

Visual Anomaly Detection for Bycatch Detection onboard Fishing Vessels

Stefan Hein Bengtson and **Malte Pedersen**

The Visual Analysis and Perception Lab, Aalborg University

The Pioneer Centre for AI

{shbe,mape}@create.aau.dk

Mathias Sogaard and **Claus Reedtz Sparrevohn**

Danish Pelagic Producers Organisation, Denmark

{ms,crs}@pelagisk.dk

Kamal Nasrollahi

Milestone Systems, Denmark

The Visual Analysis and Perception Lab, Aalborg University

KNA@milestone.dk

Abstract

With increasingly fragile marine environments, accurately documenting bycatch on fishing vessels is becoming ever more important for conducting informed and sustainable fisheries management. In this study, we investigate the feasibility of using visual anomaly detection (VAD) for bycatch detection using pre-installed *general-purpose* electronic monitoring cameras. We introduce a novel VAD dataset on which we achieve a mean AUROC of 0.79 with consistent performance across different types of bycatch, underscoring the potential of applying anomaly detection to automate bycatch detection.

Keywords: Visual Anomaly Detection, Electronic Monitoring

1 Introduction

In recent years, there has been an increase in regulatory requirements for fishing vessels to implement electronic monitoring (EM) systems, reflecting a growing emphasis on transparency and sustainable fisheries management. EM systems typically include multiple cameras strategically placed to visually cover the entire vessel, which allow authorities to conduct offline fisheries control through manual visual inspection of the EM recordings. This is not a scalable solution and only few percentages of the recordings are currently inspected.

In this paper, we explore the feasibility of using visual anomaly detection (VAD) for bycatch detection on pelagic trawlers to reduce the time needed for manual inspection. The major challenges are similar to those in other VAD domains: abnormal examples are rare and ill-defined, whereas normal samples are abundant. What differs, however, is that a target species in one fishing trip may appear as bycatch in the next. Moreover, the environment is highly variable and noisy, with changes in weather, lighting, and day–night cycles, making it a novel and interesting VAD application.

2 Methods

We compiled a dataset of images from a single EM CCTV camera mounted above the dewatering station on a pelagic trawler. The data was collected across eight trips with the target species being either herring or whiting. The camera was configured to capture 15 frames per second with a resolution of 1280x720 pixels. The remaining camera parameters are unknown, and the data contains issues like motion blur and compression artifacts due to suboptimal settings.

A training split was gathered by randomly sampling 10,000 images for each trip, for a total of 80,000 images. The evaluation split consists of manually chosen short sequences from the same eight trips, where bycatch occurs 10–80% of the time. The sequences range between ≈ 6 –33 seconds and are sampled with the full 15 FPS, resulting in ≈ 100 – 500 images per sequence. Each sequence is annotated on image-level using the binary labels: *normal* or *bycatch*.

The evaluation set is divided into two splits based on the bycatch type: *sharks* and *redfish*, as shown in Table 1. This distinction is made because these two types of bycatch differ significantly in appearance. *Sharks* are typically larger, with irregular shapes and muted grayish tones, and usually appear alone. In contrast, *redfish* are smaller, with rounder shapes and vivid reddish coloration, and multiple individuals often appear in the same image (see Figure 2).

Split	Sequences	Normal	Bycatch	Total
Redfish	29	2697	1312	4009
Sharks	9	1035	215	1250
Total	38	3732	1527	5259

Table 1: Evaluation dataset composition in terms of sequences, normal images, bycatch counts, and totals.

Five recent VAD models were evaluated: Dinomaly Guo et al. [2025], UniNet Wei et al. [2025], EfficientAd Batzner et al. [2024], Uflow Tailanian et al. [2024], and ReverseDistillation Deng and Li [2022]. The models were trained in an unsupervised manner, for a total of 10 epochs with a batch size of 32, with two exceptions: EfficientAd used a batch size of 1 and Uflow only trained for 5 epochs. All models were trained using the implementation and default settings found in Anomalib Akcay et al. [2022].

Perspective correction and cropping was applied to all images prior to feeding them into the models, as the camera was mounted at an angle to the dewatering station. The cropped images were resized to a resolution of 256x256 and normalized. Higher resolutions were initially tried but it did not appear to boost performance. No further augmentations were applied to the images. All of the trained models output anomaly maps, as shown in Figure 2, from which the max-value was extracted and used as the anomaly score for each image.

3 Results

The models are evaluated using the area under the receiver operating characteristic (AUROC) curve Fawcett [2006], which is calculated based on the anomaly scores derived from the anomaly maps. The area under the curve is used as it evaluates the performance of the models across different thresholds. The AUROC score is calculated separately for each sequence in the evaluation dataset, and reported for the different splits in Figure 1.

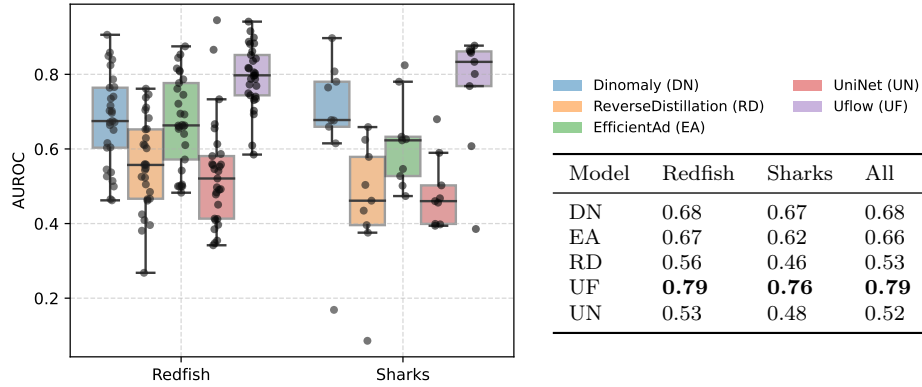


Figure 1: Boxplot: distribution of the AUROC score for each sequence across the two splits. Table: summary of the average AUROC for each model across the different data splits.

