

改进野生动物的分布外检测：非洲五大兽

Mufhumudzi Muthivhi
Institute for Artificial Intelligent Systems
University of Johannesburg
mmuthivhi@uj.ac.za

Fredrik Gustafsson
Department of Electrical Engineering
Linköping University
fredrik.gustafsson@liu.se

Jiahao Huo
Institute for Artificial Intelligent Systems
University of Johannesburg
jhao@uj.ac.za

Terence L. van Zyl
Institute for Artificial Intelligent Systems
University of Johannesburg
tvanzyl@uj.ac.za

Abstract

缓解人类与野生动物之间的冲突旨在解决这些群体之间不期而遇的冲突。计算机视觉提供了一种解决方案，可以识别可能引发冲突的个体，例如非洲五大动物的成员。然而，环境中通常包含多种不同的物种。目前最先进的动物分类模型是在封闭世界假设下训练的。即使在面对未知类别时，它们几乎总是对其预测过于自信。本研究调查了野生动物的分布外 (OOD) 检测，特别是“非洲五大”。为此，我们选择一个参数的最近类别均值 (NCM) 和一个非参数的对比学习方法作为基线，以利用流行分类编码器的预训练和投影特征。此外，我们将基线与文献中各种常见的 OOD 方法进行比较。结果表明，基于特征的方法在各种分类阈值下表现出更强的泛化能力。具体而言，使用 ImageNet 预训练特征的 NCM 在 AUPR-IN、AUPR-OUT 和 AUTC 方面分别取得了比最佳 OOD 方法更好的 2%、4% 和 22% 提升。代码可以在这里找到 <https://github.com/pxpana/BIG5OOD>

1. 介绍

人与野生动物冲突是由于人类、牲畜和野生动物之间的不必要互动而产生的。在野生动物保护区附近居住的社区往往发现自己与野生动物争夺资源，例如牲畜捕食和农作物侵袭可能升级为报复性杀戮，损害保护工作。水牛、大象、狮子、豹和犀牛被认为是非洲的旗舰物种。它们通常被称为“五大”，是主要的旅游景点。它们还表现出攻击性行为，可能导致财产损害，频繁侵入人类居住地并被困。它们的敌对性质进一步威胁其濒危状态。为减轻人与野生动物互动，常见的策略包括电动围栏、人造驱虫剂和运动传感器。最近的文献探讨了使用边缘设备来分类野生动物，并同时使用蜂鸣器或电子鞭炮来驱赶它们。[2, 21, 27] 与此同时，大规模分类模型在野生动物监测和生态研究中取得了成功 [5]

。其中最大的模型 SpeciesNet 能够分类多达 2000 种动物。这些最先进的“基础”模型是在封闭世界假设下训练的 [3, 12, 32, 36, 37, 40]。它们在训练中见过的类别表现良好，但会将未知类别错误分类为已知类别 [42]。因此，实际上，一个模型需要在一个地区的每个物种上进行训练，即使模型旨在用于有限子集。本文探讨了大五动物分类模型的训练。我们通过将不属于大五的任何动物标记为分布外 (OOD) 来处理错误检测。关于使用机器学习结合相机陷阱数据监测大五非洲标志性动物的近期研究有限。此外，OOD 检测仍然是一个开放的研究课题，探讨这些方法如何在基准之外的 OOD 野生动物数据上表现的研究更少。本研究评估了当前最先进 (SOTA) 的预训练架构在大五动物分类上的表现。然后，我们对现有的推理和基于特征的 OOD 方法进行多项实验，以检测未包含在训练数据中的动物 [17]。结果显示，基于特征的方法在 AUROC、AUPR-IN、AUPR-OUT 和 AUTC 指标上维持了最强表现。我们选择的基线方法在 OOD 检测中接近当前 SOTA，并激励在 OOD 野生动物监测中继续研究基于特征的方法。我们的贡献在于文献：

1. 提供一个使用 ImageNet 特征进行大五人格分类和 OOD 检测的模型；
2. 建立一个简单但有效的基于一致性的双头预测算法；
3. 展示了通用预训练特征在检测 OOD 样本中的优越性；并且
4. 对每个预训练分类模型与最新的 OOD 方法进行比较。

