不确定性启发 RGB-D 显著性检测

张靖， 范登平，戴玉超，

赛义德·安瓦尔、法特梅·萨利赫、 萨德格 ·阿利亚克巴里安、尼克·巴恩斯

**摘要**— 我们提出了第一个随机框架，通过从数据标记过程中学习，对RGB-D显著性检测采用不确定性。现有的 RGB-D 显著性检测模型通过在确定性学习管道之后预测单个显著性图，将此任务视为点估计问题。然而，我们认为，确定性解决方案相对不合时宜。受显著性数据标记过程的启发，我们提出了一种生成架构来实现概率RGB-D显著性表征，该体系结构利用潜在变量对标记变体进行建模。我们的框架包括两个主要模型：1）生成器模型，它将输入图像和潜在变量映射到随机显著性预测，以及2）推理模型，通过从真实或近似后验分布中采样来逐渐更新潜在变量。生成器模型是编码器-解码器显著性网络。为了推断潜在变量，我们引入了两种不同的解决方案：i）带有额外编码器的条件变分自动编码器，用于近似潜在变量的后分布;ii）交替反向传播技术，它直接从真实后分布中采样潜在变量。六个具有挑战性的RGB-D基准数据集的定性和定量结果显示了我们的方法在学习显著性图分布方面的卓越性能。源代码可通过我们的项目页面公开获得[：https://github.com/JingZhang617/UCNet。](https://github.com/JingZhang617/UCNet)

**索引项**— 不确定性、RGB-D 显著性检测、条件变分自动编码器、交替反向传播。

F

# 1 I介绍

或

BJECT级显著性检测（*即*显著物体检测）涉及将吸引人类注意力的最显眼物体与背景[2]-[9]分开。最近，由于深度信息在人类视觉系统中的重要性以及深度传感技术的普及[1]，[10]-[15]，RGB-D图像的视觉显著性检测引起了很多兴趣。使用额外的深度数据，传统的RGB-D显著性检测模型专注于通过探索RGB图像和深度数据之间的复合信息来预测RGB-D输入的单个显著性图。

RGB-D 显著性检测的标准做法是使用地面实况 （GT） 显著性映射来训练深度神经网络。通过这种方式，显著性检测通常通过学习映射函数*Y*  = *f（X*;*θ），*其中 *θ* 表示网络参数集，X 和 *Y* 是输入 RGB-D 图像对和相应的 GT 显著性映射。通常，GT显著性图是通过人类共识或由dataset创建者获得的[16]。基于大规模RGB-D数据集，基于深度卷积神经网络的RGB-D显著性检测模型[10]、[11]、[14]、[17]、[18]取得了深远进展。我们认为，RGB-D显著性检测的方式与传统管道[10]，[11]相比进展顺利，

|  |
| --- |
| * *张静是澳大利亚国立大学工程研究学院ACRV，DATA61-CSIRO。（电子邮件： zjnwpu@gmail.com）* * *范邓平在中国南开大学计算机科学系工作。（电子邮件：dengpfan@gmail.com）* * *戴玉超现现就职于中国西北工业大学电子信息学院。（电子邮件： daiyuchao@gmail.com）*• *Saeed Anwar在澳大利亚国立大学DATA61-CSIRO工作。（电子邮件： saeed.anwar@data61.csiro.au）* * *Fatemeh Saleh在澳大利亚国立大学ACRV工作。（电子邮件： fatemehsadat.saleh@anu.edu.au）* * *Sadegh Aliakbarian 在澳大利亚国立大学ACRV工作。（电子邮件： sadegh.aliakbarian@anu.edu.au）* * *Nick Barnes在澳大利亚国立大学计算机学院工作。（电子邮件： nick.barnes@anu.edu.au）* * *这项工作的初步版本出现在CVPR 2020上[1]。* * *通讯作者：范邓平。* |

[14]，[17]，[18]未能捕捉到标记GT显著性图的不确定性。

根据人类视觉感知研究[19]，视觉显著性检测在一定程度上是主观的。在标记显著性图时，每个人都可以有特定的偏好[20]（ 这在特定于用户的显著性检测[21]中已经讨论过了）。更准确地说，GT标记过程从来都不是一个确定性过程，这与类别感知任务（如语义分割[22]）不同，因为 "表"永远不会被模棱两可地标记为"Cat"，而一个注释器的显着前景可能被其他注释器定义为背景，如图1的第二行所示。

在图1中，我们介绍了GT显著性图和其他 候选显著区域（由我们基于CVAE的方法生成，将在第3.2节中介绍），这些区域可能会引起人类注意。图1显示了确定性映射（从"图像"到

"GT"）可能导致"过度自信"的模型，因为所提供的"GT"可能是biased，如图1的第二行所示。为了克服这一点，我们感兴趣的不是执行点估计，而是网络如何通过生成1的多样化显著性图实现分布估计，捕获人类注释的不确定性。Furthermore，在实践中，更可取的是

生成多个显著性图以反映人类的不确定性，而不是为后续任务预测单个显著性图。

受人类感知不确定性以及显著性图标记过程的启发，我们提出了一种生成架构，以实现概率RGB-D显著性检测，并在注释中使用潜在变量 *z* 模拟人类不确定性。该框架中包括两个主要模型：1）生成器（*即*编码器-decoder）模型，它将输入RGB-D数据和潜在变量映射到随机显著性预测;和 2） 推理模型，它逐渐刷新潜在变量。

1. 预测的多样性取决于图像的上下文，其中简单的 上下文图像将导致一致的预测，而复杂的上下文图像可能会生成不同的预测。

|  |
| --- |
| 图 1.GT与我们预测的显著性地图进行了比较。对于简单的上下文图像（第一行），我们可以生成一致的预测。当对突出区域（第二行）存在不确定性的复杂场景进行分析时，我们的模型可以产生不同的预测（"我们的CVAE样本"和"我们的ABP样本"），其中"我们的CVAE"和"我们的ABP"是我们在saliency共识模块之后的确定性预测，其中 |

将在第 3.3 节中介绍。

为了推断潜在变量，我们引入了两种不同的策略：

* 基于条件变分自动编码器 （CVAE） [23] 的模型，带有附加编码器，用于近似潜在变量的后分布。
* 一种基于交替反向传播（ABP）[24]的技术，通过基于朗格文动力学的马尔可夫链蒙特卡洛（MCMC）采样[25]，直接从真后验分布中对潜在变异le进行采样。

本文是我们会议论文UCNet [1]的扩展版本。特别是，UC-Net专注于通过CVAE生成显著性图，并增强地面实况，以模拟多样性并避免后部坍缩问题[26]。虽然UC-Net通过对此类变化进行建模显示出有希望的性能，但它仍然有许多缺点。首先，UC-Net需要工程工作（地面- 事实增强）来模拟多样性并实现稳定训练（减轻后塌陷）。在这里，我们使用一种更简单的技术来实现相同的目标，通过使用标准的KL退火策略[27]，[28]减少人为干预。图13中的实验结果清楚地说明了KL退火策略的有效性。其次，我们通过设计一种更具表现力的解码器来提高生成的显著性图的质量，该解码器受益于空间和通道注意机制[29]。第三，受[23]的启发，我们修改了UC-Net的成本函数，以减少在训练和测试时编码潜在变量的差异，这在第3节中有所阐述。

此外，基于CVAE的方法通过推理模型（或编码器）近似后验分布，并优化证据下限（ELBO）。下限只是重建损失的组成以及近似后分布和先验分布之间的分歧。如果模型更侧重于优化重建质量，则潜在空间可能无法学习有意义的表示。另一方面，如果模型更侧重于减少近似后分布和先验分布之间的背离，则模型可能会提高重建质量。此外，由于模型近似于后验分布，而不是对真实后验进行建模，因此一般而言，它可能会失去表现力。在这里，我们建议使用交替反向传播（ABP）技术[24]，直接从真后验中采样潜在变量。虽然它要简单得多，但我们的实验结果表明，ABP在生成显著性图方面取得了令人印象深刻的结果。请注意，基于 CVAE 和基于 ABP 的解决方案都可以通过将输出空间分布建模为以输入 RGB-D 图像对为条件的生成模型来生成显著性预测。与UC-Net类似，在测试阶段，引入了显著性共识模块，以模拟GT显著性地图生成的多数投票机甲动画，并最终生成一个显著性映射用于性能评估。最后，除了产生最先进的结果外，我们的实验还对m odel的不同成分进行了全面评估，并对生成的显著性图的多样性进行了广泛的研究。

我们的主要贡献是：1）我们提出了第一个受不确定性启发的概率RGB-D显著性预测模型，该模型将潜在变量*z*引入网络以表示注释中的人类不确定性; 2）我们引入了两种不同的方案来推断潜在变量，包括一个带有附加编码器的CVAE [23]框架，以近似z的后验分布和ABP [24]管道， 它通过基于朗格文动力学的马尔可夫链蒙特卡洛（MCMC）采样直接从其真实后验分布中采样潜变量[25]。它们中的每一个都可以对输出的条件分布进行建模，并在测试过程中导致不同的预测;3）在6个RGB-D显著性检测基准数据集上的广泛实验结果证明了我们提出的解决方案的有效 性。

# 2 R兴高采华 W兽人

在本节中，我们首先简要回顾现有的 RGB-D 显著性检测模型。然后，我们研究了现有的生成模型，包括变分自动编码器（VAE）[23]，[30]和生成对抗网络（GAN）[31]，[32]。我们还在本节中强调了建议的解决方案的独特性。

## 2.1 RGB-D显著性检测

根据RGB图像和深度数据之间的互补信息融合方式，现有的RGB-D显著性检测模型可以大致分为三类：早期融合模型[1]，[33]，晚期融合模型[18]，[34]和模型上的交叉水平fusi[10]–[15]，[17]，[35]-[42]。第一种解直接将RGB图像与其深度信息连接起来，形成四通道输入，并馈送到网络，获取外观信息和几何信息。[33]提出了一个早期融合模型，为RGB-D对的每个超像素生成特征，然后将其馈送到CNN以产生每个超像素的显著性。第二种方法独立处理每个模态，并且来自两种模态的预测在网络的末端融合在一起。引入了一个后期融合网络（*即*AFNet），以自适应地融合来自RGB和深度分支的预测。在类似的管道中，[18]通过完全连接的层融合了RGB和深度信息。第三种通过考虑上述两种模式的相关性，融合了每种模态的中间特征。为了实现这一目标，[35]提出了一个互补感知融合块。[17]设计了注意力感知的跨级组合块，以获得每种模式的补充信息。[11]采用流体金字塔集成框架，实现多尺度跨模式特征融合。[13]设计了一个自注意模型，以有效地融合RGB和深度信息。类似地，[12]提出了一个互补的y交互模块（CIM），用于从RGB和深度数据中选择互补表示。[14]为互补特征发现提供了一个联合学习和密集合作的融合框架。[15]引入了一个深度蒸馏器，将深度流的知识从深度流转移到RGB流，从而在测试时不使用深度数据即可实现轻量级架构。全面的调查可以在[43]中找到。

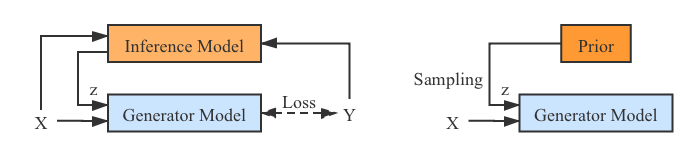
## 2.2 基于VAE或CVAE的深度概率模型

自从金马*等人*的开创性工作以来。[30] 和雷岑德*等人。*VAE及其条件对应物CVAE[23]已广泛应用于各种计算机视觉问题。典型的基于 VAE 的模型由编码器、解码器和损耗函数组成。编码器是一个权重和偏差θ的神经网络，它将输入数据点*X*映射到潜在（隐藏）表示*z。*解码器是另一个具有权重和偏差φ的神经网络，它从*z*重建数据点*X。* 为了训练VAE，需要重建损失和正则器来惩罚潜在表示的先验和后分布的分歧。基于 CVAE 的网络不是将潜在表示的先验分布定义为标准高斯分布，而是利用输入观测来调制高斯潜在变量上的先验以生成输出。

在低水平视觉中，VAE和CVAE已被应用于诸如具有清晰样本的潜在表示[45]，运动模式差异[46]，医学图像分割模型[47]以及建模图像固有的模糊性[48]等任务。同时，VAE和CVAE已经在更复杂的视觉任务中进行了探索，例如不确定的未来预测[49]，显着特征e增强[50]，人体运动预测[51]，[52]和形状引导图像生成[53]。最近，VAE和CVAE已扩展到3D域目标应用，如3D网格变形[54]和点云实例分割[55]。对于显著性检测，[56]采用VAE对图像背景进行建模，并通过重建残差将显著物体与背景分开。

## 2.3 GAN或基于CGAN的密集模型

GAN [31]及其条件对应物[32]也被用于密集的预测任务。现有的基于GAN的密集预测模型主要集中在两个方向上：1）以完全监督的方式使用GAN[57]–[61]，并将鉴别器损失视为密集预测的高阶正则化器;或2）将GANs转换为半监督场景[62]，[63]，其中鉴别器的输出可作为评估参与网络训练的无监督样本程度的指导。在显著性检测中，沿着第一个方向，[64]intr在注视预测网络中诱导鉴别器来区分预测的注视映射和基本事实。与上述两个方向不同，[65]在RGB-D显著性检测网络中采用了GAN，同时探索模态内（RGB，深度）和跨模态。[66]使用GAN作为去噪



（a） 培训管道 （b） 测试管道

图 2.培训和测试管道。在训练期间，推断出的潜在变量*z*和输入图像 *X* 被馈送到"生成器模型"，用于 s的显著性预测。在测试过程中，我们从 z 的先验分布中采样，以为每个输入图像生成不同的预测。

清除嘈杂输入图像的技术。[61]设计了一个鉴别器来区分真实的显著性图（group truth）和假的显著性图（预测），因此可以在不使用CRF[67]作为后处理技术的情况下学习结构信息。[68]采用 CycleGAN[69] 作为域自适应技术，为现有的RGB sal iency数据集生成伪近红外图像，并实现多光谱图像显著目标检测。

## 2.4 我们解决方案的独特性

据我们所知，除了我们的预定义版本[1]之外，在显著性检测中尚未利用生成模型来对注释不确定性进行建模。作为条件潜在变量模型，可以使用两种不同的解来推断潜在变量。一种是基于CVAE的[23]方法（我们在初步版本[1]中使用的方法），它使用变分推理推断潜在变量 ，另一种是基于MCMC的方法，我们建议在这项工作中使用。具体而言，我们提出了一种基于交变反向传播技术的参数负载较小的新潜在变量推理解决方案[24]。

