arXiv:2303.00871v1 [cs.CV] 2 Mar 2023

面向图像可供性分割的贝叶斯深度学习

Lorenzo Mur-Labadia、Ruben Martinez-Cantin 和 Jose J. Guerrero

摘要——可供性是机器人学中的一个基本概念，因为它将智能体可执行的动作与其感觉运动能力及所处环境联系起来。我们提出一种新颖的贝叶斯深度网络，用于检测图像中的可供性，同时量化空间层面的偶然方差和认知方差分布。我们对 Mask-RCNN 架构进行改造，利用蒙特卡洛丢弃学习概率表征。我们的结果优于当前最先进的确定性网络。我们认为，这一提升归功于编码器中更优的概率特征空间表征，以及掩码生成过程中引入的贝叶斯变异性；后者能更好地适应物体轮廓。我们还提出一种新的基于概率的掩码质量度量，用于揭示概率实例分割模型在语义和空间上的差异。我们通过比较二值掩码而非预测的边界框，改进了现有的概率检测质量指标，从而对概率分割进行更细粒度的评估。我们发现，由于相机噪声，物体轮廓处存在偶然方差，而认知方差则出现在视觉上具有挑战性的像素处。

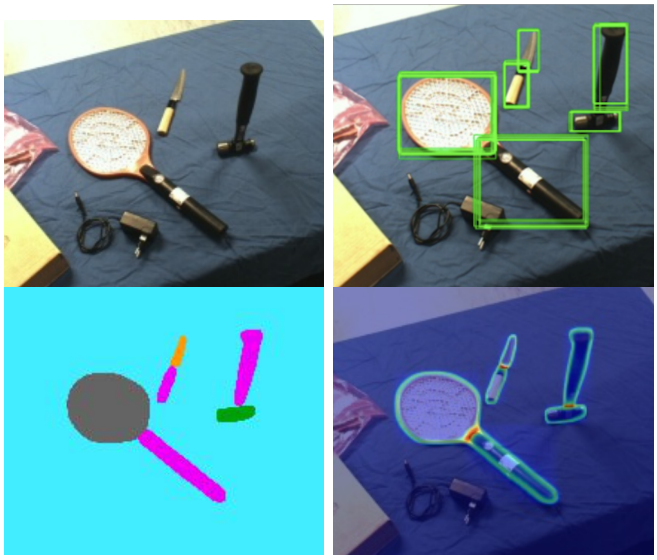

图 1. 我们根据 RGB图像（左上）检测图像中的可供性（左下）。我们的贝叶斯实例分割架构会生成掩码和类别标签的概率分布，并计算相应的认知方差和偶然方差。尽管以往方法仅限于概率检测（右上），我们的贝叶斯实例分割架构还能生成逐像素方差图（右下）。

I. 引言

心理学家J.J Gibson定义的 [1], 可供性，是环境中可用的不同行动可能性。它们关联了物体、可能采取的行动、语境以及所选行动的后续结果。从机器人学的角度来看，可供性 [2] 作为一种编码物体属性的强大工具，主要有两个原因。首先，可供性取决于感知能力、运动能力以及智能体的学习能力。其次，可供性定义中编码的行动可能性，是预测行动后果和进行未来路径规划的基础。例如，门是可打开的 或可关闭的，但这取决于它当前的状态（语境）以及智能体抓握门把手的能力（运动能力），同时也决定了物体的未来状态（行动可能性）。

用于计算机视觉的深度学习模型使机器人能够从场景中获取大量信息。然而，若要像对待其他传感器一样使用这些方法，就必须对不确定性进行估计。借助成熟的贝叶斯技术，这可以将网络的预测与先验知识或其他传感器测量值融合起来，或随时间累积信息。当前的实例分割和可供性分割模型会生成确定性预测，这些预测并非经过校准的概率，并且无法适配贝叶斯传感器融合框架。我们通过应用一种贝叶斯深度学习技术——蒙特卡洛 dropout（MC-dropout），扩展了一种广受欢迎的实例分割模型，使其能够生成概率式 检测结果。此外，我们计算了偶然方差和认知方差各自的贡献，表明空间不确定性估计能够支持更深入的可供性推理，并提取更多信息。不确定性估计对于安全关键应用非常重要。例如，如图 1 所示，我们可以发现刀柄上最不确定的部分是靠近刀刃的一端，而这也是最难以安全操作的部分。我们的主要贡献如下：

- 我们将可供性分割扩展到概率阶段，在空间层面提取逐像素的偶然方差和认知方差估计。
- 我们分析了 MC-dropout 对架构的影响，并报告了多项不确定性质量指标。
- 我们在 IIT-Aff 数据集上取得了最先进水平的性能，表明经过恰当校准的贝叶斯模型优于各自的确定性版本。
- 我们提出基于概率的掩码质量度量（PMQ），这是基于概率的检测质量度量（PDQ）的扩展，用于评估概率预测掩码。

作者隶属于Instituto de Ingenieria de Aragon（I3A）、萨拉戈萨大学（西班牙）˜ lmur@unizar.es、rmcantin@unizar.es、josechu.guerrero@unizar.es。本研究获得PID2021-125209OB-I00和TED2021-129410B-I00项目资助（MCIN/AEI/10.13039/501100011033以及NextGenerationEU/PRTR）

