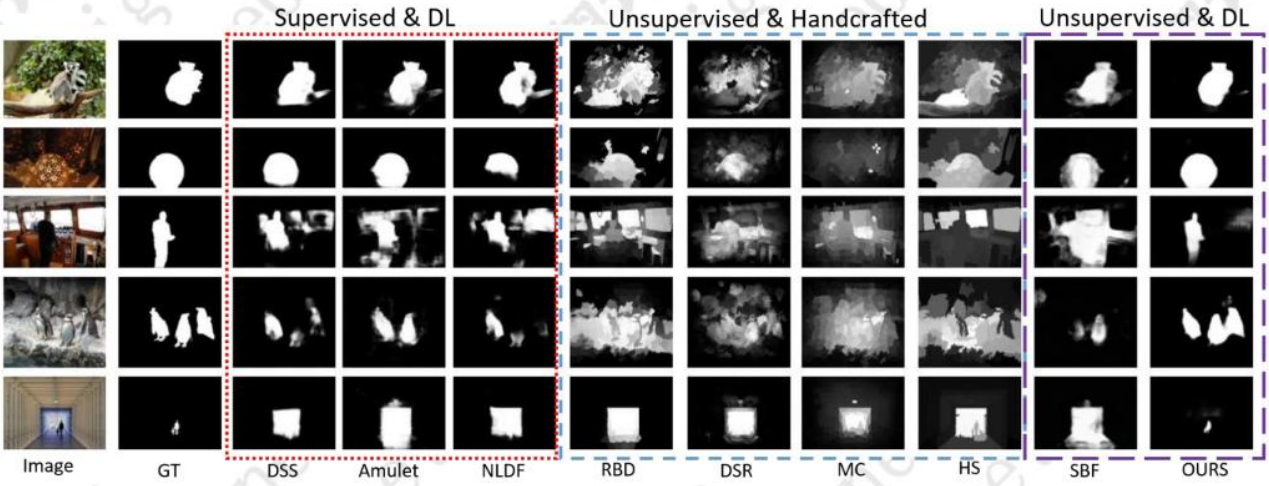
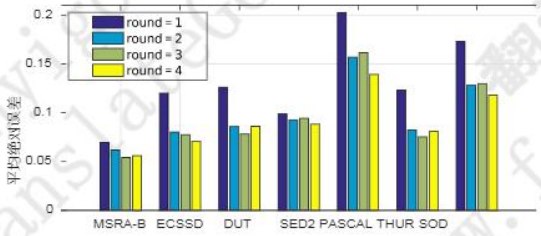


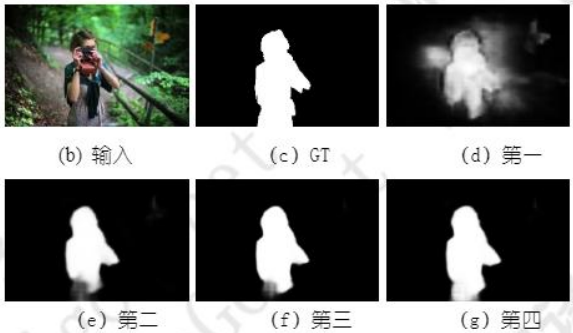
图4. 我们的方法和其他竞争方法之间的视觉比较。

捕获显着区域，而我们的方法以高精度捕捉整个显着区域。

消融研究：在本文中，我们建议迭代更新噪声建模模块和潜在显着性预测模型以实现准确的显着性检测。由于这两个模块协同工作以优化整体损失函数，所以有趣的是看看显着性预测结果如何随着更新轮的增加而发展。 在图1中，5我们举例说明了关于更新回合的性能指标（MAE）和示例显着性检测结果。 从零噪声初始化开始，我们的方法通过更新噪声建模来不断提高显着性检测的性能。 而且，只有在几轮更新之后，我们的方法才会收敛到期望的状态，如图3所示。5.



(a) 每轮7个数据集的MAE



(b) 输入 (c) GT (d) 第一

(e) 第二 (f) 第三 (g) 第四

图5. 每轮的表现 顶部：每个数据集的MAE。 底部：每个更新回合产生的示例图像，地面实况和中间结果。

5. 结论

在本文中，我们提出了一个端到端的显着学习框架，而不需要网络训练中的人注释显着图。 我们将无监督显着性学习表示为由各种高效和有效的常规无监督显着性检测方法生成的多个噪声显着图学习。 我们的框架由潜在显着性预测模块和明确的噪声建模模型组成，这些模型协同工作。 对各种基准数据集的广泛的实验结果证明了我们的方法的优越性，它不仅比传统的无监督方法有更大的优势，而且与目前最先进的深度监督显着性检测方法相比，还具有高度可比的性能。 未来，我们计划调查我们的多重显着物体检测和小型显着物体检测的具有挑战性的场景框架。 将我们的框架扩展到密集的预测任务，如语义分割[25]和单眼深度估计[18]可能是有趣的方向。