基于CVAE 的模型通过找到对数似然的ELBO来推断潜在变量，以避免MCMC，因为它在非深度学习时代太慢了。换句话说，CVAE通过找到带有额外编码器的ELBO来近似最大似然估计（MLE）。这种策略的主要内容是"后坍缩"[26]，其中潜在变量独立于网络预测，使其无法表示人类注解的不确定性。我们在[1]的初步版本中引入了"新标签生成"策略，作为避免后折叠问题的有效方法。在这个扩展版本中，我们提出了一种更简单的策略，使用KL退火策略[27]，[28]，它慢慢地将KL损耗项引入损耗函数，并具有额外的权重。 实验结果表明，利用所提供的单个GT显著性图，该简单策略可以避免后倾塌陷问题。

除了KL退火术语外，我们还引入了ABP [24]作为防止网络后部塌陷的替代解决方案。ABP 引入了基于梯度的 MCMC，并通过梯度下降反向传播更新了潜在变量，以直接训练靶向 MLE 的 network。与 CVAE 相比，ABP 直接从其真实后分布中对潜在变量进行采样，使其在推断潜在变量时更加准确。此外，A BP中没有使用辅助网络（CVAE中的附加编码器），这导致较小的网络参数负载。

我们引入了一个基于ABP的推理模型，作为基于CVAE的管道的扩展[1]。实验结果表明，两种解都能有效估计潜在变量，从而产生随机的异化预测。第 3.2 节详细介绍了这两种推理模型。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表 1  我们框架不同部分的符号摘要。   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  | 数据 |  | 发电机型号 |  | 推理模型 | 显著性共识模块 | | | *X* | RGB-D 映像对 | *哦* | 发电机参数 | *Qφ*（*z*|*X，Y* ） | 后分布 | *C* | 采样次数 | | *和*  *D* | 地面实况显著性地图  训练数据集 | *跟*  *S* | 潜在变量  发电机预测 | *Pθ*（*带*|*X*）  *Pω*（*z*|*X，Y* ） | 事先分发  ABP的后分布 |  | 二进制多数  灰色多数 | |

# 3 O小时 M奥德尔

在本节中，我们将介绍概率 RGB-D 显著性检测模型，该模型学习显著性映射的基础条件分布，而不是从 RGB-D 输入到单个显著性映射的映射函数。设为训练数据集，其中*Xi* 表示 RGB-D 输入，Y *i*表示GT 显著性映射，N 表示数据集中图像的总数。我们打算对 *Pω*（*Y*  |进行建模*X，z），*其中 *z* 是一个潜在变量，表示显著区域的固有不确定性，这也可以从人类如何注释显著对象中看出。我们的框架在训练过程中利用了两个主要组件：1）生成器模型，它将输入RGB-D *X*和latent变量*z*映射到条件预测*Pω（Y|*  *X，z*）;和 2） 一个推理模型，它推断潜在变量 *z*。在测试过程中，我们可以从学习的先验分布*Pθ（z*|中采样多个潜在变量*X*） 生成随机显著性预测。图 2 （a） 和 （b） 分别说明了模型在训练和测试期间的整个流程。具体而言，在训练期间，模型从"生成器模型"中学习显著性，并使用"推理模式 l"更新潜在变量。在测试期间，我们从潜在变量的"先验"分布中采样，以获得随机显著性预测。我们在表1中总结了本文中使用的主要符号，以便于参考。

## 3.1 发电机模型

发生器模型将*X*和d 潜在变量 *z* 作为输入，并生成随机预测 *S*  = *Pω*（*Y*  |*X，z*），其中 *ω* 是生成器模型的参数集。我们选择ResNet50 [70]作为我们的主干，它包含四个卷积块。为了扩大接受场，我们通过 DenseASPP [71] 来获得一个特征图，其中包含骨干网络每个阶段整个图像的感受场。然后，我们以自上而下的方式逐渐连接两个相邻的特征图，并将串联的特征馈送到"残余C汉内尔注意"模块[72]，以获得随机显著性图S。如图3所示，我们的生成器模型通过正确使用混合注意力机制，遵循密集预测问题（如语义分割[22]）的最新进展。最后，我们的生成器模型受益于两种类型的注意力：位置注意力模块[29]和通道注意力模块[29]。前者旨在捕获特征图的任意两个位置之间的空间依赖关系，而后者旨在捕获特征图中任意两个通道之间的通道依赖关系。我们遵循[29]聚合和融合这两个注意力模块的输出，以进一步增强特征表示。

## 3.2 推理模型

我们提出了两种不同的解决方案来推断或更新潜在变量z：1）基于CVAE的[23]管道，其中我们通过神经网络（即编码器）近似后验分布*;*和2）基于ABP [24]的策略，通过基于Langevin动力学的MCMC直接从*z*的真实后驱支流中采样[25]。

使用CVAE推断z：变分自动编码器[30]是一个有向图形模型，通常由两个基本组件组成，一个编码器将输入变量*X*映射到潜在空间*Q φ（z*|*X），*其中*z* 是一个低维高斯变量，也是一个从 *z*重建*X* 以获得*Pω（X*|*z*）。为了训练VAE，重建损失和正则器来惩罚z的先验和近似后验分布的分歧被利用为：

LVAE = *E，带*∼*Qφ（带*|*X*）[−对数*Pω*（*X*|*与*）]

(1)

## +*DKL*（*Qφ*（*带*|*X*）||*P*（*与*））*，*

其中第一项是重构损失，或预期的负对数似然，第二项是正则化器，即 Kullback-Leibler 散度 *D*KL（Q*φ*（z|*X*）||*P（z））*来减小正态分布的先验P（z）和ap近似后*量子*Q φ（z|*X*）。期望*值 Ez*∼*Qφ*（*z*|*X*）取由近似后验分布*Q* *φ（ z|*生成的潜在变量z*X*）。

与VAE不同，VAE使用从标准正态分布生成的潜在变量对边际似然（特别是P（X））进行建模，CVAE [23]将潜在变量z的先验调制为高斯分布，其参数以输入数据X为条件。条件生成模型中有三种类型的变量：条件变量、潜在变量和输出变量。在我们的显著性检测场景中，我们将输出定义为显著性预测 *Y，*并将潜在变量定义为 *z*。由于我们的输出*Y*以输入 RGB-D 数据 *X*为条件，因此我们将输入 *X* 定义为条件变量。对于潜在变量 *z，*取自高斯分布*Pθ（z*|*X），*输出变量*Y*由*Pω（Y*  |生成*X，z），*则*z* 的后验公式为*Qφ*（z|*X，Y），*给定输入-输出对*（X，Y）*的重现特征嵌入。CVAE的损失定义为：

LCVAE = *Ez*∼*Qφ*（*|X，Y* ）[−对数*Pω*（*Y* |*X，z*）]

(2)

+*λkl* ∗  *DKL*（*Qφ*（*z*|*X，Y* ）||*Pθ*（*z*|*X*）），其中*Pω*（*Y*  |*X，z）*是给定潜在变量*z*和条件变量*X*的 *P（Y）*的可能性，库尔巴克-莱布勒散度*D*KL（Qφ（z|*X，Y* ）||*Pθ*（*z*|*X*））作为正则化损耗来减小先前*Pθ*（*z*| 之间的间隙*X）* 和辅助后验 *Qφ*（*z*|*X，Y* ）。此外，为了防止第2.4节中提到的可能的后塌陷问题，我们引入了线性KL退火[27]，[28]项*λkl*作为KL损失项*DKL*的权重，其定义为*λkl* = ep */ N*  ep，其中*ep*是当前epoch，N *ep ep* 是最大纪元数。通过这种方式，在训练过程中，CVAE旨在模拟编码误差*D*KL（Qφ（z*|*下预测的条件对数似然。*X，Y* ）||*Pθ*（*z*|*X*））。在测试过程中，我们可以从先前的网络*Pθ（z*|*X*） 来获得随机预测。

如 [23] 中所述，训练时输出变量的 c onditional 自动编码可能不是在测试中进行预测的最佳选择

|  |
| --- |
| )  图5.推理模型的详细结构，其中  *K*  是 维度  潜在空间，"c1  4  K"表示  1  ×  1  卷积层  输出 通道 大小  4  ×  *K*  ，"fc"表示完全连接的层。  是网络参数集。我们使用五个卷积层  和两个完全连接的图层，用于映射输入 RGB-D 图像  *X*  (  或串联  *X*  和  *和*  对于后验网） 到统计  潜在空间：  (  *µ*  事先  *， s*  事先  )  对于PriorNet 和  (  *µ*  发布  *， s*  发布  分别用于后置网络。然后相应的潜在向量  *跟*  可以通过重新参数化技巧来实现：  *跟*  =  *µ*  +  *s*  ·  ∼N  (0  *,*  **我**  )  ,  图 3."生成器模型"的详细信息，它以图像*X* 和潜在变量 *z* 作为输入，并生成随机显著性映射 *S，*其中 "S1- where . |

S4"代表了我们骨干网的四个卷积块。"DASPP"是DenseASPP 模块 [71]，"PAM"和 "CAM"是位置注意和通道注意模块 [29]，"RCA"是 [72] 中的剩余通道注意操作。

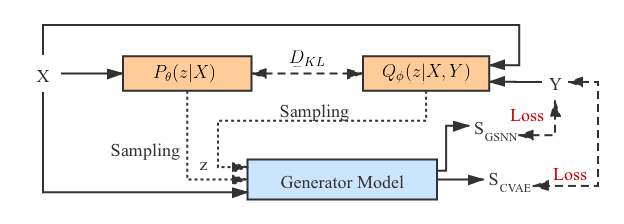


图 4.通过 CVAE 进行 RGB-D 显著性检测。"发电机模型"如图 3 所示。在训练期间，我们从后部净*z*  ∼ *Qφ*（*z*|进行采样*X，Y* ） 和先前的净 *z*  ∼ *Pθ*（*z*|*X*）分别获得预测 *SCVAE* 和 *SGSNN。*  在测试期间，S *GSNN* 是我们的预测。

时间，因为 CVAE 使用*z*  （*z*  ∼ *Q*φ（*z|*的后验*X，Y* ））表示训练阶段的重建损失，而它使用 *z（*  *z*  ∼ *Pθ*（*z*| 的先验*X*）） 在测试期间。在训练和测试中，一种缓解编码潜在变簇le 的差异的解决方案是为 KL 损失项分配更多权重（*例如，λkl*）。另一种解决方案是将后验网络设置为与前一网络相同，*即Q φ（z*|*X，Y* ） = *Pθ*（*z*|*X），*我们可以在训练和测试阶段直接从先前的网络中对潜在变量 *z* 进行采样。我们将此模型称为"高斯随机神经网络"（GSNN）[23]，目标函数为：

LGSNN = *Ez*∼*Pθ*（*z*|*X*）[−对数*Pω*（*Y* |*X，z*）]*.* （3）

我们可以结合上面的两个目标函数（LCVAE和 LGSNN）来获得一个混合目标函数：

LHybrid = *α*LCVAE + （1 − *α*）LGSNN （4）

遵循CVAE的标准实践[23]，我们设计了一个基于CVAE的RGB-D显著性检测管道，如图4所示。两个推理模型（ *Qφ*（*z*|*X，Y* ） 和 *Pθ*（*z*|*X）*）共享与图 5 中所示相同的结构，但 *Qφ*（z|*X，Y），*我们有*X*和 *Y* 的串联作为输入，而 *Pθ（z*|*X*） 接受 *X* 作为输入。让我们定义 *Pθ*（*z*|*X）*作为 PriorNet，它将输入的 RGB-D 数据 *X*映射到一个低维潜在特征空间，其中 *θ*是PriorNet的参数集。使用提供的 GT 显著性映射*Y*  ，我们定义 *Qφ*（*z*|*X，Y）*作为后置网，*具有φ* 根据方程 4，L CVAE中的 KL 发散用于测量 *Pθ（z*| *X*） 和 *Qφ*（*z*|*X，Y），*或使用*Q* φ（z*|*时丢失了多少信息*X，Y* ） 表示 *Pθ*（*z*|*X*）。另一方面，GSNN损失项LGSNN可以减轻在训练和测试期间编码潜在变量的差异。方程4中的混合损耗可以通过超参数α实现结构化输出，以平衡方程2和方程3中的两个目标函数。

用ABP推断*z：*如前所述，基于CVAE的模型的一个缺点是后坍缩问题[26]，其中模型学会忽略潜在变量，因此它变得独立于预测*Y，*因为*Q φ（z|X，Y* ） 将简单地坍缩为 *Pθ*（*z*|*X），*而 *z* 没有嵌入任何关于预测的信息。在我们的场景中，"后坍缩"现象可以被解释为潜在变量*z*未能在注释中捕获固有的人类不确定性。为此，我们提出了另一种基于交替反向传播的替代解决方案[24]。我们不是像在CVAE中那样用编码器网络近似z的后验，而是通过基于梯度的MCMC直接从其真正的后验分布中采样*z。*

引入了交替的Back-Pro分页[24]来学习发电机网络模型。它以 EM 方式更新潜在变量和网络参数。首先，给定当前参数集的网络预测，它推断出基于LANGEvin dynamics的MCMC的潜在变量，他们称之为"推断反向传播"[24]。其次，给定更新的潜在变量，使用梯度下降更新网络参数集。他们称之为"学习反向传播"[24]。在前面的变量定义之后，给定训练示例*（X，Y），*我们打算推断*z*并学习网络参数*ω，*以最小化重构误差以及与z上的先验相对应的正则化项。

作为因子分析的非线性泛化，条件生成模型旨在推广从连续潜在变量*z*到以输入图像*X*为条件的预测 *Y*的映射。与传统因子分析一样，我们将生成模型定义为：

*z* ∼ *P*（*z*） =N（0*，***I**）*，* （5）

*，*诊断（*σ*）2）*，* （6）

其中*P（z）*是*z*的先验分布。给定*X*的*Y*的条件分布为*Pω*（*Y*  |*X*） = R  *p*（*z*）*Pω*（*Y*  |*X，z）dz* 与潜在变量 *z* 积分。我们定义观测数据日志算法 1通过 Alternating 反向传播学习随机显著性

