

用于分布外的不稳定性分布

JinYoung Kim¹, DaeUng Jo², Kimin Yun^{3,4}, Jeonghyo Song¹, Youngjoon Yoo^{1,*}

¹ Department of Artificial Intelligence, Chung-Ang University, Seoul, Korea

² School of Electronics Engineering, Kyungpook National University, Daegu, Korea

³ Visual Intelligence Lab., ETRI, Daejeon, Korea

⁴ University of Science and Technology (UST), Daejeon, Korea

{ barraki7226, thd9592s } @cau.ac.kr, daeung.jo@knu.ac.kr, kimin.yun@etri.re.kr, yjyoo3312@gmail.com *

Abstract

深度神经网络的不稳定性是一种广泛使用的方法,用于分布外(OoD)检测。然而,一些方法如蒙特卡罗(MC) Dropout通常只关注模型或数据不稳定性中的一方面,未能解决OoD检测的盲目性。我们提出了自由能后验,这是一种新的框架,结合建模分布不稳定性,通过自由能解决OoD和分布外检测。我们的方法引入了以下改进:(1)一种基于Beta分布量化的自由能密度函数,使得在模糊或未建模区域附近检测粒度不稳定性;(2)一种集成在后验中的损失,允许直接检测分布外不稳定性,而无需随机采样。通过我们的方法,我们分支(RPL)框架结合,所提出的方法超越了事后能量函数,使得我们能够通过利用Beta分布的方差检测OoD区域,而无需在模型上有意且计算上高效的检测不稳定性。我们在具有挑战性的世界基准集中验证了我方法的有效性,包括Fishyscapes、RoadAnomaly和Segment-Me-If-You-Can。

1. 简介

在诸如自动驾驶等安全应用中,检测分割模型不稳定性至关重要,需要检测不稳定性以检测不可见或分布外(OoD)区域[7]。在现实情况下,一个核心挑战是数据分布的异常输入,这些输入通常导致模型自信的失败。一种广泛使用的方法是蒙特卡罗(MC) Dropout [5],它通过多次随机前向模型不稳定性。然而,它存在高的计算成本,可能无法充分地反映不稳定性[12]。

为了克服这些限制,最近的研究集中于分布不稳定性,但不稳定性旨在捕捉来自未建模或模糊数据的模糊性,而不是模型随机性[11]。后验[4]和数据深度[13]引入了使用Dirichlet和Beta分布模型不稳定性,提供了更丰富和更易解释的置信度

。在分布分割中,Siddharth等人[1]提出了一种扩展到像素级,而不仅仅是OoD检测的粒度不稳定性。然而,一些方法通常需要额外的采样、后验或OoD区域手工制定规则,这可能阻碍可扩展性和泛化。一种工作利用基于自由能的置信度量进行OoD检测。FlowEneDet [6]引入了一种基于流的密度函数,使用自由能作为似然的替代,而无需监督训练在域内和域外。在此基础上,Residual Pattern Learning (RPL) [10]通过使用比自由能高的域,实现了最先的OoD检测性能。然而,RPL依赖于固定的自由能目标和后期检测,限制了其活性,尤其是在早期阶段。

了解一些挑战,我们提出了一种合适的范例的框架。具体而言,我们提出了一种分割模型-一种可解释的基于流的后验相结合,能够在每个像素级-一种Beta分布。通过一些Beta函数,我们方差计算分布不稳定性直接度量,通过新的Beta不稳定性交叉熵和能量(BUCE)损失注入到模型过程中。一种基于方差的信心使模型能够抑制模糊的OoD区域,同时抑制已知域中由于自信的失败,而模型更稳定和模型上有依据的不稳定性感知。图1展示了整个框架。以往的工作相比,我们的方法解决了端到端的不稳定性检测和OoD分割,无需依赖采样或手工定义的规则,提供了一种更合理和高效的解方方案。我们通过在现实世界的OoD基准上运行大量实验验证了所提出框架的有效性,包括Fishyscapes [2] (Static和LostAndFound)、RoadAnomaly [9]和Segment-Me-If-You-Can (SMIYC) [3]。

