ELSEVIER

计算机视觉与图像理解
期刊主页： www.elsevier.com

SalGAN：使用对抗网络进行视觉显着性预测

Junting Pan[a], Cristian Canton-Ferrer[b], Kevin McGuinness[c], Noel E. O'Connor[c], Jordi Torres[d], Elisa Sayrol[a], Xavier Giro-i-Nieto[a,**]

[a]Universitat Politecnica de Catalunya, Barcelona 08034, Catalonia / Spain[b]Facebook AML, Seattle, WA, United States of America[c]Insight Center for Data Analytics, Dublin City University, Dublin 9, Ireland[d]Barcelona Supercomputing中心, 巴塞罗那08034, 加泰罗尼亚/西班牙

摘要

用于显着性预测的最近方法通常基于单个显着性度量来训练具有损失函数。 在使用其他显着性指标进行评估时，这可能会导致性能下降。 在本文中，我们提出了一种新的基于数据驱动度量的显着性预测方法，名为SalGAN（Saliency GAN），训练有对抗性损失函数。 SalGAN由两个网络组成：一个预测来自输入图像的原始像素的显着性图； 另一个采用第一个的输出来区分显着图是预测的还是基本事实。 通过尝试使预测的显着性图与地面实况无法区分，预计SalGAN将生成类似于地面实况的显着性图。 我们的实验表明，对抗性训练使我们的模型能够在各种显着性指标中获得最先进的表现。

1. 介绍

视觉显着性描述了图像中吸引人类注意力的空间位置。它被理解为自下而上的过程的结果，其中人类观察者在没有特定任务的情况下探索图像几秒钟。 因此，显着性预测对于诸如对象识别的各种机器视觉任务是必不可少的（Walther等， 2002）.

视觉显着性数据传统上由眼动仪收集 （Judd等人，2009a），最近点击了鼠标 （江等人， 2015）或网络摄像头（Krafka等， 2016）.使用高斯核对图像的突出点进行聚合和卷积以获得显着性图。 结果，生成灰度图像或热图以表示图像中的每个对应像素捕获人类注意力的概率。

在设计用于显着性预测的最佳损失函数方面已经进行了大量研究工作。 最先进的方法 （黄等人， 2015）采用基于显着性的指标而其他指标（P.等人， 2016; Cornia等，2016; Jetley等人， 2016; 太阳 等人， 2017）在显着图空间中使用距离。 如何选择

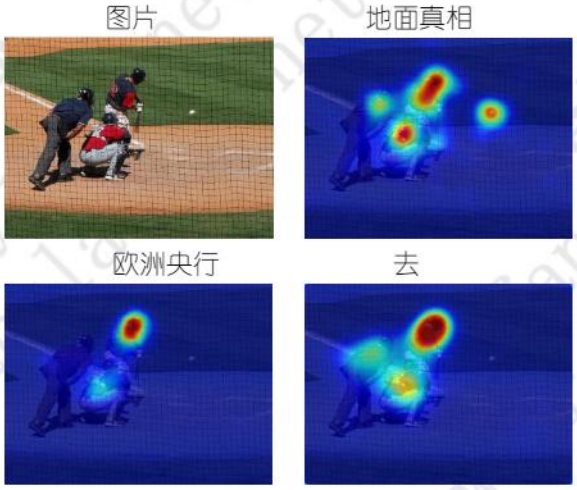

图1.显着性图生成的示例，其中所提出的系统（SalGAN）优于标准二元交叉熵（BCE）预测模型。

**通讯作者：电话：+ 34-934-015-769;
电子邮件： xavier.giro”upc.edu (Xavier Giro-i-Nieto)

或设计最好的训练损失仍然是一个悬而未决的问题。 此外，不同的显着性度量在定义显着性图的含义方面存在差异，并且与模型比较存在不一致性。 例如，已经指出模型优化的最佳度量可能取决于最终应用 (Bylinskii等，2016a).

为此，我们引入了针对视觉显着性预测的对抗性训练，而不是设计定制的损失函数，这是由生成性对抗网络引发的。（甘斯）（古德费洛 等人， 2014). 我们将提出的方法称为SalGAN。 我们专注于探索使用这种对抗性损失的好处，使得输出显着性图不能与真实的显着性图区分开。 在GAN中，训练由两个竞争代理驱动：首先，生成器合成与训练数据匹配的样本；第二，鉴别器区分直接从训练数据绘制的真实样本和由生成器合成的假样本。 在我们的例子中，该数据分布对应于真实图像对及其对应的视觉显着图。

具体而言，SalGAN使用深度卷积神经网络（DCNN）估计输入图像的显着性图。 如图所示 2 该网络最初在显着图的下采样版本上以二进制交叉熵（BCE）丢失进行训练。然后使用经过训练的鉴别器网络对该模型进行细化，以解决由SalGAN生成的显着性图和用作基础事实的真实性之间的二元分类任务。 我们的实验表明，当与单塔和单任务模型中的BCE内容丢失相结合时，对抗性训练如何允许跨不同指标达到最先进的性能。

总而言之，我们调查了视觉显着性学习中对抗性损失的引入。 通过在BCE显着性预测模型中引入对抗性损失，我们在MIT300和SALICON数据集中实现了几乎所有评估指标的最先进性能

文本的其余部分安排如下。 部分 2 回顾最先进的视觉显着性预测模型，讨论它们所依据的损失函数，它们与不同指标的关系以及它们在架构和培训方面的复杂性。 部分 3 介绍了SalGAN，我们基于卷积编码器 - 解码器架构的深度卷积神经网络，以及在其对抗训练中使用的鉴别器网络。 部分 4 描述了SalGAN的培训过程和使用的损失函数。 部分 5 包括所提出技术的实验和结果。 最后，部分 6 通过得出主要结论来结束论文。

我们的结果可以通过源代码和训练有素的模型重现 HTTPS：//imatge-upc.github。 IO /显着性-salgan-2017 /.