输入： 训练数据集

网络设置：最大epoch *Nep，*朗文步数*l，*步长*s，*学习速率*γ*

输出：网络参数集 *ω* 和推断的潜在变量

{*withand*}*Ni*=1

1：使用 ResNet50 [70] 初始化"生成器模型"的主干网以进行图像分类，以及其他具有截断高斯分布的新添加层。使用标准高斯分布初始化*zi。*

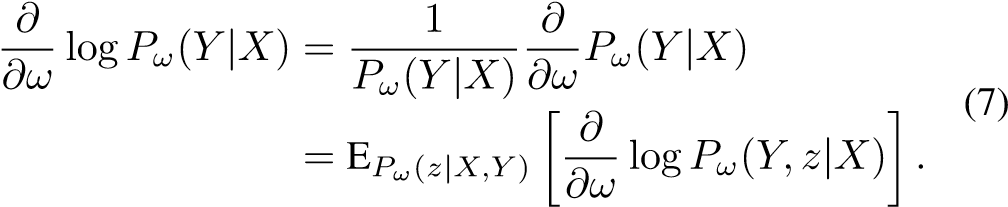
2：对于*t*  = 1*,...，Nep*  do

3：推导反向传播：对于每个 *i，*运行朗格文动力学的 l 步长来采样 *zi* ∼ *Pω*（*zi*|  *Yi，Xi）*遵循等式 8，z *i* 初始化为高斯白噪声（第一次迭代）或从上一次迭代中获得。

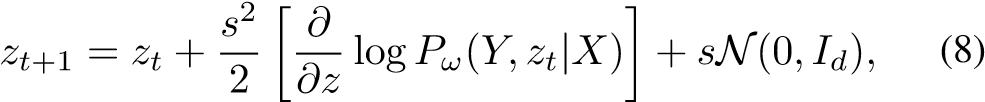
4： 学习反向传播：通过以下方式更新模型参数：

，其中*L（ω）*的梯度可以通过随机梯度下降获得。5：结束于

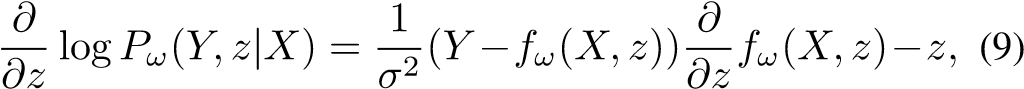
似然为 ，其中*Pω*（*Y*  |的梯度*X*） 定义为：



期望项E*Pω*（*z*|*X，Y*  ）可以通过从*Pω*（*z*|抽取样本来近似*X，Y），*然后计算蒙特卡罗平均值。此步骤对应于推断潜在变量 *z*。在ABP [24]之后，我们使用基于朗格文动力学的MCMC（一种基于梯度的蒙特卡罗方法）来sample *z，*它迭代：



跟



其中*t* 是朗格文采样的时间步长，s 是步长。通过 ABP 推断潜在变量*z*的整个流程如算法 1 所示。

分析两种推理模型：基于CVAE的[23]推理模型和基于ABP的[24]策略都可以推断出潜在变量*z，*其中前者使用额外的编码器近似于z的后验分布，而后者的解决方案通过直接从真后验分布中采样来针对MLE。如上所述，基于CVAE的解决方案可能会遭受后倾[26]的影响，其中潜在变量*z*与预测无关，使其无法表示标记的不确定性。为了防止后部或塌陷，我们采用KL退火策略[27]，[28]，让方程2中的KL损失项逐渐贡献CVAE损失函数。相反，基于ABP的解决方案没有后部坍缩问题，这导致更简单，更刺痛的训练，其中潜在变量*z*基于当前预测进行更新。在我们提出的两个解决方案中，使用推断的高斯随机变量z，我们的模型可以导致随机预测，*其中z*表示标记变体。

### 3.3 或估算

一旦学习了生成模型参数，我们的模型就可以在条件生成模型的生成过程之后从输入 X 生成预测。通过多次迭代采样，我们可以从相同的输入*X*获得多个显著性映射。为了评估生成网络的性能，我们需要估计结构化输出的确定性预测。受[23]的启发，我们的第一个解决方案是简单地平均多个预测。或者，我们可以从先验分布中获得多个 *z，*并将确定性预测定义为 *Y*  = *fω*（*X，E*（*z*）），其中 *E（z）* 是多个潜在变量的平均值。受GT显著性图的获取方式（*例如*，多数投票）的启发，我们引入了第三种解决方案，即"显著性共识模块"。

显著性共识模块： 为了准备用于显著性检测的训练数据集，要求多个注释器标记一个图像，并且显著性区域s的大多数[16]被定义为在最终GT显著性图中是显著的。

尽管GT的获取方式在显著性检测社区中是众所周知的，但没有关于将这种机制嵌入到深度显著性框架中的研究。 主要原因是 ，目前的模型将显著性检测定义为点估计问题而不是分布估计问题，最终的单显著性图 无法 进一步处理以实现"多数投票"。相反，我们设计了一个随机学习管道来获得预测的条件分布，这使得以类似于准备地面实况显著性图的方式生成确定性预测成为可能。具体来说，我们引入了显著性共识模块来计算测试阶段的大多数不同预测，如图2（b）所示。

在测试过程中，我们从PriorNet（对于基于CVAE的推理模型）对z进行采样，或者直接从标准高斯分布N（0，I）中采样z，并将其馈送到"Gener ator模型"以产生随机显著性预测，如图2（b）所示。通过*C*个不同的抽样，我们可以得到*C*预测*P*1,...*，PC*.我们同时将这些多个预测提供给显著性共识模块，以获得用于性能评估的 p 命令的共识。

给定多个预测，其中*Pc*  ∈ [0，1]，我们首先通过对*P*c执行自适应阈值 [73] 来计算预测的二进制2 版本*Pb* *c。*对于每个像素*（u，v），*我们得到一个 *C* 维特征向量*Pu，v* ∈ {0， 1}。我们定义为单通道显著性映射，表示

大多数 *Pu，v*，其定义为：

 *C*

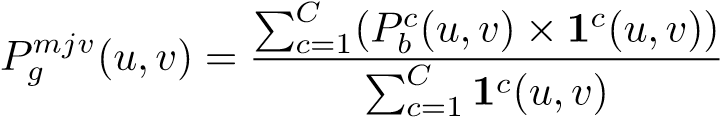
*mjv* 1*，*  X*c*=1 *Pbc*（*u，v*）*/C* ≥ 0*.*5*，*

*Pb* （*u，v*）= *C* （10）

0*，*  X*Pbc*（*u，v*）*/C <*  0*.*5*.* *c*=1

我们定义一个指标 

表示二进制预测是否与大多数预测一致。如果*Pbc*（*u，v*）= *Pbmjv（u，v*），则 **1***c（u，v）*=1。否则**，1***c*（*u，v*） =0。在显著性共识之后，我们得到一个灰色显著性图，如下所示：

*.* （11）

2. 当 GT 映射*Y*  ∈ {0，1}时，我们生成一系列二进制预测，每个预测表示来自一个显著性注释器的注释。

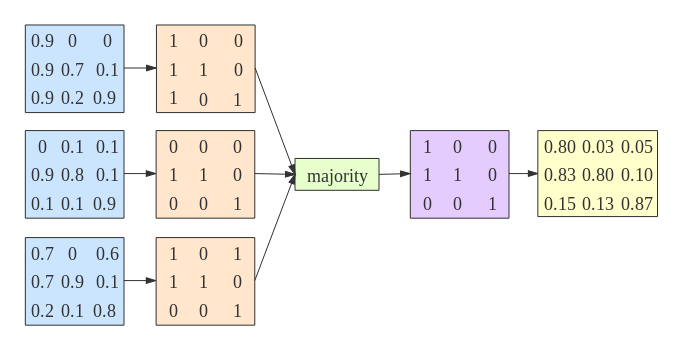


图 6.显示显著性共识模块如何工作的示例。

我们在图 6 中展示了一个*C*  = 3的玩具示例，以说明显著性共识模块的工作原理。如图 6 所示，给定三个灰度预测（以蓝色示意图），我们执行自适应阈值以获得三种不同的二进制预测（以橙色示出）。然后，我们计算一个多数矩阵（以紫色示意图），该矩阵也是二进制的，每个像素代表特定坐标的多数预测。最后，在显著性共识模块之后，我们的最终灰度预测是根据商定的像素的平均值计算的（当

， we mean in location *u，v*， 预测

同意多数）与多数矩阵，而忽略其他人。例如，坐标（1，1）中的大多数显著性为1，我们得到显著性共识模块之后的灰色预测为（0。9+0*.*7）*/*2 = 0*.*8，其中 0.9 和 0.7 是第一个和第三个预测（1， 1）中的预测。

### 3.4 损失函数

我们引入了两种不同的推理模型来更新潜在变量z：一个基于 CVAE 的模型，如图 4 所示，以及一个基于 ABP 的策略，如 Algorithm 1 所示。为了进一步突出预测的结构准确性，我们引入了平滑度损失，该假设基于以下假设：显著对象内的像素应具有相似的显著性值，并且沿着对象边缘发生明显的区分。

作为一种边缘感知损失，平滑度损失最初在[74]中引入，以鼓励视差局部平滑，并在视差梯度上受到L1惩罚。然后在[75]中采用它，通过使用图像pri or来恢复闭塞区域的光流。我们采用平滑度损失来实现高类内相似性的显著性图，在显著性对象内部具有一致的显著性预测，并且沿着对象边缘进行区分。在 [75] 之后，我们在平滑度项中将显著性映射的一阶导数定义为

L平滑 = X X Ψ（|*∂dPu，v*|*e*−*α*|*∂dIg*（*u，v*）|）*，* （12）

*u，v d*∈→−*x，*→−*y*

√

其中Ψ 被定义为 Ψ（ *s*）= *s*2 + 1*e*−6，P u，v 是位置*（u，v）*处的预测显著性映射*，Ig（u，v）*是偏导数上→−x和 →−上的图像尾*d*索引*y* 方向。我们

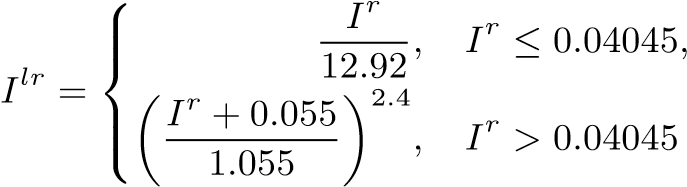
sity，在我们的实验中，按照[75]中的设置设置，设置α = 10。

我们需要计算图像在平滑度损失中的强度Ig，如等式（12）所示。为了实现这一点，我们遵循显著性保持[76]彩色图像反式形成策略，并将RGB图像*I*转换为灰度强度图像Ig，如下所示：



其中*Ilr*、I *lg*和 *ilb* 表示从原始色彩空间中移除 Gamma 函数后线性色彩空间中的颜色分量。*Ilr*通过以下方式实现：

(14)

*,*

其中*Ir*是图像*I*的原始红色通道，我们计算 *Ig* 和 *Ib* 的方式与等式 （14） 相同。

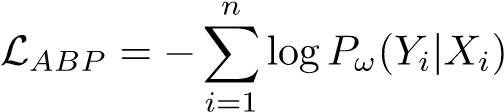
基于CVAE推理模型的损失函数： 对于 基于CVAE的 推理模型，我们在方程4中显示了其损失函数，其中负对数似然损失测量重建误差。为了保留结构信息并惩罚沿对象边界的错误预测，我们采用了结构电子感知损失[7]。结构感知损失是交叉熵损失的加权扩展，它整合了边界IOU损失[77]，以突出边界预测的准确性。

对于平滑度损失L平滑和 CVAE 损失 L混合，我们基于 CVAE 的框架的最终损失函数定义为：

L*CV* *AE*sal = LHybrid + *λ*1LSmooth*.* （15）

我们在*[0*]的范围内测试了λ 1。  1，0*.*2*,...，*0*。*9，1*.*0]，并发现 *λ*1  = 0时具有更好的性能*。*3.

ABP 推理模型的损失函数： 由于没有用于后验分布估计的额外编码器，因此 ABP 推理模型的损失函数只是负观测数据日志似然：

*，* （16）

它可以是与[7]中相同的结构感知损失，类似于基于CVAE的推理模型。

与上述平滑度损失相集成，我们得到基于ABP的显著性检测模型的损失函数为：

平滑*.* （17）

同样，我们也根据经验设置*λ 2* = 0*。*3 在我们的实验中。

# 4 EXPERIMENTAL RESULTS

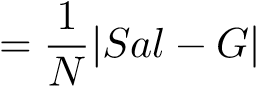
## 4.1 设置

数据集： 我们对六个数据集进行实验，包括五个广泛使用的RGB-D显著性检测数据集（即NJU2K [84]，NLPR [79]，SSB [89]，LFSD [90]，DES [81]）和一个新发布的数据集（SIP [16]）。

比较方法： 我们比较了我们的方法与第19种算法，包括10种手工制作的常规方法和9种深度RGB-D显著性检测模型。

评估指标：使用四个评估指标来评估确定性预测，即：1）平均绝对误差（MAE M）;2）平均F测量*（F β）*和最近提出的两个：3）结构测量（S测量，S α）[91]和4）平均增强对齐测量（E测量，E *ξ）[78]。*

