Object Tracking & Classification in Marine Radar

A Survey of Data Annotation Strategies

J.C. Vaught

December 5, 2025

1 Annotating Marine Radar Data

Marine radar plan position indicator (PPI) images(see Fig. ??) present distinctive challenges for annotation. In contrast to optical imagery, radar returns appear as noisy "blips" whose apparent shape and size vary with range and sensor settings. Although many modern navigation radars transmit multiple pulse lengths to cover different ranges, those per–pulse-length data are typically inaccessible to end users; only a composite PPI is provided.

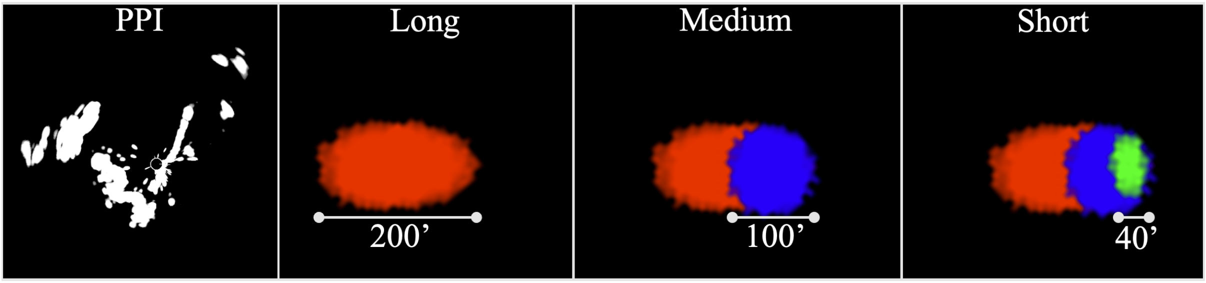

Figure 1: Plan Position Indicator (PPI) example from a Furuno DRS4D-NXT radar. Close-up echoes from the long (M1), medium (S2), and short (S1) pulses.

In practice, to provide more granular control, sequential collections of three pulse lengths are acquired and subsequently stitched during post-processing. However, this procedure introduces non-uniform resolution where nearby targets recorded with shorter pulses appear as small, sharp spots(see Fig. ??), whereas distant targets, more frequently recorded with longer pulses, appear as elongated radial and range smears(see Fig. ??). Further complicating interpretation, fast-moving targets can produce tadpole-shaped blips with a bright head and a trailing echo(see Fig. ?? from [?]), or multiple echoes(see Fig. ??).

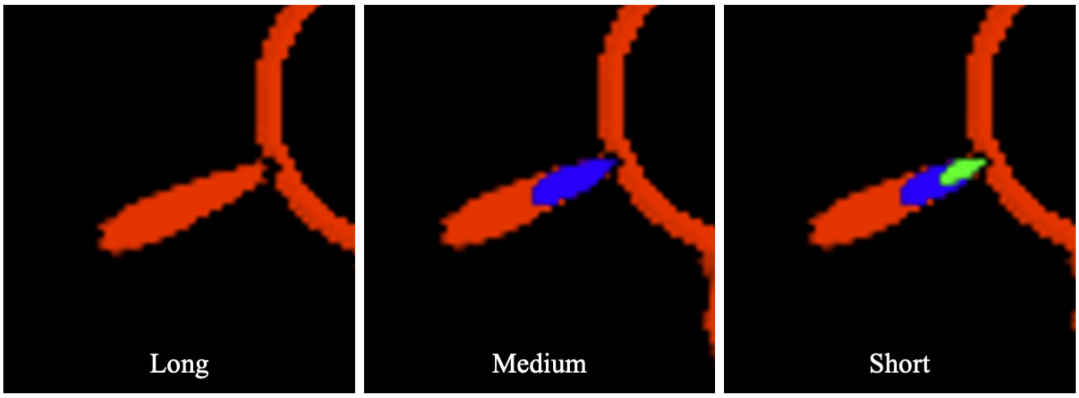

Figure 2: Close-range target (e.g., a buoy) observed with interleaved pulses on a Furuno DRS4D-NXT.

