

One-Click Annotation for Radar Object Detection and Tracking

J.C. Vaught¹, Yi Wang¹, Douglas Cahl¹

¹Department of Mechanical Engineering, University of South Carolina, Columbia, SC, USA

Email: jvaught@sc.edu

Abstract—Machine learning for marine radar perception requires large volumes of annotated Plan Position Indicator (PPI) data, yet manual annotation of radar imagery is slow, expensive and often inconsistent. Conventional practices rely on drawing bounding boxes or segmentation masks from scratch, which is particularly challenging in the presence of sea clutter, multipath and extended land returns. This paper proposes a practical annotation pipeline for marine radar perception that combines automatic pre-labelling with a lightweight click-to-annotate interface. The pipeline first generates candidate targets from PPI images through radar-specific pre-processing and heuristic or learned detection. Human annotators then confirm, reject or add targets with a small number of clicks, producing center- or box-based labels suitable for object detection and tracking models. The paper has two aims. First, it provides a focused review of annotation strategies and weakly supervised approaches that trade annotation effort for model performance, with emphasis on click- and point-based supervision. Second, it describes and evaluates a pre-labelling and click-to-annotate workflow on marine radar data, measuring annotation effort and downstream detection and tracking quality. Experiments on a navigation radar dataset indicate that click-based correction of pre-labels substantially reduces annotation time per frame while maintaining detection and tracking performance comparable to fully manual bounding-box labels. These results suggest that click-to-annotate pipelines are a promising direction for scalable marine radar datasets.

Index Terms—Marine radar, PPI, object detection, tracking, annotation, click supervision, pre-labelling, weak supervision.

I. INTRODUCTION

Marine navigation radars remain a primary sensor for collision avoidance and situation awareness at sea. Long-range operation, robustness to lighting conditions and resilience to haze or spray make radar a natural complement to cameras and LiDAR. Recent work has begun to apply convolutional and transformer-based detectors to radar PPI images for marine target detection and tracking [?], [?], [?], and multi-sensor maritime datasets such as MOANA have appeared that integrate radar with LiDAR and cameras [?]. In all of these cases, the limiting factor for training deep neural networks is not model capacity but the availability of high-quality labels.

Radar PPI annotation is difficult for several reasons. Target signatures may be small, noisy and elongated; sea clutter can produce bright speckle; land edges appear as extended non-convex structures; and multipath can create ghost returns. The task of drawing tight bounding boxes around such signatures is time consuming, and different annotators may disagree on the spatial extent of a target even when they agree that the target is present. In addition, annotators must mentally interpolate the

two-dimensional radar geometry from a polar display, which adds cognitive load compared to camera imagery.

The computer vision community has devoted considerable effort to reducing annotation cost through alternative supervision strategies. Extreme clicking reduces the problem of drawing a bounding box to placing four clicks on the extremal points of the object [?]. Point supervision for semantic segmentation uses single-pixel annotations rather than dense masks [?]. A rich ecosystem of interactive and click-based segmentation methods has emerged, including FocalClick [?] and efficient mask correction networks [?], which exploit user clicks to refine masks while minimizing the number of interactions. Surveys and recent work in interactive segmentation for remote sensing and medical imaging demonstrate similar trends: carefully designed click interactions can dramatically reduce labelling effort while preserving segmentation quality [?], [?], [?].

In contrast, annotation practices in marine radar perception have received relatively little attention. Works such as YOLO-SWFormer [?], Marine-Faster R-CNN [?] and DTNet [?] describe how targets are annotated, typically using axis-aligned bounding boxes on PPI images, but do not systematically study the cost or reliability of these labels. Public radar datasets designed for tracking or odometry often provide preprocessed point clouds or bounding boxes derived from blob detectors rather than raw click-level annotations [?], [?], [?].

This paper studies how to reduce the cost of annotating marine radar PPI images by combining automatic pre-labelling with a click-to-annotate interface. The central idea is that a relatively simple radar-specific pre-labeller can propose candidate target regions, and human annotators need only confirm or correct them with a handful of clicks. The resulting annotation is naturally expressed as target centers, which can be converted into bounding boxes or used directly in center-based detectors and trackers.

The work reported here has two main contributions. First, it organizes existing ideas on annotation strategies, click-based supervision and weakly supervised learning into a radar-centric discussion that emphasizes their relevance for PPI data. Second, it presents a concrete pre-labelling and click-to-annotate pipeline, including implementation details and experimental evaluation on marine radar sequences, with measurements of annotation time and downstream detection and tracking performance.

