frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 11 July 2025
DOI 10.3389/fmars.2025.1581778

OPEN ACCESS

EDITED BY
Johannes Karstensen,
Helmholtz Association of German Research Centres (HZ), Germany

REVIEWED BY
Keyu Chen,
Xiamen University, China
M. Elizabeth Clarke,
NOAA Fisheries, United States

*CORRESPONDENCE
Oscar Pizarro
oscar.pizarro@ntnu.no

RECEIVED 23 February 2025
ACCEPTED 18 June 2025
PUBLISHED 11 July 2025

CITATION
Doig H, Pizarro O and Williams S (2025) Training marine species object detectors with synthetic images and unsupervised domain adaptation.
Front. Mar. Sci. 12:1581778.
doi: 10.3389/fmars.2025.1581778

Training marine species object detectors with synthetic images and unsupervised domain adaptation

Heather Doig[1], Oscar Pizarro[2*] and Stefan Williams[1]

[1]Australian Centre for Robotics, University of Sydney, Sydney, NSW, Australia, [2]Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway

Visual surveys by autonomous underwater vehicles (AUVs) and other underwater platforms provide a valuable method for analysing and understanding the benthic environment. Scientists can measure the presence and abundance of benthic species by manually annotating survey images with online annotation software or other tools. Neural network object detectors can reduce the effort involved in this process by locating and classifying species of interest in the images. However, accurate object detectors often rely on large numbers of annotated training images which are not currently available for many marine applications. To address this issue, we propose a novel pipeline for generating large amounts of synthetic annotated training data for a species of interest using 3D modelling and rendering software. The detector is trained with synthetic images and annotations along with real unlabelled images to improve performance through domain adaptation. Our method is demonstrated on a sea urchin detector trained only with synthetic data, achieving a performance slightly lower than an equivalent detector trained with manually labelled real images (AP50 of 84.3 vs 92.3). Using realistic synthetic data for species or objects with few or no annotations is a promising approach to reducing the manual effort required to analyse imaging survey data.

KEYWORDS

benthic monitoring, object detection, unsupervised domain adaptation, synthetic images, benthic imaging

1 Introduction

Increasing human activity in our oceans, such as wind energy, aquaculture and mining, can lead to changes in marine communities. Regular photographic surveys of the seafloor allow scientists to measure changes in the presence and abundance of a variety of marine species that may be impacted. Detection of specific benthic species supports the study and exploration of a healthy underwater environment (Estes et al., 2021; Perkins et al., 2022; Er et al., 2023; Peng et al., 2021). Images of the seafloor can be used to detect endangered (Stuart-Smith et al., 2020), invasive (Liu et al., 2022, 2021b) and sentinel species (Perkins

et al., 2022) captured by Autonomous Underwater Vehicles (AUVs), Remote Operated Vehicles (ROVs) and other underwater platforms. Using neural network object detectors to locate and classify a species of interest can help marine scientists efficiently review the large volume of images captured during underwater surveys. Ideally, the detector is trained in a supervised manner with large amounts of data to provide high performance (Inoue et al., 2018; Zhang et al., 2022; Munir et al., 2023), but this is often not available for underwater images (Liu et al., 2020; Er et al., 2023).

A high-performing detector often relies on training the model with large numbers of images annotated by experts who can identify the species of interest. These annotations may comprise bounding boxes around the objects of interest but may also include simple point annotations, object boundaries or full pixel-level semantic labelling. For underwater scenes, there are often few labelled annotations to train supervised models such as an object detector (Xu et al., 2023). Underwater images present particular challenges for reliable object detection due to variations in the attenuation of light in water, water clarity and camera platforms. The low number of annotations may not provide a good distribution of examples to train a model to generalise for variations in light, water and camera platforms. Generating synthetic images with 3D modelling software offers a new method to alleviate these issues by providing abundant annotations with variations in simulated water, light and camera conditions increasing the ability for the detector to perform well (Zhang et al., 2022; Oza et al., 2024; Er et al., 2023).

Synthetic images have been used to address the lack of annotated data in other domains such as autonomous driving (Johnson-Roberson et al., 2017) but only in a limited manner in the underwater domain (Sans-Muntadas et al., 2022; Zwilgmeyer et al., 2021; Lin et al., 2023). 3D modelling and simulation software such as Blender[1], Unity[2] and Unreal Engine[3] are increasingly popular in providing photorealistic images and animations to supplement or replace real data (Saini et al., 2022; Ebadi et al., 2022; Peñarroya et al., 2023; Lu et al., 2023; Becktor et al., 2022; Diamanti et al., 2024). Blender has combined functionality to produce large numbers of synthetically rendered images with variety and realism for underwater scenes. Through its Python interface, it provides procedural generation of scenes and assets creating randomisation (Santos et al., 2024; Raistrick et al., 2024, 2023), add-ons to generate annotations for training machine learning models (Denninger et al., 2023; Raistrick et al., 2023) and the ability to simulate the effect of light in water (Zwilgmeyer et al., 2021; Sans-Muntadas et al., 2022).

Infinigen (Raistrick et al., 2023) is a framework based on Blender that generates natural scenes using procedurally generated objects, providing the opportunity to generate a vast range of realistic images. Creating realistic 3D models of the morphospecies or object of interest allows the generation of a tailored synthetic dataset. The 3D model can be placed in various benthic scenes with random variations of water conditions and camera configurations.

Our work proposes a new method for training an object detector for a target marine species or object by generating large volumes of realistic benthic images with Blender and Infinigen combined with domain adaptation during training. While the synthetic data provides larger volumes of annotations for supervised training, the different distributions between the synthetic and real images must be addressed to perform effectively during real-world deployment. Previous work has used image-to-image translation to bridge this gap but this requires training an additional translation network (Sans-Muntadas et al., 2022; Lin et al., 2023). In our method, we apply two domain adaptation methods to reduce the domain gap during training without image-to-image translation.

The contributions of this paper are:

- a new method to train an object detector of a target marine species or object with no manually annotated training data
- a framework to generate realistic and varied synthetic images of benthic scenes for a target underwater object with annotations for object detection and semantic segmentation
- a training pipeline for high-performing object detection in underwater images using synthetic data and unlabelled real images using state-of-the-art unsupervised domain adaptation and semi-supervised methods
- experimental validation of our method on two real datasets with images of sea urchins captured with different underwater vehicles, water conditions and locations

Our results shows that synthetic data can be used to train good-quality detectors without requiring any manually annotated real images. While we focus on object detection, our method could also be applied to semantic segmentation.