Due to this and other factors, discriminating true targets from sea clutter and land returns is often extremely challenging. For a stationary radar, strong static echoes from coastlines and large fixed objects are produced, often forming extensive, contiguous blobs. However, when annotating data for machine learning algorithms, these blobs are ill-suited for representation by a single box or point. At other times, land clutter manifests as

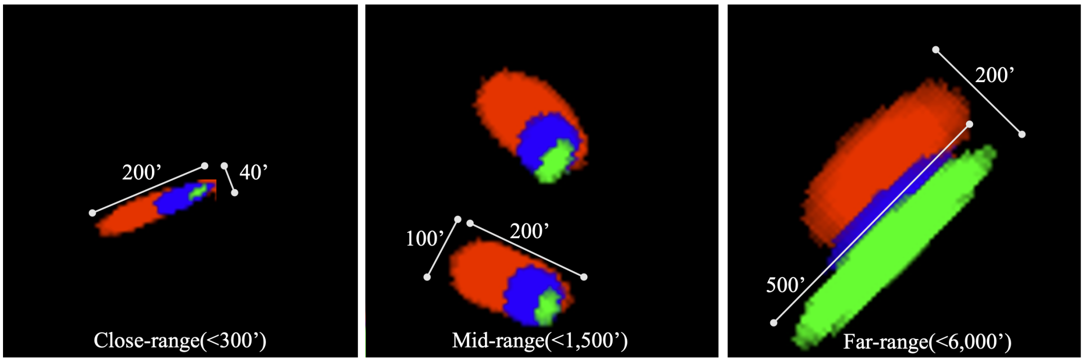

Figure 3: Distant targets exhibiting radial smearing and range spreading due to beamwidth and pulse duration. Data from a Furuno DRS4D-NXT.

irregular, random blobs that can confound standard detection methods and elevate false-positive rates in the absence of auxiliary data (e.g., GPS). When only a PPI is available, the lack of velocity or Doppler information further limits separation of moving targets from static clutter, requiring the use of temporal tracking.

These factors render conventional annotation primitives inadequate. Fixed-size bounding boxes do not consistently accommodate both near and far objects, and pixel-level segmentation can be ambiguous at the faint boundaries of returns. A detailed annotation protocol would therefore require per-echo or per-pulse labeling prior to stitching during post-processing using a segmentation approach, followed by deriving bounding boxes from the resulting masks. Yet, such a solution would require many times more effort to label as compared to the point or bounding box methods.

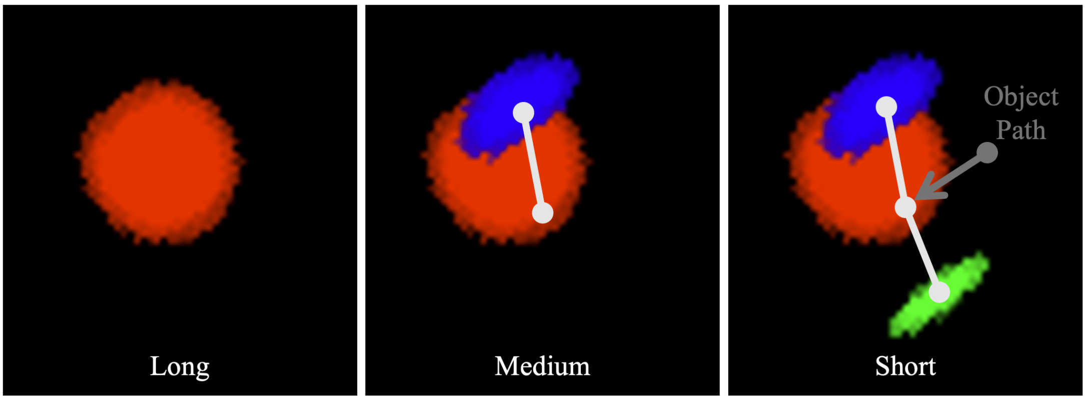

Figure 4: Moving target captured across sequential interleaved pulses (S1, S2, M1). The apparent offsets between echoes arise from the target's motion. The estimated trajectory is overlaid in white. Examples shown from a Furuno DRS4D-NXT.

These practical constraints mean that annotation is a trade-off between representational fidelity and annotation effort. Radar PPI peculiarities (multi-pulse artifacts, range-dependent smearing, motion trails, etc.) make some label types highly informative but costly, while simpler primitives (points or boxes) are cheaper but may omit useful size/shape cues. To evaluate these trade-offs in a structured way, the three aforementioned annotation strategies are assessed for their suitability for the radar challenges outlined.