2. 相关工作

多年来，显着性预测已经受到研究界的关注。 因此开创性的作品 (Itti等， 1998) 提出预测显着性图，考虑多尺度的低级特征，并将它们组合成一个显着性图。 (Harel等， 2006), 同样从低级特征图开始，引入了基于图形的显着性模型，该模型在各种图像映射上定义马尔可夫链，并将地图位置上的平衡分布视为激活和显着性值。 (Judd等人， 2009b) 提出了一个自下而上，自上而下的显着性模型，不仅基于低，中，高级图像特征。 (Borji， 2012) 将以前最佳自下而上模型的低级特征显着性图与自上而下的认知视觉特征相结合，并学习从这些特征到眼睛注视的直接映射。

与计算机视觉中的许多其他领域一样，最近提出了许多深度学习解决方案，其显着改善了性能。 例如，深度网络集成（eDN） (Vig等， 2014) 代表了一种早期的架构，它自动学习显着性预测的表示，混合来自不同层的特征图。在 (P. 等人， 2016) 比较了两个用于显着性预测训练的ebd-to-end的卷积神经网络，一个从头开始设计和训练的较轻的，以及为图像分类预先训练的第二个和更深的一个。即使在为其他目的构建数据集的预训练时，DCNN也显示出更好的结果。 DeepGaze Kümmerer 等。 (2015a) 使用知名的AlexNet提供了更深入的网络 (Krizhevsky等人，2012), 在Imagenet上使用预先训练过的权重 (邓等人，2009) 并且在顶部有一个读出网络，其输入由AlexNet的某些层输出组成。 网络输出模糊，中心偏置并使用softmax转换为概率分布。 黄等人。 (黄等人， 2015), 在所谓的SALICON网中，通过使用VGG而不是AlexNet或GoogleNet获得了更好的结果 (Szegedy等， 2015). 在他们的提案中，他们考虑了两个具有精细和粗略输入的网络，其特征图输出被连接起来。

李等人。 (李和宇， 2015) 提出了一种多分辨率卷积神经网络，该网络是在多个分辨率下以固定和非固定位置为中心的图像区域进行训练的。 可以在更高层中学习多种自上而下的视觉特征，并且还可以通过组合多个分辨率的信息来推断自下而上的视觉显着性。 他们最近的工作称为DSCLRCN，进一步发展了这些想法 刘汉 (2018), 其中所提出的模型并行地在每个图像位置上学习与显着性相关的局部特征，然后学习同时结合全局上下文和场景上下文来推断显着性。 它们结合了一个模型来有效地学习长期空间相互作用和场景上下文调制，以推断图像显着性。深凝视II (Ku ̈mmerer等， 2017) 通过将训练用于图像识别的特征与四层1x1卷积相结合，设置MIT300数据集中的最新技术水平。 当与中心偏差相结合时，DSCLRCN和DeepGaze II都在基准测试中获得了优异的结果，而在SalGAN中没有考虑到这一结果，因为结果纯粹是推理时间的结果。MLNET (Cornia等， 2016) 提出了一种结合了在DCNN的不同级别提取的特征的体系结构。 他们引入了一个受三个目标启发的损失函数：测量与基本事实的相似性，保持预测图的不变性达到最大值并重视

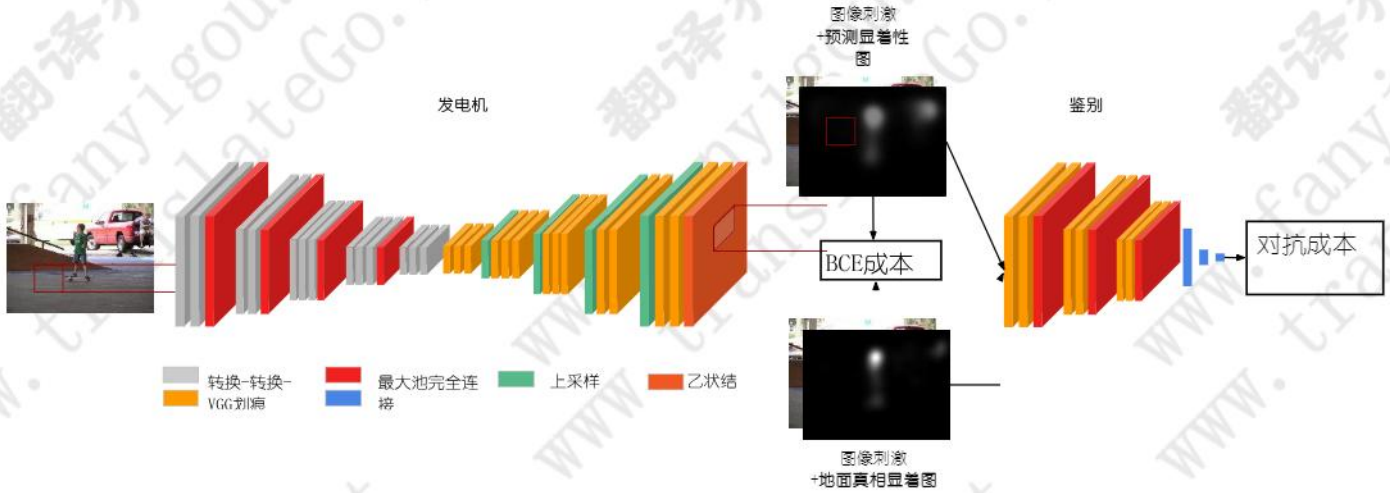

图2. 提议的显着性系统的总体架构。 显着性预测网络的输入预测给定自然图像作为输入的输出显着性图。 然后，该对显着性和图像被馈送到鉴别器网络中。 鉴别器的输出是一个分数，用于说明输入显着性图是真实的还是假的。