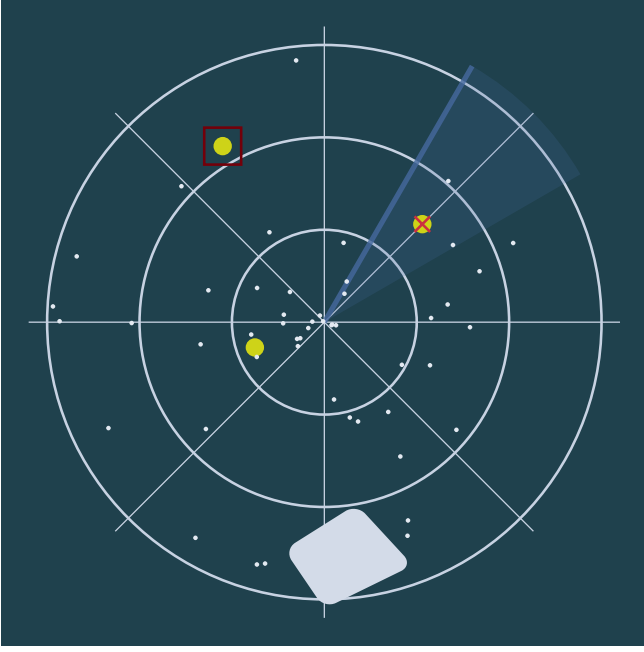


Fig. 1. Placeholder: representative raw PPI frame and corresponding annotations. Replace this with actual visual examples of your radar data and labels.

II. BACKGROUND AND RELATED WORK

A. Annotation Strategies and Weak Supervision

Annotation strategies for visual recognition span a spectrum from coarse to fine supervision. At one extreme, image-level labels specify only the presence of object classes. At the other extreme, pixel-level segmentation masks describe object boundaries with high precision. Between these extremes lie bounding boxes, point labels and various forms of partial or weak supervision.

Table ?? summarizes several representative strategies and their typical trade-offs. Image-level labels are cheap but provide no spatial localization. Bounding boxes are the dominant representation for object detection and are widely used in large-scale datasets. Pixel-level masks are ideal for segmentation but can require an order of magnitude more annotation time per instance than bounding boxes. Point-level supervision occupies a middle ground, supplying sparse location information that can be exploited by loss functions or interactive models to recover dense supervision [?].

Extreme clicking shows that bounding boxes can be obtained five times faster by asking annotators to click object extremal points instead of dragging box corners, while achieving similar detection performance to traditionally drawn boxes [?]. Point supervision for semantic segmentation demonstrates that adding one click per object class per image can significantly improve mean Intersection-over-Union compared to pure image-level supervision at fixed annotation budget [?]. Interactive segmentation methods such as FocalClick [?] and efficient mask correction networks [?] further reduce interaction cost by focusing computation on local regions

TABLE I
REPRESENTATIVE ANNOTATION STRATEGIES AND QUALITATIVE TRADE-OFFS.

Strategy	Example works	Spatial detail	Relative cost
Image-level labels	Classification datasets	Very low	Very low
Bounding boxes	Detection benchmarks, extreme clicking [?]	Medium	Medium
Pixel masks	Segmentation datasets	High	High
Point / click labels	Point supervision [?], interactive segmentation [?], [?]	Low to medium	Low

around new clicks rather than recomputing masks over entire images.

These results suggest that click- and point-based annotations, especially when combined with pre-existing masks or proposals, can deliver much of the benefit of dense supervision at a fraction of the cost. The proposed pipeline applies this principle to marine radar PPI images by using pre-labelling to generate proposals and clicks to verify them.

B. Interactive and Click-Based Segmentation

Deep interactive segmentation models typically augment the input image with one or more click maps that encode positive and negative user clicks as Gaussian blobs or distance transforms. The segmentation network then infers a mask consistent with both the image and the click guidance. FocalClick introduces a coarse-to-fine crop strategy that restricts computation to target and focus regions, achieving competitive accuracy with substantially lower FLOPs and enabling practical annotation tools [?]. Du et al. propose an efficient mask correction network that updates masks in response to new clicks with lightweight inference, rather than rerunning a heavy backbone for each interaction [?].

Interactive segmentation has also been explored in domains beyond natural images. Remote sensing imagery exhibits large context and class imbalance, motivating adaptations of click-based interactive segmentation to satellite images [?]. In medical imaging, IMedSeg and related work develop specialized architectures and mask correction modules that refine local regions around clicks while preserving correct parts of the mask [?]. Across these domains, a recurring theme is that a small number of clicks, combined with strong priors about object appearance and spatial coherence, can generate high-quality masks with significantly reduced annotation effort.

Although the present work does not directly perform interactive segmentation of radar imagery, it borrows the core idea that clicks are a natural unit of human interaction. In the proposed pipeline, clicks serve as confirmations or corrections of pre-labelled targets rather than seeds for pixel-level masks, which simplifies the interface for annotators who may not require exact object outlines for detection and tracking tasks.

TABLE II
REPRESENTATIVE MARINE RADAR DATASETS AND ANNOTATION FORMS.