2. 知识

2.1. 不稳定性后验

在数据深度中,不稳定性被分解成三部分:数据、分布和模型不稳定性。部分反映了输入、在表示和模型的性质。通过方程1中的后验框架,可以在一次前向模型中整合不稳定性。

*Corresponding author

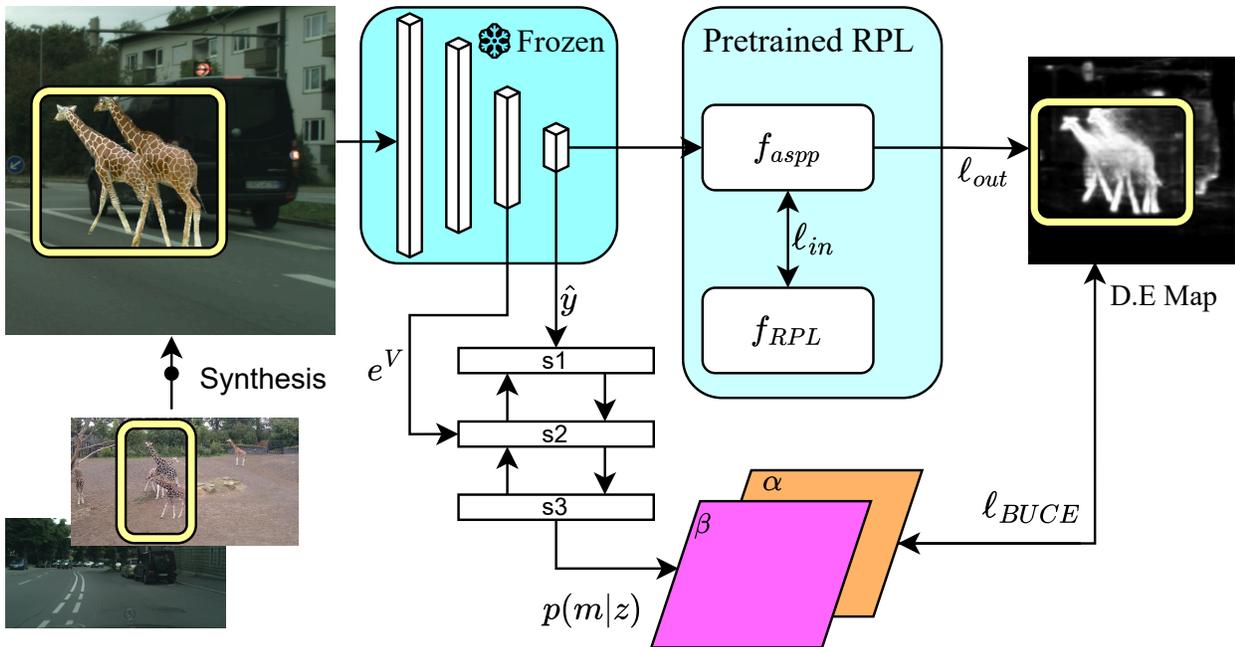


Figure 1. 所提出架构的描述。常用曝光 (OE) 本通合成点 (Cityscapes) 和点 (COCO) 像生成, 然后通过分割骨干网输入分支: RPL 架构和基于流的模型 (FlowEneDet)。流模型使用合成 \$(s_1, s_2, s_3)\$ 计算自由能表示, 其中 \$e^v\$ 表示ReLU激活的特征, \$\hat{y}\$ 通过conv1提取。这些表示被Beta后 \$(\alpha, \beta)\$。然后推出一微分 (D.E) 映射, 替换 RPL 的 \$\ell_{in}\$ 中的自由能, 以更准确的常用。

$$\mathbb{P}(\omega_c | x^*, \mathcal{D}) = \iint \underbrace{\mathbb{P}(\omega_c | \mu)}_{\text{Data}} \underbrace{\mathbb{P}(\mu | x^*, \theta)}_{\text{Distributional}} \underbrace{\mathbb{P}(\theta | \mathcal{D})}_{\text{Model}} d\mu d\theta. \quad (1)$$

这里, \$\omega_c\$ 是数据, \$x^*\$ 是输入, \$\mathcal{D}\$ 是数据集。后使用狄利克雷 (Dirichlet) 形式近似此分布, 允许进行具有置信水平的逐点。