确认。 J. 张先生感谢何明义教授给予的无比的支持和鼓励。 T. Zhang获得澳大利亚研究委员会（ARC）发现项目资助计划（项目DP150104645）的支持。 Y. Dai被中国国家青年人才计划，国家自然科学基金（61420106007,61671387）和ARC授予（DE140100180）部分资助。

参考

[1] S.Alpert, M.Galun, A.Brandt和R.Basri。 通过概率自下而上聚合和提示集成进行图像分割。 IEEE Trans。模式分析。 马赫。 Intell。, 34 (2) : 315-327, 2012年2月。5

[2] A. Borji, M. Cheng, Q. Hou, H. Jiang和J. Li。 突出物体检测：一项调查。 CoRR, abs / 1411.5878,2014。2

[3] A. Borji, M. Cheng, H. Jiang和J. Li。 突出物体检测：基准。 IEEE Trans。 Image Proc。, 24 (12) : 5706-5722,2015。2,5

[4] LC Chen, G. Papandreou, I. Kokkinos, K. Murphy, and
A. L.尤伊尔。 Deeplab：深度卷积网络的语义图像分割，无限卷积和完全连接的crfs。 IEEE Trans。 模式分析。 马赫。 INTELL,
PP (99) : 1-1, 2017。5

[5] M. Cheng, NJ Mitra, X. Huang和S. Hu。 Salientshape：图像集合中的群体显着性。 Visual Computer, 30 (4) : 443-453,2014。5

[6] M. Cheng, G. Zhang, N. Mitra, X. Huang和S.-M. 胡。基于全局对比度的显着区域检测。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第409-416页，2011。1,2

[7] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet和N. Crook。 软图像抽象的高效显着区域检测。 在Proc。 IEEE Int。 CONF。 比较。 见1529-1536页，2013年。2

[8] M. Everingham, SMA Eslami, L. Van Gool, CKI Williams, J. Winn和A. Zisserman。 帕斯卡视觉对象课挑战：回顾。 诠释。 J. Comp。 可见，111 (1) : 98-136，2015。5

[9] S. Goferman, L.Zelnik-Manor和A. Tal。 上下文感知显着性检测。 IEEE Trans。 模式分析。 马赫。 Intell。, 34 (10) : 1915-1926，2012年10月。1,2

[10] K. He, X. Zhang, S. Ren和J. Sun. 图像识别的深度残留学习。 在Proc。 IEEE会议。 比较。 可见。 帕特。
*识别，第770-778页，2016年6月。*5,7

[11] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu和PHS Torr。 深度监控短连接的显着物体检测。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第3203-3212页，2017年7月。1,2,5,6,7

[12] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Gir-shick, S. Guadarrama和T. Darrell。 Caffe：用于快速特征嵌入的卷积体系结构。 在Proc。 ACM Int。
*CONF。 多媒体，第675-678页，2014年。*5

[13] B. Jiang, L. Zhang, H. Lu, C. Yang和M. Yang。 通过吸收马尔可夫链进行显着性检测。 在Proc。 IEEE Int。
*CONF。 比较。 见1665-1672页，2013年。*3,5,6

[14] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li。 突出物体检测：区分性区域特征整合方法。 在Proc。 IEEE会议。 比较。 可见。 帕特。
*识别，第2083-2090页，2013。*2,5,6

[15] P. Jiang, H. Ling, J. Yu和J. Peng。 不明飞行物突出区域检测：唯一性，关注度和客观性。 在Proc。
*IEEE Int。 CONF。 比较。 见，第1976-1983页，2013年。*2

[16] I. Jindal, M. Nokleby和X. Chen。 用辍学规则化学习来自嘈杂标签的深度网络。 在Proc。 IEEE Int。 CONF。 数据挖掘“，第967-972页，2016年12月。3

[17] J. Kim, D. Han, Y.-W. Tai和J. Kim。 通过高维颜色变换进行显着区域检测。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第883-890页，2014。2

[18] B. Li, C. Shen, Y. Dai, A. van den Hengel和M. He。 使用深度特征和分层crfs回归的单眼图像的深度和表面法线估计。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第1119-1127页，2015年6月。8

[19] G. Li和Y. Yu。 基于多尺度深度特征的视觉显着性。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第5455-5463页，2015年6月。2, 5, 6

[20] G. Li和Y. Yu。 用于显着物体检测的深度对比学习。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第478-487页，2016年6月。2, 5, 6

[21] X. Li, H. Lu, L. Zhang, X. Ruan和M. Yang。 通过密集和稀疏重建进行显着性检测。 在Proc。 IEEE Int。 CONF。 比较。 见，第2976-2983页，2013年12月。3, 5, 6

[22] X. Li, L. Zhao, L. Wei, MH Yang, F. Wu, Y. Zhuang,
H. Ling和J. Wang。 Deepsaliency：用于显着物体检测的多任务深度神经网络模型。 IEEE Trans。
*Image Proc。，25（8）：3919-3930，2016年8月。*2, 5, 6