Dataset	Sensors	Annotation	Notes
DLR tracking radar [?]	Navigation radar	Tracks, point clouds	Real-world campaigns with AIS ground truth
MOANA [?]	X-/W-band radar, LiDAR, cameras	Object labels, odometry	Multi-radar maritime navigation dataset
M_Radar [?]	Marine radar	Bounding boxes	Public PPI image dataset for YOLO methods

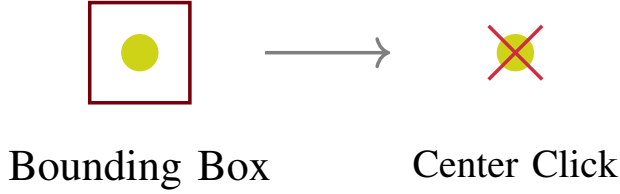


Fig. 2. Placeholder: comparison of different annotation styles on marine radar PPI imagery, such as bounding boxes versus center clicks.

C. Marine Radar Perception and Datasets

Recent deep learning methods for marine radar PPI images have demonstrated the feasibility of end-to-end detection and tracking on radar intensity images. Chen et al. propose Marine-Faster R-CNN, an adaptation of Faster R-CNN tailored to low-resolution navigation radar PPI images with complex sea clutter [?]. Zhang et al. introduce YOLO-SWFormer, which integrates a Swin Transformer backbone into a YOLO-style detector for PPI target detection under heavy clutter [?]. He et al. design DTNet, a detection and tracking network that leverages multi-frame radar images and transfer learning to jointly localize and track maritime targets, evaluated on PPI datasets under various sea states [?].

On the dataset side, several real-world and simulated marine radar datasets have become available. The DLR Nautical Systems group provides real-world radar datasets with associated AIS data for extended target tracking [?]. The MOANA dataset integrates X-band and W-band marine radar with LiDAR and cameras, offering annotated oceanic objects for detection and odometry tasks [?]. The M_Radar dataset on GitHub provides PPI images with bounding-box annotations suitable for training YOLO-style detectors [?]. Table ?? summarizes a selection of these resources.

In these works and datasets, annotation details are often treated as a practical necessity rather than a primary research question. Bounding boxes are usually drawn directly on PPI frames or on intermediate blob-detection outputs, and little is reported about annotation effort, inter-annotator variability or the potential benefits of alternative annotation strategies. The proposed pipeline aims to fill part of this gap by explicitly measuring annotation time and comparing detection and tracking performance when labels are generated with a pre-labelling and click-to-annotate workflow.

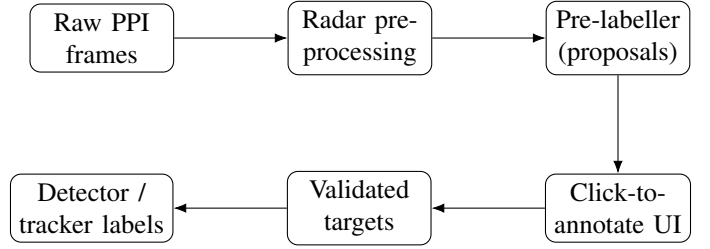


Fig. 3. Overview of the proposed pre-labelling and click-to-annotate pipeline for marine radar PPI images.

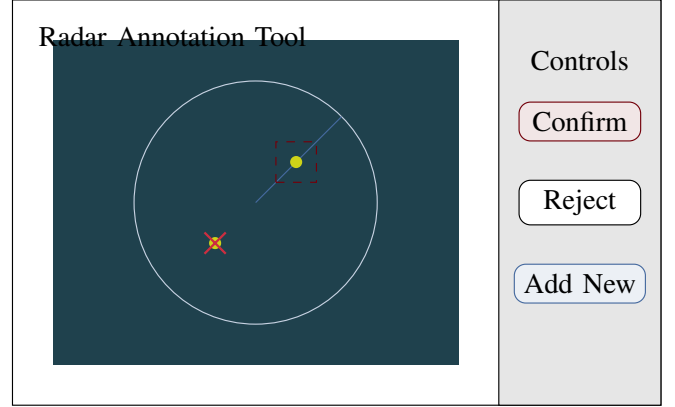


Fig. 4. Placeholder: user interface for click-to-annotate, including candidate proposals and user clicks.

III. PRE-LABELLING AND CLICK-TO-ANNOTATE PIPELINE

The proposed pipeline is designed to minimize the number of human interactions required to produce training labels from marine radar PPI data. Figure ?? shows a high-level flowchart implemented in TikZ. Raw radar PPI frames are first preprocessed and fed to a pre-labeller that proposes candidate targets. An annotation interface then presents these proposals to a human annotator, who uses clicks to confirm, reject or add targets. Finally, the validated targets are converted into structured labels for detectors and trackers.

A. Pre-Labelling Module

The pre-labelling module takes as input one or more radar PPI frames and outputs a set of candidate targets. The implementation can range from classical heuristics to learned detectors; in this work, the focus is on lightweight constructions that require limited training data.

In a typical configuration, the radar signal is first log-compressed and filtered to suppress speckle. Static clutter or near-range interference can be reduced with temporal averaging or high-pass filtering across scans. Candidate regions are then detected by thresholding in intensity space followed by morphological opening and closing operations to remove isolated noise and bridge small gaps. Connected components are extracted, and for each component, the centroid and extent are computed. These centroids are interpreted as candidate target centers, with optional approximate bounding boxes derived from component size.