The remainder of this paper is organised as follows. Section 2 presents an overview of work using synthetic training images and training object detectors with domain adaptation. Section 3 describes our method to generate synthetic training images and train an object detector with domain adaptation. Section 4 describes the application of the method to the detection of black spiny urchins in images from AUV and ROV platforms, while Section 5 provides a summary of the insights from this method followed by concluding remarks in Section 6.

2 Related work

2.1 Object detection of marine species

Underwater object detection with low numbers of annotations has been improved with enhancements to network architectures, augmentation and pre-processing strategies (Liu et al., 2023, 2020; Israk Ahmed et al., 2024). Recent work has addressed the lack of labelled data for training underwater detectors with augmentation

1 https://www.blender.org/

2 https://unity.com/

3 https://www.unrealengine.com/

using style transfer of different water types as well as a domain alignment step during training (Liu et al., 2020). Israk Ahmed et al. (2024) trained an object detector for green sea urchins using labelled images from an AUV with data augmentation, colour correction and enhancements to the detector architecture. Our method is independent of image pre-processing and detector architecture and could be applied in addition to these enhancements without the need for labelled data.

2.2 Synthetic data

Synthetic images have been generated for other domains to address the lack of labelled data using 3D modelling software such as Unity gaming engine, Unreal Engine and Blender providing photorealistic renderings of scenes (Saini et al., 2022; Ebadi et al., 2022; Diamanti et al., 2024). The Sim10k dataset (Johnson-Roberson et al., 2017) provides 10,000 images of street scenes captured from a car dashboard generated from the Grand Theft Auto video game. Mayer et al. (2016) created large-scale datasets generated from Blender to train optical flow models that performed well on real video data. Procedural generation of assets and scenes using Blender's geometry nodes and Python interface provided randomisation in synthetic images (Santos et al., 2024).

Synthetic data has also been created to train neural networks for underwater image tasks (Lin et al., 2023; Sans-Muntadas et al., 2022; Zwilgmeyer et al., 2021). OysterNet (Lin et al., 2023) created simulated oyster reefs using Blender followed by an image-to-image translation step to provide realistic water effects for a segmentation task. Sans-Muntadas et al. (2022) used simulated underwater images with a simple box structure followed by an image-to-image translation step to train a segmentation model for robotic localisation. The effect of water on light was modelled using Blender's volume shaders to simulate absorption and scattering (Sans-Muntadas et al., 2022; Zwilgmeyer et al., 2021; Diamanti et al., 2024). Our framework generates a more visually complex and varied 3D model of the scene and objects to be detected than previous work, combined with simulated water effects with domain adaptation methods that do not require a separate image-to-image translation network.

2.3 Domain adaptation

Unsupervised Domain Adaptation (UDA) methods update a model during training to reduce the domain gap between labelled source data and unlabelled target data. UDA for classification tasks commonly update feature representations to be domain-invariant at an image level using adversarial learning (Ganin et al., 2016; Tzeng et al., 2017) or other alignment strategies (Baktashmotlagh et al., 2013; Sun and Saenko, 2016). Domain adaptation object detection (DAOD) uses several approaches to reduce the domain gap including aligning feature distributions (Saito et al., 2019; Chen et al., 2018, 2021) and pseudo-labelling (Chen et al., 2022; Zhu et al., 2023; Maurya et al., 2023). We use ALDI++ (Kay et al., 2025) which is state-of-the-art for the synthetic-to-real object detection benchmark between Sim10k and Cityscapes (Cordts et al., 2016). Semi-supervised training is a related training approach to UDA as it has the same training inputs of small amounts of labelled data and larger amounts of unlabelled training data (Zhang et al., 2021). Mean Teacher (Tarvainen and Valpola, 2017) uses pseudo-labelling to reduce the domain gap and can be applied to other object detector architectures and other tasks such as semantic segmentation (Zhou et al., 2023). We have used Mean Teacher as an additional method to reduce the domain gap.

3 Method

Our method uses synthetic training data generated from a Blender-based framework followed by a training pipeline using UDA to create

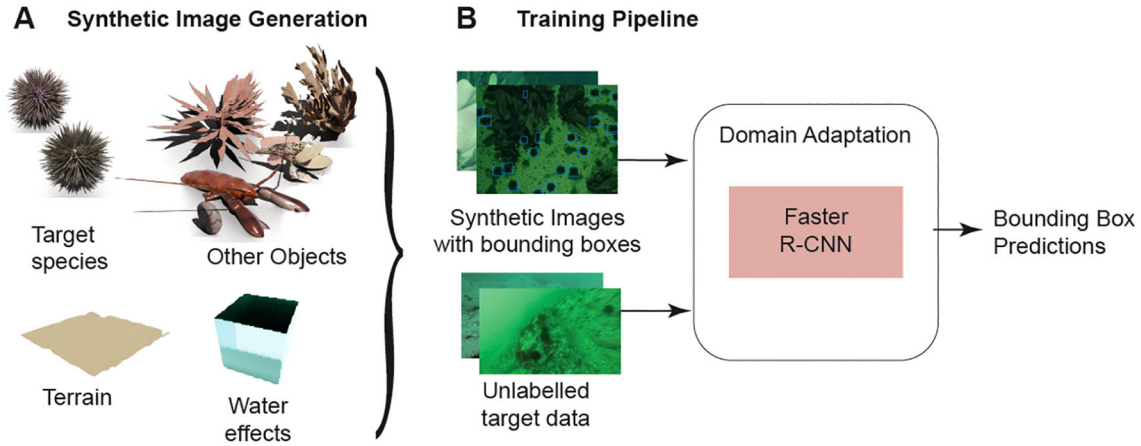

FIGURE 1
Overview of our method. **(A)** 3D models of marine objects, terrain and water are combined to create a unique scene for generating images and annotations. The training pipeline **(B)** uses the labelled synthetic data and real unlabelled target images to train the object detector with domain adaptation.

an object detector with good performance for detecting a target marine species. An overview of our method is shown in Figure 1.