2. 背景

在不同的生态系统中，利用计算机视觉技术从图像和视频识别不同的野生动物物种的兴趣日益增加。Villa et al. [37] 使用神经网络处理相机陷阱图像，以对来自 Snapshot Serengeti 数据集的动物进行分类。Willi et al. [40] 建议使用公民科学来生成塞伦盖蒂相机陷阱的注释。MegaClassifier 使用主要来自北美和欧洲物种的裁

剪 MegaDetector 图像。BioClip 在 TreeOfLife-10M 数据集上训练，该数据集包括各种动物、植物、真菌和昆虫。SpeciesNet 通过与 MegaDetector 组合在超过 60 百万张图像和大约 2,000 个类别上进行训练。然而，这些方法都无法检测到其分布之外的动物。

2.1. 分布外检测

已经提出了几种策略来区分已知类与未知类 [42]。这些方法可以分为基于推理和特征正则化的方法 [24]。推理方法使用预训练模型，同时修改预测结果。这些方法中最早的一些通过对最大 logits 或 softmax 概率进行阈值处理，被认为在 OOD 中是一种普遍的基线 [14, 15]。温度缩放已经被证明在校准预测时是有效的 [13, 19]。基于能量的方法捕捉模型预测中的整体不确定性 [8, 20]。其他策略采用一种距离度量与一个最近邻聚类算法在特征空间中配合 [18, 35]。极值理论方法记录概率中的异常值出现 [4, 25]。特征正则化方法修改网络以学习 OOD 表示。它们可以进一步分为无监督和有监督的方法。无监督方法在训练期间仅使用分布内 (ID) 数据。深度支持向量数据描述 (SVDD) 在嵌入空间中放置一个单类原型 [26]。中心损失通过为每个类分配一个原型延伸了该想法 [39]。目标是将所有 ID 数据的表示拉近到其关联的类原型。最近的方法旨在根据特征空间生成置信评分 [18, 29, 35]。后续研究在特征空间激活上拟合类条件高斯分布，以导出 ID 和 OOD 样本之间的判别决策边界 [38, 41]。特别地，Vojtíš et al. [38] 利用最近类均值 (NCM) 创建类特定的特征原型。Winkens et al. [41] 使用对比学习来学习一个更具判别性的无标签特征空间。有监督的方法在训练期间需要 OOD 数据。在野生动物学习的背景下，这种数据很难获得。因此，本研究只训练 ID 数据，并同时评估 ID 和 OOD。

3. 方法论

3.1. 方法

我们探讨了四种预训练的主干编码器 f ：SpeciesNet、MegaClassifier、BioClip 和一个基于 Vision Transformer (ViT) 的模型，该模型在 Imagenet [3, 6, 12, 32] 上进行了预训练。给定一幅图像 x ，我们通过冻结 f 来提取其特征 $z = f(x)$ 。我们训练了一个带有 ReLU 分离的两层分类头 g ，以获得预测 $y_1 \in \mathbb{R}^5$ 。我们探讨了一种参数化和非参数化的方法来确定 x 是分布内 (ID) 还是分布外 (OOD)。

对于我们提出的参数方法，我们使用最近类别均值 (NCM) 来计算验证集 \mathcal{D} 中类别 c 的平均特征向量 $\mu \in \mathbb{R}^d$ ，使得：

$$\mu_c = \frac{1}{|\mathcal{D}_c|} \sum_{x \in \mathcal{D}_c} f(x) \quad (1)$$

其中 d 是特征维度。我们通过在特征空间中找到 NCM 来对图像 x 进行分类，使得： $y_2 = \operatorname{argmin}_c \|f(x) - \mu_c\|_2$

。当图像 x 的预测类别在分类头和 NCM 上达成一致时，我们认为它是 ID；否则，它是 OOD。

我们的非参数方法同时训练一个分类 g 和一个投影头 p 。给定投影特征 $\hat{z} = p(z)$ ，我们使用对比学习损失函数，归一化温度缩放交叉熵 (NTXent)，来学习一个判别特征空间，使得：其中 \hat{z}_i 和 \hat{z}_j 是批次大小为 N 和温度项 τ 的正样本对 [7]。对于 OOD 检测，我们使用投影空间中的 k 近邻。我们在验证集上运行 KNN ($k=50$)，并选择图像 x 的多数类 y_2 。像 NCM 一样，如果分类头和 KNN 预测同一类别，则 x 为 ID。