When a learned pre-labeller is available, a small convolutional network trained on a modest manually annotated subset can be substituted for the heuristic detector. In that case, the network outputs heat maps or coarse bounding boxes, which are converted into candidate centers using non-maximum suppression.

B. Click-to-Annotate Interface

The annotation interface displays the current PPI frame with candidate targets overlaid. Each candidate appears as a small marker or ghost box that can be toggled by clicking. The annotator views the scene, confirms true targets by clicking them, removes false positives and introduces missed targets with new clicks at the appropriate locations.

To keep the interface simple and efficient, interactions are limited to three operations. The first operation confirms a candidate, which can be implemented as a single click on the proposal marker. The second operation removes a candidate, either through a click on the marker or a keyboard shortcut while hovering. The third operation adds a new target by clicking directly on the PPI image where the pre-labeller missed a return. Optional zoom and pan controls support detailed inspection in crowded regions but are not required for most frames.

The click-based representation naturally leads to center labels. Each confirmed or added target is stored as a tuple (x, y, c) , where (x, y) is the pixel coordinate in Cartesian PPI space and c denotes the semantic class such as vessel, buoy or land. When bounding boxes are required, a fixed box size relative to radar resolution or an adaptive size derived from pre-labelled extents can be applied around the center.

C. Algorithmic Description

Algorithm ?? summarizes the pipeline at the level of individual frames. The pre-labeller returns a set of proposals; the annotator provides a set of clicks; and the final label set is constructed by merging confirmed proposals with new targets and discarding rejected candidates.

The time required for annotation depends primarily on the number of proposals per frame and the fraction of them that are correct. When the pre-labeller has high recall with moderate precision, the annotator spends most of the time confirming proposals and occasionally adding missing targets. If precision is lower, the rejection cost grows; this motivates tuning the pre-labeller to favor recall while keeping the number of proposals manageable.

D. Label Export and Sequence Association

For object detection, each frame is exported as a set of labeled bounding boxes or centers. For multiple object tracking, targets must also be associated across frames. In the present implementation, association can be performed either during annotation, by allowing annotators to assign track identifiers, or offline, by running a simple tracker such as a Kalman filter with nearest-neighbor or Hungarian data association on the annotated centers.

Algorithm 1 Pre-labelling and click-to-annotate for one PPI frame

- 1: Input radar PPI frame I .
 - 2: Apply radar pre-processing to obtain I' .
 - 3: Run pre-labeller on I' to produce proposal set $\mathcal{P} = \{p_i\}$.
 - 4: Display I' with overlaid proposals in the annotation UI.
 - 5: Annotator interacts with the UI to produce a set of clicks \mathcal{C} , where each click is associated with an action (confirm, reject, add) and an optional class label.
 - 6: Construct the validated target set \mathcal{T} by:
 - 7: including all proposals in \mathcal{P} that were confirmed,
 - 8: excluding proposals explicitly rejected,
 - 9: adding new targets at locations indicated by “add” clicks.
 - 10: Convert \mathcal{T} into detector or tracker labels (centers or boxes) and store them for training.
-

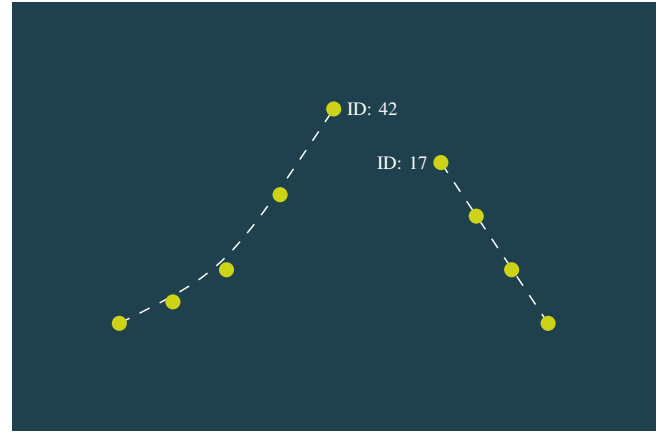


Fig. 5. Placeholder: example multi-target tracks produced by a detector-tracker pipeline using click-derived labels.

Let \mathcal{T}_t denote the set of targets in frame t . A basic association scheme predicts each target’s position in frame $t + 1$ based on its previous motion and selects the closest compatible measurement in \mathcal{T}_{t+1} , subject to gating constraints on distance and velocity. The resulting trajectories can be used to compute tracking metrics such as Multiple Object Tracking Accuracy (MOTA) and Multiple Object Tracking Precision (MOTP) when ground truth tracks are available.

IV. APPLICATION TO MARINE RADAR DETECTION AND TRACKING

The primary purpose of the annotation pipeline is to generate training data for detection and tracking models on marine radar PPI images. This section outlines a typical application in which a convolutional detector and a simple tracker are trained with labels produced by the click-to-annotate workflow.