为了提高可操作性, 使用一化流输入在空, 其中狄利克雷 \$\alpha\$ 定义如下:

$$\mathbb{E}_{p \sim \text{Dir}(\alpha^{(i)})}[p_c] = \frac{\beta_c^{\text{prior}} + N \cdot P(c | z^{(i)}; \phi) \cdot P(z^{(i)}; \phi)}{\sum_c \beta_c^{\text{prior}} + N \cdot P(z^{(i)}; \phi)}. \quad (2)$$

这里, \$P(c | z^{(i)}; \phi)\$ 表示在特征 \$z^{(i)}\$ 的置信率, 而 \$N\$ 是一置信度放大因子。

在二元分割中, Dirichlet 分布化 Beta 分布, 其中 \$\alpha\$ 和 \$\beta\$ 代表包含和不含的数据。期望的含率是:

$$\mathbb{E}[p_{in}] = \frac{P_{in}(z^{(i)}; \phi)}{P_{in}(z^{(i)}; \phi) + P_{out}(z^{(i)}; \phi)} \quad (3)$$

相的不定性通过塔方差行量化:

$$\text{Var}(p) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \quad (4)$$

2.2. 基于能量的流模型

我采用基于流的密度模型, 其中似乎被解一自由能。一化流提供了一活的机制, 可以通过一系列可逆的数据分布映射到易于处理的空, 而更精确且具有影响力的密度。

定通一量 \$z\$, 我在高斯分布下其密度建模。相的似然函数是:

$$\log \mathcal{N}(u | \mu, \Sigma) = \sum_{d=1}^D \text{diag}(U)_d - \frac{1}{2} \|U(z - \mu)\|_2^2, \quad (5)$$

其中, \$U\$ 是逆方差矩阵 \$\Sigma^{-1}\$ 的 Cholesky 分解, \$\mu\$ 是高斯均值。二次偏均的情况, 而角考虑到了每度中到的尺度。在点下, 低能域于熟悉的、高置信度的输入, 而高能域表明超出分布 (OoD) 的本。

解分割任务尤其有用, 因一些任务需要密集的、空分辨的不定性。通特征上的能量, 可以不定或模糊的域。

我通斯算件似然, 以支持判:

$$p_\theta(m | z) = \frac{p_\theta(z | m)p(m)}{\sum_{d=1}^D p_\theta(z | m = d)}, \quad (6)$$

其中, $p_\theta(z|m)$ 是在 m 下 z 的似然, 而 $p(m)$ 是先验。这使得基于相似似然行每像素密度建模和不确定性可能。

而言, 基于能量的流模型通常提供一基于角式似然的角理解不常, 而充了后。

3. 提出的方法

在本中, 我出了一用于分割中棒的分布外 (OoD) 的方法。整个架由主要成部分成: (1) 基于流的后, 利用 Beta 分布的不定性; (2) 基于自由能的分割模型, 具有一差分支 (RPL), 通一的目整合不稳定性感知失函。

3.1. 不稳定性表述

1 展示了架。定一入像, 模型首先通一准的分割骨干算自由能分。然后, 些特征被一基于流的, 以 Beta 分布的 α 和 β 。具, 定像素 i 的在表示 $z^{(i)}$, 流出一率向量 $z^{(i)} = [z_{in}, z_{out}]$, 被 Beta 。

$$\begin{aligned}\alpha^{(i)} &= 1 + \log(1 + e^{z_{in}^{(i)}}), \\ \beta^{(i)} &= 1 + \log(1 + e^{z_{out}^{(i)}}).\end{aligned}\quad (7)$$

些中, 我通 Beta 分布的均算期望的点置信分 $\bar{p}^{(i)}$:

$$\bar{p}^{(i)} = \mathbb{E}[p^{(i)}] = \frac{\alpha^{(i)}}{\alpha^{(i)} + \beta^{(i)}}.\quad (8)$$

像素 i 的硬 $y^{(i)}$ 通点和外点分分行化或使用 $\arg \max$ 。

$$y^{(i)} = \begin{cases} 1, & \text{if } \bar{p}^{(i)} \geq \tau \\ 0, & \text{otherwise.} \end{cases}\quad (9)$$