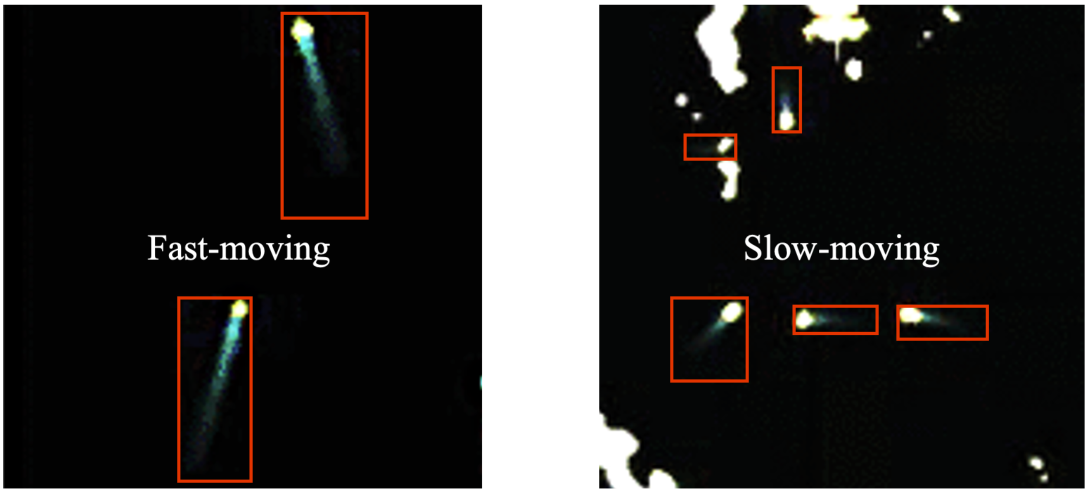

Figure 5: Data from Ma *et al.*'s 2024 paper showing comparison between slow and fast moving objects trails.

2 Annotation approaches

For training models on standard RGB images and radar PPI frames, three annotation paradigms are routinely used: axis-aligned bounding boxes, pixel-wise masks, and center-point labels. Boxes are quick to draw and remain common across maritime radar datasets, including recent PPI-based detection work and fusion datasets that adopt box supervision [?, ?, ?].

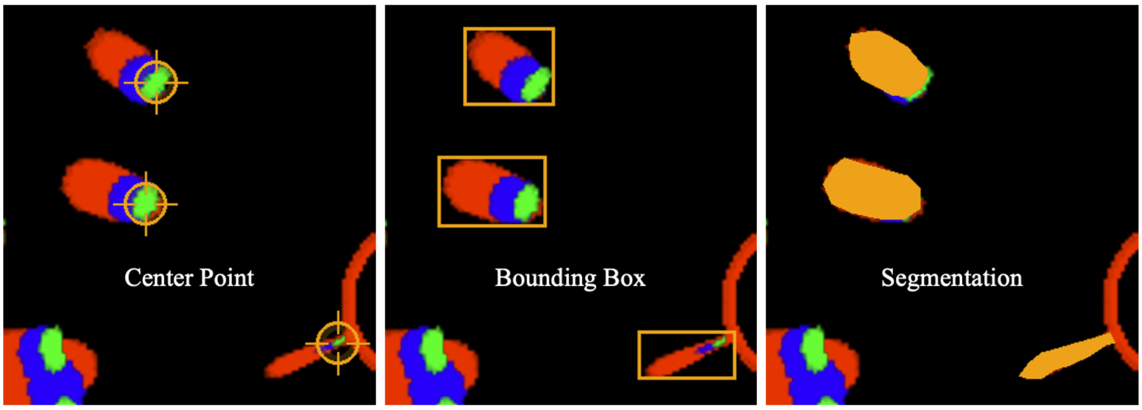

Figure 6: A single radar PPI frame annotated with three paradigms: axis-aligned box, pixel mask, and center-point.

Pixel-wise instance or semantic masks deliver the most accurate localization and enable shape- or area-dependent downstream measures, but they are substantially slower to produce than boxes; for reference, COCO-style polygon masks have been reported to require on the order of a minute per instance on average [?].