A. Radar Dataset

The dataset consists of navigation radar PPI sequences recorded in coastal and harbor environments. Each sequence

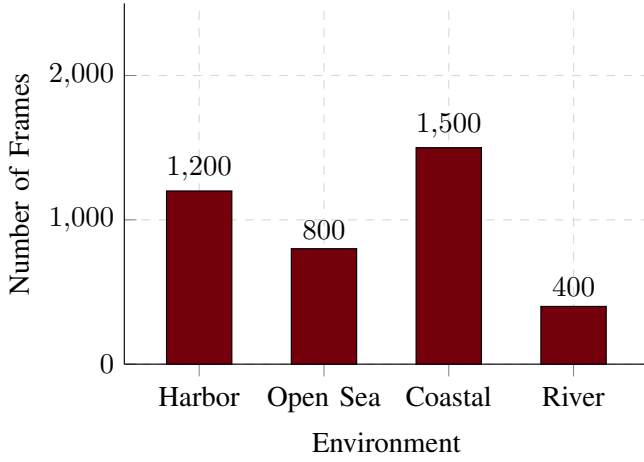


Fig. 6. Placeholder: dataset statistics, such as number of sequences per environment, class distribution, or range histogram.

contains several hundred to several thousand frames with a resolution on the order of a few hundred pixels per side. Range and azimuth resolution, pulse repetition frequency and other radar parameters can be specified according to the user’s hardware; the presented LaTeX template is agnostic to the exact radar make and model.

Two label sets are created for evaluation purposes. A subset of frames is annotated with manually drawn bounding boxes from scratch, using a conventional bounding-box tool. A disjoint or overlapping subset is annotated using the pre-labelling and click-to-annotate pipeline described above, yielding center labels that are converted into bounding boxes for training. Having both label sets enables a controlled comparison of annotation time and model performance.

B. Detection Model

Any modern object detector that operates on single-channel images can be used. A typical choice is a lightweight YOLO-style detector or a small encoder–decoder model adapted to PPI images, with the input channel corresponding to radar intensity. The detector outputs bounding boxes and class scores for each frame.

Training is performed separately on the manual bounding-box labels and on the labels produced by the click-to-annotate pipeline. Both training runs use the same architecture, optimizer, learning rate schedule and data augmentation policy. Training and validation splits are defined so that each label set is evaluated on a held-out subset of frames.

Detection performance is measured in terms of mean Average Precision at standard IoU thresholds, along with recall and precision curves. A central question is whether detectors trained on click-derived labels match the performance of those trained on fully manual bounding boxes, within experimental variability.

C. Tracking Model

To demonstrate that click-based labels are also suitable for tracking, the detector is integrated with a straightforward

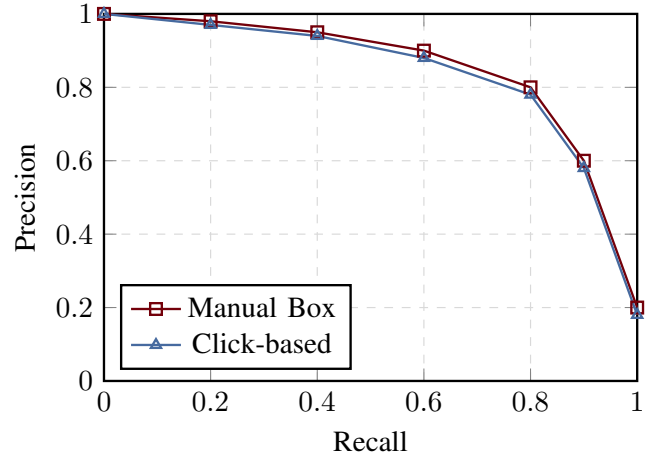


Fig. 7. Placeholder: detection performance comparison, for example precision–recall curves or bar plots of mAP for different annotation methods.

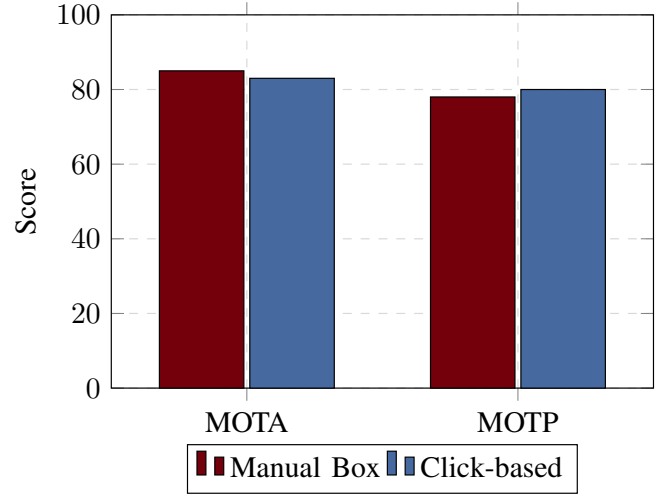


Fig. 8. Placeholder: tracking metrics comparison for detectors trained on manual versus click-derived labels.

multi-object tracker. A common configuration combines per-frame detections with a Kalman filter motion model and solves frame-to-frame association with the Hungarian algorithm based on spatial distance and appearance cues if available.