3.1 Synthetic data generation

Synthetic training data is generated to provide a large volume of annotated data to train an object detector to address the lack of annotated real images. Images and annotations are generated using the Blender-based framework, Infinigen (Raistrick et al., 2023), with our enhancements for the benthic environment. The framework generates large numbers of realistic images taken from a simulated camera rig from a vehicle like an AUV or ROV, following a typical survey path. Blender provides physically-based rendering (PBR) (Cornetto and Suway, 2019) of light which can be combined with Blender's Absorption Volume and Scattering Volume shaders to simulate absorption and scattering of light underwater. Figure 2 shows examples of scenes with and without the water effects provided by the Volume shaders. Bounding box annotations, semantic segmentation masks and depth maps are also generated by the framework.

Based on Blender, Infinigen generates realistic natural scenes using procedural generation, allowing an infinite variety of natural terrain and assets to increase the variety of training data (Raistrick et al., 2023). Our method enhances Infinigen with a simulated camera rig that follows a 'mow-the-lawn' survey track over the scene, similar to paths used in AUV missions. The camera features vehicle-mounted lighting, lens distortion based on real AUV camera calibration (Williams et al., 2012), sensor noise and motion blur. Updated natural and new man-made assets have also been added (see Supplementary Material for examples). Blender v3.6.0 has been used with Infinigen v1.6.6. The code for enhancements for benthic scenes is available on GitHub[4].

To simulate the effect of water on light, we have used Blender's Volume Absorption shader and Volume Scattering shader. This is based on previous work that uses the 'Principled Volume Shader' (Zwilgmeyer et al., 2021; Sans-Muntadas et al., 2022; Diamanti et al., 2024), which combines the Absorption and Scattering shaders but does not allow all variables to be set. Using the separate Volume Absorption and Volume Scattering shader allows density to be defined for each shader giving more control over the amount of absorption and scattering instead of using one value for the density of both. Figure 2 gives examples with two different sets of parameters for the absorption and scattering shaders at the upper and lower ranges of values used. For each scene, a random value between the lower and upper thresholds is set for each parameter.

3.2 Object detector training

Our method uses the two-stage object detection architecture Faster R-CNN. The detector is trained using both labelled synthetic images and unlabelled real images from underwater surveys. Labelling provides a tight-fitting bounding box around the species or object of interest with a classification label. For this work, we have one target species being classified.

The domain gap between the generated synthetic data and the real images is reduced using domain adaptation with unlabelled real images. Following the naming conventions from domain adaptation methodologies, the labelled training data is called the *source*, and the unlabelled data is called the *target*. A state-of-the-art UDA method called ALDI++ by Kay et al. (2025) has been used to reduce the domain gap between the synthetic images and the target images. In addition, a semi-supervised training method called Mean Teacher by Tarvainen and Valpola (2017) has also been used. Mean Teacher is a flexible semi-supervised approach that can be applied to other detection architectures and is a state-of-the-art method for semi-supervised semantic segmentation (Tarvainen and Valpola, 2017). These approaches to reducing the domain gap have the advantage of not needing to train an additional network for image-to-image translation as used in Lin et al. (2023) and Sans-Muntadas et al. (2022).

If some labelled target data is available, the synthetic images can be used to augment the labelled target dataset. In this scenario, the object detector can be trained without domain adaptation. This method can be applied to any object detection architecture.

4 Experiment and results

We have demonstrated our method on the detection of black spiny urchins and performed evaluation on two publicly available datasets of images taken from AUVs and ROVs.

4.1 Datasets

Synthetic data A dataset of 2,406 synthetic images with 23,248 annotations of urchins was generated using the enhanced Infinigen framework. Images were generated from an ROV setup with variations in the angle of the camera and altitudes and an AUV-like configuration with a downward-looking camera at around 2m altitude. The scenes provide a variety of benthic terrain and marine assets such as kelp, seaweed, seashells and fish. Water absorption and scattering were varied in each scene produced. Figure 3 provides examples of the generated synthetic images with bounding box annotations. The synthetic images and annotations are available at https://huggingface.co/datasets/hdoi5324/SyntheticUrchin.

Target data Two target datasets with black spiny urchins have been used to test the ability to train a good-quality detector with synthetic data. Each dataset has a training and test split to provide evaluation with unseen data. Table 1 provides the numbers of images and annotations in each dataset. The first dataset is UDD (Underwater Detection Dataset) (Liu et al., 2021a) with images captured by divers and ROVs in an open-sea farm with sea urchins, sea cucumbers and scallops. Only the sea urchin annotations are used.

The second urchin dataset, called IMOS AUV, is from AUV surveys off the coast of Tasmania, Australia, performed as part of the Integrated Marine Observing System[5] (IMOS). The IMOS AUV