* MAE M： MAE估计显著性图*Sal*和地面实况*G*之间的近似度。它提供估计图和 GT 图之间一致性的直接估计值。MAE 定义为：

（18）

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表 2  在六个RGBD显著性数据集上，对十个领先的手工制作特征模型和九个深度模型进行基准测试。↑ & ↓ 分别表示较大和较小越好。在这里，我们采用平均值*Fβ*和平均值 *Eξ* [78]。评估工具[：https://github.com/DengPingFan/D3NetBenchmark。](https://github.com/DengPingFan/D3NetBenchmark)   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | | 基于功能的手工模型 | | | | | | | | | Deep 型号 | | | | | | | | | 我们 | | | | 度量 | | LHM CDB DESM GP CDCP ACSD LBE DCMC MDSF SE | | | | | | | | | DF AFNet CTMF MMCI PCF TANet CPFP DMRA JL-DCF | | | | | | | | | UC-Net CVAE ABP | | | |  | | [79] [80] | [81] | [82] | [83] | [84] | [85] | [86] | [87] | [88] | [33] | [34] | [18] | [36] | [35] | [17] | [11] | [10] | [14] | [1] |  |  | | *Sα*  ↑  *NJU2K* [84] *β*  ↑  *F*  *Eξ* ↑ | | .514 .632 .328 .498  .447 .572 | .665 .550  .590 | .527 .357  .466 | .669 .595  .706 | .699 .512  .594 | .695 .606  .655 | .686 .556  .619 | .748 .628  .677 | .664 .583  .624 | .763 .653  .700 | .822 .827  .867 | .849 .779  .846 | .858 .793  .851 | .877 .840  .895 | .879 .841  .895 | .878 .850  .910 | .886 .873  .920 | .902 .885  .935 | .897  .886  .930 | .902 .893  .937 | .900 .889 .937 | |  | M ↓ | .205 .199 | .283 | .211 | .180 | .202 | .153 | .172 | .157 | .169 | .140 | .077 | .085 | .079 | .059 | .061 | .053 | .051 | .041 | .043 | .039 | .039 | | *SSB* [89] | *Sα*  ↑ *Fβ*  ↑  *Eξ* ↑ | .562 .615  .378 .489  .484 .561 | .642  .519  .579 | .588  .405  .508 | .713  .638  .751 | .692  .478  .592 | .660  .501  .601 | .731  .590  .655 | .728  .527  .614 | .708  .611  .664 | .757  .617  .692 | .825  .806  .872 | .848  .758  .841 | .873  .813  .873 | .875  .818  .887 | .871  .828  .893 | .879  .841  .911 | .835  .837  .879 | .903  .873  .936 | .903  .884  .938 | .898  .878  .935 | .904  .886  .939 | |  | M ↓ | .172 .166 | .295 | .182 | .149 | .200 | .250 | .148 | .176 | .143 | .141 | .075 | .086 | .068 | .064 | .060 | .051 | .066 | .040 | .039 | .039 | .037 | | *DES* [81] | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ | .578 .645  .345 .502  .477 .572 | .622  .483  .566 | .636  .412  .503 | .709  .585  .748 | .728  .513  .613 | .703  .576  .650 | .707  .542  .631 | .741  .523  .621 | .741  .618  .706 | .752  .604  .684 | .770  .713  .809 | .863  .756  .826 | .848  .735  .825 | .842  .765  .838 | .858  .790  .863 | .872  .824  .888 | .900  .873  .933 | .931  .907  .959 | .934  .919  .967 | .937  .929  .975 | 0940.928.975 | |  | M ↓ | .114 .100 | .299 | .168 | .115 | .169 | .208 | .111 | .122 | .090 | .093 | .068 | .055 | .065 | .049 | .046 | .038 | .030 | .021 | .019 | .016 | .016 | | *荷兰皇家* 共和国 [79] | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ | .630 .632  .427 .421  .560 .567 | .572  .430  .542 | .655  .451  .571 | .727  .609  .782 | .673  .429  .579 | .762  .636  .719 | .724  .542  .684 | .805  .649  .745 | .756  .624  .742 | .806  .664  .757 | .799  .755  .851 | .860  .740  .840 | .856  .737  .841 | .874  .802  .887 | .886  .819  .902 | .888  .840  .918 | .899  .865  .940 | 099.000  .951 | .920 .891  .951 | .917 .893  .952 | .919 .891 .952 | |  | M ↓ | .108 .108 | .312 | .146 | .112 | .179 | .081 | .117 | .095 | .091 | .079 | .058 | .056 | .059 | .044 | .041 | .036 | .031 | .024 | .025 | .025 | .024 | | *法律空间发展* [90] | *Sα*  ↑  *Fb* ↑  *Eξ* ↑ | .557 .520  .396 .376  .491 .465 | .722  .612  .638 | .640  .519  .584 | .717  .680  .754 | .734  .566  .625 | .736  .612  .670 | .753  .655  .682 | .700  .521  .588 | .698  .640  .653 | .791  .679  .725 | .738  .736  .796 | .796  .756  .810 | .787  .722  .775 | .794  .761  .818 | .801  .771  .821 | .828  .811  .863 | .847  .845  .893 | .862  .848  .894 | .864  .855  .901 | 0.868 .857  .904 | .866  .859 .903 | |  | M ↓ | .211 .218 | .248 | .183 | .167 | .188 | .208 | .155 | .190 | .167 | .138 | .134 | .119 | .132 | .112 | .111 | .088 | .075 | .070 | .066 | .065 | .065 | | *SIP* [16] | *Sα*  ↑  *Fβ* ↑ *Eξ* ↑ | .511 .557 .287 .341  .437 .455 | .616 .496  .564 | .588 .411  .511 | .595 .482  .683 | .732 .542  .614 | .727 .572  .651 | .683 .500  .598 | .717 .568  .645 | .628 .515  .592 | .653 .465  .565 | .720 .702  .793 | .716 .608  .704 | .833 .771  .845 | .842 .814  .878 | .835 .803  .870 | .850 .821  .893 | .806 .811  .844 | .880 .873  .918 | .875  .867  .914 | .883 .877  .927 | .876 .863  .921 | |  | M ↓ | .184 .192 | .298 | .173 | .224 | .172 | .200 | .186 | .167 | .164 | .185 | .118 | .139 | .086 | .071 | .075 | .064 | .085 | .049 | .051 | .045 | .049 |   图 7.六个测试数据集（NJU2K、SSB、DES、NLPR、LFSD 和 SIP）上的电子测量和 F 测量曲线。最好在屏幕上观看。 |

其中 *N* 是像素总数。

* F测量*F β：*它本质上是一个基于区域的相似性指标，它估计精度和召回率的非线性相关性。我们使用[0，255]中的变化阈值提供平均F测量值，如图7所示。此外，我们在图9[10]之后的图9中显示了直接精度召回率曲线，以进行全面评估。
* S-measure *S*α：MAE和F-measure指标都忽略了重要的结构信息评估，而行为视觉研究表明，人类视觉系统对场景中的s结构高度敏感[91]。因此，我们另外包括结构测量（Smeasure [91]）。S测度结合了区域感知*（Sr）*和物体感知（S*o）*结构相似性作为它们

|  |
| --- |
| 图 8.我们的方法和c操作方法的预测的视觉比较。请注意，我们的最终预测是使用建议的 |

"显著性 共识 模块"（参见第 3.3 节）。

最终结构指标：

*Sα* = *α* ∗ *So* + （1 − *α*）∗ *Sr，* （19）

其中*α*  ∈ [0*，*1] 是一个余额参数，默认设置为 0.5。

* E-measure *Eξ*：E-measure是最近在二进制地图评估领域提出的增强对齐测量[78]。该测量基于认知视觉研究，该研究将局部像素值与一个项中的图像级平均值相结合，联合捕获图像级统计数据和局部像素匹配信息。

实现细节：我们使用PyTorch训练模型，并使用在 ImageNet 上预先训练的 ResNet50 [70] 参数初始化"生成器模型"的编码器。在图3中"发电机模型"的"DASPP"模块中，我们使用四种不同的膨胀速率尺度：6，12，18，24，与[71]相同，并将所有中间通道大小设置为*M*  = 32。对于这两个推理模型，我们将潜在变量的维数设置为 *K*  = 3。新层的围 hts 用N（0， 0初始化*。*01），并且偏差设置为常量。我们使用动量为 0.9 的 Adam 方法，并在最大 epoch 的 80% 后将学习率降低 10%。基本学习速率初始化为 5e-5。整个训练时间约为 9 小时，训练批量大小为 5，在配备 NVIDIA GeForce RTX GPU 的 PC 上的最大 epoch 为 100。对于输入图像大小352×352，CVAE模型和ABP模型的推理时间平均分别为0.06s和0.05s。

## 4.2 与最先进方法的比较

定量比较：我们在表2中报告了我们的方法（使用推理模型）和竞争方法的性能，其中"CVAE"是我们的框架，CVAE作为推理模型，"ABP"表示通过交替反向传播更新潜在变量*z*的模型。表2中的结果显示了CVAE和ABP的优势，它们在所有数据集上始终如一地实现了最佳性能。具体而言，在DES [81]和LFSD [90]datasets上，与表2中的"深度模型"相比，我们的方法实现了大约1%的S测量，E测量和F测量性能提升，MAE降低了5%。此外，与我们的初步版本"UC-Net"[1]相比，我们观察到性能有所提高，这表明了所提出结构的有效性。我们还在图7中显示了竞争方法和我们的E测量和F测量曲线。我们观察到，我们的方法不仅产生稳定的电子测量和F测量曲线，而且还产生最佳性能。我们还将图9中的PR曲线显示为额外的性能评估。请注意，我们的解决方案（UC-Net，CVAE和ABP）主要关注概率模型，这是我们的主要贡献，我们直接将输入层中的RGB和深度连接起来，以引导早期融合模型。通过执行JL-DCF [14]和DMRRA [10]等多模态融合，可以实现有效的模型，其中使用跨级融合策略来实现有效的几何（从深度）和外观（来自RGB图像）的通信融合，这可能会导致相对更好的性能，如NLPR数据集的表2所示。然而，我们观察到，与早期聚变模型相比，跨水平融合模型可能导致更多的参数。*例如*，对于 CVAE、ABP、UC-Net[1] 和 JL-DCF [14]，模型和 JL-DCF 的总参数编号分别为 62m、58m、31m 和 143m（其中"m"表示百万）。与JL-DCF相比，较低的参数负载进一步显示了我们模型的卓越性能，这也使得将我们的模型应用于基于器件的应用成为可能。此外，我们注意到"CVAE"和"ABP"之间存在性能差距。CVAE 模型在测试期间从先验分布中对潜在变量 z 进行采样，而 ABP 模型直接对*z*进行采样，以标准正态分布为 fr。潜在变量的不同分布导致不同的模型性能。由于潜在变量是随机的（尽管它在测试期间遵循固定分布），因此在不同的采样迭代中将获得不同的性能。

为了进一步评估所提出的方法，我们计算了8个尖端RGB显著性检测模型的性能。

|  |
| --- |
| 图 9.六个测试数据集上的精度召回率 （PR） 曲线，其中 x 轴是召回率 ，y 轴是精度。最好在屏幕上观看。 |

表 3

竞争的 RGB 显著性检测模型的性能和我们在 RGBD 显著性数据集上的性能，其中在使用 RGB 显著性模型进行测试时不使用深度数据。我们采用均值*Fβ* 和均值 *Eξ*。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 度量 | | AFBNet NLDF PiCANet RAS DGRL CPD SCRN F3Net | | | | | | | | 断续器 |
|  | | [92] | [77] | [93] | [94] | [95] | [96] | [9] | [7] | 我们 |
| *Sα*  ↑  *NJU2K* [97] *β*  ↑  *F*  *Eξ* ↑ | | .862  .835  .888 | .813  .783  .848 | .864  .818  .869 | .754  .744  .800 | .767  .716  .804 | .875  .852  .903 | .879  .863  .912 | .861  .837  .890 | .902  .893  .937 |
|  | M ↓ | .064 | .091 | .072 | .115 | .107 | .056 | .052 | .061 | .039 |
| *SSB* [89] | *Sα*  ↑ *Fβ*  ↑  *Eξ* ↑ | .893  .865  .918 | .859  .831  .893 | .896  .844  .899 | .828  .820  .871 | .824  .781  .865 | .902  .880  .928 | .902  .881  .928 | .891  .868  .921 | .898  .878  .935 |
|  | M ↓ | .045 | .062 | .053 | .076 | .073 | .040 | .041 | .043 | .039 |
| *DES* [81] | *Sα*  ↑ *Fβ*  ↑  *Eξ* ↑ | .879  .845  .893 | .828  .758  .831 | .883  .822  .872 | .806  .762  .823 | .833  .753  .849 | .894  .870  .907 | .907  .885  .927 | .880  .845  .892 | .937  .929  .975 |
|  | M ↓ | .035 | .058 | .039 | .060 | .054 | .029 | .026 | .030 | .016 |
| *荷兰皇家* 共和国 [79] | *Sα*  ↑ *Fβ*  ↑  *Eξ* ↑ | .881  .816  .896 | .847  .782  .876 | .876  .789  .870 | .853  .810  .888 | .840  .767  .873 | .893  .844  .914 | .894  .846  .920 | .884  .838  .912 | .917  .893  .952 |
|  | M ↓ | .042 | .052 | .051 | .049 | .053 | .034 | .036 | .035 | .025 |
| *法律空间发展* [90] | *Sα*  ↑ *Fβ*  ↑  *Eξ* ↑ | .817  .784  .838 | .777  .756  .806 | .827  .778  .825 | .673  .672  .727 | .782  .759  .817 | .836  .811  .856 | .827  .800  .847 | .835  .810  .857 | .868  .857  .904 |
|  | M ↓ | .094 | .121 | .103 | .162 | .117 | .088 | .088 | .089 | .065 |
| *SIP* [16] | *Sα*  ↑ *Fβ*  ↑  *Eξ* ↑ | .876  .847  .911 | .795  .752  .840 | .851  .806  .866 | .718  .696  .766 | .682  .606  .744 | .870  .859  .910 | .866  .861  .903 | .866  .850  .905 | .883  .877  .927 |
|  | M ↓ | .055 | .100 | .073 | .121 | .138 | .053 | .057 | .055 | .045 |

RGB-D测试数据集3，并与我们基于"CVAE"的模型进行了比较。结果示于表3中，进一步说明了所提出框架的优越性能。

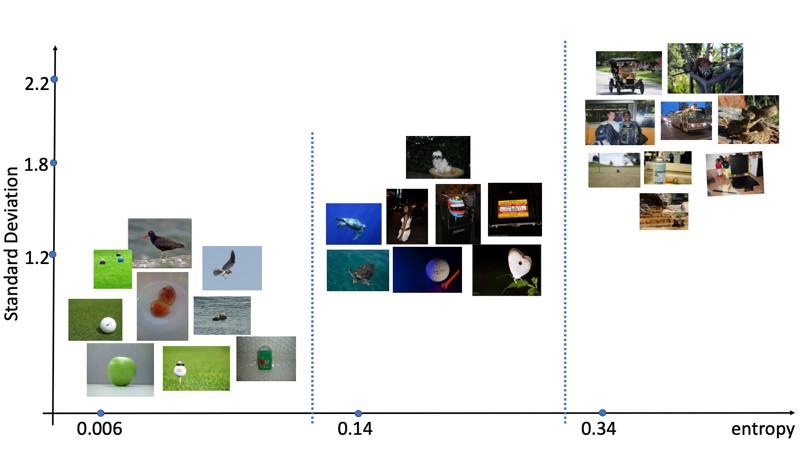
定性比较： 在图8中，我们展示了五个示例，将我们的方法与六个RGB-D显著性 检测模型进行比较。这些图像中的显著对象可以是大（第五行）、小（第二行）或复杂背景（第一、第三、第四和第五）