Center-point labels reduce the interaction to a single click per object and therefore minimize per-instance time; center-based detectors and radar–vision fusion methods naturally leverage such point supervision [?]. Comparative studies show that box and point supervision trade small drops in pixel-level fidelity for large savings in annotation effort; for example, point-based schemes around "extreme clicking" measure approximately 7 seconds per object and are about five times faster than traditional bounding boxes, which themselves were measured at roughly 35 seconds per box in large-scale annotation protocols [?].

Point-style supervision has also been shown to cut the cost of producing training signals for segmentation compared to full masks, with recent work reporting about five-fold speedups for point annotations relative to mask annotation while still enabling competitive instance segmentation when combined with appropriate learning objectives [?]. Where mask-quality supervision is desired but budgets are constrained, box-supervised instance segmentation provides a practical bridge by learning high-quality masks from only box labels [?].

In marine radar PPI specifically, clutter and multi-pulse effects amplify these tradeoffs. Instance segmentation captures target extent and boundary details most faithfully and has been demonstrated for ships on PPI imagery [?]. Bounding boxes remain a robust, low-cost default that are widely used to label radar echoes for detection and tracking [?, ?, ?].

Center-point labels are the fastest to collect and integrate cleanly with center-based formulations for detection and sensor fusion [?]. A practical workflow may be to adopt boxes or points for the bulk of frames and reserve masks for a curated subset, optionally leveraging box-supervised or point-supervised objectives to recover mask-quality performance at a fraction of the labeling cost [?, ?].

Annotation type	Typical cost profile	Example use on radar
Bounding boxes	Moderate cost; widely adopted	Used in PPI ship detection and radar–vision fusion datasets [?, ?, ?]
Pixel-wise masks (polygons)	Highest cost; about a minute per instance in large-scale practice [?]	Pixel-level ship segmentation on PPI imagery [?]
Center-point labels	Lowest cost; about 7s per object and roughly 5x faster than B-boxes [?]	Center-based radar–camera fusion and detectors [?]

Table 1: Cost–utility sketch for common annotation paradigms in radar PPI contexts, with representative evidence and examples.

2.1 Bounding Box Annotations

Bounding boxes are widely used in maritime datasets and papers as a straightforward way to localize radar targets. For example, the WHUT-MSF Vessel dataset labeled vessels in radar PPI images with bounding boxes(see Fig. ??) using tools such as DarkLabel[?]; Radar3000 similarly drew boxes around ship blips using LabelImg and assigned a ship class[?].

Relative to per-pixel masks, boxes are simple and fast to annotate and provide approximate location and size for each target, which aligns with detectors that directly predict box coordinates (e.g., YOLO, Faster R-CNN)[?]. However, axis-aligned boxes can be a poor geometric fit(see Fig. ??) for elongated or smeared echoes at long range; enlarging a box to cover faint extremities risks overlap with nearby returns and lower training IoU.

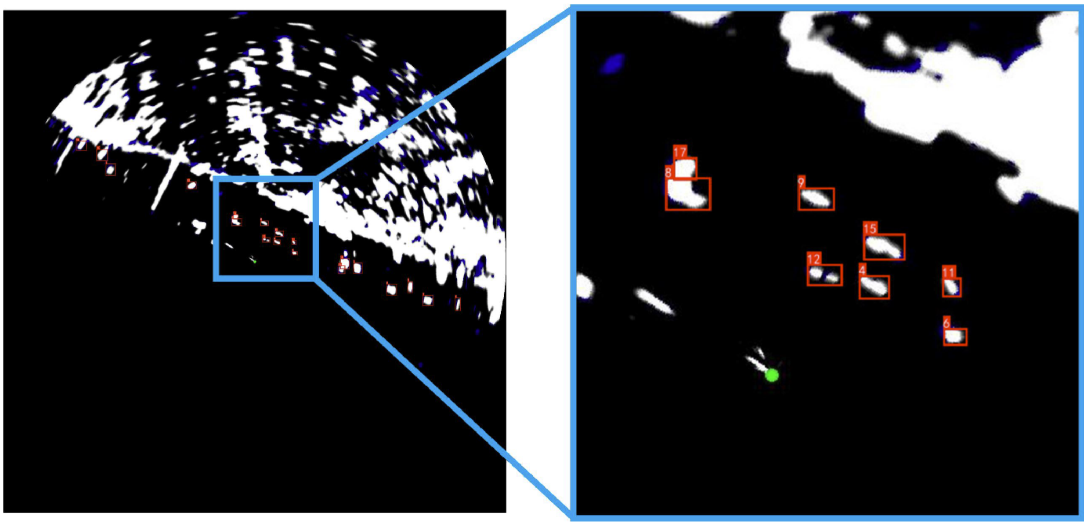

Figure 7: Example of axis-aligned bounding boxes over radar "blips" (cf. WHUT-MSFVessel[?]). Helps visualize geometric mismatch for elongated returns.