4 https://github.com/hdoi5324/infinigenBenthic

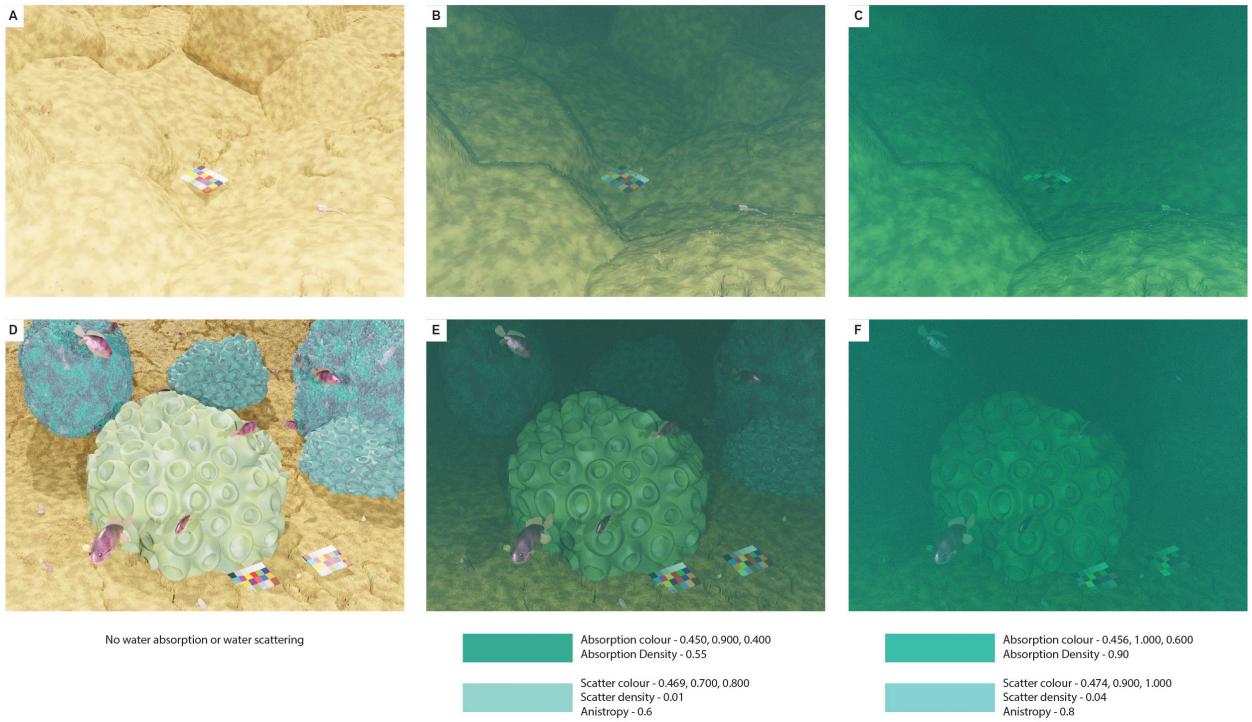

FIGURE 2
Synthetic images showing two scenes with no absorption or scattering **(A, D)** then two different combinations of absorption and scattering **(B, C, E, F)**. The absorption and scattering parameters are shown below the images with colour shown in hue, saturation, value (HSV).

dataset has training images from 2009 and test images from 2011 to investigate whether the model can generalise to images from a different time period. The AUV images and annotations are available at the underwater image repository Squidle+[6].

4.2 Evaluation

To validate our method, we follow other UDA methodologies (Deng et al. (2021); Chen et al. (2021); Kay et al. (2025); Chen et al. (2022)), by comparing performance of models trained with our method to an *oracle*, which is a reference upper limit of performance from a fully-supervised model trained with available labelled training target data. The base model refers to a model trained without any domain adaptation. A separate test split of the target data is used for evaluation using the average precision for detection with an intersection over union (IoU) of 50 or AP50. Performance is measured using the model from the final iteration of training. Due to the scarcity of labelled training data, there is no separate validation dataset for selecting the best-performing model from an earlier iteration.

4.3 Training setup and results

In this section, we describe the setup for the object detector training. We use a two-stage object detector, Faster-RCNN, which has state-of-the-art performance for UDA (Kay et al., 2025; Chen et al., 2018, 2021). Our implementation is based on ALDI++ and the Detectron2 framework (Kay et al., 2025; Wu et al., 2019). In all experiments we use the same training parameters as Detectron2 Faster-RCNN training with changes noted below.

We use a total batch size of 16 with two GPUs and an initial learning rate of 0.02. The total iterations are 9000, which is one-tenth of the iterations used by Detectron2 due to the smaller size of the datasets for this research. We use a ResNet-50 with FPN backbone (Lin et al., 2017) initialised with weights from pre-training on the ImageNet dataset (Deng et al., 2009). Weak augmentation uses a horizontal and vertical flip, and in addition, strong augmentation uses random adjustment to brightness, contrast and saturation, and random blurring. The training was performed on Nvidia P100 GPUs.

Results are provided in Figure 4 for base models trained only with synthetic data, models trained with synthetic data and unlabelled target data using Mean Teacher and ALDI++ and finally, the fully-supervised oracle model. Base models are trained with weak augmentation and strong augmentation. Training with strong augmentation uses a student-teacher configuration updating the weights of the teacher model using an exponential moving average (EMA) of the student model's weights based on Laine and

5 https://imos.org.au/
6 https://squidle.org/

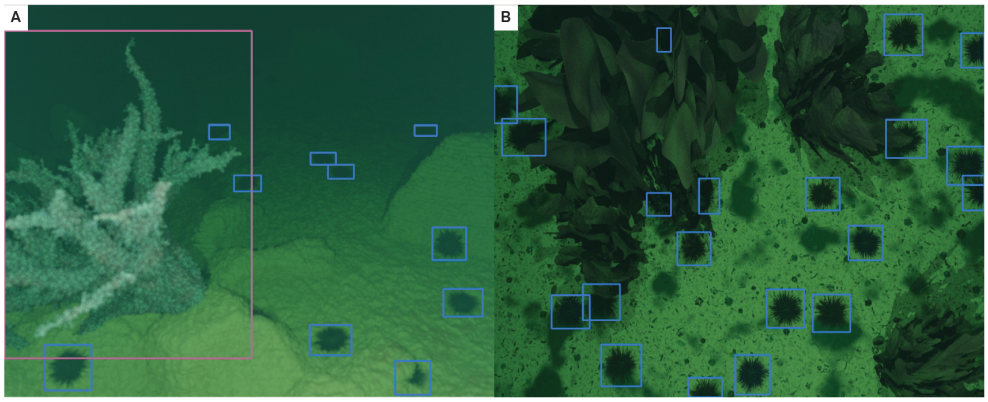

FIGURE 3
Examples of synthetically generated images from two underwater scenes with urchins, coral and kelp with bounding box annotations (blue for urchins, pink for coral). **(A)** shows a forward-looking view and **(B)** shows a downward-looking view.

Aila (2017) (Strong Augmentation with EMA). Mean Teacher and ALDI++ start training with the base model with strong augmentation with EMA. The oracle is also trained with strong augmentation with EMA. Table 2 shows the AP50 for each model trained for the UDD and IMOS AUV datasets. Figure 5 provides annotated examples from each dataset.

4.4 Augmenting labelled target data with synthetic data

The labelled synthetic data can also be used to augment labelled target data to create a larger and more varied training dataset. Synthetic data was added to a subset of labelled real data and trained using strong augmentation with EMA. The model from the final iteration of training is used for evaluation. The results were averaged from three training runs due to the variation when training with small amounts of data (Gao et al., 2022). Figure 6 shows AP50 for training with increasing amounts of labelled target data with and without synthetic data with the standard deviation shown around the line. The UDD datasets benefited the most from the addition of synthetic data achieving higher performance than the oracle. The IMOS AUV dataset benefited from the synthetic augmentation data only when there were less than 200 labelled target images in the training dataset.

TABLE 1 Dataset counts for images and annotations.

Dataset	Train		Test	
	Images	Annotations	Images	Annotations
Synthetic	2,406	23,248	n/a	n/a
UDD	1,707	10,652	400	2,940
IMOS AUV	1,462	6,595	220	650

n/a, not applicable.