具有高地面实况固定概率的像素。 实际上，选择合适的损失函数已成为可以改善结果的问题。 因此，另一个有趣的贡献 （黄等人， 2015) 基于可区分的指标最小化损失函数，例如NSS，CC，SIM和KL差异来训练网络（参见Riche等人。 (2013) 和 Kümmerer等。 (2015b) 用于定义这些指标。 可以在中找到对指标的全面比较 （Bylinskii等人， 2016a)). In （黄等人， 2015) KL分歧给出了最好的结果。 （Jetley等人， 2016) 通过将显着性映射模型视为广义伯努利分布，还测试了基于概率距离的损失函数，例如X2散度，总变差距离，KL散度和Bhattacharyya距离。发现Bhattacharyya距离给出了最好的结果。

在我们的工作中，我们提出了一种采用不同方法的网络架构。 通过将高级别对抗性损失结合到传统的显着性预测训练方法中，所提出的方法以明显的余量实现了MIT300和SALICON数据集中的最新性能。

3. 建筑

SalGAN的训练是两个相互竞争的卷积神经网络的结果：一个显着图的生成器，它是SalGAN本身，以及一个鉴别器网络，旨在区分真正的显着图和SalGAN生成的图。 本节提供有关两个模块的结构，所考虑的损失函数以及开始对抗训练之前的初始化的详细信息。 数字 2 显示了系统的体系结构。

3.1. 发电机

生成器网络SalGAN采用卷积编码器 - 解码器架构，其中编码器部分包括最大池层减少了特征映射的大小，而解码器部分使用上采样层，然后使用卷积滤波器构造与输入具有相同分辨率的输出。

网络的编码器部分在架构上与VGG-16相同 （西蒙尼安和齐瑟曼， 2015)，省略最终的池和完全连接的层。 使用在用于对象分类的ImageNet数据集上训练的VGG-16模型的权重来初始化网络 （邓等人， 2009). 在用于显着性预测的训练期间，仅修改VGG-16中的最后两组卷积层，而较早的层保持固定不同于原始VGG-16模型。 我们在训练期间修复权重以节省计算资源，即使可能会牺牲一些性能损失。

解码器架构的结构与编码器的结构相同，但是层的顺序相反，并且池化层被上采样层替换。 同样，ReLU非线性用于所有卷积层，并且是最终的
添加具有S形非线性的11个卷积层以产生显着性图。 解码器的权重随机初始化。 网络的最终输出是与输入图像大小相同的显着图。

SalGAN的实施细节见表 1.

3.2. 鉴别

表 2 给出了鉴别器的体系结构和层配置。 简而言之，网络由六个3x3内核卷积组成，其中散布着三个汇集层（2），然后是三个完全连接的层。 卷积层都使用ReLU激活，而完全连接的层采用tanh激活，最后一层除外，它使用S形激活。

层	深度	核心	迈	垫	激活conv1
1 -	64	1×1	1	1	重读
conv1_2	64	3×3	1	1	重读
池1		2×2	2	0	-
conv2_1	128	3×3	1	1	重读
conv2_2	128	3×3	1	1	重读
pool2		2×2	2	0	-
conv3_1	256	3×3	1	1	重读
conv3_2	256	3×3	1	1	重读
conv3_3	256	3×3	1	1	重读
pool3		2×2	2	0	-
conv4_1	512	3×3	1	1	重读
conv4_2	512	3×3	1	1	重读
conv4_3	512	3×3	1	1	重读
池4		2×2	2	0	-
conv5_1	512	3×3	1	1	重读
conv5_2	512	3×3	1	1	重读
conv5_3	512	3×3	1	1	重读
conv6_1	512	3×3	1	1	重读
conv6_2	512	3×3	1	1	重读
conv6_3	512	3×3	1	1	重读
upsample6		2×2	2	0	-
conv7_1	512	3×3	1	1	重读
conv7_2	512	3×3	1	1	重读
conv7_3	512	3×3	1	1	重读
upsample7		2×2	2	0	-
conv8_1	256	3×3	1	1	重读
conv8_2	256	3×3	1	1	重读
conv8_3	256	3×3	1	1	重读
upsample8		2×2	2	0	-
conv9_1	128	3×3	1	1	重读
conv9_2	128	3×3	1	1	重读
upsample9		2×2	2	0	-
conv10_1	64	3×3	1	1	重读
conv10_2	64	3×3	1	1	重读
产量	1	1×1	1	0	乙状结肠

表1. 发电机网络的架构。

层	深度	核心	迈	垫	激活conv1_1
-	3	1×1	1	1	重读
conv1_2	32	3×3	1	1	重读
池1		2×2	2	0	-
conv2_1	64	3×3	1	1	重读
conv2_2	64	3×3	1	1	重读
pool2		2×2	2	0	-
conv3_1	64	3×3	1	1	重读
conv3_2	64	3×3	1	1	重读
pool3		2×2	2	0	-
fc4	100	-	-	-	双曲正切
fc5	2	-	-	-	正切
fc6	1	-	-	-	乙状结肠

表2. 鉴别器网络的体系结构。

4. 训练

SalGAN中的滤波器权重已经过感知损失的训练（约翰逊等人，2016）结合内容和对抗性损失。内容丢失遵循经典方法，其中预测的显着性图以像素方式与来自地面实况的对应的显着性图进行比较。对抗性损失取决于鉴别器对生成的显着性图的实际/合成预测。

4.1. 内容丢失

以每像素为基础计算内容损失，其中将预测显着性图的每个值与来自地面实况图的对应对等体进行比较。鉴于我的形象
尺寸$N = W \times H$，我们将显着性图S表示为vec-

概率的tor，其中S_j是像素I_j的概率

迷恋。内容丢失功能 $L(S, \hat{S})$ 定义在预测显着图$\hat{S}$与其对应的基础事实S之间。

首先考虑的内容损失是均方误差（MSE）或欧几里德损失，定义如下：

$$L_{MSE}(S, \hat{S}) = \frac{1}{N}\sum_{j}^{N}(S_j - \hat{S}_j)^2。 \quad (1)$$

在我们的工作中，MSE被用作基线参考，因为它已被直接采用或在其他最先进的视觉显着性预测解决方案中有一些变化（P等人，2016; Cornia 等人，2016）.