II. 相关工作

深度学习模型能够从场景中捕获大量信息，因此推动了机器人感知的发展。在可供性检测任务中，研究者提出了多种不同架构的方法：Nguyen 等人 [3]采用编码器-解码器，从深度潜在空间中学习可供性； [4] 使用条件随机场对特征图进行后处理，以细化类别边界；Affordance-Net [5] 通过增加反卷积层来修改 Faster R-CNN 目标检测器，从而对 RoI 特征图进行上采样； [6] 中的作者则通过采用更强大的骨干网络并引入多重对齐来提升性能。最后， [7]提出将标准目标检测模型用于可供性感知任务。他们的结果表明，与将任务拆分为先预测物体、再提取其不同可供性的做法相比，不经过中间目标检测而直接预测可供性效率更高。

另一方面，贝叶斯深度学习能够通过大型神经网络量化不确定性。由于后验分布无法通过解析方法计算，已有多种近似方法尝试解决这一问题。MC-dropout 将后验分布近似为在推理阶段启用丢弃层时多次 M 前向传播的样本分布 [8]。另一方面，深度集成 [9] 通过随机初始化神经网络参数并打乱数据集，训练由 M个模型组成的集成。Kendall 和 Gal[8]将贝叶斯神经网络引入语义分割和深度回归任务中的不确定性估计。他们指出，偶然不确定性出现在物体轮廓及远处区域，而认知不确定性则出现在具有挑战性的像素上，这些像素对应的输入样本与训练数据分布不同。对于目标检测，基于采样的方法会对M 个不同检测模型的输出求平均，以提取不确定性分布的估计，并用于多个方面：检测到的车辆边界框中的空间不确定性估计反映了激光雷达感知中的复杂环境噪声 [10]，认知方差可排除训练数据集中未出现的目标类别所产生的假阳性检测[11] 而 Bayes-OD [12] 则通过用贝叶斯方法处理非极大值抑制组件，避免信息损失。尽管大多数研究局限于目标检测，基于采样的实例分割 [13]提出将多个掩码检测结果分组，从而在检测到的物体周围生成精细的掩码。然而，其空间不确定性估计将其简化为单个值，丢失了所有像素级分布信息。

III. 贝叶斯可供性分割

基于 Caselles 等人先前的工作， [7],我们采用实例分割模型（Mask-RCNN）来执行可供性分割任务。Mask-RCNN [14], 是在 Faster R-CNN 基础上开发的基于区域的卷积神经网络，它会为每个实例返回类别标签、边界框和带有置信度分数的二值掩码。第一阶段的

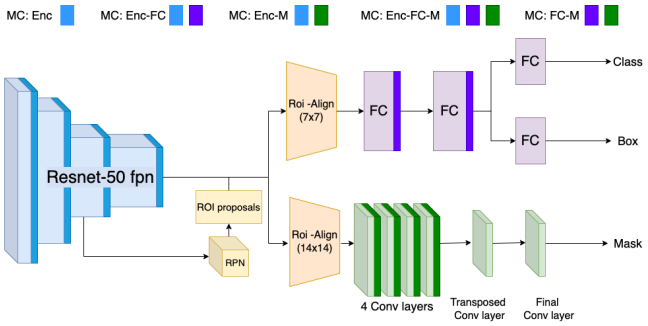

图 2. Mask-RCNN 中 MC-dropout 的不同配置。我们将丢弃层分别置于骨干网络四层模块的末端（浅蓝色）、掩码分支四个卷积层的激活层之后（绿色），以及边界框分支中两个具有 1024 个通道的全连接层的激活层之后（紫色）。

Faster R-CNN 是一个编码器，后接区域建议网络，该网络根据编码器特征生成一组区域建议（RP）。它会预测$b = \{x_1, y_1, x_2, y_2\}$ 包含物体的潜在边界框，然后通过Softmax层分别对每个边界框中的物体进行分类，输出一个 p_c $c + 1$ 类别概率向量。感兴趣区域对齐层会固定所提议感兴趣区域（RoI）的大小，并细化每个 RP 的边界框和物体类别。随后，掩码头应用若干卷积层，为每个 RoI 生成二值掩码 Υ ，并在像素级对图像进行分割。我们还保留 Sigmoid掩码向量，并对其进行归一化以获得概率掩码向量 p_Υ（x）。在推理过程中，Mask-RCNN 的每次前向传播都会生成一组独立的检测结果 $S = \{D_1, ..., D_k\}$，其中每个检测结果均为 $D_i = \{b, p_c, \Upsilon, p_\Upsilon$（$x$）$\}$。

我们的概率实例分割分为三个阶段。首先，我们对融入中间丢弃层的 Mask-RCNN 进行改造，利用 MC-dropout 生成大量D_i 检测结果。然后，我们采用合并策略，将 $\mathcal{O}_i$ 观测中的 D_i 检测结果分组，其中 N 合并后的检测结果为单个观测 $\mathcal{O}_i = \{D_1, ..., D_N\}$提供概率分布。最后，我们从概率分布中提取不确定性估计和确定性指标，以评估预测结果的质量。

