

LUNA: Localizing Unfamiliarity Near Acquaintance for Open-set Long-Tailed Recognition

Jiarui Cai¹, Yizhou Wang¹, Hung-Min Hsu¹, Jenq-Neng Hwang¹, Kelsey Magrane², Craig Rose³ ¹University of Washington, ²Pacific States Marine Fisheries Commission (PSMFC), ³FishNext Research



Introduction

- No authentic collected open long-tailed dataset to evaluate the OLTR methodology.
- Most open-set methods decouples the open set challenges with long-tailed distribution.
- Most long-tail recognition methods ignore the open set issues.
- > Our Goal is to answer:
- (1) Whether the input is novel;
- (2) If no, which class it is; if yes, what is the novelty level of the new class concerning the pretrained acquaintance classes.
- > Our Achievements are:
- (1) The first natural OLTR dataset in a finegrained domain: we collect a new wellannotated real Marine Species open longtailed (MS-LT) dataset.
- (2) A distribution-aware loss: It centralizes the deep features of the head classes, while preserving the classification accuracy of the tails, resulting in more distinctive features.
- (3) A quantitative measurements of novelty under the long-tailed distribution.
- (4) LUNA significantly outperforms the sota methods by 4-6% on the closed set and in average 4% improvement of the F-measure under the open-set setting.

The Proposed Method



- Weighted Center Loss
- Distribution-aware weights are applied on each class separately for accurate long-tailed recognition and clustering.

$$L_{wc} = \frac{1}{2} \sum_{i} \lambda_{i} ||x_{i} - c_{j}||^{2} = \frac{1}{2} \sum_{i} (\frac{\widetilde{n}_{j}}{\max_{c} \widetilde{n}_{c}} + 1) ||x_{i} - c_{j}||^{2}$$

- LUNA Factor
- Inside each cluster, the points are classified into core and D = boundary points by its sub local reachability density (D)
- Define the second boundary points as the points which belong to other clusters but regard this cluster as their second-best choice by the distance to its center point.

LUNA factor of p_i

$$= \min(|1 - COF_i|, |1 - BOF_i|) + |1 - (\frac{1}{M}\sum_{j}^{M} sBOF_j)/sBOF_i| + (1 - conf_i)$$

Results

Dataset	Model	Closed-set				Open-set			
		Many	Medium	Few	Overall	Many	Medium	Few	F-measure
ImageNet-LT (ResNet-10)	Base Model [15]	40.9	10.7	0.4	20.9	40.1	10.4	0.4	0.295
	Lifted Loss [29]	35.8	30.4	17.9	30.8	34.8	29.3	17.4	0.374
	Focal Loss [23]	36.4	29.9	16.0	30.5	35.7	29.3	15.6	0.371
	Range Loss [39]	35.8	30.3	17.6	30.7	34.7	29.4	17.2	0.373
	OpenMax [3]	-	-	-	-	35.8	30.3	17.6	0.368
	FSLwF [12]	40.9	22.1	15.0	28.4	40.8	21.7	14.5	0.347
	OLTR [25]	43.2	35.1	18.5	35.6	41.9	33.9	17.4	0.474
	IEM [43]	48.9	44.0	24.4	43.2	46.1	42.3	20.1	0.525
	LUNA (Ours)	51.8	48.6	26.2	46.6	48.2	44.7	23.6	0.579
Places-LT (ResNet-152)	Base Model [15]	45.9	22.4	0.4	27.2	45.9	22.4	0.4	0.366
	Lifted Loss [29]	41.1	35.4	24	35.2	41.0	35.2	23.8	0.459
	Focal Loss [23]	41.1	34.8	22.4	34.6	41.0	34.8	22.3	0.453
	Range Loss [39]	41.1	35.4	23.2	35.1	41.0	35.3	23.1	0.457
	OpenMax [3]	-	-	-		41.1	35.4	23.2	0.458
	FSLwF [12]	43.9	29.9	29.5	34.9	38.1	19.5	14.8	0.375
	OLTR [25]	44.7	37	25.3	35.9	44.6	36.8	25.2	0.464
	IEM [43]	46.8	39.2	28.0	39.7	48.8	42.4	28.9	0.486
	LUNA (Ours)	48.7	42.4	30.2	42.1	48.1	41.6	29.0	0.491
MS-LT (ResNet-32)	Base Model [15]	56.1	35.1	8.0	35.7	56.1	35.1	11.4	0.537
	Lifted Loss [29]	53.2	42.3	12.6	38.0	53.0	42.2	12.4	0.549
	Focal Loss [23]	57.3	44.6	18.5	42.1	57.0	42.8	15.4	0.576
	Range Loss [39]	55.8	43.8	15.7	40.5	55.8	43.6	15.6	0.575
	OpenMax [3]	-	-	-	-	54.2	44.9	12.8	0.564
	OLTR [25]	57.8	49.8	28.6	46.8	56.7	45.3	23.6	0.603
	LUNA (Ours)	61.2	56.6	34.6	52.0	60.4	51.8	30.4	0.657

Marine Species Dataset

 $\frac{1}{k}\sum_{i}d_{i}$



 $\Sigma^{N_{sB}}_{..} D_p / D_{sB}^{\iota}$

NSR

Average D of boundary points

Dofp

Average D of sb points

sBOF =