了捕捉分布不稳定性, 我的方法利用基于流的后每像素的 Beta 分布。具, 定一自由能表示中衍生的入特征 z , 后出逐像素的 α 和 β :

$$\begin{aligned}\alpha(\omega) &= 1 + \log P(z(\omega)|c), \\ \beta(\omega) &= 1 + \log Q(z(\omega)|c).\end{aligned}\quad (10)$$

里, P 和 Q 表示基于流的、以上下文特征 c 的率映射, ω 表示空位置。然后我使用 Beta 分布算方差:

$$\text{Var}(p(\omega)) = \frac{\alpha(\omega)\beta(\omega)}{(\alpha(\omega) + \beta(\omega))^2(\alpha(\omega) + \beta(\omega) + 1)}.\quad (11)$$

最后, 了鼓模型强不定的 OoD 域, 方差被直接注入到 RPL 架的基于能量的失中:

$$\mathcal{L}_{\text{out}} = \sum_{\omega \in \Omega} \max(-\mathbf{m}(\omega) \cdot \text{Var}(p(\omega)), 0),\quad (12)$$

其中 $\mathbf{m}(\omega) \in \{0, 1\}$ 是一指示 OoD 像素的二掩。我注意到, 由于 α 和 β 是由流模型生成的, 最小化 \mathcal{L}_{out} 通方差梯度更新, 而播回流。

$$\nabla_{\theta_{\text{flow}}} \mathcal{L}_{\text{out}} = -\frac{\partial \text{Var}(p)}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial \theta_{\text{flow}}} - \frac{\partial \text{Var}(p)}{\partial \beta} \cdot \frac{\partial \beta}{\partial \theta_{\text{flow}}}.\quad (13)$$

鼓流 OoD 域生成具有更高方差的 Beta。由于 Beta 分布的方差在 $\alpha \approx \beta$ 和都小到最大, 模型通少 OoD 域分布的尖性式表示不稳定性。适有助于在点和外点域之的行健的分, 增强 RPL 早期段的定性。

了在督信中引入不稳定性, 我了一一的失函, 函使用塔后考点点置信度和常点模糊性。提出的塔不稳定性感知交叉 (BUCE) 失由三部分成: 我首先算的 \hat{y} 普通生成的目 \tilde{y} 之的典交叉失: 了模糊或未知域的自信, 我基于 UCE 失 [8] 引入了基于塔的交叉不稳定性失 (BUCE)。模型置信度塔后建模的不定性之的不匹配。在里, $\psi(\cdot)$ 是 digamma 函, $\mathbb{E}[\text{Beta}(\alpha, \beta)]$ 表示塔分布的微分:

方程式 ?? 中的第一提供了基于置信度的交叉自信。在模糊域中鼓更高的不稳定性, 助模型避免度自信的。正化于在界外 (OoD) 界附近表示分布不稳定性尤其重要。中, 我 \mathcal{L}_{UCE} 乘以 10^{-7} 以定性。我通使用二元交叉强制 Beta 分布的方差二元 OoD 目的一致性:

$$\text{Var}^{(i)} = \frac{\alpha^{(i)}\beta^{(i)}}{(\alpha^{(i)} + \beta^{(i)})^2(\alpha^{(i)} + \beta^{(i)} + 1)}.\quad (14)$$

$$\mathcal{L}_{\text{var}} = \text{BCE}(\text{Var}^{(i)}, y^{(i)}),\quad (15)$$

(最版) 最的 BUCE 失定:

$$\mathcal{L}_{\text{BUCE}} = \mathcal{L}_{\text{ce}} + \lambda_1 \cdot \mathcal{L}_{\text{uce}} + \lambda_2 \cdot \mathcal{L}_{\text{var}},\quad (16)$$

其中 λ_1, λ_2 是控制不稳定性相影响的超。

4. 。

在本中, 我首先解置, 在不同的基准上比我方法, 同分析在 OoD 情下 Beta 分布 α 和 β 之的系。

我遵循 RPL 架的准。模型在 Cityscapes 据集 (2975 幅像/500 幅像)