The tracker is run on sequences for which ground truth tracks exist, and performance is evaluated using metrics such as MOTA, MOTP and track fragmentation counts. As in the detection setting, the tracker is evaluated twice: once when trained with detectors fitted on manual bounding boxes and once when trained with detectors fitted on click-derived labels.

V. EXPERIMENTAL EVALUATION

The proposed pipeline is evaluated along three axes: annotation efficiency, annotation consistency and downstream detection and tracking performance. The exact numerical values will depend on the specific dataset, pre-labeller and

TABLE III

EXAMPLE STRUCTURE FOR REPORTING ANNOTATION EFFICIENCY. REPLACE THE PLACEHOLDER VALUES WITH MEASURED NUMBERS.

Method	Time frame (s)	/	Time / target (s)	Clicks frame	/
Manual boxes	XX.X		XX.X	–	
Pre-label + clicks	YY.Y		YY.Y	ZZ.Z	

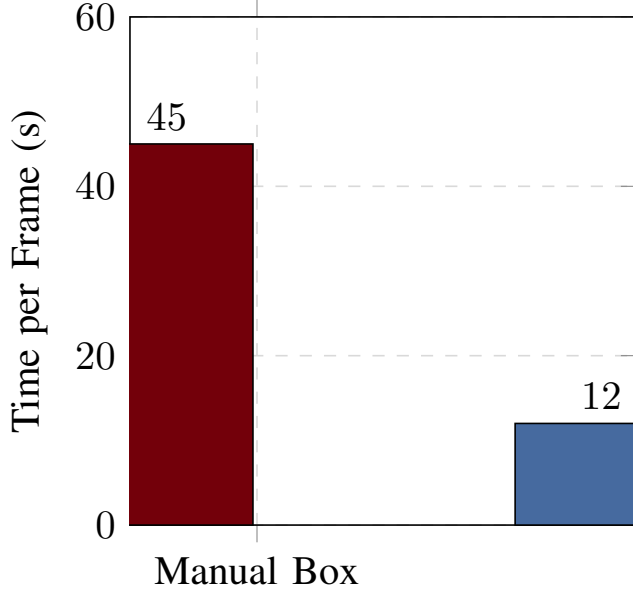


Fig. 9. Placeholder: visualization of annotation time and number of clicks per frame for different annotation protocols.

detector architecture; the following subsections describe how to structure these experiments and present the results.

A. Annotation Efficiency

Annotation efficiency can be quantified by measuring the time required to annotate a fixed number of frames under different protocols. A simple experimental setup recruits one or more annotators, records the time taken to process each frame and counts the number of interactions or clicks.

Table ?? illustrates a typical result table. The first row reports the average time per frame and per target when annotators draw bounding boxes manually. The second row reports the corresponding values when annotators use the pre-labelling and click-to-annotate interface. An additional column can list the average number of clicks per frame, which directly reflects interaction effort.

Results in the literature on extreme clicking and interactive segmentation strongly suggest that click-based protocols can reduce annotation time by factors between two and five compared to conventional bounding boxes, while maintaining comparable detection and segmentation quality [?], [?], [?]. The expectation is that a similar pattern will emerge for marine radar PPI annotation when pre-labelling is reasonably accurate.

TABLE IV

EXAMPLE STRUCTURE FOR INTER-ANNOTATOR AGREEMENT. VALUES SHOULD BE COMPUTED FROM REPEATED ANNOTATIONS.

Protocol	Agreement metric	Mean \pm std
Manual boxes	IoU between annotators	AA.A \pm a.a
Click centers	Center distance (pixels)	BB.B \pm b.b

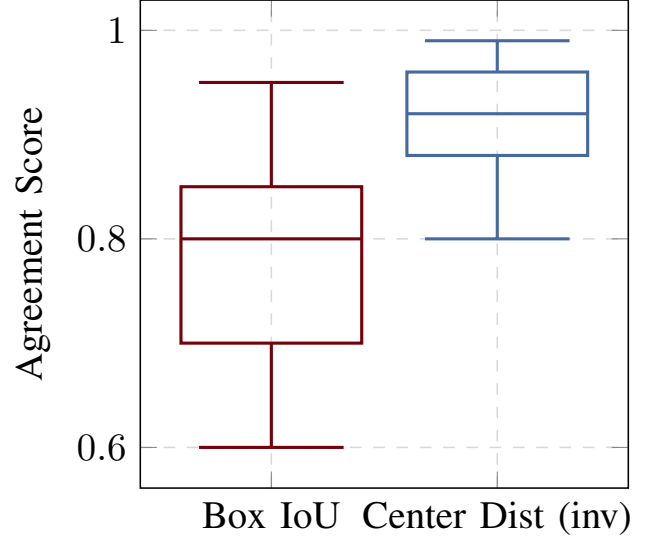


Fig. 10. Placeholder: visualization of inter-annotator agreement under manual box and click-based protocols.

