Code available on GitHub!
https://github.com/melanie wille/uod-class-disparities

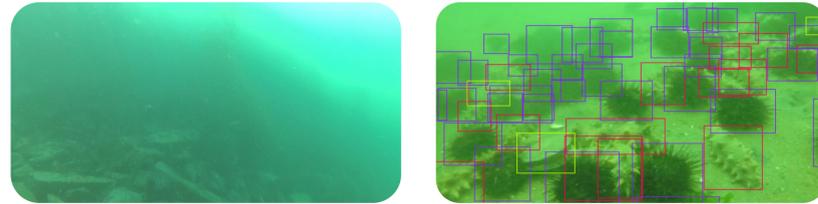# Are All Marine Species Created Equal?
Performance Disparities in Underwater Object Detection

Melanie Wille   Tobias Fischer   Scarlett Raine

WACV 2026 TUCSON, AZ 3/6 – 3/10

Motivation

The underwater domain presents unique challenges for accurate object detection, including:

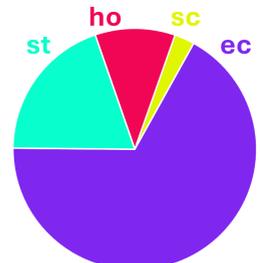degraded image quality   and   small, clustered targets.

Not every species is detected equally well, previous works show:

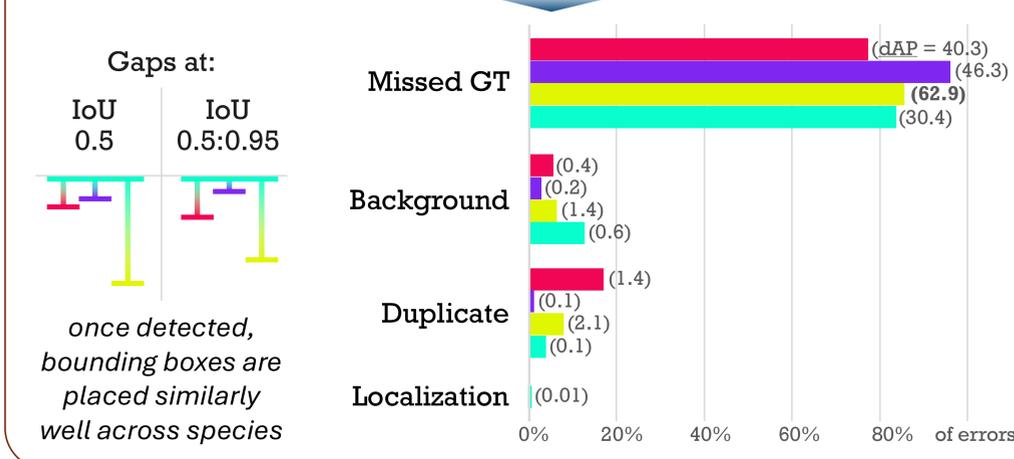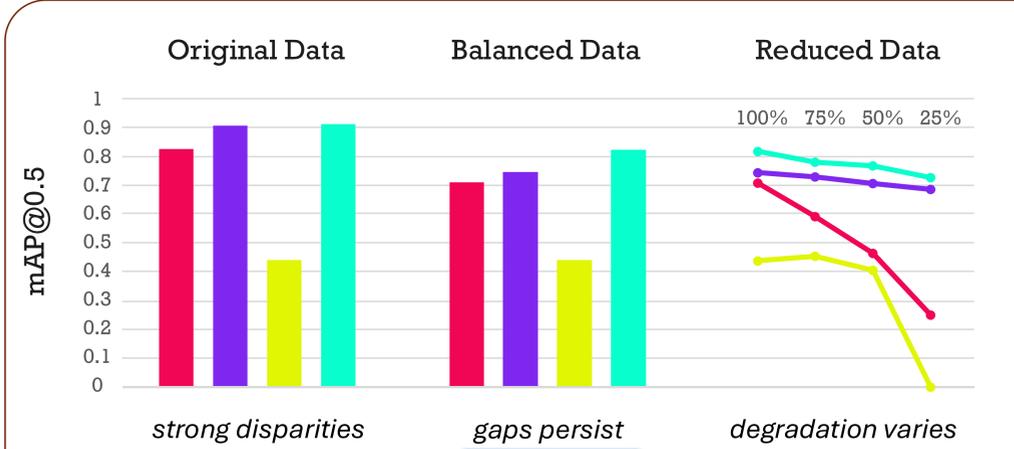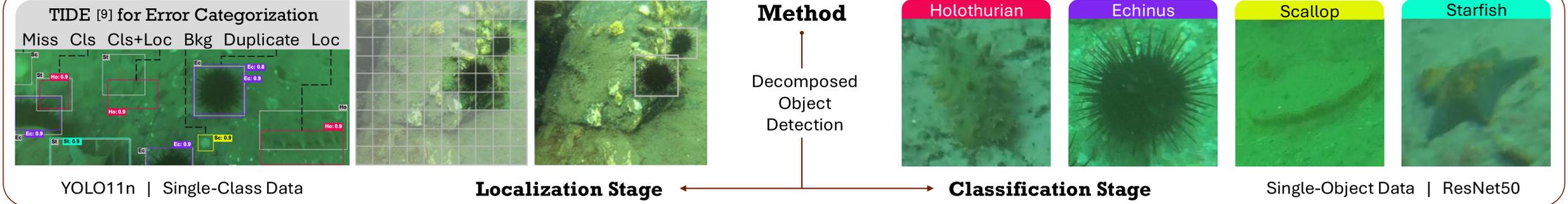| AP:	ho	ec	sc	st	mAP	comment
[1]	58.6	69.1	41.3	65.3	58.6	"smallest no"
[2]	59.5	68.5	41.6	64.9	58.6	
[3]	68.2	75.2	56.3	76.7	69.1	
[4]	81.0	84.1	57.6	85.2	77.0	"scarcity"
[5]	85.3	93.9	64.0	95.2	84.6	"scarcity"
[6]	91.1	93.3	73.3	95.3	88.4	
[7]	79.6	93.0	65.8	93.1	82.9	
[8]	84.7	92.6	70.8	93.3	85.4	