Without Doppler, motion trails appear, and a box covering the trail can be much larger than the object's true extent, potentially biasing learning. Oriented boxes(see Fig. ??) mitigate some issues by aligning with object angle (e.g., in SAR ship detection)[?, ?], but rotated labels are more complex and not universally supported.

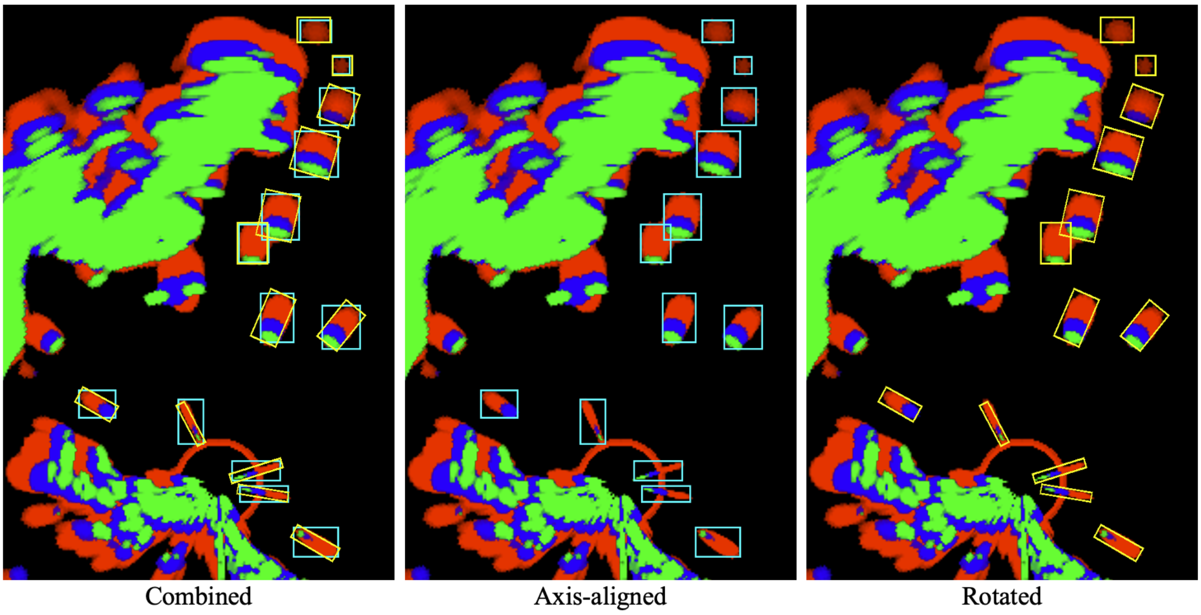

Figure 8: Comparison of axis-aligned versus rotated boxes on elongated echoes. Highlights geometric fit differences and potential label complexity.

Multi-pulse imaging also changes apparent size with range; a single-scale box representation can encourage learning pulse artifacts rather than true size, and although multi-scale features help[?], challenges remain. In practice, boxes remain a feasible baseline[?, ?] but often require refinement steps (e.g., clustering or a second stage) and alternative handling for large static regions such as coastlines[?].

2.2 Pixel-wise Segmentation Annotations

Segmentation labels each pixel of an object's echo. There are two types of segmentation: semantic segmentation, where every pixel gets a class label, e.g., ship, buoy, land, or background), and instance segmentation, where each object's pixels get a unique ID or mask. As the most detailed form of annotation, segmentation can provide a rich training signal and support precise localization.

A segmentation mask captures the detailed shape and extent of a radar return. Models trained with segmentation can learn both the presence of a target and aspects of its size/shape, which is valuable for extended targets.

For example, Blackman et al. trained a fully convolutional network for pixel-wise target detection and size estimation on simulated radar images[**?**], estimating radial length and outperforming traditional CFAR in detecting extended targets[**?**]. An instance segmentation approach, MrisNet, was proposed to segment ships in challenging marine radar scenes[**?**].