A. MC-Dropout

基于采样的技术会生成 M 样本 $\mathbb{S} = \{S_1, ..., S_m\}$，其中每个样本 $S_j = \{D_1, ..., D_k\}$ 由一组检测结果 D_i 组成。通过 MC-dropout，我们将观测的真实后验分布近似为构成该观测的 N 不同检测结果的样本分布。这是变分推断的一种特殊情况：通过近似伯努利分布的变分参数 [15]，使近似后验分布与真实后验分布之间的 Kullback-Leibler散度最小化。受 Rodriguez等人 [16], 的启发，我们创建了不同的配置，并比较丢弃层不同布置方式的影响，以衡量贝叶斯组成部分在各个部分产生的影响，如图 2 所示。

B. 合并物体掩码

计算出所有检测结果 D_i后，我们将它们分组，形成场景中可供性的观测 $\mathcal{O}_i$ 。遵循 Miller 等人 [17], 的方法，我们将观测定义为一组检测结果，其中各检测结果之间具有较高的（空间和语义）亲和度。他们比较了不同亲和度指标和合并策略的性能，结果表明，基本顺序算法方案（BSAS）与检测到的掩码 Υ 之间的交并比（IoU）相结合时，性能最佳。我们只对标签类别相同的样本进行分组，从而避免合并掩码非常接近的重叠物体，并确保观测属于同一类别。与以往受限于目标检测 [18], [17], 的方法相比，在我们的概率实例分割中，比较逐像素掩码而非边界框的优势在于：两个显著重叠的掩码代表同一物体的可能性，远低于两个重叠边界框代表同一物体的可能性；尤其是在物体密集排列或形状不规则时，更是如此。

C. 空间不确定性估计

为了从概率分布中获得离散输出，我们对多个检测到的边界框坐标 b、类别概率 p_c 和掩码概率 $p_\Upsilon(x)$求平均值，从而得到单个观测到的边界框$\mathcal{O}(b)$、类别概率 $\mathcal{O}(p_c)$和掩码 $\mathcal{O}(p_\Upsilon(x))$。预测分布的方差是认知不确定性与偶然不确定性之和。认知不确定性$\sigma_e(x)$与模型知识有关，会随着数据集扩大并涵盖更多变异性而降低。它表示模型参数上的概率分布。偶然不确定性 $\sigma_a(x)$反映观测中的噪声，例如相机运动或物体边界带来的噪声。由于它是数据分布固有的，因此无法通过收集更多数据来降低。它编码了测试数据中相应输入的变异性。以往的研究 [13], [18], [17]忽略了这两种不确定性之间的差异，并将空间不确定性简化为单个标量值。遵循 [8],[19],，我们提出一种统一方法，将空间不确定性 $\sigma_{sp}(x)$计算为认知方差与偶然方差之和。我们将概率物体检测扩展到新颖的概率实例分割，因为它使用公式 1，从检测到的掩码输出 $p_\Upsilon(x)$中提取像素级的细粒度不确定性。

$$\sigma_{sp}(x) = \underbrace{\frac{1}{N}\sum_{n=1}^{N} \text{diag}(p_n) - p_n^{\otimes^2}}_{aleatoric} + \underbrace{\frac{1}{N}\sum_{n=1}^{N}(p_n - \bar{p})^{\otimes^2}}_{epistemic} \quad (1)$$

其中，$\frac{1}{N}\sum_{n=1}^{N} p_n$ 是 $\bar{p}$ 检测 $p^{\otimes^2}$ $\bar{p} =$ () 的掩码概率向量，是概率向量的均值；该运算符将向量与其转置相乘 $p^{\otimes^2} = p \cdot p^T$，而 N 是构成该观测的检测结果数量。请注意，作为未来的工作，我们可以通过将 p_n 替换为 n 检测的类别向量 p_c ，扩展此公式以获得语义不确定性。

IV. 实验设置

我们使用随机梯度下降（SGD）优化器训练模型，ResNet-50 和 ResNeXt-101 的学习率分别为 10^{-2} 和 10^{-5}。我们降低学习率，以避免梯度爆炸。按照以往的流程 [7],，我们将预测头的数量调整为功能类别数量加上背景类别 $c+1$，并使用在 COCO 数据集 [20]上预训练的骨干网络版本。两种情况下，我们都在训练开始时进行线性预热，在前 500 个训练轮次中，学习率分别从 10^{-3} 和 10^{-6}逐步升高。对于采用交并比和同标签条件的 BSAS 聚类算法，我们按照其他研究的做法 [17], 将形成一次观测所需的最小检测数量设为 2，Soft max 分数设为 0.5，空间亲和度阈值设为 0.5。

A. 数据集

我们在 IIT-AFF 数据集 [4], 上进行实验。这是一个机器人数据集，包含 8,835 张真实场景图像。与受控场景或合成数据集相比， [21], [22], IIT-Aff 数据集包含真实环境中的标注，并提供上下文信息，且具有很大的变异性。它涵盖 7 种不同的功能类别（容纳、切割、展示、发动、抓握、击打、捶打、支撑、*w- grasp*），代表 10 种不同物体（瓶子、碗、杯子、电钻、锤子、刀、显示器、平底锅、球拍、锅铲）的常见操作能力。