1. RGB 显著性模型在 RGB 显著性训练集上进行训练 ，并在 不使用深度的 RGB-D 测试集上进行测试。

行）。特别是对于第一行中的示例，背景很复杂，部分背景与突出的前景具有相似的颜色和纹理。大多数竞争方法（AFNet [34]，CPFP [11]和DMRRA [10]）未能正确分割精确的显着前景，而我们的方法则通过提出的两种推理模型中的每一个都实现了更好的显着物体检测 。对于最后一行的图像，存在一个对象（*即*绿色玩具）从其背景中强烈突出，而深度图可以在一定程度上降低这种高对比度区域的显着性。所有竞争方法 （特别是DCMC [86]，SE [88]，AFNet [34]，CPFP [11]）错误地检测出部分背景区域是显着的，而我们的准确预测进一步表明了我们解决方案的有效性。 通过图8中的所有结果，我们可以看到我们方法的优越性。

概率分布评估：作为概率网络，我们的模型可以生成合理显著性映射的分布，而不是对每个输入图像进行单个确定性预测。我们认为，对于具有简单背景的图像，应该产生一致的预测，而对于背景混乱的复杂图像，我们希望我们的模型能够捕获显著性图中的不确定性，从而可以生成不同的预测。为了评估模型的性能，按照主动学习管道 [98]，我们首先生成 *B*  = 100 个简单和困难的样本。为了实现这一目标，我们首先采用了三种不同的传统显著性模型（RBD [99]、MR [100] 和 GS [101]，它们跻身于基于 RGB的传统手工特征显著性模型前六名 [73]）。我们分别将它们命名为 *f*1、f 2 和 *f*3。 给定训练数据集 *D*中的图像*Xi* 4，我们计算其对应的显著性映射 *f*1（*X*i）、f 2（*Xi）*和 *f*3（*X*i）。我们选择熵作为图像复杂性的我。然后，我们将 X i 的平均显著性映射定义为 *Pi*  = （*f*1（*Xi*）+*f*2（*Xi*）+*f*3（*Xi*））*/*3。我们将图像的复杂性定义为任务驱动（用于显著性检测）。然后给予

1. 我们仅使用 RGB 数据。



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 图 10.结构化输出生成，其中"我们的 CVAE 样本"和"我们的 CVAE"分别是样本和确定性预测。  表 4  评估指标的分布（均值和方差），其中方差"2.3"表示"2.3e-06"。   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  | 新锐2K | 单边 带 | 一些 | 断续器 | 排水系统 | 啜 | | 静力学 | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | | 断续器 | .902 .893 .937 .039 | .898 .878 .935 .039 | .937 .929 .975 .016 | .917 .893 .952 .025 | .868 .857 .904 .065 | .883 .877 .927 .045 | | 断续器 | .905 .898 .939 .038 | .905 .887 .940 .037 | .936 .927 .973 .016 | .921 .896 .952 .025 | .864 .846 .899 .063 | .880 .872 .922 .046 | | 断续器 | 2.3 2.1 1.9 2.6 | 2.0 2.4 1.7 2.5 | 2.3 2.1 2.0 2.4 | 2.5 2.4 2.1 2.9 | 2.6 2.4 2.2 3.1 | 2.0 2.0 1.7 2.3 | | 总部基地 | .900 .889 .937 .039 | .904 .886 .939 .037 | .940 .928 .975 .016 | .919 .891 .952 .024 | .866 .859 .903 .065 | .876 .863 .921 .049 | | 马巴勃 | .903 .891 .941 .036 | .903 .890 .938 .035 | .937 .922 .970 .017 | .920 .897 .950 .025 | .868 .860 .912 .066 | .870 .867 .928 .046 | | 瓦博普 | 2.4 2.3 2.2 2.9 | 2.5 2.1 2.3 2.8 | 2.5 2.3 2.1 2.5 | 2.4 2.2 2.3 1.9 | 2.3 2.2 1.8 2.4 | 2.6 2.3 2.0 2.6 | |

图 11.通过分析熵和标准偏差进行图像分布。

一个地真显著性映射*Yi* 和均值显著性映射 *Pi，*我们将前景熵定义为： −*Pi*  log*Pi*。

然后，我们将平均熵定义为复杂性度量，并选择熵最小的B图像作为简单样本，选择熵最大的B图像作为困难样本（B  = 100）。我们从先验分布中对*Sn* = 5次进行采样，并计算每组的方差。具体来说，对于图像对*Xi，*使用*Sn*迭代采样，我们得到它的预测。我们计算这些 *Sn*不同预测的相似性，并将其视为预测多样性评估。我们在图11中显示了图像的熵和标准偏差。

当我们对潜在空间的每次采样迭代生成随机预测时，我们进一步分析了我们使用的四个评估指标的分布。具体而言，我们计算这些指标的均值和方差，并在表4中显示结果，其中"CVAE"和"ABP"是我们在表2中报告的性能，"MCVAE"，"VCVAE"，"MABP"，"VABP"分别是"CVAE"和"ABP"模型的平均值和变异值。为了获得指标的均值和方差，我们在测试期间从先前分布中多次采样（在本文中为五次）。然后，我们计算每次采样迭代的四个指标，并得到多个指标的平均值和方差，如表 4 所示。每个指标的相对稳定的平均值表示我们预测的有效性。同时，现有测试数据集中的大多数图像都有一个简单的背景，因此我们的模型没有得到充分的探索，因为简单的背景表明我们的任务中的不确定性较小（每个指标的方差较小）。我们认为，为每个图像提供具有不同注释的新训练数据集可以进一步探索我们的模型。

推理时间5  Comparison：我们在表 5 中总结了比较方法的基本信息，包括它们的代码类型和推理时间。表5表明，我们方法的推理时间6与竞争方法相当，这进一步说明我们的模型可以在不牺牲推理时间的情况下进行概率预测。

## 4.3 结构化输出生成

作为一个生成网络，我们引入了一个潜在变量 *z* 来模拟人类注释的不确定性。我们进一步展示了我们的模型生成 structured 输出的示例，如图 10 所示。图10中的"我们的CVAE样本"表示我们的方法与CVAE推理模型的三个随机样本，

1. 传统的基于手工特征的方法在aCPU上实现，深度RGB-D显著性预测模型基于GPU，因此我们报告前者的CPU时间和后者的GPU时间。
2. 我们报告的推理时间表示来自PriorNet的一个随机采样的预测。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表 5  现有方法的代码类型和推理时间。M =马特拉布。Pt = PyTorch.Tf = Tensorflow。   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 方法 | LHM [79] | 国开行 [80] | 德斯曼 [81] | 普通合伙人 [82] | 疾控中心 [83] | 加拿大难解运会 [97] | 磅 [85] | 直流电 [86] | MDSF [87] | | 时间 | 2.13 | 0.60 | 7.79 | 12.98 | 60.00 | 0.72 | 3.11 | 1.20 | 60.00 | | 代码类型 | M | M | M | M&C++ | M&C++ | C++ | M&C++ | M | C++ | | 方法 | 东南 [88] | 自由归化板 [33] | AFNet [34] | 断续器 [18] | MMCI [36] | 聚氯乙烯 [35] | CPFP [11] | 我们的总部基地 | 我们的简历 | | 时间 | 1.57 | 10.36 | 0.03 | 0.63 | 0.05 | 0.06 | 0.17 | 0.05 | 0.06 | | 代码类型 | M&C++ | M&C++ | 断续器 | 咖啡 | 咖啡 | 咖啡 | 咖啡 | 铂 | 铂 |   表 6  评估我们模型中不同组件的影响，以及替代结构。我们呈现均值*Fβ* 和均值 *Eξ*。   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  | 新锐2K | 单边 带 | 一些 | 断续器 | 排水系统 | 啜 | | 方法 | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | | 中间 | .897 .888 .933 .042 | .895 .880 .934 .041 | .931 .920 .968 .018 | .916 .887 .950 .026 | .854 .843 .888 .073 | .873 .863 .914 .048 | | 晚 | .890 .875 .929 .046 | .891 .866 .931 .042 | .929 .909 .970 .020 | .907 .877 .947 .028 | .839 .828 .887 .076 | .870 .853 .916 .051 | | AveP | .900 .892 .936 .040 | .897 .877 .934 .040 | .935 .924 .970 .017 | .914 .890 .951 .025 | .857 .842 .899 .067 | .880 .876 .926 .046 | | 有 | .901 .890 .927 .040 | .892 .875 .930 .040 | .929 .921 .971 .018 | .914 .884 .950 .026 | .855 .843 .892 .068 | .880 .874 .926 .046 | | 国安 | .900 .887 .935 .040 | .894 .873 .930 .041 | .931 .919 .971 .018 | .913 .885 .949 .026 | .852 .834 .894 .070 | .871 .864 .916 .051 | | CVAE S | .900 .890 .932 .040 | .894 .876 .931 .041 | .936 .927 .974 .016 | .914 .891 .949 .026 | .856 .843 .897 .068 | .877 .867 .920 .048 | | 我们 | .893 .881 .933 .042 | .885 .876 .930 .044 | .931 .921 .966 .017 | .914 .878 .950 .027 | .853 .845 .898 .069 | .882 .868 .924 .047 | | 那 | .900 .891 .936 .041 | .894 .876 .930 .040 | .935 .921 .970 .018 | .913 .891 .950 .025 | .851 .833 .887 .075 | .876 .856 .916 .051 | | 呵呵 | .897 .886 .934 .042 | .902 .882 .937 .038 | .930 .917 .970 .019 | .919 .892 .950 .024 | .850 .834 .888 .074 | .870 .856 .915 .052 | | 无 KLA | .900 .890 .932 .041 | .893 .870 .931 .040 | .932 .923 .972 .017 | .913 .887 .948 .027 | .854 .841 .893 .069 | .881 .872 .923 .046 | | z0 | .899 .890 .933 .040 | .897 .882 .931 .039 | .934 .927 .967 .018 | .916 .886 .947 .026 | .864 .846 .899 .067 | .871 .867 .920 .048 | | z1 | .900 .891 .930 .041 | .901 .880 .935 .039 | .936 .927 .969 .017 | .917 .890 .950 .025 | .866 .850 .900 .067 | .870 .868 .920 .049 | | 我们的简历 | .902 .893 .937 .039 | .898 .878 .935 .039 | 937 .929 .975 .016 | 0917.893 .952 .025 | 0868.857 .904 .065 | .883 .877 .927 .045 | | 我们的总部基地 | 0900.889 .937 .039 | .904 .886 .939 .037 | 940.928 .975 .016 | 0919.891 .952 .024 | 0866.859 .903 .065 | .876 .863 .921 .049 |   （a） 早期聚变模型 （b） 中聚变模型 （c） 晚期聚变模型 |

图 12.不同融合方案的详细网络结构。

和"我们的总部基地样品"是具有总部基地战略的样品。"我们的CVAE"和"我们的总部基地"是我们框架的确定性预测，通过我们的"显著性共识模块"获得的上述两个推理模型。图10表明，两个推理e模型都可以产生合理的随机预测，并且"显著性共识模块"（"我们的CVAE"和"我们的ABP"）之后的最终确定性预测与提供的GT一致，这验证了我们的潜在变量和 "显著性共识模块"的有效性。

## 4.4 烧蚀研究

我们进一步分析了本节中提出的框架，包括生成网络相关策略，损失函数，替代深度数据（特别是HHA [102]）以及防止网络后部崩溃的solutio n。我们在表 6 中显示了性能。请注意，除非另有说明，否则我们在以下实验中使用基于CVAE的推理模型。不同的融合方案：潜在变量*z*可以通过三种不同的方式融合到网络工作中：早期融合（在输入层中），中间融合（在瓶颈网络中）或后期融合（在输出层之前）。我们提出了一个早期聚变模型，如图12（a）所示。我们进一步设计了一个中间聚变模型和一个晚期fusion模型，分别如图12（b）和（c）所示。各型号的性能如表6"中间"所示，并

"迟到"。对于中间融合模型，将骨干网的第四组（*例如*S4）的最后一个卷积层馈送到1×1 co卷积层以获得*M*  = 32维特征图，然后将其映射到*K（*潜在变量z的维数*）* ） 具有全连接层 （"fc"） 的维度特征向量。为了避免后倾[26]，受[51]的启发，我们混合（"Mixup"）特征向量和*z*通道;因此，网络无法区分确定性分支和概率分支的特征。然后，我们扩展空间维度中的混合特征向量，并将其馈送到另一个 1 × 1 卷积层，以获得与 S4 具有相同维度的特征图 S4'，并在图 3 中将 S4' 替换为 S4'。对于后期融合模型，"发生器模型"表示图3中最后一个"RCA"模块之前的发生器模型。我们在空间角币中展开*z，*并将其与确定性特征连接起来。我们在这里也执行类似于中间融合模型的"混合"。然后，我们将混合特征图馈送到一个"RCA"模块和"DASPP"模型，以实现预测*S。*我们观察到中间融合模型（"中间"）和晚期融合模型（"晚期"）的性能略差。主要原因是，当潜在变量馈送到网络的开头时，可以获得强的非线性表示，这也与"Middle"优于"Late"的结果一致。

分析 z维数的影响*：z*的尺度可能会影响网络性能和预测的多样性。在本文中，我们将 z 的维数设置为 3。我们进一步对z的维数在[3，32]的范围内进行实验，

|  |
| --- |
| 图 13.使用我们基于 CVAE 的模型 （"CVAE"）、基于 ABP 的模型 （"ABP"） 和基于 CVAE 的模型进行多个预测的平均方差 |

没有KL退火术语（"w/o KLA"）。最好在屏幕上观看。

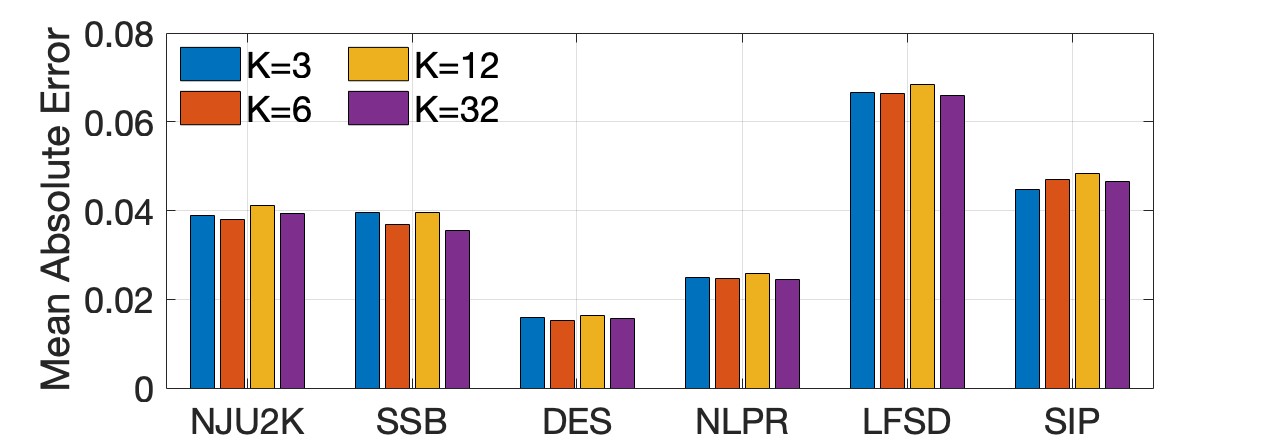


图 14.潜在变量的维度分析。

并在图 14 中显示了我们的模型在六个基准 RGB-D 显著性数据集上的平均绝对误差。我们观察到z的不同维度的性能相对稳定。z维数相对稳定的性能regard表明，网络的容量足够大，可以在输入中采取不同程度的随机性。此外，由于只存在几个相当困难的样本，并且z的较低维度足以捕获labeling的变体。