基于MSE的解决方案旨在最大化峰值信噪比（PSNR）。这些工作倾向于过滤输出中的高空间频率，有利于这种模糊的轮廓。MSE对应于计算预测显着性与地面实况之间的欧几里德距离。

将地面实况显着图标准化，以使每个值在[0,1]范围内。因此，显着性值可以被解释为观察者参与特定像素的概率的估计。因此，很容易在最后一层使用softmax对预测进行多项分布。然而，显然，可以参加多于一个像素，使得将每个预测值视为独立于其他像素更合适。因此，我们建议对最终层中的每个输出应用元素方式的sigmoid，以便像素方式的预测可以被认为是独立二元随机变量的概率。这种设置中的适当损失是二元交叉熵，它是所有像素中各个二元交叉熵（BCE）的平均值：

$$\frac{1}{N}$$

$L_{欧洲央行} = -\frac{1}{N}\sum_{j=1}^{N}(S_j \log(\hat{S}j) + (1-S_j)\log(1-\hat{S}j))$。

(2)

4.2. 对抗性损失

生成对抗网络（GAN） （Goodfellow等人，2014）通常用于生成具有真实统计特性的图像。这个想法是同时适合两个参数函数。 这些函数中的第一个，称为生成器，经过训练，可以将样本从简单的分布（例如高斯分布）转换为更复杂的样本

分布（例如自然图像）。 训练第二个函数，鉴别器，以区分真实分布和生成样本的样本。 训练在使用生成和实际样本训练鉴别器之间交替进行，并且通过保持鉴别器权重恒定并且通过鉴别器反向传播误差来训练生成器以更新生成器权重。

显着性预测问题与上述情况有一些重要的区别。 首先，目标是拟合一个确定性函数，预测图像中的真实显着性值，而不是来自随机噪声的真实图像。 因此，在我们的情况下，生成器（显着性预测网络）的输入不是随机噪声而是图像。 其次，显着图对应的输入图像是必要的，因为目标不仅是使两个显着图变得难以区分，而且条件是它们都对应于相同的输入图像。 因此，我们将图像和显着性图包括为鉴别器网络的输入。 最后，当使用生成对抗性网络生成逼真的图像时，通常没有可比较的基本事实。 然而，在我们的例子中，可以获得相应的地面实况显着性图。当更新生成函数的参数时，我们发现使用损失函数，该函数是来自鉴别器的误差和交叉熵相对于地面实况的组合，提高了对抗训练的稳定性和收敛速度。 对抗训练期间显着性预测网络的最终损失函数可以表述为：

$$L = \alpha \cdot L_{\text{欧洲央行}} + L(D(I, \hat{S}), 1)。 \quad (3)$$

其中L是二元交叉熵损失，1是实际样本的目标类别，0是假（预测）样本的类别。 这里，我们优化L（D（I，S），1）而不是最小化L（D（I，S），0），它提供更强的梯度，类似于 （Goodfellow等人， 2014）。 D（I，$\hat{S}$）是愚弄鉴别器网络的概率，因此当欺骗鉴别器的机会较低时，与显着性预测网络相关联的损失将增加得更多。 在鉴别器的培训期间，没有内容丢失，丢失功能是：

$$L_D = L(D(I, S), 1) + L(D(I, \hat{S}), 0)。 \quad (4)$$

在列车时间，我们首先通过仅使用BCE训练15个时期来引导显着性预测网络功能，BCE是针对下采样输出和地面实况显着性计算的。 在此之后，我们添加鉴别器并开始对抗训练。 鉴别器网络的输入是大小为256 192 4的RGBS图像，其包含源图像通道和（预测或地面实况）显着性。

我们使用批量大小为32的SALICON训练集对15,000张图像上的网络进行训练。在对抗训练期间，我们在每次迭代（批量）后交替进行显着性预测网络和鉴别器网络的训练。 我们使用L2权重正则化（即权重衰减）时
训练发生器和鉴别器（$\lambda = 1 \times 10^{-4}$）。
我们使用AdaGrad进行优化，初始学习率为3×10^{-4}。

	酱汁 ↑	AUC_B ↑	NSS ↑	直流↓	IG中↓
行	0.752	0.825	2.473	0.761	0.712
BCE/2	0.750	0.820	**2.527**	**0.764**	0.592
BCE/4	**0.755**	**0.831**	2.511	0.763	0.825
BCE/8	0.754	0.827	2.503	0.762	**0.831**

表3. 在SALICON验证评估的（15个时期）下采样显着图的影响。 BCE / x指的是在256×192的显着图上的1 / x的下采样因子。

5. 实验

提出了用于视觉显着性预测的SalGAN模型，并从不同角度进行了比较。 首先，评估使用BCE和下采样显着图的影响。 其次，从定量和定性的角度来衡量和讨论对抗性损失的收益。 最后，将SalGAN的性能与已发表的作品进行比较，以将其性能与当前最新技术进行比较。 旨在找到SalGAN的最佳配置的实验使用SALICON数据集的训练和验证分区运行 （江等人， 2015）. 这是一个大型数据集，通过从上下文中的Microsoft公共对象（MS-CoCo）数据集中收集总共20,000个图像的鼠标点击来构建 （林等人，2014）. 我们已经将这个数据集用于我们的实验，因为它是可用于视觉显着性预测的最大数据集。 除了SALICON，我们还会在MIT300上展示结果，MIT300是提交量最大的基准。