B. 指标

遵循以往的流程， [7], [6], [5], 我们使用F_β^w 分数 [23] 来比较确定性预测与以往基线方法的性能，并展示我们相较于最先进水平的改进。该指标评估前景图，并对像素误差进行加权，以修正三个假设：依赖性、插值和同等重要性。在公式中， P_w 和 R_w 分别对应加权精确率和召回率，超参数 $\beta = 1$、 $\sigma^2 = 4$、 $\alpha = (0.5)\ 5$ 与以往工作中的设置相同。

$$F_\beta^w = (1+\beta^2) \cdot \frac{P_w \times R_w}{\beta^2 \cdot P_w + R_w} \quad (2)$$

PDQ 分数 [24] 用于评估概率目标检测器的不确定性估计。我们提出的新型概率掩码质量（PMQ）指标用于衡量概率实例分割任务的质量。我们通过比较掩码输出而非边界框预测，将 PDQ 适配到这一新任务。PMQ 分别根据类别标签向量 p_c 和掩码 Υ计算检测器的标签 Q_l质量与空间 Q_s 质量。与平均精度均值（mAP）分数 [20],相比，PMQ 不依赖交并比阈值来筛选检测结果数量，而是利用空间与标签不确定性质量，计算检测结果与真实标注之间的最优匹配。

	ED RGB [3]	BB 卷积神经网络 [4]	AN VGG 16 [5]	AN SE154 [6]	Mask R-CNN R50 [7]	贝叶斯 Mask-RCNN R50	Mask R-CNN Rx101	贝叶斯 Mask-RCNN Rx101
容纳	66.4	75.6	79.6	81.9	83.6	86.4	85.8	**86.5**
cut	60.7	69.9	75.7	79.8	84.7	82.0	**86.4**	84.6
展示	55.4	72.0	77.8	82.5	86.3	89.5	89.9	**90.6**
发动	56.3	72.8	77.5	83.2	88.9	89.4	90.5	**90.9**
抓握	59.0	63.7	68.5	77.1	71.1	77.1	77.5	**78.2**
hit	60.8	66.6	70.8	79.2	92.3	93.4	94.3	**95.4**
捶打	54.3	64.1	69.6	78.2	**81.8**	80.6	80.5	80.7
支撑	55.4	65.0	69.8	78.9	86.7	85.5	85.4	**88.7**
w-抓取	50.7	67.3	71.0	80.6	83.7	85.8	84.7	**86.6**
平均值	57.6	68.6	73.4	80.2	84.4	85.7	86.1	**86.9**

表I 数据集上的可供性感知（F_β^w* 100 分数（个））。我们将贝叶斯模型和确定性模型与以往工作进行了比较，并取得了最佳性能。我们将性能提升归因于贝叶斯模型更好的泛化能力，以及潜在空间中 ResNeXt-101 编码器更高的建模能力。

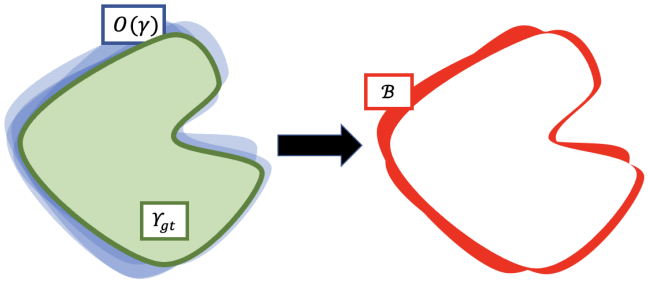

图3。我们扩展的 PMQ 将背景评估区域$\mathcal{B}$ 重新定义为分组检测到的掩码 O（Υ）与真实掩膜 Υ_{gt}之间的差异。

$$Q_s(\Upsilon_{gt}, \Upsilon_h) = \exp(-(L_{FG}(\Upsilon_{gt}, \Upsilon_h)) + (L_{BG}(\Upsilon_{gt}, \Upsilon_h)) \quad (3)$$

我们提出的新型空间质量 Q_s 是掩码层级前景损失 L_{FG} 与背景损失 L_{BG}的指数值之和，如公式 3 所示。当真实掩膜 Υ_{gt} 中的所有像素的空间概率都为 1，且其余像素的概率均为 0 时，该指标得分为 1。前景损失 L_{FG}通过将像素概率取负对数来计算，会惩罚真实分割区域内概率较低的像素。与 PDQ 中使用边界框计算高斯热图、从而丢失边界检测结果的做法不同，我们使用掩码 $\mathcal{O}$（Υ）生成细粒度热图Υ_h。分配给每个像素 x 的值，是该像素被检测到 $\Upsilon > 0.5$ 的次数除以样本总数 M。因此，如图 3 所示，背景评估 $\mathcal{B}$ 是将分组后的检测到的掩码 $\mathcal{O}$（Υ）与真实掩膜 Υ_{gt}进行比较所得的差异。

最后，我们展示两个不确定性误差指标——平均校准误差（ACE）和稀疏化误差曲线下面积（AUSE），用于报告贝叶斯模型的校准情况。尽管期望校准误差 [25], [26] 是分割模型校准中事实上的指标，我们仍遵循 [27] 报告 ACE；该指标针对目标检测模型进行了调整。ACE 分数计算每个分箱的准确率与置信度之间的绝对差值，对每个分箱赋予相同权重，从而避免受到目标检测器典型的低置信度预测的偏差影响。最后，作为标准校准指标，我们还以布里尔分数 [29] 报告稀疏化误差曲线下面积（AUSE） [28]，它提供了不确定性的相对度量。