确定性预测生成：如 § 3.3 中所述，可以使用三种不同的解决方案来生成性能评估的确定性预测，包括 1） 平均多个预测;2）平均多个潜在可变参数;3）提出的显著性共识模块。我们评估了其他确定性推理解的性能，并在表 6"AveP"和"AveZ"中显示了性能，分别表示平均预测解和平均*z* 解。我们ob与所提出的显著性共识模块相比，"AveP"和"AveZ"的性能相似。"AveP"和"AveZ"的类似性能表明，两种传统的确定性预测生成解对于显著性预测任务都非常有效。"我们的"的更好性能表明了所提出的解决方案的有效性。

损失函数的有效性：由于*Q φ（z|*的不一致*X，Y* ） 和 *Pθ*（*z*|*X）*分别用于训练和测试阶段，该模型在训练和测试期间可能有所不同。为了减轻编码潜在变量的差异，并在训练和测试期间实现类似的网络行为，我们引入了高斯随机神经网络（GSNN）和混合损失函数，如方程4所示。为了测试我们的网络在仅使用等式2中的CVAE损失或等式3中的GSNN损失的情况下的表现，我们训练了两个额外的模型，并分别将性能显示为"CVAE S"和"GSNN"。我们看到性能明显下降，每次损失都使用在其ow n上。同时，尽管这两个模型的性能比所提出的解决方案更差，但与竞争方法相比，我们仍然观察到性能始终更好。与"我们的"相比，"CVAE S"和"GSNN"的性能下降，以及"CVAE S"和"GSNN"与竞争方法相比性能的更好，都表明了所提出的生成模型在显著性检测方面的有效性。

平滑度损失： 我们将平滑度损失引入损失函数，以对 pr 编辑的结构设置约束。为了评估平滑度损失的贡献，我们将其从损失函数中删除，并将性能显示为"NoS"。较低的性能表示平滑度损失的有效性。此外，如等式12所示，平滑度loss以 显著性预测和灰度图像作为输入，也可以解释为自监督 正则化器。

结构感知损失 *与* 交叉熵损失： 与[7]类似，我们使用结构感知损失而不是广泛使用的 交叉输入opy 损失来惩罚沿物体边缘的预测，因此我们可以实现结构保持显著性预测。为了证明我们的模型也可以很好地处理基本的交叉熵损失，我们设计了另一个使用交叉熵损失而不是结构感知损失的模型，并将性能显示为"CE"。我们注意到"CE"在"LFSD"和"SIP"数据集上的性能明显下降。对于"LFSD"和"SIP"数据集，存在与背景共享相似颜色的显着前景区域，这使得 基于交叉熵的模型在这些情况下无效。虽然结构感知损失可能会因不正确的结构信息而惩罚预测，使其对困难的图像有效。

HHA*与*深度：HHA [102]是一种广泛使用的技术，它将深度数据编码为三个通道：horizontal disparity，h 8 above ground，以及像素的局部表面法线与推断的重力方向产生的一个ngle。HHA广泛用于RGB-D密集模型[18]，[106]以获得更好的特征e表示。为了测试HHA是否也适用于我们的场景，我们

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表 7  与六个基准RGB显著性数据集上最先进的RGB显著性检测模型进行比较。我们采用均值*Fβ* 和均值 *Eξ*。   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  | 淋浴 | 云服务器 | 有 | 香港大学 | 星期四 | 软件中心 | | 方法 | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | *Sα*  ↑ *Fβ*  ↑ *Eξ* ↑ M ↓ | | 断续器 [95] | .846 .790 .887 .051 | .902 .898 .934 .045 | .809 .726 .845 .063 | .897 .884 .939 .037 | .816 .727 .838 .077 | - - - - | | 皮卡 [93] | .842 .757 .853 .062 | .898 .872 .909 .054 | .817 .711 .823 .072 | .895 .854 .910 .046 | .818 .710 .821 .084 | .801 .332 .810 .133 | | 荷兰国防军 [77] | .816 .757 .851 .065 | .870 .871 .896 .066 | .770 .683 .798 .080 | .879 .871 .914 .048 | .801 .711 .827 .081 | .816 .319 .837 .106 | | 巴斯恩 [103] | .876 .823 .896 .048 | .910 .913 .938 .040 | .836 .767 .865 .057 | .909 .903 .943 .032 | .823 .737 .841 .073 | .841 .359 .864 .092 | | AFNet [92] | .867 .812 .893 .046 | .907 .901 .929 .045 | .826 .743 .846 .057 | .905 .888 .934 .036 | .825 .733 .840 .072 | .700 .062 .684 .115 | | MSNet [104] | .862 .792 .883 .049 | .905 .886 .922 .048 | .809 .710 .831 .064 | .907 .878 .930 .039 | .819 .718 .829 .079 | - - - - | | 断续器 [9] | .885 .833 .900 .040 | .920 .910 .933 .041 | .837 .749 .847 .056 | .916 .894 .935 .034 | .845 .758 .858 .066 | .838 .363 .859 .099 | | 小板 [105] | 0890.861 .925 .034 | .919 .923 .943 .036 | .839 .770 .865 .052 | .920 .913 .953 .028 | 0768.863 .064 | .835 - - .103 | | 我们的简历 | 088.860 .927 .034 | .921 .926 .947 .035 | 0.839 .773 .869 .051 | .921 .919 .957 .026 | 0.848 .765 .862 .064 | .849 .369 .872 .089 | | 我们的总部基地 | .890 .864 .931 .034 | .915 .918 .941 .037 | 0.843 .770 .864 .050 | .917 .913 .949 .027 | 0.849 .773 .869 .066 | .842 .365 .868 .091 | |

将深度替换为 HHA，性能显示为"HHA"。我们观察到使用HHA而不是原始深度数据实现了类似的性能。使用HHA的模型旨在获得更好的深度表示，因为原始深度通常是高质量的。proposed随机模型为网络引入了随机性，这也可以作为一种去噪技术来提高模型的鲁棒性，这也与[107]中的观察结果一致。

无KL退火训练：如§ 2.4中所述，we引入了KL退火策略，以防止基于CVAE的模型可能存在的后塌陷问题。为了测试此策略的贡献，我们只需删除KL退火项，并将方程2中KL损耗项的权重设置为第一个epoch的1。该实验的性能显示为"无KLA"。尽管六个基准RGB-D显著性数据集的性能没有清楚地显示KL退火的影响（因为我们生成了确定性预测），但我们观察到它高度符合预测的多样性，如图13所示，它显示了RGB-D测试集上多个预测的平均方差。具体来说，我们在测试期间执行随机抽样的五次迭代，并计算这五个差异预测的方差。我们在图 13 中显示了方差图的均值。此外，我们将基于 CVAE 和基于 ABP 的模型的平均方差分别显示为"CVAE"和"ABP"。图13清楚地表明，我们提出的两种解决方案都可以生成比"w/o KLA"更多样化的预测，从而导致比"w/o KLA"更大的方差。潜变量 *z*的不同分布：作为一个基于 CVAE [23] 的常规框架，我们假设潜变量*z* 遵循高斯分布，并设计先验净和后净，以获得潜变量 z 的均值和标准差。在测试期间，我们从之前的网络中对 z 进行采样，以生成每个输入图像对的预测。为了测试我们的模型如何使用不同的分布来测试 z 的性能，我们只需将*z* 定义为遵循均匀分布，并在表 6 "z0" 和 "z1" 中显示性能，其中*z*  = 0 和 *z*  = 1（在基于 CVAE 的框架内）。z0（和 z1）的性能与我们的性能"CVAE"i 的可比性能使用潜在变量的不同分布来表示模型的有效性。

不同质量深度图的模型性能： 为了解释我们的模型在不同质量的深度图下的表现，我们通过定义平滑度误差来评估现有RGBD测试数据集的深度q度，如[75]所示。

[75]中定义了平滑度损失，以评估深度和RGB图像对齐，其中较小的平滑度损失表示更好的深度图。然后，我们将其用作深度质量估计的 s tandard。NJU2K [84]、SSB [89]、DES [81]、NLPR [79]、LFSD [90] 和 SIP [16] 的平滑度误差分别为 0.026、0.028、0.028、0.021、0.031、0.016，表明 SIP 数据集具有最高质量的深度图， 而 LFSD 数据集具有最低质量的深度图。我们在表2（"CVAE"和"ABP"）中的结果表明，对于高质量的深度图，模型性能更好。[108]中解释了深层网络对噪声不健壮的主要原因。同时，depth质量只是可能影响最终显著性预测的一个因素;其他因素， *例如*，图像的复杂性，标签的不确定性，也可能影响最终的预测。

## 4.5 概率 RGB 显著性检测

我们提出了一种基于RGB-D显著性检测网络的生成模型，并将其扩展到RGB显著性检测，以测试所提出框架的灵活性，并在表7中显示性能。我们使用DU TS训练数据集[109]训练我们的模型（"Ours CVAE"和"Ours ABP"），并在六个广泛使用的基准上评估我们的方法和竞争方法的性能：（1）DUTS测试数据集（5，019张图像）;（2） 云服务器 [110] （1，000 张图像）;（3） DUT [100] （5，168 张图像）;（4） 香港独立调查 [111] （4，447幅图像）;（5） THUR [112] （6，232 张图像） 和 （6） SOC 测试 [113] （1，200 张图像）。与基于RGB-D的显著性检测框架类似，我们使用相同的网络结构进行RGB显著性检测，除了输入图像*X*是RGB数据而不是RGB-D图像对。具体而言，对于RGB-D显著性检测，我们将输入层中的RGB、深度和潜在变量*z连接起来*，然后将其馈送到3×3卷积层，得到三通道特征图。对于RGB显著性检测，我们直接将RGB和潜在变量连线，并馈送到另外3个×3卷积层，得到三通道特征图。我们在表7中显示了我们的概率RGB显著性检测模型的性能，包括"我们的CVAE"和"我们的ABP"，分别代表基于CVAE的模型和基于ABP的模型。我们网络（"Ours CVAE"或"Ours ABP"）始终如一的更好性能说明了我们模型的灵活性，这可以为RGB-D显著性检测和RGB显著性检测带来新的基准性能。

# 5 CONCLUSION 和 FUTURE WORK

受地面实况注释中人类不确定性的启发，我们提出了第一个不确定性启发的RGB-D显著性检测模型。与通常将显著性确定性视为点估计问题的现有方法不同，我们提出学习显著性图的分布，并提出生成学习管道以产生随机显著性预测。此外，我们引入了两种不同的推理模型：1）基于CVAE的推理e模型，其中额外的编码器来近似潜在变量z的真实后验分布;以及2）基于ABP的推理模型，直接从其基于梯度的MCMC的真实后验分布中采样*z。*  在我们的公式下，我们的模型能够生成多个预测，表示人类注释的不确定性。通过提议的显著性共识模块，我们能够按照与地面实况注释生成过程类似的管道生成准确的显著性预测。对六个标准和具有挑战性的RGB-D基准的定量和定性评估证明了我们的方法在学习显著性图分布方面的优势。

同时，我们深入研究了生成模型，并对潜在变量、损失函数和不同融合方案进行了深入分析，将*z*引入网络。此外，我们将解决方案扩展到RGB显著性检测。在不改变网络结构的情况下（我们只将输入从RGB-D数据更改为RGB数据），与上一个RGB显著性模型相比，我们实现了最先进的性能。

引入了两种不同的推理模型来学习所提出的生成网络，如图2（a）所示。根据我们的经验，CVAE-based和基于ABP的推理模型都可以导致学术显著性预测，如图13所示。由于在 CVAE 模型中使用了额外的编码器，因此与基于 ABP 的解决方案相比，它会产生更多的网络参数。另一方面，当我们 通过运行基于Langevin Dynamics的MCMC的几个步骤来更新潜在变量时，如方程8所示，这导致相对较长的训练时间。此外，由于我们对每个训练图像只有一个注释，因此我们无法解释潜在变量的每个维度的表示 ，这与GAN[31]不同，GANs[31]可以控制潜在变量以生成不同属性的图像。在我们的任务中，由于每个训练图像的单一注释，我们不能假设潜在变量的每个维度都是独立的。在这种情况下，我们的潜在变量通常对注释的不确定性进行建模，而我们无法控制在潜在空间中每次迭代采样时生成哪个预测。一个新的训练数据集，带有多个le 注释，突出显示每个训练图像的不同显著性属性，可以进一步探索我们的模型。

将来，我们希望将我们的方法扩展到其他密集预测任务，*例如*视频显著目标检测[114]，单目depth估计[115]，立体匹配[116]，[117]，光流估计[118]，[119]*等*。此外，我们计划捕获具有多个人类注释的新数据集，以进一步模拟显著性感知中人类不确定性的统计数据。

# 一个知识

本研究得到了国家自然科学基金委员会（61871325、61671387、61620106008、61572264）、国家重点研发计划（2018AAA0102803）和天津市自然科学基金（17JCJQJC43700）的部分支持。