5.1. 非对抗性训练

比较了内容丢失部分MSE和BCE中提出的两个内容损失，以确定我们后来评估对抗性训练影响的基线。 表的两个第一行 4展示了从MSE到BCE的简单变更如何在所有指标中带来持续改进。 这种改进表明，将显着性预测视为多重二元分类问题比将其视为标准回归问题更为合适，尽管目标值不是二元的。 最小化交叉熵等同于最小化预测分布和目标分布之间的KL分歧，如果两个预测目标都被解释为概率，则这是合理的目标。

基于与MSE相比较优越的基于BCE的损失，我们还探讨了计算内容丢失对显着性映射的下采样版本的影响。 该技术在训练和测试时间减少了所需的计算资源，如表所示 3，它不仅不会降低性能，而且实际上可以改善性能。 鉴于此结果，我们选择在显着性地图上训练SalGAN下采样1/4，这在我们的架构中对应于显着性
地图64×48。

5.2. 对抗性增益

在估计方程中超参数α的值之后引入了对抗性损失 3 通过最大化最一般的度量，信息增益（IG）。 如图所示 3，

搜索是在对数标度上进行的，我们在α= 0.005时获得了最佳性能。

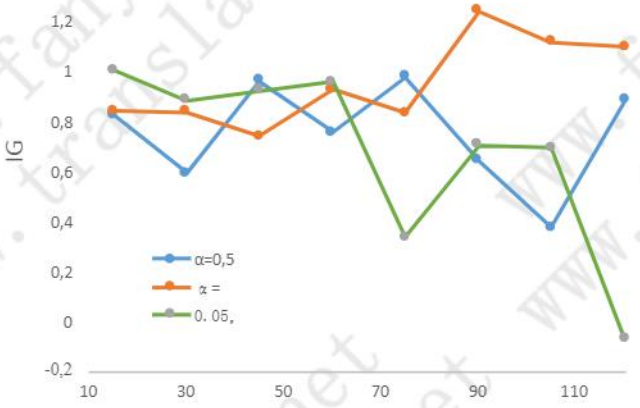

图3. SALICON验证集信息在不同时期数上的不同超参数 α 的增益。

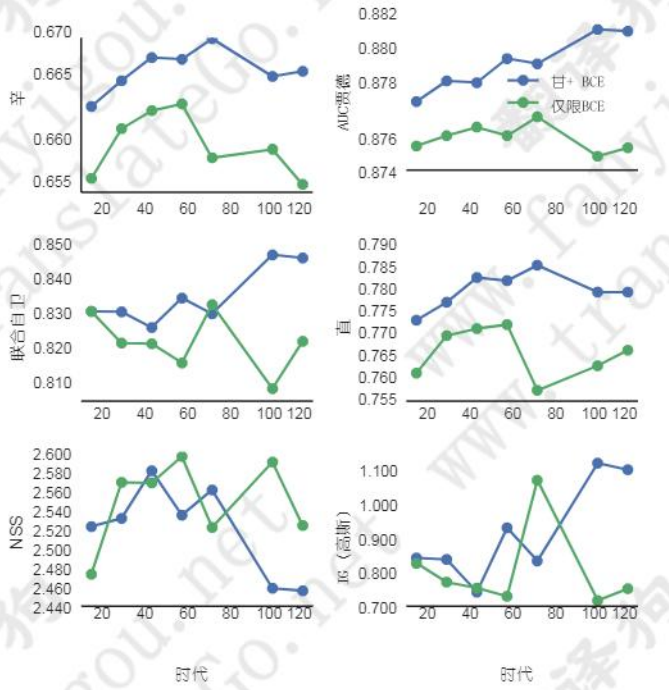

图4. 不同时期的对抗性+ BCE与BCE的SALICON验证集准确度指标。 由于趋势与AUC Judd的趋势相同，因此省略了AUC改组。

	酱汁 ↑	AUC B ↑	NSS ↑	直流	IG
MSE	0.728	0.820	1.680	0.708	0.628
欧洲央行	0.753	0.825	2.562	0.772	0.824
BCE/4	0.757	0.833	**2.580**	0.772	1.067
GAN/4	**0.773**	**0.859**	2.560	**0.786**	**1.243**

表4. 通过非对抗性（MSE和BCE）和对抗性训练获得的时期的最佳结果。 BCE / 4和GAN / 4指的是下采样显着图。 在SALICON验证中评估显着性图。

在图中比较了针对超参数α的不同值的SalGAN的信息增益（IG） 3. 寻找最佳超参数α的搜索是在对数尺度上进行的，并且我们在α= 0.005时获得了最佳性能。

通过使用BCE作为内容丢失和68的特征图来评估通过将对抗性损失引入感知损失而获得的收益。表中的第一行结果 4 指通过训练SalGAN定义的基线，仅BCE内容丢失15个时期。 之后，考虑两个选项：1）仅基于BCE（第2行）的训练，或2）引入对抗性损失（第3和第4行）。

数字 4 随着时期数量的增加，比较用于训练的验证集合准确度度量与单独的BCE相比的组合GAN和BCE损失。 在AUC指标（Judd和Borji）的情况下，当单独使用BCE时，增加时期数不会导致显着改善。 然而，通过进一步培训，BCE / GAN的总体损失将继续提高绩效。 在100和120个时期之后，对于六个指标中的五个，GCE / BCE的总损失显示出相对于BCE的显着改善。