3.2. 实验设置

我们采用了来自 PyTorch-OOD 库中的常见 OOD 基准 [17]。我们使用 Pytorch Lightning 来简化我们的训练和测试过程 [10]。每个骨干架构都配备了一个分类及/或投影头进行训练。我们使用带有线性缩放学习率 0.005 和余弦预热调度的加权 Adam 优化器。我们将所有训练限制为 100 个周期，监测验证性能。

3.2.1. 数据集

Table 1. 分布类内和类外的图像数量

In-Distribution		Out-of-Distribution	
Species	Images #	Species	Images #
Buffalo	48,184	Impala	5,000 for each
Elephant	54,453	Zebra	
Lion	15,584	Cheetah	
Leopard	6,487	Giraffe	
Rhino	7,434	Wildebeest	
		Hippo	
Training	79,285	-	-
Validation	26,428	Validation	9,000
Test	26,428	Test	21,000

我们使用 LILA BC 生物学上的多个野生动物数据集创建了非洲五大野生动物的一个子集 [1]。我们使用 Megadetector 裁剪图像中的动物，使用的置信度为 0.2。我们使用 K-means 聚类将相似的图像分组在一起，以避免训练-测试分割之间的簇重叠。我们采用相同的方法获取六个额外的物种进行 OOD 测试。表 1 展示了在我们实验中使用的 ID 和 OOD 动物。这些 OOD 物种因其在视觉上与 ID 类别相似而被选中，作为一个具有挑战性的开放世界问题。例如，在相机陷阱中，长颈鹿和猎豹的花纹在特定光照条件下与豹子相似。同样，角马、黑斑羚和河马的体型在不同条件下可能分别与水牛、狮子和犀牛相似。大象和斑马经常在相同栖息地中共同出现，导致训练数据中可能的信息泄露。表 1 展示了每个物种的计数。训练、验证和测试的分割分别为 60%、20% 和 20%。分割按照类别标签分层以保持类之间的比例。在准确性指标上进行评估时，我们从 OOD 类别中取 30% 的图像作为验证集，以找到非参数方法的最佳阈值。图像被调整大小并裁剪到 224×224 。

3.2.2. 度量标准

我们采用了四种常用的 OOD 指标。接收者操作特征曲线下面积 (AUROC) 衡量模型在所有可能的分类阈值下区分 ID 和 OOD 样本的能力 [11]。AUROC 为 1 代表一个完美的模型，而 0.5 则表示一个随机分类器。精确召回曲线下面积 (AUPR) 检查精确率和召回率之间的权衡 [23]。AUPR-IN 将 ID 样本作为正类，而 AUPR-OUT 偏向于 OOD 样本。阈值曲线下面积 (AUTC) 避免对 OOD 分数进行二值化，而是通过惩罚 ID 和 OOD 样本之间的分离不良来考虑它们的分布 [16]。

我们在各种主干架构之后测试了一个可训练的线性层基线，其次是一个非线性和一个线性层。第三，由于图像的有限性，我们对犀牛的图像进行了强大的数据增强。我们选择了一个非线性基线，其具有强大的数据增强和对犀牛的过采样，因为其在每类准确性指标上表现更为强劲。我们使用了十二种现有的 OOD 解决方案，包括参数和非参数方法。基于特征的方法包括 DeepSVDD 和 Center Loss，基于推理的方法包括 MaxSoftmax、EnergyBased 和 MaxLogit。我们将这些方法与我们提出的两个新基线方法进行比较。

Table 2. 微调后的预训练主干在分布内物种上的分类性能

Species	Species-Net	Mega-Classifier	BioClip	ImageNet
Elephant	.475	.884	.827	.904
Buffalo	.799	.882	.889	.951
Lion	.084	.843	.761	.905
Leopard	.322	.902	.939	.958
Rhino	.114	.294	.335	.529
F1 Score (Macro Ave.)	.376	.749	.761	.862
F1 Score (Weighted Ave.)	.487	.827	.815	.900