V. 结果

首先，我们将模型与以往工作进行了比较，结果表明，当贝叶斯模型经过正确校准时，其性能优于确定性版本。随后，我们对丢弃层的分布及其对 F_β^w、PMQ、ACE 和 AUSE 分数的影响进行了比较研究，并展示了指标随样本数量 M 的变化。最后，我们展示了一些定性结果，以及贝叶斯模型生成的方差图。

A. 可供性感知

我们在表I中将我们在 IIT-Aff 数据集上获得的结果与先前工作的 F_β^w 分数进行了比较。对于 ResNeXt-101 MC Enc-FC 配置，我们取得了最佳表现，F_β^w达到 86.9%，比此前的最先进水平提高了 +2.5 个百分点。除切割 和捶打 类别外，我们在每个功能类别上都取得了最佳表现。我们将这一提升归因于 ResNeXt-101 编码器更强的建模能力，以及通过贝叶斯累积 M 预测对掩码轮廓进行更精细的优化。

我们在表 II 中使用新颖的 PMQ 评估概率可供性分割，该指标揭示了检测器的语义性能和空间性能。概率版本与确定性版本在 PMQ 指标上的差异，大于 F_β^w score 所报告的差异。例如，确定性 R50 的得分为 84.4 %F_β^w，R50 Enc-FC 的得分为 85.9 F_β^w （表I），而它们的 PMQ 分别为 10.8 和 34.7（表 II）。新颖的 PMQ 突显了贝叶斯检测模型相较于确定性基线的优势：对 M 掩码取平均值后，能够更好地近似真实标注，并

	F_w^b (%) (↑) 确定性 ($M = 1$)	F_w^b (%) (↑) 贝叶斯 ($M = 24$)	AUSE (↓)	ACE (↓)	PMQ (%)(↑)	$pPMQ$ (%)(↑)	Q_s (%)(↑)	Q_l (%)(↑)	FG (%)(↑)	BG (%)(↑)
R50 基线	84.4	-	-	-	10.8	13.7	4.2	94.2	28.7	16.5
R50 MC 编码器	85.6	84.9	0.150	0.00410	35.7	46.6	28.9	94.3	58.5	52.3
R50 MC 编码器-全连接层	85.7	85.9	0.132	0.00357	34.7	57.2	40.5	95.2	56.5	71.5
R50 MC 编码器-M	85.3	84.1	0.133	0.00385	31.3	49.8	33.4	92.7	58.9	58.2
R50 MC Enc-FC-M	84.5	85.7	0.134	0.00356	35.5	55.1	40.6	93.4	**64.6**	63.3
R50 MC 全连接层-M	83.1	81.8	0.188	0.00453	30.8	45.4	29.6	93.6	56.5	53.3
Rx101 基线	86.1	-	-	-	12.5	15.6	4.9	**97.6**	32.9	17.8
Rx101 MC 编码器	**86.8**	86.3	0.134	0.00359	36.7	44.8	27.6	95.0	62.5	44.4
Rx101 MC 编码器-FC	86.6	**86.9**	0.129	**0.00309**	**37.7**	**66.8**	**52.8**	95.5	61.2	**78.4**
Rx101 MC 编码器-M	85.6	84.4	0.128	0.00377	32.6	53.9	38.8	93.1	60.2	60.5
Rx101 MC 编码器-FC-M	86.4	86.6	**0.122**	0.00328	36.7	55.5	41.3	93.0	**64.6**	59.6
Rx101 MC FC-M	85.5	84.3	0.154	0.00410	31.1	48.4	34.4	94.2	61.4	54.1

表 II 不确定性误差（AUSE、ACE）与可供性感知指标（F_β^w、PMQ）的比较。 Q_s = 空间质量，Q_l =标签质量，FG = 前景质量（$exp(-L_{FG})$），BG = 背景质量（$exp(-L_{BG})$）

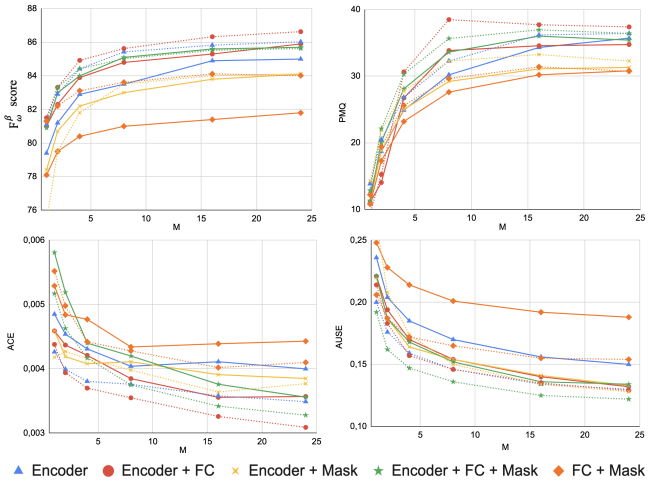

图 4. 不同配置下可供性感知指标和不确定性指标随前向传播 M 的变化。左上： F_w^β。右上：PMQ。左下：ACE。右下：AUSE。随着样本 M 数量增加，各项指标逐渐收敛。虚线表示 Rx101 版本，实线表示 R50 版本。