5 Discussion

Our method trained an object detector for a marine species without any manually labelled training data. Performance using synthetic data with Mean Teacher adaptation was within 13.9% and 8.6% of the oracle performance for UDD and IMOS AUV datasets, respectively (See Table 2). Strong augmentation with student-teacher training updating weights with EMA (Strong Augmentation with EMA in Table 2) gave the largest increase in AP50, followed by domain adaptation methods. Strong augmentation is also part of the Mean Teacher and ALDI++ methods. These results may be further improved with other strong augmentation strategies (e.g., mixup (Zhang et al., 2018)) or improvements to image pre-processing and detector architecture (Israk Ahmed et al., 2024).

Both domain adaptation methods successfully improved performance with the Mean Teacher method producing a slightly better result than ALDI++. Ideally, when training a neural network, a separate validation dataset is used for hyperparameter tuning and model selection. In our case, we have no labelled target data preventing any model selection. We have used the model from the final training iteration for evaluation. Being able to measure performance without a labelled validation dataset could lead to further increases in performance as seen in domain adaptation for classification (Hu et al., 2023).

TABLE 2 AP50 for urchin detection trained with synthetic source data and unlabelled target data.

Training method	UDD	IMOS AUV
Base model - Weak Augmentation	16.0	55.3
Base model - Strong Augmentation with EMA	53.0	83.0
ALDI++ (Kay et al., 2025)	68.5	80.8
Mean Teacher (Tarvainen and Valpola, 2017)	75.0	84.3
Oracle - Strong Augmentation with EMA	86.9	92.3

Also shows base models trained with weak and strong augmentation and the fully supervised oracle model. Highest performing model is shown in bold.

10.3389/fmars.2025.1581778

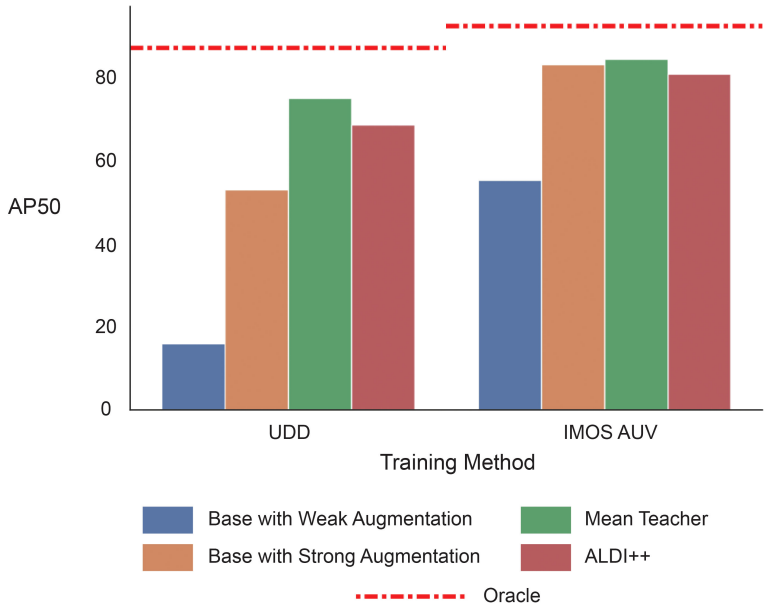

FIGURE 4
AP50 for urchin detection for UDD and IMOS AUV datasets using different training methods. The base models are only trained with synthetic data, while Mean Teacher and ALDI++ also use unlabelled target data. The oracle is the performance from the fully supervised model for reference.

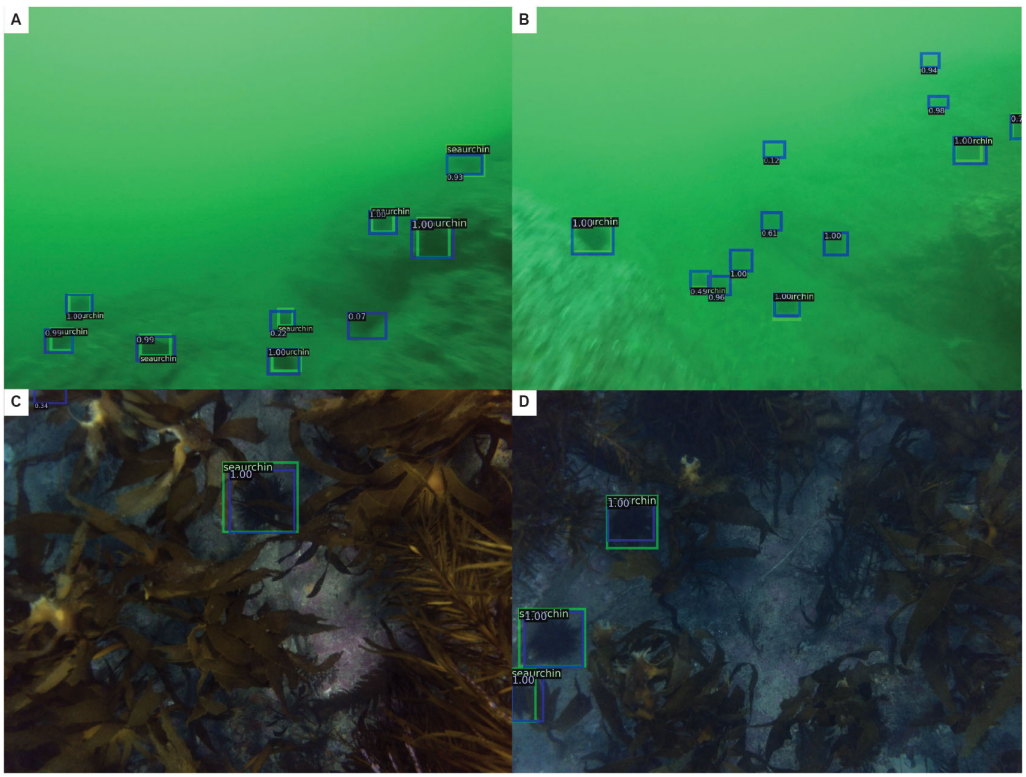

FIGURE 5
Examples of the target images with ground truth shown with a green border with the label 'seaurchin' and predictions in blue with the probability. **(A, B)** are from the UDD dataset, and **(C, D)** are from the IMOS AUV dataset.