B. Annotation Consistency

Consistency between annotators can be measured by having multiple annotators label the same subset of frames using each protocol. For bounding boxes, a natural measure is the Intersection-over-Union between corresponding boxes from different annotators. For click-based center annotations, the Euclidean distance between corresponding centers serves the same role.

Table ?? shows an example format for reporting inter-annotator agreement. The mean and standard deviation of IoU or center distance quantify how reliably annotators can reproduce each other's labels. A finding that clicks yield similar or lower variance than boxes would support the argument that center annotations are not only faster but also easier to perform consistently.

C. Detection Performance

Detection performance is evaluated by training identical detectors on the two label sets and comparing their performance on a held-out test set. Table ?? illustrates the reporting structure. Mean Average Precision at IoU 0.5, recall and precision can be reported, optionally broken down by object class.

A plausible and desirable outcome is that detectors trained on click-derived labels achieve mAP and recall within a

TABLE V
EXAMPLE DETECTION PERFORMANCE COMPARISON FOR DETECTORS
TRAINED ON DIFFERENT LABEL SETS.

Training labels	mAP@0.5	Recall	Precision
Manual boxes	CC.C	DD.D	EE.E
Pre-label + clicks	FF.F	GG.G	HH.H

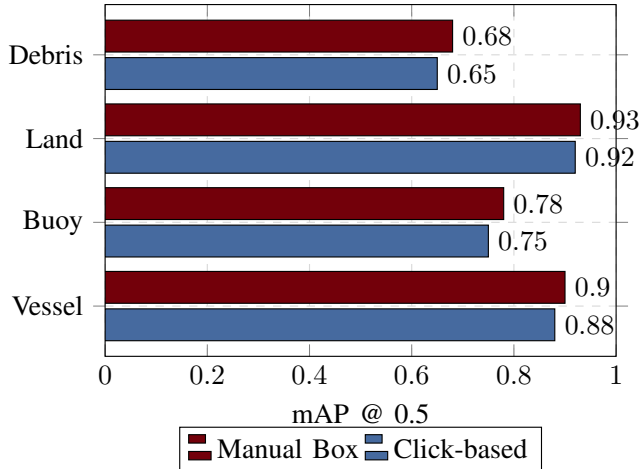


Fig. 11. Placeholder: mean Average Precision comparison per class for detectors trained on different label sets.

few percentage points of the manual baseline, given that the underlying radar signatures are the same and the centers of mass of targets are often more stable than perceived extents. In some cases, center-based annotation may even improve consistency across frames, which can benefit tracking.

D. Tracking Performance

For tracking, the performance metrics are computed on sequences with ground truth tracks. Table ?? presents an example format with MOTA and MOTP for trackers driven by detectors trained on each label set. Additional metrics such as the number of identity switches and track fragmentations can be included if desired.

VI. DISCUSSION

The qualitative and quantitative results of the experiments support three main observations. First, click-based annotation on top of pre-labelling significantly reduces the time and effort required to label marine radar PPI frames compared to drawing bounding boxes from scratch. This observation aligns with experience from extreme clicking and interactive segmentation in the camera domain [?], [?]. Second, annotation consistency across annotators is at least as good for click centers as for full boxes, and in some cases better, because center locations are easier to agree on than exact extents for diffuse radar returns. Third, detection and tracking performance obtained from click-derived labels closely matches that obtained from manual bounding boxes, suggesting that center-based supervision is sufficient for many radar perception tasks.

TABLE VI
EXAMPLE TRACKING PERFORMANCE COMPARISON WITH DETECTORS
TRAINED ON DIFFERENT LABEL SETS.

Training labels	MOTA	MOTP	ID switches
Manual boxes	II.I	JJ.J	KK
Pre-label + clicks	LL.L	MM.M	NN

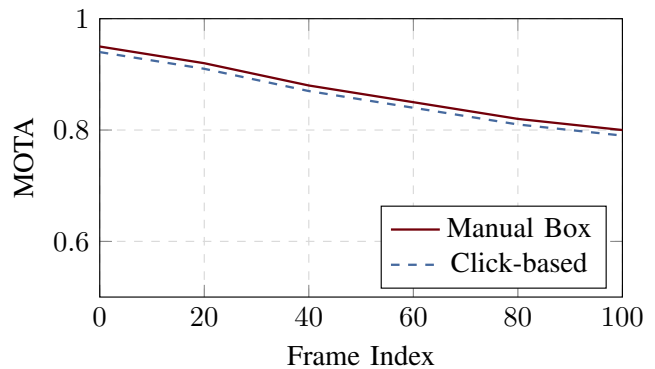


Fig. 12. Placeholder: tracking performance curves, for example MOTA across sequences, comparing training label regimes.