获得更高的 Q_s 质量。此外， Q_s 和 Q_l质量表明，我们的可供性分割模型更有可能在物体所在的位置而非我们能用该物体做什么方面出错。这表明空间不确定性对可供性分割的重要性，并指出这一领域仍有潜在的改进空间。

B. 模型分析

首先，我们使用网格搜索算法为两个编码器设置丢弃率 $p_d = 0.5$ 。过高的丢弃率（如 $p_d = 0.7$ ）会引入过多变异性，而$p_d = 0.1$ 则会使模型非常接近确定性行为。

a) 特征提取模块的重要性：编码器在潜在空间中的建模能力表现出显著差异。R50 基线（43.9 M 个可训练参数）达到 84.4 % F_β^w，而 Rx101 基线得分为 86.1 % F_β^w（107.1 M 个参数），如表I所示。PMQ 指标也反映了这一差异：在所有情况下，Rx101 配置的表现都优于各自对应的 R50 版本。例如，Rx101 MC Enc-FC 的 PMQ 为 37.7，pPMQ 为 66.8，52.8Q_s 和 95.5 Q_l；而相同的 R50 版本则为 34.7 PMQ、57.2 pPMQ、40.5 Q_s 和 95.2 Q_l。同样，如表 II 和图 4 所示，Rx101 模型的不确定性误差指标更低，表明其校准效果更好，且其估计概率更接近实际值。

b) 丢弃层分布的影响：丢弃层的位置会因丢弃效果不同而决定模型的性能。当我们将丢弃置于骨干网络中时，它会对网络中最大的一部分进行正则化，并促使骨干网络在潜在空间中学习概率表示，从而得到校准效果更好、误差指标更低的模型（表 II）。在骨干网络中启用丢弃的模型（MC Enc、MC Enc+FC、MC Enc+M、MC Enc+FC+M）性能优于在此阶段未启用丢弃的配置（MC FC-M）。FC-M 版本在所有指标上表现最差（Rx101 骨干网络上的 84.3 % F_β^w、0.154 AUSE、0.00410 ACE、31.1 % PMQ），因为架构中占比最大的部分（骨干网络）仍具有确定性，因此模型学到的是确定性表示；但在架构的末端，我们又要求其产生变异性，导致特征空间中的不确定性未能得到适当建模。

另一方面，在 FC-Mask 层中加入丢弃，可引入贝叶斯预测所需的变异性，进而实现不确定性估计，并展现相同配置下贝叶斯行为与确定性行为之间的差异。结果表明，在 FC 层中使用丢弃，足以将变异性传递至最终的掩码预测。在 FC 分支中使用丢弃，会使类别标签和边界框 b的定位产生变异性，而这种变异性随后会传递至掩码分支。例如，R50 MC 编码器-全连接层 M 的得分从 84.5 提高到 85.7（% F_w^β）。此外，仅在掩码分支的卷积层中设置丢弃层，会对 F_w^β 和 PMQ 得分产生不利影响，如表 II 所示。合理配置丢弃能更好地校准贝叶斯模型，并将这一改进体现在 PMQ 和 F_w^β 得分上。