4. 结果与讨论

4.1. 分布内 (ID)

表 2 中的数据表明，在评估所有考虑的物种的准确性时，ImageNet 预训练的骨干模型比其他在动物数据集上预训练的模型更为优越。这个结果表明，更广泛范围内涉及有生命和无生命物体的通用化模型在 OOD 检测中更为有效。尽管对犀牛数据进行了过采样，但每个模型的准确性都低于 60%，但与之前的研究相当。ID 准确性的结果与最近关于野生动物相机陷阱分类的研究一致 [30, 37]。我们的模型在水牛、狮子和豹子方面显著优于这些研究。对于大象，我们的结果与之前的研究相当，结果在 85% 和 93.5% 之间。

在表格 3 中，结果显示表现最好的方法是使用预训练的 ImageNet 骨干网络的我们提出的 NCM 基准。使用不同的 OOD 方法时，ImageNet 骨干网络在 OOD 和 ID 类别之间的二分类表现最佳。有趣的是，与使用 ImageNet 的更通用模型相比，预训练在野生动物物种

上的骨干在 OOD 检测中的表现不佳。EnergyBased 与 ImageNet 结合产生了最高的 AUROC 和准确率，分别为 0.751 和 0.728，紧随其后的是 MaxLogit。然而，当评估 AUPR-IN 和 AUPR-OUT 时，NCM、Center Loss、SHE 和 ReAct 实现了最高的结果。此性能表明，由于 logit 方法直接建模类别边界的不确定性，它们在全局可分性方面表现出色。通过利用嵌入空间，使用特征的方法在精确-召回权衡 (AUPR) 中表现更佳，并可以应对不平衡数据集。NCM 和对比损失的 AUTC 分数分别在使用 ImageNet 时最低，为 0.306 和 0.367。此外，AUTC 不将 OOD 视为二分类问题。相反，AUTC 衡量的是模型在所有可能的决策阈值下如何有效地保持 ID 和 OOD 分布之间的分离。NCM 的有效性源于其类别条件均值对齐。对比学习因其专注于优化实例区分而显著，因为这创建了可以很好地推广到各种阈值选择的稳健的决策边界。当处理真实世界的多类别分类问题时，这种表现优势尤其重要，因为最优阈值可能未知或者必须适应变化的数据分布。补充材料中的表格 4 反映了表现最佳的 OOD 模型的性能。

本研究旨在评估和改进当前最先进的检测非洲五大动物的方法，同时尽量减少 OOD 个体的误分类。我们的研究表明，使用 ImageNet 数据集进行预训练的模型在野生动物分类方面优于那些在野生动物数据集上训练的模型。此外，我们的 OOD 测试表明，现有的 OOD 检测方法在使用 ImageNet 预训练的主干网络时表现更好。这些结果表明，通用化特征对于野生动物检测更为有效。此外，我们提出了一个利用多头输出的简单基线，结合了分类器和对比损失头的匹配。

References

- [1] Labeled information library of alexandria: Biology and conservationn. <https://lila.science/>.
- [2] Vikas Bavane, Arti Raut, Swapnil Sonune, AP Bawane, and PM Jawandhiya. Protection of crops from wild animals using intelligent surveillance system. *International Journal of Research in Advent Technology (IJRAT)*, pages 2321–9637, 2018.
- [3] Sara Beery, Dan Morris, and Siyu Yang. Efficient Pipeline for Camera Trap Image Review.
- [4] Abhijit Bendale and Terrance E Boulton. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1563–1572, 2016.
- [5] Tanya Berger-Wolf, Daniel Rubenstein, Charles Stewart, Jason Holmberg, Jason Parham, Sreejith Menon, J.P. Crall, Jon Van Oast, Emre Kiciman, and Lucas Joppa. *Wildbook: Crowdsourcing, computer vision, and data science for conservation*. 2017.
- [6] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021.
- [7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning

Table 3. 每种方法的分布外性能。每个模型的前三种方法被突出显示。整体最佳用粗体显示。

Metrics	AUROC \uparrow				AUPR-IN \uparrow				AUPR-OUT \uparrow				AUTC \downarrow			
	Species-Net	Mega-Classifier	BioClip	ImageNet	Species-Net	Mega-Classifier	BioClip	ImageNet	Species-Net	Mega-Classifier	BioClip	ImageNet	Species-Net	Mega-Classifier	BioClip	ImageNet
inference-based																
MaxSoftmax [14]	.523	.556	.635	.744	.590	.609	.658	.747	.458	.481	.539	.669	.492	.474	.444	.408
MaxLogit [15]	.536	.541	.683	.750	.589	.591	.675	.701	.478	.472	.635	.700	.499	.496	.478	.463
Temp'Scaling [13]	.475	.478	.445	.540	.564	.550	.520	.582	.444	.420	.417	.464	.500	.500	.454	.496
OpenMax [4]	.526	.530	.663	.724	.574	.579	.673	.719	.473	.566	.679	.663	.497	.498	.459	.460
KNN [35]	.557	.607	.659	.737	.619	.656	.721	.797	.482	.555	.584	.632	.498	.497	.517	.441
EnergyBased [20]	.537	.531	.694	.751	.587	.585	.679	.725	.480	.463	.625	.705	.499	.497	.480	.465
Entropy [31]	.525	.555	.635	.747	.585	.609	.659	.748	.463	.478	.536	.674	.493	.476	.438	.395
SHE [44]	.528	.500	.526	.494	.778	.779	.771	.777	.469	.721	.454	.412	.498	.502	.498	.424
ReAct [34]	.533	.534	.695	.732	.584	.584	.681	.733	.477	.466	.658	.667	.499	.497	.476	.469
DICE [33]	.518	.528	.681	.749	.573	.580	.670	.727	.463	.467	.643	.696	.497	.498	.483	.465
feature-based																
Deep SVDD [26]	.500	.503	.596	.597	.779	.563	.642	.634	.571	.441	.548	.559	.500	.500	.499	.499
Center Loss [39]	.531	.514	.647	.762	.627	.582	.714	.807	.712	.688	.709	.745	.497	.491	.451	.446
inference and feature based																
GROOD [38]	.544	.597	.693	.824	.603	.645	.573	.889	.469	.515	.764	.682	.482	.461	.411	.322
proposed baselines																
NCM Agreement	.518	.531	.594	.693	.654	.673	.747	.809	.632	.629	.642	.734	.482	.469	.406	.306
Contrastive Loss	.547	.546	.580	.634	.725	.698	.751	.792	.591	.621	.619	.683	.453	.454	.420	.367
NCM Agreement Score *	.542	.532	.651	.801	.597	.587	.676	.858	.477	.467	.554	.700	.490	.492	.475	.408
Contrastive Agreement Score *	.588	.553	.620	.702	.630	.604	.624	.712	.518	.478	.571	.653	.500	.480	.500	.500

* Recent results using the agreement score described in the supplementary material

of visual representations. In *International conference on machine learning*, pages 1597–1607. PmlR, 2020.

- [8] Xuefeng Du, Zhaoning Wang, Mu Cai, and Sharon Li. Towards unknown-aware learning with virtual outlier synthesis. In *International Conference on Learning Representations*, 2022.
- [9] Hossein Estiri and Shawn N Murphy. Semi-supervised encoding for outlier detection in clinical observation data. *Computer methods and programs in biomedicine*, 181: 104830, 2019.
- [10] William Falcon and The PyTorch Lightning team. PyTorch Lightning, 2019.
- [11] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.
- [12] Tomer Gadot, tefan Istrate, Hyungwon Kim, Dan Morris, Sara Beery, Tanya Birch, and Jorge Ahumada. To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images. *IET Computer Vision*, 18(8):1193–1208, 2024.
- [13] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pages 1321–1330. PMLR, 2017.
- [14] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2017.
- [15] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joseph Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. In *Proceedings of the 39th International Conference on Machine Learning*, pages 8759–8773. PMLR, 2022.
- [16] Galadrielle Humblot-Renaux, Sergio Escalera, and Thomas B Moeslund. Beyond auroc & co. for evaluating out-of-distribution detection performance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3881–3890, 2023.
- [17] Konstantin Kirchheim, Marco Filax, and Frank Ortmeier. Pytorch-ood: A library for out-of-distribution detection based on pytorch. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4351–4360, 2022.
- [18] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31, 2018.
- [19] Shiyu Liang, Yixuan Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *International Conference on Learning Representations*, 2018.
- [20] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33:21464–21475, 2020.
- [21] Akanksha Mishra and Kamlesh Kumar Yadav. Smart animal repelling device: Utilizing iot and ai for effective anti-adaptive harmful animal deterrence. In *BIO Web of Conferences*, page 05014. EDP Sciences, 2024.
- [22] Seon Ho Oh, Sang Won Ghyme, Soon Ki Jung, and Geon-