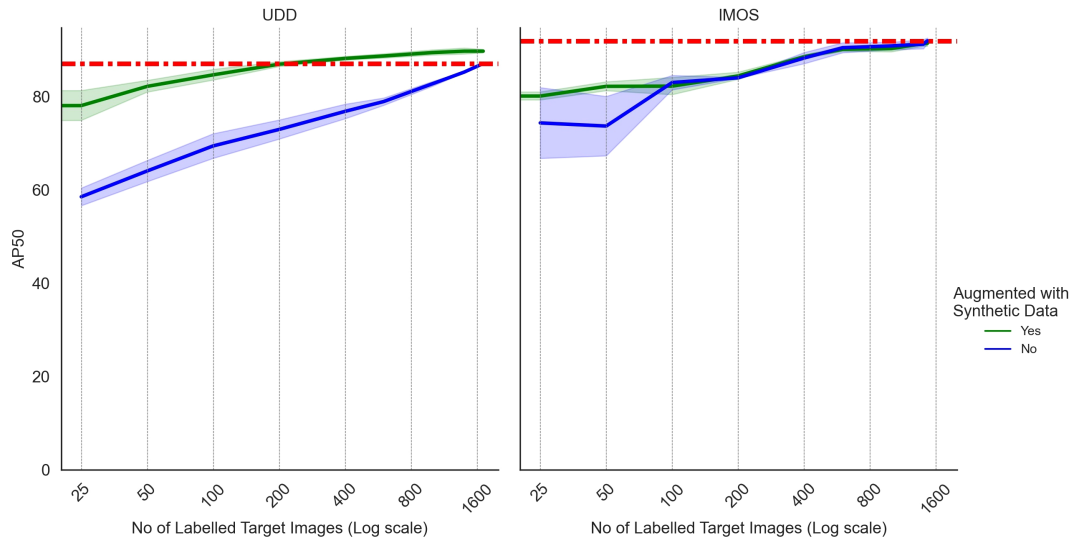

FIGURE 6
Augmenting labelled target data with synthetic data: AP50 when using synthetic data to augment a subset of labelled target data for UDD and IMOS AUV datasets. Number of labelled target images shown with $\log_2$ scale to emphasise the differences at lower numbers of labelled images where we see the most significant differences between the performance. Models are trained with strong augmentation with EMA, and performance is measured on the model from the last iteration of training. Results are averaged from three training runs and the standard deviation is shown in lighter colour around the line. The oracle performance is shown by the dashed orange line.

While scarce labelled target data could be used for model selection during training, it could also be used as training data augmented with synthetic data. Augmentation with synthetic data was beneficial for both datasets when training with less than 100 labelled images, as shown in Figure 6. For the UDD dataset, training data augmented with synthetic data also produced better performance with any amount of labelled target data. The UDD images had more cases of blurry images and low clarity, turbid water with a strong yellow-green tinge (See Figure 5). The IMOS AUV images were generally better lit with and had little blurring. For the UDD dataset, the improvement from synthetic data augmentation may have continued with increasing amounts of labelled target data as the synthetic data behaved like a colour space augmentation (Shorten and Khoshgoftaar, 2019), training the model to ignore colour.

The species or objects used in the underwater scenes could be generated based on physical examples. 3D Blender models of marine species and man-made structures have been generated for volumetric measurements and simulations (Zhang et al., 2023; Adamczak et al., 2019; Diamanti et al., 2024). Scanning images or a physical object to generate a 3D model in Blender could extend the variety and availability of realistic assets to place within an underwater scene. Using these models for the procedural generation of assets could provide a range of realistic variations.

6 Conclusion

Our method has trained a high-performing detector of marine species in images from photographic surveys using generated synthetic labelled data combined with domain adaptation during training. The synthetic images are generated with a flexible framework based on Infinigen with variations in water conditions and camera configurations. New marine species or man-made objects could be added by creating new procedural models based on existing assets or scans of images or physical objects. Blender's modelling capability would allow for a variety of marine species to be created, including rare species with few real images available. Future research into addressing low annotations in benthic images could investigate performing model selection without a labelled validation dataset and how synthetic data can be a successful augmentation strategy.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

HD: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. OP: Resources, Supervision, Writing – review & editing. SW: Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

The object detection training was performed on resources provided by Sigma2 - the National Infrastructure for High-Performance Computing and Data Storage in Norway. The IMOS dataset was sourced from Australia's Integrated Marine Observing System (IMOS) – IMOS is enabled by the National Collaborative Research Infrastructure Strategy (NCRIS). It is operated by a consortium of institutions as an unincorporated joint venture, with the University of Tasmania as Lead Agent.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2025.1581778/full#supplementary-material

References

Adamczak, S. K., Pabst, A., McLellan, W. A., and Thorne, L. H. (2019). Using 3D models to improve estimates of marine mammal size and external morphology. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00334

Baktashmotlagh, M., Harandi, M. T., Lovell, B. C., and Salzmann, M. (2013). "Unsupervised domain adaptation by domain invariant projection," in *Proceedings of the IEEE International Conference on Computer Vision*. (Sydney, Australia: IEEE) 769–776. doi: 10.1109/ICCV.2013.100

Becktor, J., Schöller, F. E. T., Boukas, E., Blanke, M., and Nalpantidis, L. (2022). Bolstering maritime object detection with synthetic data. *IFAC-PapersOnLine* 55, 64–69. doi: 10.1016/j.ifacol.2022.10.410

Chen, M., Chen, W., Yang, S., Song, J., Wang, X., Zhang, L., et al. (2022). "Learning domain adaptive object detection with probabilistic teacher," in *Proceedings of the 39th International Conference on Machine Learning*, (Baltimore, Maryland, USA) Vol. 162. 3040–3055.

Chen, Y., Li, W., Sakaridis, C., Dai, D., and Van Gool, L. (2018). "Domain adaptive faster R-CNN for object detection in the wild," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Salt Lake City, UT, USA: IEEE) 3339–3348. doi: 10.1109/CVPR.2018.00352

Chen, Y., Wang, H., Li, W., Sakaridis, C., Dai, D., and Van Gool, L. (2021). Scale-aware domain adaptive faster R-CNN. *Int. J. Comput. Vision* 129, 2223–2243. doi: 10.1007/s11263-021-01447-x

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., et al. (2016). "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Las Vegas, NV, USA: IEEE) 3213–3223. doi: 10.1109/CVPR.2016.350

Cornetto, A. D., and Suway, J. (2019). Validation of the cycles engine for creation of physically correct lighting models. *SAE Tech. papers*. doi: 10.4271/2019-01-1004

Deng, J., Dong, W., Socher, R., Li, L., Li, K., and Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition (IEEE)*. (Miami, FL, USA: IEEE) 248–255. doi: 10.1109/CVPR.2009.5206848