对抗训练无法提高性能的单一指标是标准化扫描路径显着性（NSS）。 其原因可能是GAN训练倾向于产生更平滑且更分散的显着性估计，这更好地匹配真实显着性图的统计特性，但可能增加误报率。 如中所述 (Bylinskii等，2016a)，NSS对这种误报非常敏感。 误报增加的影响取决于最终的应用。 在显着图用作乘法注意模型的应用中（例如，在检索应用中，空间特征是重要加权的），误报通常不如假阴性重要，因为前者包括更多干扰物，后者删除可能有用特征。 还要注意NSS是可区分的，因此当对特定的ap-很重要时，可能会直接进行优化。

折叠。

5.3. 与最先进的技术进行比较

SalGAN在表中进行比较 5 来自最先进的几种其他算法。该比较基于SALICON组织者和MIT300基准测试对测试数据集进行的评估，测试数据集的基本事实不公开。 这两个基准提供了互补功能：虽然SALICON是一个包含5,000个测试图像的更大的数据集，但MIT300吸引了更多研究人员的参与。 在这两种情况下，使用SALICON数据集的训练（10,000）和验证（5,000）分区中包含的15,000个图像训练SalGAN。 请注意，虽然两个数据集都旨在捕获视觉显着性，但数据采集不同，因为SALICON基础事实是基于众包鼠标点击生成的，而MIT300是使用眼动追踪器在有限和受控制的用户组上构建的。 表 5 将SalGAN与其他使用过 SALICON 和 MIT300 数据集的当代作品进行比较。SalGAN在两个数据集中都呈现出非常有竞争力的结果，因为它在至少一个度量标准中改进或等于所有其他模型的性能。

SALICON（测试）	AUC-J ↑	Sim ↑	EMD ↓	AUC-B ↑	sAUC ↑	CC ↑	NSS ↑	KL ↓
DSCLRCN （刘和汉， 2018）	-	-	-	0.884	0.776	0.831	3.157	-
去	-	-	-	**0.884**	**0.772**	**0.781**	**2.459**	-
毫升网（Cornia等， 2016）	-	-	-	(0.866)	(0.768)	(0.743)	2.789	-
SalNet （潘等人， 2016）	-	-	-	(0.858)	(0.724)	(0.609)	(1.859)	-
MIT300	AUC-J ↑	Sim ↑	EMD ↓	AUC-B ↑	sAUC ↑	CC ↑	NSS ↑	KL ↓
人类	0.92	1.00	0.00	0.88	0.81	1.0	3.29	0.00
深凝视II Kümmerer等。 （2017）	(0.84)	(0.43)	(4.52)	(0.83)	0.77	(0.45)	(1.16)	(1.04)
DSCLRCN （刘和汉， 2018）	0.87	0.68	2.17	(0.79)	0.72	0.80	2.35	0.95
SALICON （黄等人， 2015）	0.87	(0.60)	(2.62)	0.85	0.74	0.74	2.12	0.54
去	**0.86**	**0.63**	**2.29**	**0.81**	**0.72**	**0.73**	**2.04**	**1.07**
PDP （Jetley等人， 2016）	(0.85)	(0.60)	(2.58)	(0.80)	0.73	(0.70)	2.05	0.92
毫升网（Cornia等， 2016）	(0.85)	(0.59)	(2.63)	(0.75)	(0.70)	(0.67)	2.05	(1.10)
深深的凝视我 （Kümmerer等， 2015a）	(0.84)	(0.39)	(4.97)	0.83	(0.66)	(0.48)	(1.22)	(1.23)
SalNet （潘等人， 2016）	(0.83)	(0.52)	(3.31)	0.82	(0.69)	(0.58)	(1.51)	0.81
BMS （张和Sclaroff, 2013）	(0.83)	(0.51)	(3.35)	0.82	(0.65)	(0.55)	(1.41)	0.81

表5. SalGAN与SALICON（测试）和MIT300基准测试的其他最先进解决方案的比较。 括号中的值对应于比SalGAN更差的性能。

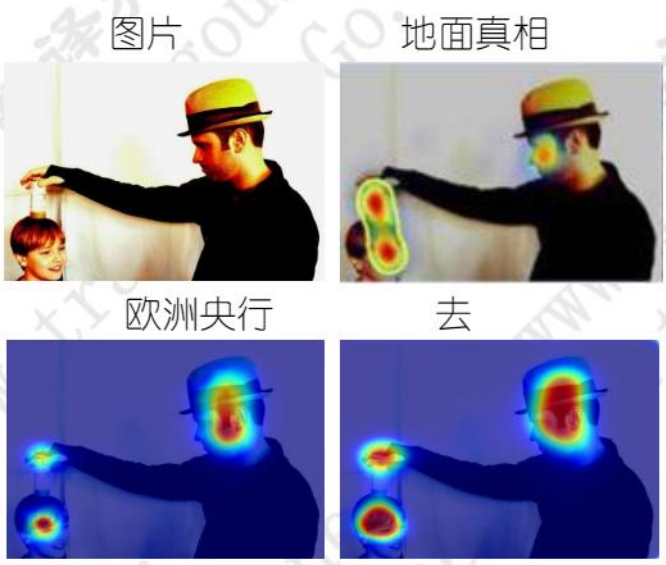

图5. 来自MIT300的示例图像，其包含通常由计算模型遗漏的显着区域（以黄色标记）和由SalGAN估计的显着性图。

5.4. 定性结果

通过观察产生的显着性图，从定性的角度探讨了对抗性训练的影响。 数字 5 显示了MIT300数据集中的一个示例，突出显示在 (Bylinskii等， 2016b) 作为现有显着性算法的特殊挑战。 左侧图像中以黄色突出显示的区域是算法通常会遗漏的区域。 在这个例子中，我们看到SalGAN成功地检测到魔术师经常错过的手和男孩的脸是显着的。

数字 6 说明了对抗性训练对生成的显着性图的统计特性的影响。 显示了交叉熵训练（左）和对抗训练（右）的显着图的两个特写部分。 仅在BCE上进行培训会生成显着性地图，而这些地图可能在当地