- Woo Kim. Early wildfire detection using convolutional neural network. In *International workshop on frontiers of computer vision*, pages 18–30. Springer, 2020.
- [23] David MW Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*, 2020.
- [24] Ryne Roady, Tyler L Hayes, Ronald Kemker, Ayesha Gonzales, and Christopher Kanan. Are open set classification methods effective on large-scale datasets? *Plos one*, 15(9): e0238302, 2020.
- [25] Ethan M Rudd, Lalit P Jain, Walter J Scheirer, and Terrence E Boult. The extreme value machine. *IEEE transactions on pattern analysis and machine intelligence*, 40(3): 762–768, 2017.
- [26] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pages 4393–4402. PMLR, 2018.
- [27] Sahane Pradnya Sambhaji, Salunke Nikita Sanjiv, Shirsath Vitthal Somnath, Shukla Shreyas Sanjay, and AR Panhalkar. Early warning system for detection of harmful animals using iot. *International Journal of Advance Research and Innovative Ideas in Education*, 5(3):2395–4396, 2019.
- [28] Matthew D Scherrek and Brian D Rigling. Open set recognition for automatic target classification with rejection. *IEEE Transactions on Aerospace and Electronic Systems*, 52(2): 632–642, 2016.
- [29] Vikash Sehwal, Mung Chiang, and Prateek Mittal. Ssd: A unified framework for self-supervised outlier detection. *arXiv preprint arXiv:2103.12051*, 2021.
- [30] Karoline Seljebotn and Isah A. Lawal. Machine learning tool for wildlife image classification. In *Proceedings of the 2024 9th International Conference on Machine Learning Technologies*, page 127–132, New York, NY, USA, 2024. Association for Computing Machinery.
- [31] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
- [32] Samuel Stevens, Jiaman Wu, Matthew J Thompson, Elizabeth G Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M Dahdul, Charles Stewart, Tanya Berger-Wolf, Wei-Lun Chao, and Yu Su. BioCLIP: A vision foundation model for the tree of life. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19412–19424, 2024.
- [33] Yiyu Sun and Yixuan Li. Dice: Leveraging sparsification for out-of-distribution detection. In *European conference on computer vision*, pages 691–708. Springer, 2022.
- [34] Yiyu Sun, Chuan Guo, and Yixuan Li. React: Out-of-distribution detection with rectified activations. *Advances in neural information processing systems*, 34:144–157, 2021.
- [35] Yiyu Sun, Yifei Ming, Xiaojin Zhu, and Yixuan Li. Out-of-distribution detection with deep nearest neighbors. In *International Conference on Machine Learning*, pages 20827–20840. PMLR, 2022.
- [36] Alexandra Swanson, Margaret Kosmala, Chris Lintott, Robert Simpson, Arfon Smith, and Craig Packer. Snapshot serengeti, high-frequency annotated camera trap images of 40 mammalian species in an african savanna. *Scientific data*, 2(1):1–14, 2015.
- [37] Alexander Gomez Villa, Augusto Salazar, and Francisco Vargas. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological informatics*, 41:24–32, 2017.
- [38] Tomáš Vojtš, Jan Šochman, Rahaf Aljundi, and Jiří Matas. Calibrated out-of-distribution detection with a generic representation. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 4509–4518. IEEE, 2023.
- [39] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *Computer vision—ECCV 2016: 14th European conference, amsterdam, the netherlands, October 11–14, 2016, proceedings, part VII 14*, pages 499–515. Springer, 2016.
- [40] Marco Willi, Ross T Pitman, Anabelle W Cardoso, Christina Locke, Alexandra Swanson, Amy Boyer, Marten Veldhuis, and Lucy Fortson. Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, 10(1):80–91, 2019.
- [41] Jim Winkens, Rudy Bunel, Abhijit Guha Roy, Robert Stanforth, Vivek Natarajan, Joseph R LedSAM, Patricia MacWilliams, Pushmeet Kohli, Alan Karthikesalingam, Simon Kohl, et al. Contrastive training for improved out-of-distribution detection. *arXiv preprint arXiv:2007.05566*, 2020.
- [42] JingKang Yang, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. Generalized out-of-distribution detection: A survey. *International Journal of Computer Vision*, 132(12):5635–5662, 2024.
- [43] William J Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950.
- [44] Jinsong Zhang, Qiang Fu, Xu Chen, Lun Du, Zelin Li, Gang Wang, Shi Han, Dongmei Zhang, et al. Out-of-distribution detection based on in-distribution data patterns memorization with modern hopfield energy. In *The Eleventh International Conference on Learning Representations*, 2022.