By learning at the pixel level, such models can delineate ship "spots," including long or short trails, and have reported strong precision and recall in heavy clutter[**?**]. These results indicate that segmentation labels support learning the fine-grained structure of radar returns, which is helpful when objects have irregular shapes or when distinguishing targets from clutter.

Another benefit is that segmentation naturally handles large/static objects like land. The entire landmass region can be labeled as a separate class mask. The model can then learn to classify those large blobs as "land" and not confuse them with ships. This is more direct than attempting to cover a coast with many small boxes or points – or leaving it entirely unlabeled.

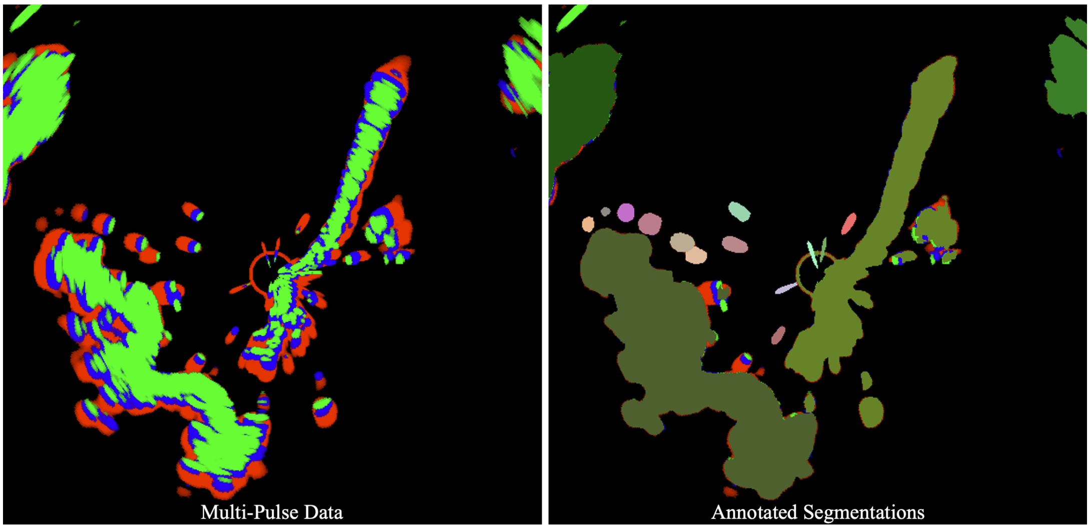

Figure 9: Segmentation overlay illustrating pixel-level masks for vessels and land.

However, the chief cost of this annotation method is the annotation effort. Pixel-wise labeling is labor-intensive, especially for noisy radar images. Deciding the exact boundary of a radar target can be subjective (e.g., whether to include a faint tail). Consistency across annotators is difficult and often requires cross-referencing additional information (AIS or consecutive frames, as in WHUT-MSFVessel[**?**]). With many frames, progress can be painfully slow without semi-automatic tools.

Modeling also becomes heavier; segmentation architectures (UNet, FCN, Mask R-CNN, etc.) are typically more complex and slower than simple detectors. Given the stationary nature of shore-based radars and the frequent need for high accuracy, segmentation can, nevertheless, be justified when resources permit. Hybrid workflows (detector proposals followed by segmentation) are also used, though mask labels are still required for training.

Overall, segmentation yields the most information and can effectively capture radar target shapes[?, ?]. Additionally, it is possible to derive bounding boxes from segmentation and possibly even center points if segmenting each echo sweep individually. Segmentation is well suited when extended targets and precise localization are critical (e.g., measuring target size or handling land clutter), but it comes with significant annotation and computational costs. When the goal is only target position and class, segmentation is often more detailed than necessary.