[23] Y.Li, X. Hou, C.Koch, JM Rehg和ALYuille。 显着物体分割的秘密。 在Proc。 IEEE会议。
*比较。 可见。 帕特。 识别，第280-287页，2014。*5

[24] T. Liu, J. Sun, N.-N. 郑, X.唐和H.-Y. 岑。 学习检测显着的对象。 在Proc。 IEEE会议。
*比较。 可见。 帕特。 识别，第1-8页，2007年。*5

[25] Z. Lu, Z. Fu, T. Xiang, P. Han, L. Wang, and X. Gao。 从弱和嘈杂的标签中学习语义分割。 IEEE Trans。 模式分析。 马赫。 Intell。，39（3）：486-500，2017年3月。3,8

[26] Z. Luo, A. Mishra, A. Achkar, J. Eichel, S. Li和P.-M. Jodoin。 用于显着物体检测的非局部深度特征。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，2017年7月。2, 5, 6, 7

[27] X. Shen和Y. Wu。 通过低秩矩阵恢复的显着物体检测统一方法。 在Proc。 IEEE会议。
*比较。 可见。 帕特。 识别，第853-860页，2012。*2

[28] L. Wang, L. Wang, H. Lu, P. Zhang和X. Ruan。 循环完全卷积网络的显着性检测。 在Proc。 欧元。 CONF。 比较。 见，第825-841页，2016。2, 5, 6

[29] T. Wang, A. Borji, L. Zhang, P. Zhang和H. Lu。 用于检测图像中的显着对象的分阶段细化模型。 在Proc。 IEEE Int。 CONF。 比较。 可见，2017。2, 6

[30] J. Whitehill, T.fan Wu, J.Bergsma, JR Movellan和PL Ruvolo。 谁的投票数应该更多：从未知专家的贴标机中最优化整合标签。 在Proc。 进阶
*神经Inf。 处理。 Syst。，第2035-2043页。 2009年。*3

[31] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudi-nov, R. Zemel和Y. Bengio。 显示，参加并讲述：神经图像标题生成与视觉注意力。 在Proc。
诠释。
*CONF。 马赫。 学习，第37卷，第2048-2057页，2015年。*1

[32] Q. Yan, L. Xu, J. Shi和J. Jia。 分层显着性检测。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第1155-1162页，2013。3,5

[33] C. Yang, L. Zhang, H. Lu, X. Ruan和M. Yang。 通过基于图形的多方排名进行显着性检测。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第3166-3173页，2013。5

[34] J. Yao, J. Wang, I. Tsang, Y. Zhang, J. Sun, C. Zhang, and R. Zhang。 从质量嵌入的噪声图像标签深度学习。 ArXiv电子版，2017年11月。3

[35] D. Zhang, J. Han和Y. Zhang。 融合监督：面向深部显着物体探测器的无监督学习。 在Proc。 IEEE Int。 CONF。 比较。 可见，2017年10月。1, 2, 3, 5, 6, 7

[36] G.-X. 张，M.-M. Cheng，S.-M. 胡和RR马丁。 保持图像大小的形状保留方法。 计算机图形学论坛，28（7）：1897-1906，2009。1

[37] J. Zhang, Y. Dai, B. Li和M. He。 注意规模：深度多尺度显着物体检测。 在Proc。 诠释。 CONF。 关于数字图像计算：技术和应用，第1-7页，2017年11月。2

[38] J. Zhang, Y. Dai和F. Porikli。 通过集成多级线索来实现深度显着物体检测。 在Proc。 IEEE Winter Conference on Applications of Computer Vision，第1-10页，2017年3月。2

[39] J. Zhang, B. Li, Y. Dai, F. Porikli和M. He。 用于显着物体检测的集成深层和浅层网络。 在 *PROC。 IEEE Int。 CONF。 Image Process，第1537-1541页，* 2017年9月。2

[40] P. Zhang, D. Wang, H. Lu, H. Wang和X. Ruan。 护身符：为显着物体检测汇总多级卷积特征。 在Proc。 IEEE Int。 CONF。 比较。 可见，2017年10月。1, 2, 5, 6, 7

[41] P. Zhang, D. Wang, H. Lu, H. Wang, and B. Yin。 学习不确定的卷积特征用于精确显着性检测。 在Proc。 IEEE Int。 CONF。 比较。 可见，2017年10月。2, 5, 6

[42] R. Zhao, W. Ouyang, H. Li, and X. Wang。 多情境深度学习的显着性检测。 在Proc。 IEEE会议。 比较。 *可见。 帕特。 识别，第1265-1274页，2015。*1, 2, 5, 6

[43] W. Zhu, S. Liang, Y. Wei和J. Sun. 强大的背景检测显着优化。 在Proc。 IEEE会议。 比较。 可见。 帕特。 识别，第2814-2821页，2014。1, 2, 3, 5, 6

[44] W. Zou和N. Komodakis。 Harf：用于显着对象检测的层次关联丰富特征。 在Proc。 IEEE Int。 CONF。 比较。 Vis。，第406-414页，2015年12月。6