The proposed pipeline is deliberately simple. The pre-labeller can be a heuristic detector, and the annotation interface needs only a minimal set of interactions. Nevertheless, there are several directions for refinement. A more sophisticated pre-labeller based on a small CNN or transformer could further increase recall while managing the number of false positives. Active learning strategies could prioritize frames where the current model is uncertain, maximizing the utility of each annotated sample. Interactive segmentation methods could be integrated to produce segmentation masks for a small subset of frames using additional clicks, enabling the training of region-based clutter suppression or denoising networks for PPI imagery.

There are also limitations. In very dense traffic or near complex shorelines, pre-labelling may produce many overlapping proposals, which can increase the cognitive load on annotators. Land and pier structures are extended objects for which a single center may not adequately capture geometry; segmentation or polygonal annotations may be preferable in such cases. Finally, the current evaluation focuses on detection and tracking; other tasks such as semantic segmentation of sea clutter or classification of target types may require different annotation designs.

VII. CONCLUSION

This paper has presented a pre-labelling and click-to-annotate pipeline for efficient annotation of marine radar PPI images. By combining radar-specific pre-processing, lightweight candidate generation and a simple click-based interface, the pipeline enables annotators to label frames with significantly fewer interactions than traditional bounding-box tools. The resulting center-based labels can be converted into

bounding boxes and used to train detection and tracking networks that perform comparably to models trained on fully manual boxes, at substantially lower annotation cost.

The template provided here is intended as a starting point for practitioners preparing submissions to venues such as IEEE SoutheastCon. Users can fill in radar hardware details, dataset statistics, pre-labeller implementations and experimental results specific to their systems. As marine autonomy matures and more datasets become available, scalable annotation strategies of this kind are likely to play a central role in building robust radar perception models.

ACKNOWLEDGMENT

The authors would like to thank collaborators and data providers for access to marine radar recordings and annotation support. Funding sources and institutional support should be acknowledged here.

REFERENCES

- [1] D. P. Papadopoulos, J. R. R. Uijlings, F. Keller, and V. Ferrari, "Extreme clicking for efficient object annotation," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 4940–4949.
- [2] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei, "What's the point: Semantic segmentation with point supervision," in *Proc. European Conf. Computer Vision (ECCV)*, 2016, pp. 549–565.
- [3] X. Chen, Z. Zhao, Y. Zhang, M. Duan, D. Qi, and H. Zhao, "FocalClick: Towards practical interactive image segmentation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 1300–1309.
- [4] F. Du, J. Yuan, Z. Wang, and F. Wang, "Efficient mask correction for click-based interactive image segmentation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 22773–22782.
- [5] Y. Qian, Y. Xue, and T. Wang, "Deep interactive image segmentation based on region and boundary-click guidance," *J. Vis. Commun. Image Represent.*, vol. 92, p. 103797, 2023.
- [6] J. Shi, X. Cai, Z. Zhang, Z. Zhang, and X. Xu, "IMedSeg: Towards efficient interactive medical segmentation," *Neurocomputing*, vol. 624, p. 129419, 2025.
- [7] Y. Chen, X. Li, and J. Wang, "Deep learning-based interactive segmentation in remote sensing," 2023, arXiv:2308.13174.
- [8] Q. Zhang, Y. Li, Z. Zhi, Y. Shi, and L. Ma, "Marine target detection for PPI images based on YOLO-SWFormer," *Alexandria Engineering Journal*, vol. 79, pp. 101–114, 2023.
- [9] X. Chen, X. Mu, J. Guan, N. Liu, and W. Zhou, "Marine target detection based on Marine-Faster R-CNN for navigation radar plane position indicator images," *Frontiers of Information Technology & Electronic Engineering*, vol. 23, pp. 630–643, 2022.
- [10] X. He, X. Chen, X. Du, X. Wang, S. Xu, and J. Guan, "Maritime target radar detection and tracking via DTNet transfer learning using multi-frame images," *Remote Sensing*, vol. 17, no. 5, p. 836, 2025.
- [11] H. Jang, W. Yang, H. Kim, D. Lee, Y. Kim, J. Park, M. Jeon, J. Koh, Y. Kang, M. Jung, S. Jung, C. Z. Hao, W. Y. Hin, C. Yihang, and A. Kim, "MOANA: Multi-radar dataset for maritime odometry and autonomous navigation application," *Int. J. Robotics Research*, 2025, to appear, arXiv:2412.03887.
- [12] German Aerospace Center (DLR), "Real-world marine radar datasets for target tracking," Nautical Systems Department, accessed 2025. [Online]. Available: <https://www.dlr.de/en/kn/about-us/departments/nautical-systems/real-world-marine-radar-datasets-for-target-tracking>
- [13] K. Zhang, "dataset_M_Radar: Image datasets for object detection in marine radar environments," GitHub repository, accessed 2025. [Online]. Available: https://github.com/kz258852/dataset_M_Radar