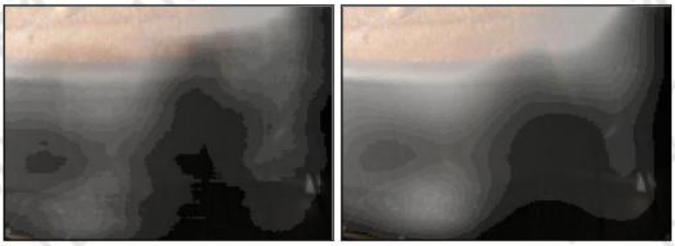

图6. BCE损失与联合BCE /对抗性损失培训结果的特写比较。 左：来自网络的显着性图，训练有BCE损失。 右：来自拟议的对抗训练的显着性图。

与基本事实一致，通常不太平滑并且具有复杂的水平集。另一方面，对抗训练产生更平滑和更简单的水平集。 最后，图 7 显示了一些定性结果，比较了培训与BCE和BCE / Adversarial的结果与SALICON验证集图像的基本事实。

6. 结论

据我们所知，这是第一个提出基于对抗的显着性预测方法的工作，并且已经展示了如何通过简单的编码器 - 解码器在深度卷积神经网络上进行对抗性训练以实现最先进的性能建筑。 基于BCE的内容丢失被证明对于初始化显着性预测网络是有效的，并且作为稳定对抗性训练的正则化术语。 我们的实验表明，与单独进行交叉熵的进一步训练相比，对抗训练改善了所有一杆显着性指标。

值得指出的是，尽管我们在本文中使用基于VGG-16的编码器 - 解码器模型作为显着性预测网络，但所提出的对抗性训练方法是通用的，并且可以应用于提高其他显着性模型的性能。

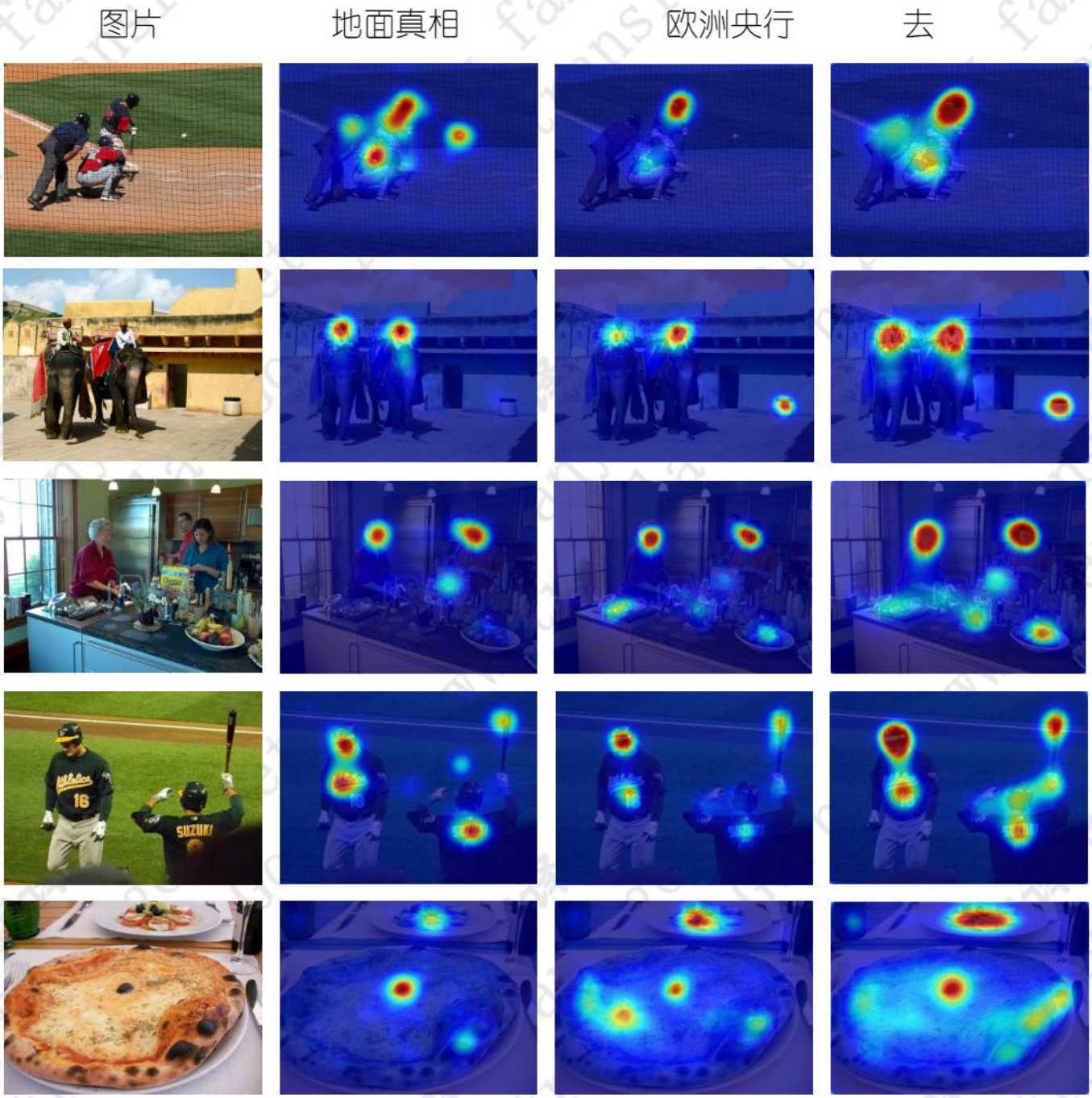

图7. SALICON验证集上SalGAN的定性结果。 SalGAN很好地预测了那些被BCE模型遗漏的高显着区域。 BCE模型的显着性图在一些显着区域中非常局部化，当突出区域的数量增加时它们往往会失败。