改进野生动物的分布外检测：非洲五大兽

Supplementary Material

5.

附录 A

5.1. 每类准确率

在本节中，我们为论文包含补充结果。表格 4 显示了每种方法各自的类别准确性表现。为了确定一张图片是分布内 (ID) 还是分布外 (OOD)，我们使用 Youden's J 统计量 [43] 类似于 [9, 22, 28] 获得最优阈值。Youden's J 通过测量接收者操作特性 (ROC) 曲线与表示随机分类器的对角线之间的最大垂直距离来量化二分类器的判别能力。正式地，Youden's J 被公式化为：

$$J = \max_{\tau} (\text{Sensitivity}(\tau) + \text{Specificity}(\tau) - 1) \quad (2)$$

其中， τ 是阈值，Sensitivity 是真阳性率，Specificity 是真阴性率。最优阈值 τ^* 对应于最大化 J 的 τ 值。

5.2. 一致性分数

通过在最近类均值 (NCM) 和对比学习方法之间引入一致性分数，我们观察到性能的提升。我们为 n 类定义了两个向量 $v_1, v_2 \in \mathbb{R}^n$ 。 v_1 捕捉基于推理的预测，而 v_2 编码基于特征的信息。特别地， v_2 是从分类头获得的 softmax 值向量。对于我们的 NCM 策略， v_2 是图像 x 的特征到每个类均值 μ_c 的距离。对于对比学习， v_2 是在给定 x 特征的情况下每个类的最近邻居数量。我们通过 k 对 v_2 进行归一化，以将其转换为五个目标类中的概率分布。然后一致性分数由以下公式给出：

$$\frac{1 - H(v_1 \odot v_2 + \epsilon)}{\log(n)} \times (1 - \text{JS}(v_1, v_2)) \quad (3)$$

，其中 H 是熵，JS 是由以下方程给出的詹森-香农方程：

$$H(p) = - \sum_{i=1}^n p_i \log p_i \quad (4)$$

，并且

$$\text{JS}(v_1, v_2) = \frac{1}{2} [\text{D}_{\text{KL}}(v_1 || m) + \text{D}_{\text{KL}}(v_2 || m)] \quad (5)$$

，其中 \odot 是逐元素乘法操作， $\epsilon \ll 1$ 是一个很小的值以防止除以零， $\text{D}_{\text{KL}}(\cdot || \cdot)$ 是库尔贝克-莱布勒 (KL) 散度， $m = \frac{1}{2}(v_1 + v_2)$ 是中点分布。