# REFERENCES

1. 张俊杰Fan， Y. Dai， S. Anwar， F. Sadat Saleh， T. Zhang， and N. Barnes，"Uc-net： Uncertainty inspired rgb-d saliency detection via conditional variational autoencoders，"在*IEEE Conf。计算机.Vis. Pattern*  *Recog.*，2020年。
2. L. Itti， C. Koch， 和 E.Niebur，"用于快速场景分析的基于salie ncy的视觉注意力模型"，IEEE *T. Pattern Anal. Mach。英特尔.*《第20卷》，第11期，第1254-1259页，1998年。
3. R·阿钱塔赫马米、F·埃斯特拉达和S.Susstrunk，"频率调谐显著区域检测"，IEEE *Conf。计算机.Vis. Pattern*  *Recog.*，第1597-1604页，2009年。
4. 王伟， 沈军， M.-M.Cheng和L. Shao，"用于显着对象检测的迭代和协作自上而下和自下而上的推理网络"，在*IEEE Conf上。计算机.Vis. Pattern*  *Recog.*， 2019.
5. J. Zhang， T. Zhang， Y. Dai， M. Harandi， and R. Hartley， "Deep unsupervised saliency detection： A multiple noisy labeling perspective"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，第9029-9038页，2018年。
6. Y. Liu， Q. Zhang， D. Zhang， and J. Han， "Employ deep part-object relation for salient object detection"，Int. *Conf.计算机.Vis.*， 2019.
7. S. W. Jun Wei和Q. Huang，"F3net：Fusion， feedback and focus for salient object detection"，载于 *AAAI Conf. Art。 英特尔.*，2020年。
8. J. Zhang， X. Yu， A. Li， P. Song， B. Liu， and Y. Dai，"Weaklysupervised显着对象检测通过涂鸦注释"，在*IEEE Conf。计算机.Vis. Pattern*  *Recog.*，2020年。
9. Z. Wu， L. Su， and Q. Huang， "用于边缘感知显著物体检测的堆叠交叉细化网络"，载于 *Int. Conf.计算机.Vis.*， 2019.
10. Y. Piao， W. Ji， J. Li， M. Zhang， and H. Lu， "用于显著性检测的深度诱导多尺度循环注意网络"，载于 *Int. Conf. 计算机.Vis.*， 2019.
11. 赵俊旭， 曹勇， D.-P.范，M.-M.程旭东Li和L. Zhang，"用于rgbd显着对象检测的对比先验和流体金字塔积分"，IEEE *Conf。计算机.Vis. Pattern*  *Recog.*， 2019.
12. M. Zhang， W. Ren， Y. Piao， Z. Rong， and H. Lu， "Select， Supplement and Focus for rgb-d saliency detection"，IEEE *Conf.公司.Vis. Pattern*  *Recog.*，2020年。
13. N. Liu，N. Zhang和J. Han，"学习rgb-d显著性检测的选择性自注意"，在*IEEE Conf。计算机.Vis. Pattern*  *Recog.*，2020年。
14. 傅凯范，G.-P.Ji， and Q. Zhao，"Jl-dcf： Joint learning and densely-cooperative fusion framework for rgb-d 显着对象检测"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，2020年。
15. Y. Piao， Z. Rong， M. Zhang， W. Ren， and H. Lu， "A2dele： Adaptive and Attentive depth distiller for efficient rgb-d 显着 object detection"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，2020年。
16. 范德佩、林志、张志、朱明、明明Cheng，"重新思考RGB-D显著对象检测：模型，数据集和大规模基准测试*"，IEEE T. Neural*  *Netw。学习。系统。*，2020年。
17. H. Chen和Y. Li，"用于RGB-D显着对象检测的三流注意力感知网络"，IEEE *T. Image Process。*，第2825-2835页，2019年。[18] J. Han， H. Chen， N. Liu， C. Yan， and X. Li， "基于CNN的RGB-D显著性检测通过跨视图传输和多视图融合，" *IEEE T.赛博因.*，第3171-3183页，2018年。
18. O. Le Meur 和 T.Baccino，"比较扫描路径和显著性图的方法：优点和缺点"，*行为研究方法*，第45卷，第1期，第251-266页，2013年。
19. J.M. Henderson和T. R. Hayes，" 意义地图揭示的场景中基于意义的注意力指导"，《 *自然人类 行为》，*第1卷，第10期，第743-747页，2017年。
20. L. Itti和C. Koch，"一种基于显著性的搜索机制，用于视觉注意力的显性和隐性转移，"Vis. *Res。*《第40卷，第10期》，第1489–1506页，2000年。
21. J. Long，E. Shelhamer和T. Darrell，"用于语义分割的完全卷积网络"，在*IEEE Conf。计算机.Vis. Pattern*  *Recog.*，第3431-3440页，2015年。
22. K. Sohn，H. Lee和X. Yan，"学习结构化输出表示u唱深度条件生成模型"，载于 *Adv. Neural Inform。过程。系统。*，第3483-3491页，2015年。
23. T. Han， Y. Lu， S. Zhu， and Y. Wu， "Generator Network的交替反向传播"，载于 *AAAI Conf. Art。 英特尔.*， 02 2017.
24. R.M. Neal *et al.*，"Mcmc using hamiltonian dynamics"，*马尔可夫链蒙特卡洛*手册，第2卷，第11期，第2页，2011年。
25. J. He， D. Spokoyny， G. Neubig和T. Berg-Kirkpatrick，"变分自动编码器中的滞后推理网络和后坍缩"，载于 *Int. Conf. Learn。Represent.*， 2019.
26. C. K. Sø nderby， T.莱科Maalø e， S. r. K.Sø nderby， 和

O. Winther，"梯形变分自动编码器"，载于 *Adv. Neural Inform。过程。系统。* （D. D. Lee， M. Sugiyama， U. V. Luxburg，I. Guyon， and R. Garnett， eds.）， pp. 3738–3746， 2016.

1. I. 希金斯， L. 马修， A. 帕尔， C. 伯吉斯， X.格洛罗特， M.鲍特维尼克、S·穆罕默德和A.Lerchner，"beta-vae：使用约束变分框架学习基本视觉概念"，载于*Int. Conf. Learn。代表。*， 2017.
2. J. Fu， J. Liu， H. Tian， Y. Li， Y. Bao， Z. Fang， and H. Lu， "Dual attention network for scene segmentation"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，第3146-3154页，2019年。
3. D. P. Kingma 和 M. Welling，"自动编码变分贝叶斯"，载于 *Int. Conf. Learn。代表。*， 2013.
4. I. 古德费罗普热-阿巴迪， 米尔扎， 徐斌， D.沃德-法利奥泽尔、A·库尔维尔和Y.Bengio，"生成对抗网络"，载于*Adv. Neural Inform。过程。系统。*，第2672-2680页，2014年。
5. M. Mirza和S. Osindero，"条件生成对抗网"，CoRR，vol. abs/1411.1784，2014。
6. L. Qu， S. He， J. Zhang， J. Tian， Y. Tang， and Q. Yang， "RGBD显著目标检测通过深度融合 *"，IEEE T. Image Process。*，第26卷，第5期，第2274-2285页，2017年。
7. N. Wang和X. Gong，"用于RGB-D显著物体检测的自适应融合*"，arXiv：1901.01369，2019。*
8. H. Chen和Y. Li，"用于RGB-D显著目标检测的渐进互补性感知融合网络"，IEEE *Conf。计算机.Vis. Pattern*  *Recog.*，第3051-3060页，2018年。
9. H. Chen，Y. Li和D. Su，"用于RGB-D显着目标检测的多模态融合与多尺度多路径和跨模态相互作用的多模态融合"，Pattern  *Recognit。*，第86卷，第376-385页，2019年。
10. 傅凯范，G.-P.Ji， Q. Zhao， J. Shen， and C. Zhu， "Siamese network for rgb-d 显著目标检测及超越" *arXiv* *预印本 arXiv：2008.12134*， 2020.
11. 范德佩，翟勇， A.Borji，J. Yang和L. Shao，"BBS-Net：RGBD显着对象检测与分叉骨干战略网络"，欧洲*会议。计算机.Vis.*， 2020.
12. 翟勇范，杨俊杰Borji，L. Shao， J. Han， and L. Wang， "用于 rgb-d 显著对象检测的分叉骨干策略"，arXiv*电子版*，pp.arXiv–2007， 2020.
13. W. Ji， J. Li， M. Zhang， Y. Piao， and H. Lu， "Accurate rgb-d 显著性物体检测通过协作学习"，载于 *Eur. Conf.计算机.Vis.*， 2020.
14. Y. Pang， L. Zhang， X. Zhao， and H. Lu， "用于rgb-d 显著目标检测的分层动态过滤网络"，*载于 Eur. Conf.计算机.Vis.*， 2020.
15. 张志斌， 林志军， 徐军，金伟， S.-P.卢和D.-P.Fan，"用于 rgb-d 显著目标检测的双边触手网络*"，IEEE T. 图像处理。*， 2021.
16. 周天平范，M.-M.Cheng， J. Shen， and L. Shao， "RGB-D 显著物体检测： A Survey"， *Comput.媒体*， 2021.
17. D. J. Rezende，S. Mohamed和D. Wierstra，"深生成模型中的Sto chastic反向传播和近似推理"，载于 *Int. Conf. Mach. Learn。*，第1278-1286页，2014年。
18. I.古拉贾尼， K. 库马尔， F. 艾哈迈德， A. A. 泰加， F.Visin，D. Vazquez和A. Courville，"Pixelvae：自然听觉图像的潜在变量模型"，载于*Int. Conf. Learn。代表。*， 2016.
19. X. Yan， A. Rastogi， R. Villegas， K. Sunkavalli，E.谢赫特曼哈达普Yumer和H. Lee，"MT-VAE：学习运动变换以生成多模态人类动力学"，载于*Eur. Conf。计算机.见*，第276-293页，2018年。
20. C. F. 鲍姆加特纳， K.C. Tezcan， K. Chaitanya， A.M. 霍特克， ̈

U. J. Muehlematter， K.斯科卡特， A. S. 贝克尔， O.多纳蒂和E.Konukoglu，"Phiseg：捕获医学图像分割中的不确定性"，载于*MICCAI，*第119-127页，2019年。

1. S. Kohl， B. 罗梅拉-帕雷德斯， C.迈耶， J.德 福， J.R. 莱德萨姆， K. 迈尔-海因， S.M. A. Eslami，D. Jimenez Rezende， and O. Ronneberger，"用于分割模糊图像的概率u-net"，载于 *Adv. Neural Inform。过程。系统。*，第6965-6975页，2018年。
2. J. W阿尔克多尔施Mulam和M. Hebert，"不确定的未来：使用变分自动编码器从静态图像中预测"，在*Eur. Conf。计算机.沃金什.*，第835-851页，2016年。
3. A. Abid 和 J. Y. Zou，"对比变分自动编码器增强了 salient 特征"，CoRR ，vol. abs/1902.04601，2019。
4. S.阿利亚克巴里安， F. S. 萨利赫， M. 萨尔茨曼， L.Petersson和S. Gould，"用于多样化人体运动预测的随机条件反射方案"，IEEE *Conf。计算机.Vis. Pattern*  *Recog.*，2020年。
5. S.阿利亚克巴里安， F. S. 萨利赫， M. 萨尔茨曼， L.Petersson和S. Gould，"通过cpp-vaes对良好的潜在变量进行采样：具有先前条件后验的Vaes"，arXiv*预印本arXiv：1912.08521，2019。*
6. P. Esser，E. Sutter， 和 B.Ommer，"用于条件外观和形状生成的变分u-net"，在*IEEE Conf中。计算机.Vis. Pattern*  *Recog.*，第8857-8865页，2018年。
7. 谭婧， 高磊， Y.-K.Lai和S. Xia，"用于变形3d网格模型的变分自动编码器"，IEEE *Conf。计算机.Vis. Pattern*  *Recog.*， 2018.
8. L. Yi， W. Zhao， H. Wang， M. Sung， and L. J. Guibas，"Gspn： Generative shape proposal network for 3d instance segmentation in point cloud"，载于 *Eur. Conf.* *计算机.Vis.*， 2019.
9. B. Li， Z. Sun， and Y. Guo，"Supervae： Superpixelwise variational autoencoder for salient object detection"，载于*AAAI Conf. Art。英特尔.*，第8569-8576页，2019年。
10. P.吕克，C. 库普里，S。 Chintala和J. Verbeek，"使用对抗网络的语义分割"，载于 *Adv. Neural Inform。过程。系统。 沃克什.*
11. 张某，朱某 张娟， N. Zhang， P. Li， 和 L. Wang，

"Seggan：生成对抗网络的语义分割"，2018年*IEEE第四届多媒体大数据国际会议（BigMM），*第1-5页，2018年。

1. Y. Xue，T. Xu， H. Zhang， R. Long， and X. Huang， "Segan： Adversarial network with multi-scale *l*1 loss for Medical Image segmentation"， *Neuroinformatics，*vol. 16， 06 2017.
2. H. Yu和X. Cai，"基于条件生成对抗网络的显著性检测"，第 *九届图形与图像处理国际会议*，p。 253， 04 2018.
3. Y. Tang 和 X. Wu，"使用级联卷积神经网络和对抗性学习的显著对象检测 *"，IEEE T. Multimedia，*第 21 卷，第 9 期，第 2237–2247 页，2019 年。
4. W.-C. Hung， Y.-H.蔡英文 廖玉玲林和M.-H.Yang，"用于半监督语义分割的Adve rsarial learning"，载于 *Brit. Mach. Vis. Conf。*， 2018.
5. N.Souly，C. Spampinato和M. Shah，"使用生成对抗网络的半监督语义分割"，在*Int. Conf。* *计算机.见*，第5689-5697页，第201页7。
6. J. Pan， C. Canton， K. McGuinness， N. E. O'Connor， J. Torres， E. Sayrol， and X. a. Giro-i Nieto，"Salgan： Visual s显着性预测与生成对抗网络"，在*IEEE Conf。计算机.Vis. Pattern*  *Recog.沃克什.*， 2017.
7. B. Jiang， Z. Zhou， X. Wang， and J. Tang，"cmsalgan： Rgb-d 显著对象检测与交叉视图生成对抗网络"，IEEE *T. Multimedia，2019*年。
8. P. Mukherjee， M. Sharma， M. Makwana， A. P. Singh， A. Upadhyay， A. Trivedi， B. Lall， and S. Chaudhury， "DSAL-GAN： 基于生成对抗网络的显著性预测"，CoRR ， vol. abs/1904.01215， 2019.
9. P. Krahenb̈ uhl和V.Koltun，"具有高斯边缘电位的全连接 ̈ crfs中的高效推理"，载于*Adv. Neural Inform。过程。系统。*（J.肖-泰勒Zemel，P. L. Bartlett， F. Pereira， and K. Q. Weinberger， eds.）， pp. 109–117， 2011.
10. 宋姗姗，余志祥，苗志，方俊，郑国，马云，王淑娟，"对抗域适应的多光谱显著物体检测"， *载于AAAI Conf. Art。 英特尔.*，第12023-12030页，2020年。
11. J. Zhu，T. Park，P. Isola和A. A. Efros，"使用周期一致的对抗网络进行不成对的图像到图像翻译"，在 *Int. Conf。 计算机.见*，第2242-2251页，2017年。
12. K. He， X. Zhang， S. Ren， and J. Sun， "Deep residual learning for image recognition"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，第770-778页，2016年。
13. M. Yang，K. Yu，C. Zhang，Z. Li和K.Yang，"Denseaspp用于街景中的语义分割"，IEEE *Conf。计算机.Vis. Pattern*  *Recog.*，第3684-369页，2018年第2期。
14. Y. Zhang， K. Li， K. Li， L. Wang， B. Zhong， and Y. Fu， "Image Superresolution using very deep residual channel attention networks"，载于*Eur. Conf.计算机.Vis.*， 2018.
15. A.博尔济Cheng， H. Jiang， and J. Li， "显著物体检测： Abenchmark"， *IEEE T. Image Process.*，第24卷，第12期，第5706-5722页，2015年。
16. C. 戈达尔，O.M.奥达和G. J.Brostow，"具有左右一致性的无监督单目深度估计"，在*IEEE Conf中。计算机.Vis. Pattern*  *Recog.*，第6602-6611页，2017年。
17. Y. Wang， Y. Yang， Z. Yang， L. Zhao， P. Wang， and W. Xu， "Occlusion aware unsupervised Learning of optical flow"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*， 2018.
18. S. A.C. Yohanandan，A. G. Dyer， D. Tao， and A. Song， "低分辨率灰度图像中的显著性保留 "，欧洲 *会议。 计算机.*

*螺 杆。*， 2018.