Deng, J., Li, W., Chen, Y., and Duan, L. (2021). "Unbiased mean teacher for cross-domain object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Nashville, TN, USA: IEEE) 4089–4099. doi: 10.1109/CVPR46437.2021.00408

Denninger, M., Winkelbauer, D., Sundermeyer, M., Boerdijk, W., Knauer, M., Strobl, K. H., et al. (2023). BlenderProc2: A Procedural Pipeline for Photorealistic Rendering. Journal of Open Source Software 8, 4901. doi: 10.21105/joss.04901

Diamanti, E., Yip, M., Stahl, A., and Ødegård, O. (2024). Advancing data quality of marine archaeological documentation using underwater robotics: from simulation environments to realWorld scenarios. *J. Comput. Appl. Archaeology* 7, 153–169. doi: 10.5334/jcaa.147

Ebadi, S. E., Dhakad, S., Vishwakarma, S., Wang, C., Jhang, Y.-C., Chociej, M., et al. (2022). PSP-HDRI+: A synthetic dataset generator for pre-training of human-centric computer vision models. *arXiv preprint arXiv:2207.05025*. doi: 10.48550/arXiv.2207.05025

Er, M. J., Chen, J., Zhang, Y., and Gao, W. (2023). Research challenges, recent advances, and popular datasets in deep learning-based underwater marine object detection: A review. *Sensors* 23. doi: 10.3390/s23041990

Estes, M., Anderson, C., Appeltans, W., Bax, N., Bednarsek, N., Canonico, G., et al. (2021). Enhanced monitoring of life in the sea is a critical component of conservation management and sustainable economic growth. *Mar. Policy* 132. doi: 10.1016/j.marpol.2021.104699

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., et al. (2016). Domainadversarial training of neural networks. *J. Mach. Learn. Res.* 17, 2096–2030. doi: 10.48550/arXiv.1505.07818

Gao, B. B., Chen, X., Huang, Z., Nie, C., Liu, J., Lai, J., et al. (2022). "Decoupling classifier for boosting few-shot object detection and instance segmentation," in *Advances in Neural Information Processing Systems*, (Red Hook, NY, USA: Curran Associates, Inc) Vol. 35.

Hu, D., Liang, J., Liew, J. H., Xue, C., Bai, S., and Wang, X. (2023). "Mixed samples as probes for unsupervised model selection in domain adaptation," in *Advances in Neural Information Processing Systems*, (Red Hook, NY, USA: Thirty-seventh Conference on Neural Information Processing Systems) Vol. 36.

Inoue, N., Furuta, R., Yamasaki, T., and Aizawa, K. (2018). "Cross-domain weakly-supervised object detection through progressive domain adaptation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Salt Lake City, UT, USA: IEEE) 5001–5009. doi: 10.1109/CVPR.2018.00525

Israk Ahmed, M., Peña-Castillo, L., Vardy, A., and Gagnon, P. (2024). Improving detection and localization of green sea urchin by adding attention mechanisms in a convolutional network. *J. Ocean Technol.* 19, 81–97.

Johnson-Roberson, M., Barto, C., Mehta, R., Sridhar, S. N., Rosaen, K., and Vasudevan, R. (2017). "Driving in the Matrix: Can virtual worlds replace human-generated annotations for real world tasks?," in *Proceedings - IEEE International Conference on Robotics and Automation*. (Singapore: IEEE) 746–753. doi: 10.1109/ICRA.2017.7989092

Kay, J., Haucke, T., Stathatos, S., Deng, S., Young, E., Perona, P., et al. (2025). Align and distill: unifying and improving domain adaptive object detection. *Trans. Mach. Learn. Res.*

Laine, S., and Aila, T. (2017). "Temporal ensembling for semi-supervised learning," in *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*. (Toulon, France: OpenReview.net)

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. (Honolulu, HI, USA: IEEE) 936–944. doi: 10.1109/CVPR.2017.106

Lin, X., Sanket, N., Karapetyan, N., and Aloimonos, Y. (2023). "OysterNet: Enhanced oyster detection using simulation," in *2023 IEEE International Conference on Robotics and Automation (ICRA) (IEEE)*. (London, United Kingdom: IEEE) 5170–5176. doi: 10.1109/ICRA48891.2023.10160830

Liu, M., Jiang, W., Hou, M., Qi, Z., Li, R., and Zhang, C. (2023). A deep learning approach for object detection of rockfish in challenging underwater environments. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1242041

Liu, J., Kusy, B., Marchant, R., Do, B., Merz, T., Crosswell, J., et al. (2021b). The CSIRO crown-of-thorn starfish detection dataset. *arXiv preprint arXiv:2111.14311*. doi: 10.48550/arXiv.2111.14311

Liu, C., Li, H., Wang, S., Zhu, M., Wang, D., Fan, X., et al. (2021a). "A dataset and benchmark of underwater object detection for robot picking," in *2021 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2021*. doi: 10.1109/ICMEW53276.2021.9455997

Liu, H., Song, P., and Ding, R. (2020). "Towards domain generalization in underwater object detection," in *Proceedings - International Conference on Image Processing, ICIP*. (Abu Dhabi, United Arab Emirates: IEEE) 1971–1975. doi: 10.1109/ICIP40778.2020.9191364

Liu, C., Wang, Z., Wang, S., Tang, T., Tao, Y., Yang, C., et al. (2022). A new dataset, poisson GAN and aquaNet for underwater object grabbing. *IEEE Trans. Circuits Syst. Video Technol.* 32, 2831–2844. doi: 10.1109/TCSVT.2021.3100059

Lu, Q., Jing, Y., and Zhao, X. (2023). Bolt loosening detection using key-point detection enhanced by synthetic datasets. *Appl. Sci. (Switzerland)* 13. doi: 10.3390/app13032020

Maurya, J., Ranipa, K. R., Yamaguchi, O., Shibata, T., and Kobayashi, D. (2023). "Domain adaptation using self-training with mixup for one-stage object detection," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. (Waikoloa, HI, USA: IEEE) 4178–4187. doi: 10.1109/WACV56688.2023.00417

Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., et al. (2016). "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Las Vegas, NV, USA: IEEE) 4040–4048. doi: 10.1109/CVPR.2016.438

Munir, M., Khan, M., Sarfraz, M., and Ali, M. (2023). Domain adaptive object detection via balancing between self-training and adversarial learning. *IEEE Trans. Pattern Anal. Mach. Intell.* doi: 10.1109/TPAMI.2023.3290135