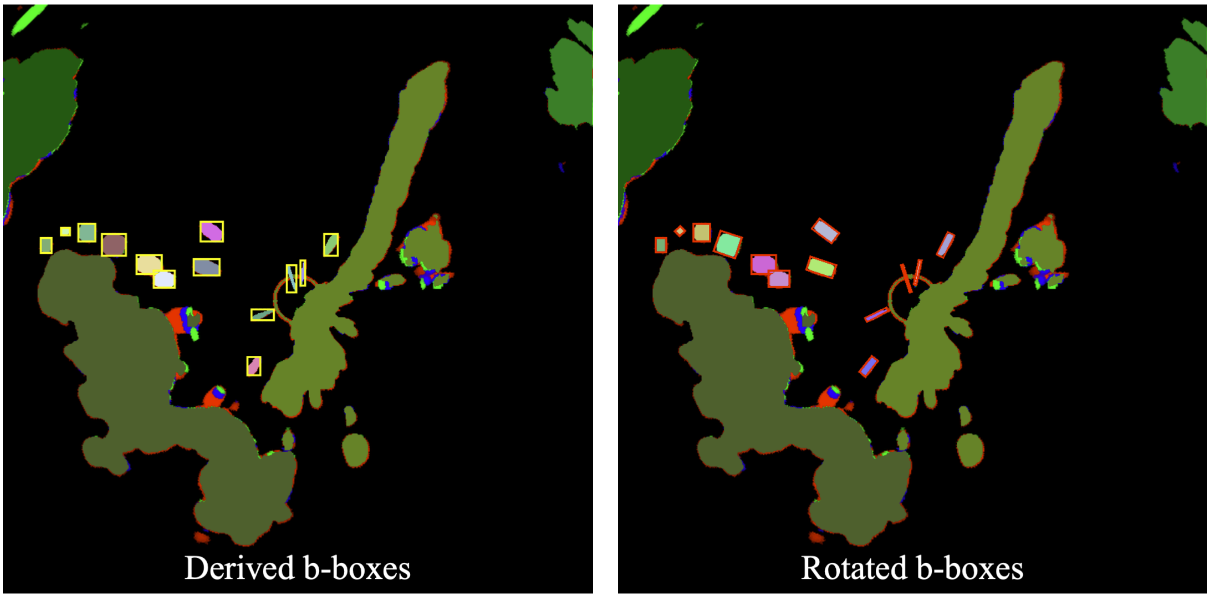

Figure 10: Example of Derived Bounding Boxes From Segmented Annotations

2.3 Center-Point (Keypoint) Annotations

Center-point annotation labels each object with a single coordinate (often at its center or 'Ground Truth' if available), treating detection as keypoint localization. This minimalist annotation – one dot plus a class – has gained popularity via anchor-free detectors such as CenterNet[?].

Labeling is significantly faster than drawing boxes or masks. In shore-based marine radar, objects (ships, buoys) often appear as compact clusters where choosing a center is intuitive. This also sidesteps shape, since regardless of elongation or irregularity, a representative point (e.g., centroid or brightest pixel) is marked, avoiding dilemmas about box size or mask extent. In multi-pulse scenarios, the true target location is typically at the core of the echo; center labels thus focus learning on the core rather than the full smear.

Indeed, some multimodal datasets have used point annotations for radar. For example, the Pohang Canal dataset aligned LiDAR and radar and provided point-wise

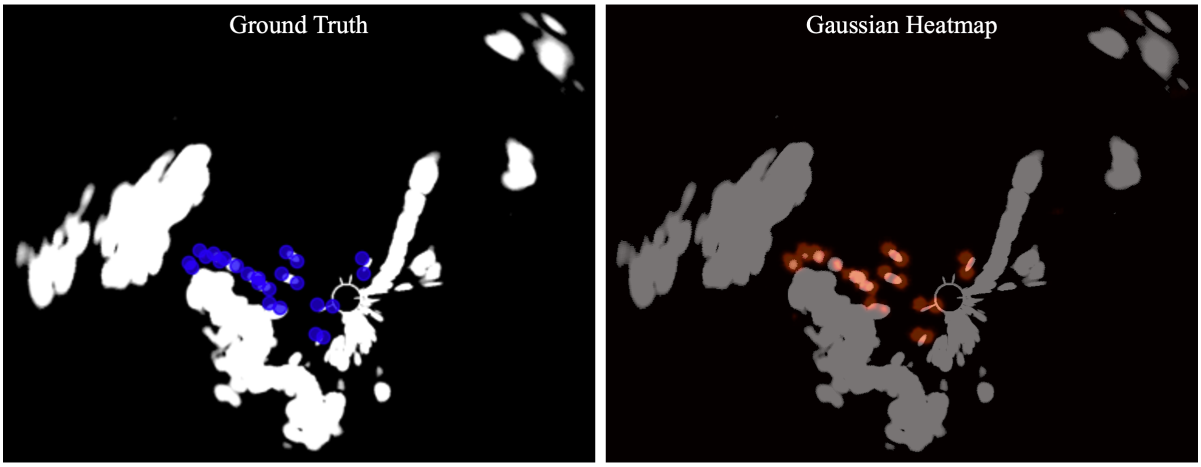

Figure 11: Center-point labeling and corresponding Gaussian heatmap targets as used in CenterNet-style detectors.

annotations by marking the centroid of each radar detection corresponding to a LiDAR-labeled object[**?**]. This indicates that representing radar targets as points is a viable strategy to indicate presence and location.