1. Z. Luo， A. Mishra， A. Achkar， J. Eichel， S. Li， and P.-M.Jodoin，"用于显著对象检测的非局部深特征"，在*IEEE Conf中。计算机.Vis. Pattern*  *Recog.*，第6609-6617页，2017年。
2. 范德平， 龚春， 曹毅， 任斌， M.-M.Cheng和A. Borji，"二进制前景地图评估的增强对齐度量"，载于*Int. Jt. Conf。阿提夫.英特尔.*，第698-704页，2018年。
3. H. Peng， B. Li， W. Xiong，W. Hu， and R. Ji，"Rgbd显著对象 detection： a benchmark and algorithms"， in *Eur. Conf.计算机.见*，第92-109页，2014年。
4. 梁芳，段磊， 马伟， Y.Qiao，Z. Cai， and L. Qing， "Stereoscopic saliency model using contrast and depth-guide-background prior"， *Neurocomputing，*vol. 275， pp. 2227–2238， 2018.
5. Y. Cheng， H. Fu， X. Wei， J. Xiao， and X. Cao， "Depth Enhanced saliency Detection Method"， in *ACM ICIMCS，*pp. 23–27， 2014.
6. J. Ren， X. Gong， L. Yu， W. Zhou， and M. Ying Yang， "Exploiting Global Priors for rgb-d saliency detection"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.沃克什.*，第25-32页，2015年。
7. C. Zhu，G. Li，W. Wang和R. Wang，"使用中心暗通道先验的创新显着物体检测"，在 *Int. Conf。 计算机.沃金 什.*， 2017.
8. R. Ju， L. Ge， W. Geng，T. Ren， and G. Wu， "Depth saliency based on anisotropic center-surround difference"，载于 *IEEE Int. Conf. Image Process。*，第1115-1119页，2014年。
9. D. 冯，N.巴恩斯，S.You和C. McCarthy，"用于RGB-D显着对象检测的本地背景外壳"，在*IEEE Conf中。计算机.Vis. Pattern*  *Recog.*，第2343-2350页，2016年。
10. R. Cong， J. Lei， C. Zhang， Q. Huang， X. Cao， and C. Hou， "基于深度置信度分析和多线索融合的立体图像显著性检测 *"，IEEE Signal Process。莱特。*，第23卷，第6期，第819-823页，2016年。
11. H. Song， Z. Liu， H. Du， G. Sun， O. Le Meur， and T. Ren，"通过多尺度判别显著性融合和引导学习进行深度感知显著目标检测和分割"，IEEE *T. Image Process。*，第26卷，第9期，第4204-4216页，2017年。
12. J. Guo， T. Ren， and J. Bei， "通过显著性演化对rgb-d 图像的显著性对象检测"，载于 *Int. Conf. Multimedia and Expo，*第 1–6 页，2016 年。
13. 牛彦耿，X. Li和F. Liu，"利用立体视觉进行显著性分析"，IEEE *Conf。计算机.Vis. Pattern*  *Recog.*， 页码.454–461， 2012.
14. N. Li， J. Ye， Y. Ji， H. Ling， and J. Yu， "Saliency detection on light field"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，第2806-2813页，2014年。
15. 范德普，M.-M.Cheng， Y. Liu， T. Li， and A. Borji，"Structure-measure： A new way to evaluation foreground maps"， in *Int. Conf.计算机.Vis.*， pp. 4548–4557， 2017.
16. M. Feng，H. Lu和E. Ding，"用于边界感知显着对象检测的细心反馈网络"，IEEE *Conf。计算机.Vis. Pattern*  *Recog.*， 2019.
17. 刘楠、韩俊、M.-H.Yang，"PiCANet：Learning Pixel-wise Contextual Attention for Saliency Detection"，载于*IEEE Conf。计算机.Vis. Pattern*  *Recog.*，第3089-3098页，2018年。
18. S. Chen， X. Tan， B. Wang， and X. Hu， "用于显著物体检测的反向注意"，载于 *《欧洲会议》。 计算机.Vis.*， 2018.
19. 王天， 张磊， 王淑， 卢汉， 杨国， X.阮和A.Borji，"全局检测，本地优化：一种显著性检测的新方法"，在*IEEE Conf上。计算机.Vis. Pattern*  *Recog.*，第3127-3135页，2018年。
20. Z. Wu，L. Su和Q. Huang，"Cascaded Partial decoder用于快速准确的显着物体检测"，在*IEEE Conf中。计算机.Vis. Pattern*  *Recog.*， 2019.
21. R. Ju， L. Ge， W. Geng，T. Ren， and G. Wu，"基于各向异性中心-环绕差的深度显著性"，载于 *IEEE Int. Conf. Image Process。*，第1115-1119页，2014年。
22. B. Settles，"主动学习文献调查"，技术代表，威斯康星大学麦迪逊分校计算机科学系，2009年。
23. W. Zhu， S. Liang， Y. Wei， and J. Sun， "来自鲁棒背景检测的显著性优化*"，IEEE Conf.计算机.Vis. Pattern*  *Recog.*，第2814-2821页，2014年。
24. C. Yang， L. Zhang， H. Lu， X. Ruan，and M. Yang， "Saliency detection via graph-based manifold ranking"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，第3166-3173页，2013年。
25. Y. Wei， F. Wen， W. Zhu， and J. Sun， "Geodesic saliency using background priors"， in *Eur. Conf.计算机.见*，第29-42页，2012年。
26. S. Gupta，R. Girshick，P.Arbelaez和J. Malik，"从RGB-D图像中学习丰富的特征，用于对象检测和分割"，在欧洲 *会议。 计算机.见*，第345-360页，2014年。
27. 秦槿，张志，黄春，高春，M.德汉，M.Jagersand，"Basnet：边界感知显著对象检测"，IEEE *Conf。*

*计算机.Vis. Pattern*  *Recog.*， 2019.

1. 吴荣、冯敏、关伟、王东、陆海、E.丁，"一种相互学习的具有交织多重视觉的显著对象检测方法"，载于*IEEE Conf。计算机.Vis. Pattern*  *Recog.*， 2019.
2. J. Wei， S. Wang， Z. Wu， C. Su，Q. Huang， and Q. Tian， "Label decpling framework for salient object detection"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，2020年。
3. D. Du， L. Wang， H. Wang， K. Zhao， and G. Wu， "Translate-to-recognition networks for rgb-d scene recognition"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，第11836-11845页，2019年。
4. C.M. Bishop，"用噪声训练 等同于 吉洪诺夫 正则化"，《 *神经计算》，*第7卷，第1期，第108-116页，1995年。
5. C. 张， S. Bengio， M. Hardt， B. 雷赫特和O. Vinyals，"理解深度学习需要重新思考泛化"，载于 *Int. Conf. Learn。Represent.*， 2017.
6. L. Wang， H. Lu， Y. Wang， M. Feng， D. Wang， B. Yin， and X. Ruan，"学习用图像级监督检测显著物体"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，第136-145页，2017年。
7. Q. Yan， L. Xu， J. Shi， and J. Jia， "Hierarchical saliency detection"，IEEE *Conf.计算机.Vis. Pattern*  *Recog.*，第1155-1162页，2013年。
8. G. Li和Y. Yu，"基于多尺度深度特征的视觉显著性*"，IEEE Conf。计算机.Vis. Pattern*  *Recog.*，第5455-5463页，2015年。
9. M.-M. Cheng， N. J. Mitra， X. Huang， and S.-M. Hu， "突出形状：图像集合中的群体显著性"，The *Vis.*  *Comput.*，第30卷，第4期，第443-453页，2014年。
10. 范德普，M.-M.程建军刘少华Gao，Q. Hou和A. Borji，"杂波中的突出物体：将突出物体检测带到地面上"，在 *Eur. Conf. 计算机.Vis.*， pp. 186–202， 2018.
11. D. Fan，W. Wang，M. Cheng和J. Shen，"将更多的注意力转移到视频显着对象检测上"， *在2019年IEEE / CVF计算机视觉和模式识别会议（CVPR）上*，第8546-8556页，2019年。
12. Y. Zhong， Y. Dai， and H. Li， "Single Mix Image的立体计算"，载于 *《欧洲计算机视觉会议（ECCV）论文集*》，2018年9月。
13. Y. Zhong， H. Li， and Y. Dai， "Open World stereo Video match with deep rnn"，载于*欧洲计算机视觉会议 （ECCV） 的* *Proceedin*gs，第 101–116 页，2018 年。
14. X. Cheng， Y. Zhong， M. Harandi， Y. Dai， X. Chang， H. Li， T. Drummond， and Z. Ge， "Hierarchical Neural Architecture search for deep stereo matching"， *Advances in Neural Information* Processing *Systems*， vol. 33， 2020.
15. Y. Zhong， P. Ji， J. Wang， Y. Dai， and H. Li， "用于静止或动态场景的无监督深 部极 流"，载于 *IEEE计算机视觉和模式识别会议论文集*，第12095-12104页，2019年。
16. J. Wang， Y. Zhong， Y. Dai， K. Zhang， P. Ji， and H. Li， "位移不变 匹配成本学习用于精确光流估计"，《 *神经信息处理系统进展*》，第33卷，2020年。

**张静** 目前是澳大利亚国立大学电气、能源和材料工程研究学院的博士生，ACRV，DATA61-CSIRO（澳大利亚堪培拉）。她的主要研究兴趣包括显著性检测，弱监督学习g，生成模型。她获得了DICTA 2017的最佳学生论文奖，APSIPA ASC 2017的最佳深度/机器学习论文奖以及IEEE CVPR 2020的最佳论文奖提名。

**范邓平于2019** 年获得南开大学博士学位 。他加入了人工智能研究所（IIAI），

2019年，发表TPAMI、CVPR、ICCV、ECCV等20篇顶级期刊和会议论文。他的研究兴趣包括计算机视觉、深度学习和显著性检测，特别是在共显著目标检测、RGB显著目标检测、RGB-D显著目标检测和视频显著性对象检测方面。

他获得了IEEE CVPR 2019的最佳论文决赛入围奖，IEEE CVPR 2020的最佳论文奖提名。他曾担任MICCAI2020 Wshp（OMIA7）的区域主席。

**戴**玉超目前是西北工业大学（NPU）电子信息学院的教授。他分别于2005年、2008年和2012年获得中国西安西北工业大学信号和信息处理专业的学士学位、硕士学位和博士学位。他是澳大利亚堪培拉澳大利亚国立大学工程研究学院的ARC DECRA研究员。

他的研究兴趣包括来自mo的结构-

tion，多视图几何，低级计算机视觉，深度学习，压缩感知和优化。他获得了IEEE CVPR 2012最佳论文奖，DSTO对图像的最佳基础贡献

DICTA 2014 年处理论文奖，最佳算法奖

CVPR 2017的NRSFM挑战赛，最佳学生论文奖

DICTA 2017，APSIPAASC 2017最佳深度/机器L论文奖，IEEE CVPR 2020最佳论文奖提名。他曾在CVPR，ICCV，ACM MM，ACCV，WACV等担任区域主席。

**赛义德·安华** 是

CSIRO-Data61，澳大利亚，澳大利亚国立大学兼职讲师。他于2019年获得澳大利亚国立大学的博士学位，以及由英国赫瑞瓦特大学，西班牙Girona大学和法国勃艮第大学联合提供的Erasmus Mundus Vision and Robotics（Vibot）硕士学位。他目前的研究兴趣包括计算机视觉，模式识别，深度学习 -

ing，机器学习，图像分析和恢复，优化，多媒体处理，医疗系统，汽车感知，汽车

导航、智能交通

**Fatemeh Saleh** 于2020年获得澳大利亚国立大学的博士学位。在攻读博士学位期间，她加入了Qualcomm AI Research，担任研究实习生，从事自我超维视频表示学习。她现在是澳大利亚机器人视觉中心和澳大利亚国立大学工程与计算机科学学院的博士后研究员。她目前的研究兴趣包括随机问题的生成e建模，视频理解，多对象跟踪 -

和场景理解。

**Sadegh Aliakbarian** 是澳大利亚国立大学的博士生。在攻读博士学位期间，他加入了高通人工智能研究公司，担任研究实习生 ，从事人体运动预测方面的工作。然后，他加入了 FiveAI， 担任研究实习生，从事对抗性机器学习方面的工作。他目前的研究兴趣包括生成模型，人体运动预测，动作分析，场景理解和对抗机器学习。

**Nick Barnes** 是澳大利亚国立大学（ANU）电气，能源和材料工程研究学院的副教授。他分别于1994年和1999年在墨尔本大学获得机器人视觉 B.Sc（荣誉）学位和工程博士学位。1999年，他是意大利热那亚热那亚大学LIRA实验室的客座研究员，并在墨尔本大学担任终身讲师，直到2003年。然后，他在NICTA，一个ICT 卓越中心，

从2003年到2016年，他担任高级首席研究员，并担任计算机视觉研究小组的执行负责人。他和

CSIRO从2016年到2019年，在那里他领导了计算机视觉研究

群。他拥有CVPR，机器人和系统科学，IROS，MICCAI Wshp计算机辅助端部复制，DICTA的最佳论文奖项和提名。他在假肢视觉视觉处理领域拥有多项专利，这为Bionic Vision Technologies公司的创建做出了贡献。他的研究兴趣包括弱监督密集预测，3D视觉和假肢视觉的计算机视觉。