c) 随样本数量的变化 M：我们在图 4 中展示了 %F_w^β、PMQ、AUSE 和

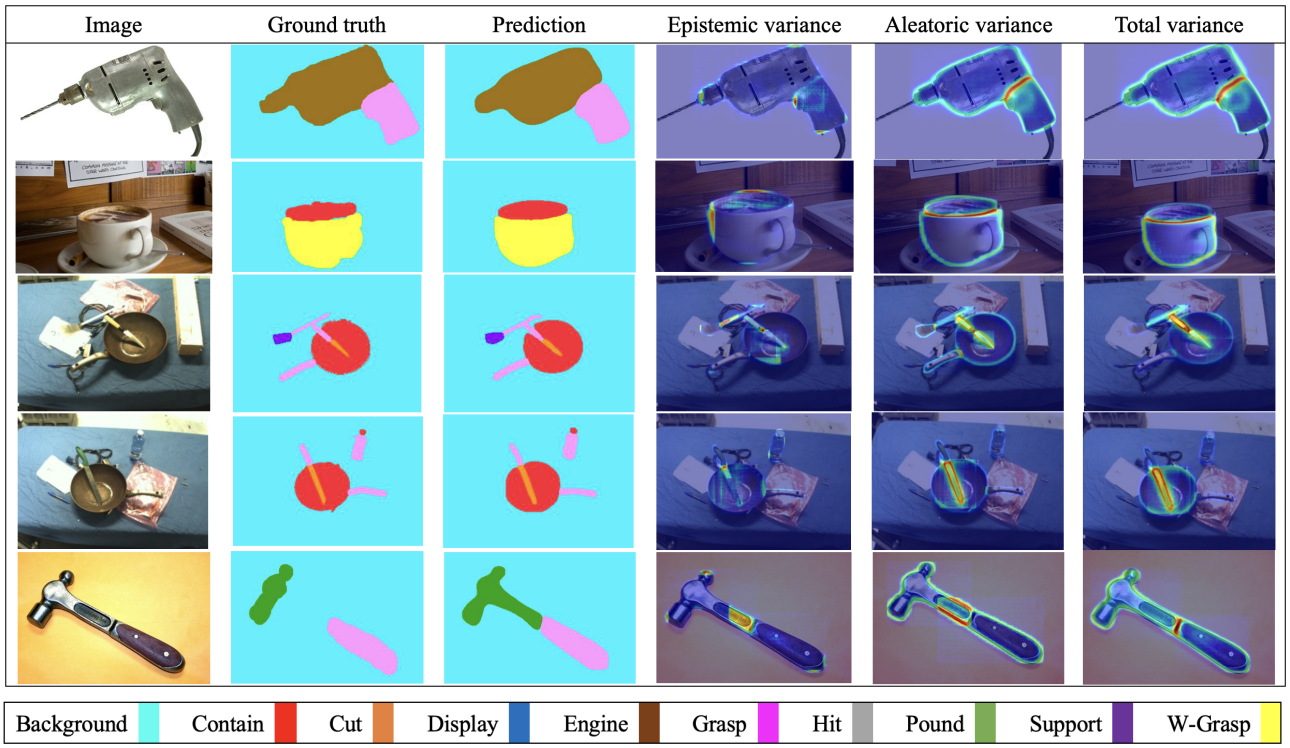

图 5. 定性结果。贝叶斯 Rx101 预测的可供性掩码非常平滑，与真实标注高度吻合。由于存在遮挡、光照伪影或模糊区域，具有挑战性的像素会出现认知方差；而物体轮廓处的噪声较多，因此会出现偶然方差。

ACE 指标随平均样本数 M 的变化。首先，我们观察到，随着平均模型数 M 的增加，预测效果有所提升，但当 $M > 16$ 时，提升速度会减缓。在贝叶斯模型中，随着 M 增加，预测的平均值能够弥补单个样本以及对 p_d 个神经元进行随机丢弃所导致的损失。根据其余结果，Rx101 版本（虚线）的指标变化优于各自的 R50（实线）配置。

C. 定性结果

我们在图 5 中展示了部分预测得到的可供性掩码，以及认知不确定性和偶然不确定性的分布。观察推断出的掩码可以发现，它们与真实标注高度吻合，且没有出现任何像素伪影。由于对每次检测的 M 样本进行平均，最终掩码非常平滑，并覆盖整个可供性区域。该方法适用于任何类型的场景，无论物体数量多少，也不受与观察者距离的影响。正如 Q_l 与Q_s 之间的差异所示，不存在语义错误的检测，但模型在掩码的空间分布方面遇到了更多困难。空间方差图也向观察者展示了相关信息，并为理解模型推理提供了洞见。由于噪声更多且模型难以生成清晰的轮廓，偶然方差出现在物体的轮廓处，并在两个轮廓的交汇处取得最高值。认知方差出现在分布外样本或视觉上具有挑战性的像素中。例如，最后一行的锤子手柄展示了一个可供性分割失败的示例，尽管认知不确定性显示出更高的数值，这是由于难以确定该区域是否可用于捶打。

六、结论

我们提出了一种对实例分割模型 Mask-RCNN 的新颖扩展，用于执行贝叶斯概率可供性分割任务。结果表明，丢弃层的不同配置会产生不同的校准效果和性能，其性能超越了确定性版本，并在 IIT-Aff 数据集上达到最先进水平。在骨干网络中设置丢弃层，促使模型学习概率潜在空间，从而更好地建模不确定性估计；而解码器阶段则产生变异性，以更好地细化掩码轮廓。此外，我们还获得了像素级的空间认知方差和偶然方差。我们改进了以往的工作，这些工作要么将不确定性简化为单一数值，要么不区分偶然方差和认知方差的贡献。我们还提出了一种 PMQ 指标，用于比较概率掩码并计算空间质量。我们希望这项工作能启发未来的可供性推理研究。