Oza, P., Sindagi, V. A., Vs, V., and Patel, V. M. (2024). Unsupervised domain adaptation of object detectors: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 46, 4018–4040. doi: 10.1109/TPAMI.2022.3217046

Peñarroya, P., Pugliatti, M., Ferrari, F., Centuori, S., Topputo, F., Vetrisano, M., et al. (2023). CubeSat landing simulations on small bodies using blender. *Adv. Space Res.* 72, 2971–2993. doi: 10.1016/j.asr.2022.07.044

Peng, F., Miao, Z., Li, F., and Li, Z. (2021). S-FPN: A shortcut feature pyramid network for sea cucumber detection in underwater images. *Expert Syst. Appl.* 182, 115306. doi: 10.1016/j.eswa.2021.115306

Perkins, N. R., Monk, J., Soler, G., Gallagher, P., and Barrett, N. S. (2022). Bleaching in sponges on temperate mesophotic reefs observed following marine heatwave events. *Climate Change Ecol.* 3. doi: 10.1016/j.ecochg.2021.100046

Raistrick, A., Lipson, L., Ma, Z., Mei, L., Wang, M., Zuo, Y., et al. (2023). "Infinite photorealistic worlds using procedural generation," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Vancouver, BC, Canada: IEEE) 12630–12641. doi: 10.1109/CVPR52729.2023.01215

Raistrick, A., Mei, L., Kayan, K., Yan, D., Zuo, Y., Han, B., et al. (2024). "Infinigen indoors: photorealistic indoor scenes using procedural generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Seattle, WA, USA: IEEE)

Saini, N., Bonetto, E., Price, E., Ahmad, A., and Black, M. J. (2022). Airpose: Multi-view fusion network for aerial 3d human pose and shape estimation. *IEEE Robotics Automation Lett.* 7, 4805–4812. doi: 10.1109/LRA.2022.3145494

Saito, K., Ushiku, Y., Harada, T., and Saenko, K. (2019). "Strong-weak distribution alignment for adaptive object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Long Beach, CA, USA: IEEE) 6956–6965. doi: 10.1109/CVPR.2019.00712

Sans-Muntadas, A., Skaldebo, M. B., Nielsen, M. C., and Schjolberg, I. (2022). Unsupervised domain transfer for task automation in unmanned underwater vehicle intervention operations. *IEEE J. Oceanic Eng.* 47(2), 312–321. doi: 10.1109/JOE.2021.3126016

Santos, A. N. P., Magboo, M. S. A., and Magboo, V. P. C. (2024). "Procedural modeling for sustainable urban development and planning: A blender plugin for 3D modeling of philippine cities (Springer nature Singapore)," in *Proceedings of the 4th International Conference on Advances in Computational Science and Engineering*. (Singapore: Springer Nature Singapore) 81–97.

Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *J. Big Data* 6. doi: 10.1186/s40537-019-0197-0

Stuart-Smith, J., Edgar, G., Last, P., Linardich, C., Lynch, T., Barrett, N., et al. (2020). Conservation challenges for the most threatened family of marine bony fishes (handfishes: Brachionichthyidae). *Biol. Conserv.* 252, 108831. doi: 10.1016/j.biocon.2020.108831

Sun, B., and Saenko, K. (2016). "Deep CORAL: Correlation alignment for deep domain adaptation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 443–450. doi: 10.1007/978-3-319-49409-835

Tarvainen, A., and Valpola, H. (2017). "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in Neural Information Processing Systems*. (Red Hook, NY, United States: Curran Associates Inc.) 1196–1205.

Tzeng, E., Hoffman, J., Saenko, K., and Darrell, T. (2017). "Adversarial discriminative domain adaptation," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. (Honolulu, HI, USA: IEEE) 2962–2971. doi: 10.1109/CVPR.2017.316

Williams, S. B., Pizarro, O. R., Jakuba, M. V., Johnson, C. R., Barrett, N. S., Babcock, R. C., et al. (2012). Monitoring of benthic reference sites: Using an autonomous underwater vehicle. *IEEE Robotics Automation Magazine* 19, 73–84. doi: 10.1109/MRA.2011.2181772

Wu, Y., Kirillo, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). Detectron2. Available online at: https://github.com/facebookresearch/detectron2. (Accessed July 1, 2024).

Xu, S., Zhang, M., Song, W., Mei, H., He, Q., and Liotta, A. (2023). A systematic review and analysis of deep learning-based underwater object detection. *Neurocomputing* 527, 204–232. doi: 10.1016/j.neucom.2023.01.056

Zhang, H., Luo, G., Li, J., and Wang, F. Y. (2022). C2FDA: coarse-to-fine domain adaptation for traffic object detection. *IEEE Trans. Intelligent Transportation Syst.* 23, 12633–12647. doi: 10.1109/TITS.2021.3115823

Zhang, H., Moustapha Cisse, Y., and Lopez-Paz, D. (2018). "mixup: Beyond empirical risk minimization," in *The International Conference on Learning*, (Vancouver, BC, Canada: OpenReview.net) Vol. 904. doi: 10.48550/arXiv.1710.09412

Zhang, Y., Zhang, H., Deng, B., Li, S., Jia, K., and Zhang, L. (2021). Semi-supervised models are strong unsupervised domain adaptation learners. *arXiv preprint arXiv:2106.00417*. doi: 10.48550/arXiv.2106.00417

Zhang, C., Zhou, H., Christiansen, F., Hao, Y., Wang, K., Kou, Z., et al. (2023). Marine mammal morphometrics: 3D modeling and estimation validation. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1105629

Zhou, H., Jiang, F., and Lu, H. (2023). SSDA-YOLO: Semi-supervised domain adaptive YOLO for cross-domain object detection. *Comput. Vision Image Understanding* (Detroit, MI, USA: IEEE) 229. doi: 10.1016/j.cviu.2023.103649

Zhu, Y., Guo, P., Wei, H., Zhao, X., and Wu, X. (2023). "Disentangled discriminator for unsupervised domain adaptation on object detection," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 5685–5691. doi: 10.1109/IROS55552.2023.10341878

Zwilgmeyer, P., Yip, M., Teigen, A., Mester, R., and Stahl, A. (2021). "The VAROS synthetic underwater data set: Towards realistic multi-sensor underwater data with ground truth," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (Montreal, BC, Canada: IEEE) 3722–3730. doi: 10.1109/ICCVW54120.2021.00415