Table 1. 在Fishyscapes和SMIYC数据集上的比较。最佳结果用粗体表示。

Method	Fishyscapes - Static			Fishyscapes - L & F			SMIYC - Anomaly			SMIYC - Obstacle		
	FPR ↓	AuPRC ↑	AUROC ↑	FPR ↓	AuPRC ↑	AUROC ↑	FPR ↓	AuPRC ↑	AUROC ↑	FPR ↓	AuPRC ↑	AUROC ↑
MCD(RN101) [10]	25.20	19.43	92.12	30.74	6.62	90.97	57.75	45.94	81.89	11.52	48.22	97.46
SE(RN101) [14]	25.22	19.68	92.12	29.59	7.04	91.19	58.18	45.33	81.77	11.92	48.41	97.43
EnE(RN101)	15.81	36.61	96.24	26.61	12.56	94.31	54.32	43.25	82.94	15.20	52.01	97.31
Ours + DE(RN101) [10]	13.23	56.34	97.26	26.68	11.56	94.19	48.09	56.55	86.98	11.08	52.42	98.01

Table 2. RoadAnomaly 数据集。所有方法都基于DeepLabv3+ 架构,最佳结果用粗体表示。

Method	FPR ↓	AuPRC ↑	AuROC ↑
SE (ResNet101)	73.93	19.24	69.49
MCD (ResNet101)	73.25	19.20	69.66
ENE (ResNet101)	74.13	19.15	70.56
Ours + DE (ResNet101)	71.96	19.76	70.73
Ours + DE (MobileNet)	72.74	18.82	71.19
GMMSeg	47.90	34.42	84.71
PEBAL	44.58	45.10	87.63
RPL+CoroCL	17.74	71.60	95.72
RPL+CoroCL+DE (Ours)	18.49	75.49	95.83

Table 3. RPL数据集在Fishyscapes和SMIYC数据集上的FPR和AUPRC性能比较。

Method	Fishyscapes - Static		Fishyscapes - L & F	
	FPR ↓	AUPRC ↑	FPR ↓	AUPRC ↑
RPL + SE	6.59	89.53	16.88	67.36
PEBAL	1.52	92.08	4.76	58.81
DenseHybrid	4.17	76.23	5.09	69.79
RPL	0.85	92.46	4.76	70.61
RPL + DE (Ours)	1.03	93.77	3.53	71.28
Method	SMIYC - Anomaly		SMIYC - Obstacle	
	FPR ↓	AUPRC ↑	FPR ↓	AUPRC ↑
RPL + SE	22.36	86.21	0.57	92.59
PEBAL	36.74	53.10	7.92	10.45
DenseHybrid	52.65	61.08	0.71	89.49
RPL	7.18	88.55	0.09	96.91
RPL + DE (Ours)	6.87	89.48	0.11	96.29

上运行,使用来自COCO的数据集的46751个常暴露(OE)样本,确保有足够重。我们在四个OoD分割基准上比较所提出的方法:Fishyscapes(静态, LostAndFound)、RoadAnomaly、Segment-me-If-You-Can (SMIYC)数据集,包括LostAndFound的一个物理子集SMIYC-L F的常规和障碍痕迹。所有方法都使用我们的DeepLabV3+骨干网络,输入尺寸调整700×700以提高效率。我们报告包括FPR95, AUROC, AP和AUPRC,用于评估可分性、鲁棒性和OoD定位能力。我们提出的方法在四个OoD分割基准上表现最佳,如在表1、表2和表3中所示。所有方法都基于相同的骨干网络DeepLabV3+和ResNet101 (RN101),以确保公平性。在表1中,我们比较了代表性的方法,如Shannon Entropy (SE)、Monte Carlo Dropout (MCD)和Free Energy (EnE)。我们的方法(我们的方法+差值)在大多数数据集上的FPR和AuPRC方面超过了所有基准。例如,在Fishyscapes-Static上,我们的方法实现了最低的FPR (13.23)和最高的AuPRC (56.34),显著超过了MCD (FPR 25.20, AuPRC 19.43)和EnE (FPR 15.81, AuPRC 36.61)。同样,在SMIYC-Anomaly基准上,我

我们的方法实现了最佳的FPR (48.09)和AuPRC (56.55),表明在开放数据集下的鲁棒性得到了提高。表2显示了在RoadAnomaly数据集上的结果。与强基准如RPL+CoroCL相比,我们的方法实现了最佳的AuPRC (75.49)并保持了低FPR (18.49),这体现了我们基于Beta的方差地提供了有意义的监督。