参考文献

[1] J. J. Gibson，《视觉知觉的生态学方法：经典版》。Psychology Press，2014年。[2] L. Jamone、E. Ugur、A. Cangelosi、L. Fadiga、A. Bernardino、J. Piater 和 J. Santos-Victor，“心理学、神经科学与机器人学中的可供性：综述”，IEEE认知与发展系统汇刊，第10卷，第1期，第4–25页，2016年。[3] A. Nguyen、D. Kanoulas、D. G. Caldwell 和 N. G. Tsagarakis，“利用卷积神经网络检测物体可供性”，载于2016年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2016年，第2765–2770页。[4] A. Nguyen、D. Kanoulas、D. G. Caldwell 和 N. Tsagarakis，“基于卷积神经网络和稠密条件随机场的物体可供性检测”，载于2017年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2017年，第5908–5915页。[5] T.-T. Do、A. Nguyen 和 I. Reid，“AffordanceNet：一种用于物体可供性检测的端到端深度学习方法”，载于2018年IEEE机器人与自动化国际会议（ICRA）。IEEE，2018年，第5882–5889页。[6] C. N. D. Minh、S. Z. Gilani、S. M. S. Islam 和 D. Suter，“学习可供性分割：一项探索性研究”，载于2020年数字图像计算：技术与应用（DICTA）。IEEE，2020年，第1–8页。[7] H. Caselles-Dupr´e、M. Garcia-Ortiz 和 D. Filliat，“标准物体分割模型是否足以学习可供性分割？”，arXiv预印本 arXiv:2107.02095，2021年。[8] A. Kendall 和 Y. Gal，“贝叶斯深度学习用于计算机视觉时需要考虑哪些不确定性？”，arXiv预印本 arXiv:1703 04977，2017年。.[9] F. K. Gustafsson、M. Danelljan 和 T. B. Schon，“评估用于鲁棒计算机视觉的可扩展贝叶斯深度学习方法”，载于 IEEE/CVF 计算机视觉与模式识别会议研讨会论文集，2020年，第318–319页。[10] D. Feng、Z. Wang、Y. Zhou、L. Rosenbaum、F. Timm、K. Dietmayer、M. Tomizuka 和 W. Zhan，“标签并不完美：推断目标检测中的空间不确定性”，IEEE智能交通系统汇刊，2021年。[11] D. Miller、N. S¨underhauf、M. Milford 和 F. Dayoub，“利用不确定性识别视觉目标检测中的开放集错误”，IEEE机器人与自动化快报，第7卷，第1期，第215–222页，2021年。[12] A. Harakeh、M. Smart 和 S. L. Waslander，“BayesOD：一种用于深度目标检测器不确定性估计的贝叶斯方法”，载于2020年IEEE机器人与自动化国际会议（ICRA）。IEEE，2020年，第87–93页。[13] D. Morrison、A. Milan 和 E. Antonakos，“利用丢弃采样进行不确定性感知实例分割”，载于机器人视觉概率目标检测挑战赛论文集（CVPR 2019研讨会），美国加利福尼亚州长滩，2019年，第16–20页。[14] K. He、G. Gkioxari、P. Doll´ar 和 R. Girshick，“Mask R-CNN”，载于 IEEE计算机视觉国际会议论文集，2017年，第2961–2969页。[15] Y. Gal 和 Z. Ghahramani，“将丢弃作为贝叶斯近似：在深度学习中表征模型不确定性”，载于国际机器学习会议。PMLR，2016年，第1050–1059页。[16] J. Rodr´ıguez-Puigvert 、 R. Mart´ınez-Cant´ın 和 J. Civera，“用于监督学习单视图深度的贝叶斯深度神经网络”，IEEE机器人与自动化快报，第7卷，2572，2022年2，2565 no. pp. –。.[17] D. Miller、F. Dayoub、M. Milford 和 N. S¨underhauf，“评估目标检测中基于采样的不确定性技术的合并策略”，载于2019年国际机器人与自动化会议（ICRA）。IEEE，2019年，第2348–2354页。[18] D. Miller、L. Nicholson、F. Dayoub 和 N. S¨underhauf，“在开放集条件下利用丢弃采样实现鲁棒目标检测”，载于2018年IEEE机器人与自动化国际会议（ICRA）。IEEE，2018年，第3243–3249页。[19] Y. Kwon、J.-H. Won、B. J. Kim 和 M. C. Paik，“利用贝叶斯神经网络进行分类不确定性量化：在生物医学图像分割中的应用”，《计算统计与数据分析》，第142卷，文章编号106816，2020年。[20] T.-Y. Lin、M. Maire、S. Belongie、J. Hays、P. Perona、D. Ramanan、P. Doll´ar 和 C. L. Zitnick，“Microsoft COCO：上下文中的常见物体”，载于欧洲计算机视觉会议。Springer，2014年，第740–755页。[21] S. Thermos、G. T. Papadopoulos、P. Daras 和 G. Potamianos，“深度可供性基础的感觉运动物体识别”，载于*IEEE*计算机视觉与模式识别会议论文集，2017年，第 6167–6175 页。[22] A. Myers、C. L. Teo、C. Ferm¨uller 和 Y. Aloimonos，“基于几何特征的工具部件可供性检测”，载于*2015*年 *IEEE*国际机器人与自动化会议（*ICRA*）。IEEE，2015年，第 1374–1381 页。[23] R. Margolin、L. Zelnik-Manor 和 A. Tal，“如何评估前景图？”，载于*IEEE*计算机视觉与模式识别会议论文集，2014年，第 248–255 页。[24] D. Hall、F. Dayoub、J. Skinner、H. Zhang、D. Miller、P. Corke、G. Carneiro、A. Angelova 和 N. S¨underhauf，“概率物体检测：定义与评估”，载于*IEEE/CVF*冬季计算机视觉应用会议论文集，2020年，第 1031–1040 页。[25] C. Guo、G. Pleiss、Y. Sun 和 K. Q. Weinberger，“现代神经网络的校准”，载于国际机器学习会议。PMLR，2017年，第 1321–1330 页。[26] M. P. Naeini、G. Cooper 和 M. Hauskrecht，“使用贝叶斯分箱获得校准良好的概率”，载于第二十九届*AAAI*人工智能会议，2015年。[27] L. Neumann、A. Zisserman 和 A. Vedaldi，“松弛softmax：用于安全行人检测的高效置信度自动校准”，2018年。[28] E. Ilg、O. Cicek、S. Galesso、A. Klein、O. Makansi、F. Hutter 和 T. Brox，“光流的不确定性估计与多假设网络”，载于欧洲计算机视觉会议（*ECCV*）论文集，2018年，第 652–667 页。[29] G. W. Brier 等，“以概率形式表达的预测验证”，《月度天气评论》，第 78 卷，第 1 期，第 1–3 页，1950年。