致谢

UPC的图像处理小组得到了TEC2016-75976-R项目的支持，该项目由西班牙的Ministeria de Economia y Competitividad和欧洲区域发展基金（ERDF）资助。 本材料基于爱尔兰科学基金会在Grant No 15 / SIRG / 3283下的支持。 我们非常感谢NVIDIA公司对这项工作中使用的GPU捐赠的支持。

参考

Borji, A., 2012。提升自下而上和自上而下的显着性视觉特征，参见：IEEE计算机视觉和模式识别会议（CVPR）。

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., Durand, F., 2016a。 不同的评估指标告诉我们有关显着性模型的信息？ ArXiv预印本：1610.01563。

Bylinskii, Z., Recasens, A., Borji, A., Oliva, A., Torralba, A., Durand, F., 2016b。 显着性模型应该放在哪里？在：欧洲计算机视觉会议（ECCV）。

Cornia, M., Baraldi, L., Serra, G., Cucchiara, R., 2016。用于显着性预测的深层多层网络，在：国际模式识别会议（ICPR）。

Deng, J., Dong, W., Socher, R., Li, LJ, Li, K., Fei-Fei, L., 2009。ImageNet：一个大规模的分层图像数据库，在：IEEE计算机视觉会议和模式识别（CVPR）。

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014。 Generative adversarial nets， in：神经信息处理系统的进展，第2672-2680页。

Harel, J., Koch, C., Perona, P., 2006。基于图形的视觉显着性，在：神经信息处理系统（NIPS）中。

Huang, X., Shen, C., Boix, X., Zhao, Q., 2015。Salicon：通过适应深度神经网络减少显着性预测中的语义鸿沟，参见：IEEE国际计算机视觉会议（ICCV）。

Itti, L., Koch, C., Niebur, E., 1998。一种基于显着性的视觉注意模型，用于快速场景分析。 IEEE模式分析和机器智能交易（PAMI），1254-1259。

Jetley, S., Murray, N., Vig, E., 2016。通过概率分布预测的端到端显着性映射，在：IEEE计算机视觉和模式识别会议（CVPR）。

Jiang, M., Huang, S., Duan, J., Zhao, Q., 2015。Salicon：Saliency in context, in：IEEE Conference on Computer Vision and Pattern Recognition（CVPR）。 Johnson, J., Alahi, A., Fei-Fei, L., 2016。实时样式转移和超分辨率的感知损失，见：欧洲计算机视觉会议（ECCV）。

Judd, T., Ehinger, K., Durand, F., Torralba, A., 2009a。 学习预测人类的外观，参见：计算机视觉，2009年IEEE第12届国际会议，IEEE。 第2106-2113页。

Judd, T., Ehinger, K., Durand, F., Torralba, A., 2009b。 学习预测人类的外观，参见：IEEE国际计算机视觉会议（ICCV）。

Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., Torralba, A., 2016。Eye tracking for everyone, in：IEEE Conference on Computer Vision and模式识别（CVPR）。

Krizhevsky, A., Sutskever, I., Hinton, GE, 2012。使用深度卷积神经网络的ImageNet分类，在：神经信息处理系统的进展，第1097-1105页。

Ku¨mmerer, M., Theis, L., Bethge, M., 2015a。 DeepGaze I：使用ImageNet培训的特征图提升显着性预测，参见：国际学习表示会议（ICLR）。

Ku¨mmerer, M., Theis, L., Bethge, M., 2015b。 信息理论模型比较统一了显着性指标。 美国国家科学院院刊（PNAS）112,1604-16059。

Ku¨mmerer, M., Wallis, TS, Gatys, LA, Bethge, M., 2017。了解对固定预测的低级和高级贡献，参见：2017 IEEE计算机视觉国际会议，第4799-4808页。

Li, G., Yu, Y., 2015。基于多尺度深度特征的视觉显着性，参见：IEEE计算机视觉和模式识别会议（CVPR）。

Lin, TY, Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Doll'r, P., Zitnick, L., 2014。Microsoft COCO：上下文中的共同对象，在：欧洲计算机视觉会议（ECCV）。

Liu, N., Han, J., 2018。一种用于显着性检测的深空间上下文长期递归卷积网络。 IEEE Transactions on Image Processing 27,3264-3274。

Pan, J., Sayrol, E., Giro'-i Nieto, X., McGuinness, K., O'Connor, NE, 2016。用于显着性预测的浅层和深层卷积网络，参见：IEEE计算机视觉与模式会议认可（CVPR）。

Riche, N., M., D., Mancas, M., Gosselin, B., Dutoit, T., 2013。Saliency and human fixations。 最新和研究比较指标，参见：IEEE国际计算机视觉会议（ICCV）。

Simonyan, K., Zisserman, A., 2015年。用于大规模图像识别的非常深的卷积网络，在：国际学习表示会议（ICLR）。

Sun, X., Huang, Z., Yin, H., Shen, HT, 2017。一种用于有效显着性预测的综合模型，在：AAAI，第274-281页。

Szegedy, C., Liu, W., Jia, Y., Sermanet, PRS, Anguelov, D., Erhan, D., Van-houcke, V., Rabinovich, A., 2015。更深入地了解卷积，在： IEEE计算机视觉和模式识别会议（CVPR）。

Vig, E., Dorr, M., Cox, D., 2014。用于自然图像中显着性预测的分层特征的大规模优化，在：IEEE计算机视觉和模式识别会议（CVPR）。

Walther, D., Itti, L., Riesenhuber, M., Poggio, T., Koch, C., 2002。对象识别的注意选择 - 一种温和的方式。 计算机科学讲义2525,472-479。

Zhang, J., Sclaroff, S., 2013。显着性检测：一种布尔映射方法，参见：IEEE国际计算机视觉会议（ICCV）。