The results suggest that Uflow generally outperforms the other models, across both splits, but especially for the *sharks*-split. Furthermore, it is observed that both Uflow and Dinomaly performs equally well across the two splits, whereas the other models performs worse on the *shark*-split. This suggests that some models possibly rely more on distinct colors and have difficulties with objects displaying muted colors similar to the background, such as the sharks.

Examples of the anomaly maps produced by the trained models can be seen in Figure 2, along with the overall AUROC score for the evaluation sequence the image originates from. The anomaly maps are generally less ambiguous with clear responses to the presence of bycatch for high AUROC scores and vice versa. In the highlighted examples, Uflow appears to produce anomaly maps with stronger and more concentrated responses than the rest of the models. Dinomaly also produce strong and concentrated responses, but only for half of the depicted examples. Other models, such as ReverseDistillation and especially EfficientAd, also have strong responses but they are more scattered and also occurs in areas of the image with no bycatch. Finally, UniNet exhibits a weak response to the presence of bycatch and the only strong response (top right corner of the image in the last row) does not align with the location of the bycatch in the image.

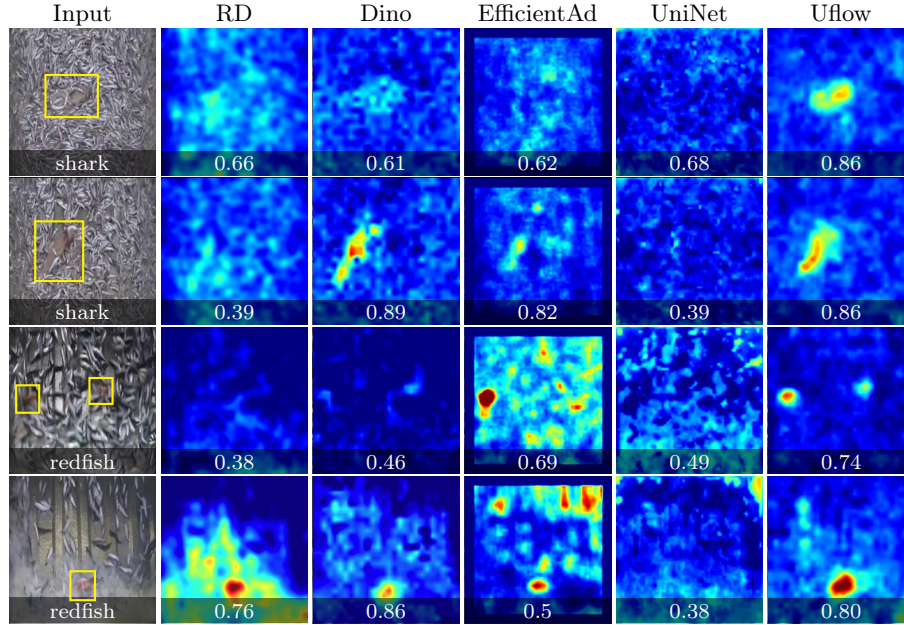


Figure 2: Examples of anomaly maps across different models. The overlaid score on each image indicates the AUROC for the sequence from which it originates.

4 Conclusions

Automated bycatch detection using recordings from *general-purpose* EM cameras is challenging; however, our results suggest that visual anomaly detection can be a promising approach. The best-performing model, Uflow Tailanian et al. [2024], achieved an average AUROC of 0.79 across 38 sequences while maintaining consistent performance across bycatch types. Other models often struggled with sharks, likely due to their muted coloration.

Qualitative inspection of the anomaly maps revealed that a high anomaly responses do not always correspond to actual bycatch. This was particularly evident for ReverseDistillation Deng and Li [2022] and EfficientAd Batzner et al. [2024]. These observations suggest that evaluation may benefit from a pixel-level approach that accounts for the spatial distribution of anomaly responses, rather than relying solely on a binary image-level normal/abnormal classification. Additionally, the method used to interpret the anomaly maps may influence the apparent performance of the VAD models. In this study, each anomaly map was reduced to a single value by taking its maximum response. While this simplification facilitates analysis and comparison across images, it may affect the results negatively by discarding potentially important spatial information.

Overall, this work highlights the potential of using VAD for bycatch detection and offers suggestions for improvement, including pixel-level evaluation and spatially informed interpretation of anomaly maps.

References

- S. Akcay, D. Ameln, A. Vaidya, B. Lakshmanan, N. Ahuja, and U. Genc. Anomalib: A deep learning library for anomaly detection. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1706–1710. IEEE, 2022.
- K. Batzner, L. Heckler, and R. König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 127–137, 2024.
- H. Deng and X. Li. Anomaly detection via reverse distillation from one-class embedding. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9727–9736, 2022.
- T. Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8): 861–874, 2006.
- J. Guo, S. Lu, W. Zhang, F. Chen, H. Li, and H. Liao. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 20405–20415, 2025.
- M. Tailanian, A. Pardo, and P. Musé. U-flow: A u-shaped normalizing flow for anomaly detection with unsupervised threshold. *Journal of Mathematical Imaging and Vision*, May 2024. ISSN 1573-7683.
- S. Wei, J. Jiang, and X. Xu. Uninet: A contrastive learning-guided unified framework with feature selection for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2025.