IoU 0.5-0.95
IoU 0.5

insufficiently explained, persistent under-performance of the 'Scallop' class in the Detecting Underwater Objects (DUO) dataset.

Research Gap: Class-Specific Performance Gaps

Data Quantity   vs   Species-Inherent Characteristics

ho   sc
st   ec

Are all species equally easy to spot?

imbalanced training data in DUO

Method

TIDE [9] for Error Categorization
Miss Cls Cls+Loc Bkg Duplicate Loc

YOLO11n | Single-Class Data

Localization Stage ←——→ **Classification Stage**

Decomposed Object Detection

Single-Object Data | ResNet50

Holothurian   Echinus   Scallop   Starfish

Original Data   Balanced Data   Reduced Data

mAP@0.5

100% 75% 50% 25%

strong disparities   gaps persist   degradation varies

Gaps at:
IoU 0.5   IoU 0.5:0.95

once detected, bounding boxes are placed similarly well across species

Missed GT	(dAP = 40.3) (46.3) **(62.9)** (30.4)
Background	(0.4) (0.2) (1.4) (0.6)
Duplicate	(1.4) (0.1) (2.1) (0.1)
Localization	(0.01)

0% 20% 40% 60% 80% of errors

Results

Non-quantitative, species-specific factors are critical drivers of performance gaps.

Class disparities arise in localization, not classification.

Foreground-background distinction is the main bottleneck.

Tradeoff for minority classes: imbalanced data prioritizes precision, while balanced data prioritizes recall.

Inter-class dependencies: models learn through negative examples of other species.

Results are confirmed through architectural ablations (SSD, MobileNet, ViT) and RUOD as additional dataset.

Original Data   Balanced Data   Reduced Data

Precision

30% data of
holothurian
echinus
scallop
starfish

-8% -4% +4% +8%

Recall

holothurian
echinus
scallop
starfish

-8% -4% +4% +8%

high scallop precision: commercial applications

high scallop recall: conservation monitoring

scallop performance is influenced by training samples from other classes

References: Datasets from Liu et al., ICME 2021 (DUO) and Fu et al., Neurocomputing 2023 (RUOD). Poster images from DUO.
[1] Liu et al., ICME 2021
[2] Xu et al., Neurocomputing 2023
[3] Dai et al., Pattern Recognition 2024
[4] Yuan et al., IET Computer Vision 2023
[5] Zhao et al., JMSE 2023
[6] Feng et al., Ecological Informatics 2024
[7] Zhao et al., PLOS One 2024
[8] Li et al., Symmetry 2025
[9] Bolya et al., ECCV 2020

Melanie Wille
Queensland University of Technology
melaniecarolin.wille@hdr.qut.edu.au