5.3. 局限性

SpeciesNet 的权重是 Tensorflow 格式的。我们必须使用 ONNX 将其转换为 PyTorch 格式。我们无法保证转换过程的稳定性。

Table 4. 每种方法的每类准确性表现。表现最佳的前三种方法已被突出显示。

Species	Elephant				Buffalo				Lion				Leopard				Rhino				Out-Of-Dist'			
	Species-Net	Mega-Classifier	BioClip	ImageNet	Species-Net	Mega-Classifier	BioClip	ImageNet	Species-Net	Mega-Classifier	BioClip	ImageNet	Species-Net	Mega-Classifier	BioClip	ImageNet	Species-Net	Mega-Classifier	BioClip	ImageNet	Species-Net	Mega-Classifier	BioClip	ImageNet
OOD Methods																								
Baseline	.475	.884	.827	.904	.799	.882	.889	.951	.084	.843	.761	.905	.322	.902	.939	.958	.114	.294	.335	.529	-	-	-	-
inference-based																								
MaxSoftmax	.237	.273	.451	.645	.383	.247	.620	.793	.026	.358	.509	.713	.248	.716	.920	.946	.047	.044	.142	.156	.613	.753	.721	.718
MaxLogit	.000	.177	.532	.683	.001	.139	.676	.813	.000	.327	.511	.727	.121	.706	.924	.950	.000	.027	.165	.227	.396	.716	.673	.699
Temp'Scaling	.131	.037	.801	.677	.327	.077	.878	.663	.023	.092	.724	.562	.221	.321	.938	.849	.058	.038	.329	.351	.613	.908	.040	.493
OpenMax	.198	.388	.462	.620	.667	.307	.712	.799	.078	.499	.440	.672	.293	.782	.916	.945	.044	.107	.172	.240	.506	.554	.658	.696
KNN	.036	.098	.390	.563	.137	.090	.526	.525	.010	.079	.436	.518	.001	.014	.837	.892	.041	.037	.236	.360	.641	.637	.748	.769
EnergyBased	.000	.689	.497	.752	.000	.520	.644	.862	.000	.605	.483	.785	.079	.837	.921	.951	.000	.331	.164	.300	.403	.668	.701	.693
Entropy	.208	.371	.461	.650	.279	.314	.635	.797	.023	.406	.511	.714	.227	.720	.921	.946	.034	.064	.143	.156	.718	.658	.713	.716
SHE	.156	.001	.300	.632	.496	.005	.545	.782	.048	.002	.346	.618	.268	.410	.906	.927	.028	.003	.192	.393	.538	.986	.722	.726
ReAct	.000	.459	.524	.818	.000	.293	.656	.894	.000	.472	.301	.689	.077	.786	.910	.944	.000	.103	.264	.666	.694	.688	.643	.677
DICE	.000	.208	.405	.685	.000	.091	.568	.818	.000	.344	.423	.755	.082	.719	.950	.754	.000	.030	.155	.251	.283	.692	.700	.713
feature-based																								
Deep SVDD	.558	.442	.558	.747	.829	.489	.538	.609	.368	.442	.374	.570	.576	.291	.625	.837	.259	.152	.254	.339	.001	.522	.491	.471
Center Loss	.511	.275	.537	.630	.837	.389	.639	.811	.187	.454	.357	.602	.494	.810	.894	.915	.244	.094	.172	.196	.000	.603	.698	.732
inference and feature based																								
GROOD	.476	.378	.420	.630	.455	.403	.509	.693	.420	.414	.483	.663	.409	.503	.553	.784	.406	.393	.402	.530	.554	.683	.853	.893
proposed baselines																								
NCM Agreement	.171	.174	.610	.699	.378	.364	.686	.831	.015	.606	.655	.820	.296	.817	.891	.928	.071	.196	.194	.523	.654	.621	.509	.887
Contrastive Loss	.488	.456	.686	.844	.651	.409	.688	.856	.078	.488	.591	.808	.490	.849	.910	.944	.137	.034	.232	.419	.562	.544	.599	.723
NCM Agreement Score	.337	.333	.460	.607	.546	.271	.642	.642	.053	.361	.544	.522	.279	.691	.871	.870	.060	.132	.060	.286	.444	.564	.714	.872
Contrastive Agreement Score	.346	.390	.438	.634	.526	.584	.520	.699	.011	.043	.462	.748	.224	.456	.901	.916	.007	.092	.143	.303	.525	.589	.691	.649