最后,在表格3中,我们进行了一项消融实验,比较了RPL的不同变体。与我们的RPL或RPL+SE相比,我们的RPL+DE一致性地改进了AP和AUPRC,尤其是在像SMIYC-Anomaly (AP 89.48 对比 57.99)和Obstacle (AUPRC 96.29 对比 22.10)困难的数据集中,这体现了我们的分布不确定性的有效性。

如图??所示,我们比较了不同不确定性的方法。我们提出的方法在Fishyscapes数据集上的推广效果更好。可视化结果表明,我们的方法在突出场景中的常规物体更加有效,同时抑制了已知区域的检测。通过符合分布的不确定性,我们的方法生成了更精确和更精确的常规边界,与基线方法相比,能更好地识别出分布外(OoD)像素。增强的检测能力在我们的模型如何识别OoD物体域背景中尤为重要,这对于自动驾驶至关重要。

在本文中,我们提出了一项新的方法,通过结合基于流的模型FlowEneDet和Beta后处理,利用自由能捕捉分布的不确定性。该架构可以有效地识别未知数据域的不确定性。此外,我们将我们的方法集成到最新的RPL架构中,使用Beta后处理输出的方差作为原性的监督信号。该集成使模型能基于分布的不确定性而不是固定的能量,更健康地指在OoD域的检测。未来的工作包括将我们的不确定性整合到多模态的基于能量的模型中,以验证其在OoD检测中的普遍适用性。

5. 致

本工作由中国政府(科技部信息通信部,MSIT)资助的信息通信科技委员会(IITP)项目(RS-2021-II211341, 人工智能研究生院项目(中央大学)和RS-2022-II220124,自我提升能力感知能力的人工智能技术)。

References

- [1] S. Ancha, P. R. Osteen, and N. Roy. Deep evidential uncertainty estimation for semantic segmentation under out-of-distribution obstacles. In 2024 IEEE International Conference on Robotics and Automation (ICRA), pages 6943–6951. IEEE, 2024.

- [2] H. Blum, P.-E. Sarlin, J. Nieto, R. Siegwart, and C. Cadena. The fishyscapes benchmark: Measuring blind spots in semantic segmentation. arXiv preprint arXiv:1904.03215, 2019.
- [3] R. Chan, K. Lis, S. Uhlemeyer, H. Blum, S. Honari, R. Siegwart, P. Fua, M. Salzmann, and M. Rottmann. Segmentmeifyoucan: A benchmark for anomaly segmentation. In Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2021.
- [4] B. Charpentier, D. Zügner, and S. Günnemann. Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts. *Advances in neural information processing systems*, 33:1356–1367, 2020.
- [5] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *ICML*, 2016.
- [6] D. Gudovskiy, T. Okuno, and Y. Nakata. Concurrent misclassification and out-of-distribution detection for semantic segmentation via energy-based normalizing flow. In *Uncertainty in Artificial Intelligence*, pages 745–755. PMLR, 2023.
- [7] A. Kendall and Y. Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NeurIPS*, 2017.
- [8] S. Landgraf, M. Hillemann, K. Wursthorn, and M. Ulrich. U-ce: Uncertainty-aware cross-entropy for semantic segmentation. arXiv preprint arXiv:2307.09947, 2023.
- [9] K. Lis, K. Nakka, P. Fua, and M. Salzmann. Detecting the unexpected via image resynthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2152–2161, 2019.
- [10] Y. Liu, C. Ding, Y. Tian, G. Pang, V. Belagiannis, I. Reid, and G. Carneiro. Residual pattern learning for pixel-wise out-of-distribution detection in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1151–1161, 2023.
- [11] A. Malinin and M. Gales. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31, 2018.
- [12] J. Mukhoti and Y. Gal. Evaluating bayesian deep learning methods for semantic segmentation. arXiv preprint arXiv:1811.12709, 2018.
- [13] M. Sensoy, L. Kaplan, and M. Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018.
- [14] W. Zhao, J. Li, X. Dong, Y. Xiang, and Y. Guo. Segment every out-of-distribution object. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3910–3920, 2024.