Another benefit is straightforward integration with tracking. For position tracking, a single point aligns with common filters (e.g., a Kalman filter on the centroid). Output complexity is reduced to one coordinate per object rather than multiple box coordinates or a mask.

A limitation is the lack of direct supervision for object extent. The model does not inherently learn apparent size, which can make classification more difficult when size/shape cues would be helpful. For example, distinguishing a large ship from a small buoy may require relying on intensity patterns or learned features around the center rather than explicit extent labels.

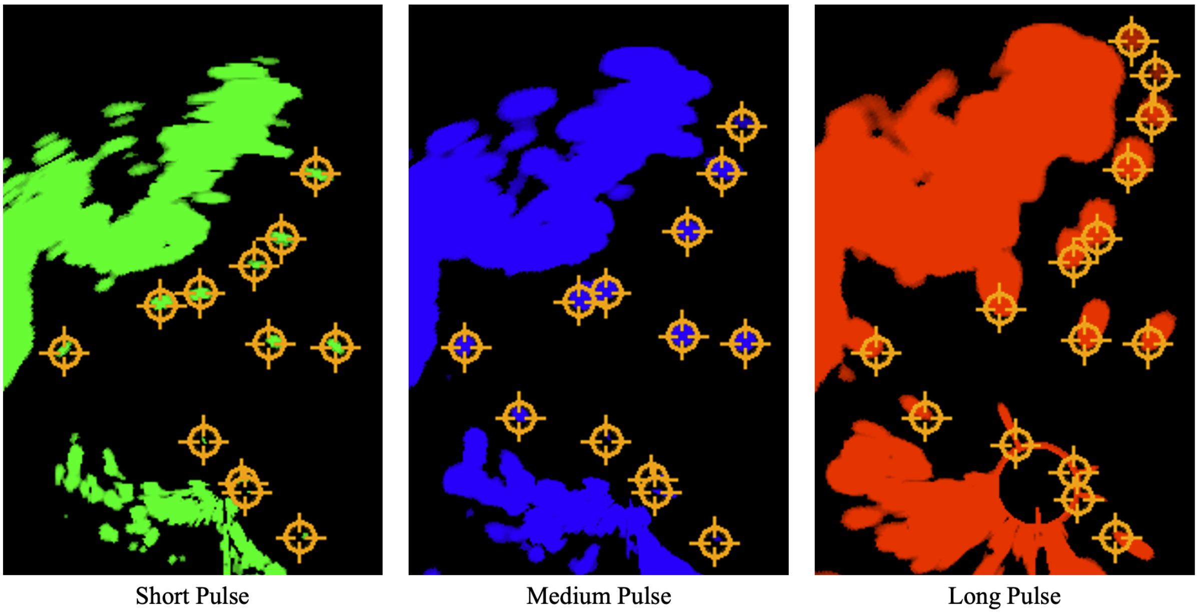

Figure 12: Comparison of center-point labeling for long, medium, and short pulses

Another challenge is the training objective. Networks cannot directly output a variablelength list of coordinates; common practice is to predict a heatmap in which each target center appears as a peak. Ground-truth targets are constructed as 2D Gaussian blobs centered at labeled points, and training uses a dense, pixel-wise loss (often a logistic focal loss) to encourage high response at true centers and low response elsewhere[?].

CenterNet, for instance, produces an output heatmap (downsampled from the input) with a channel for each class, and uses a modified focal loss to encourage high response at true centers and low elsewhere[?]. Prediction peaks that match ground truth centers are treated as detections. This formulation addresses extreme class imbalance between the few center pixels and the many background pixels on the heatmap[?]. In simpler terms, center-point detection functions like blob detection: the loss penalizes both missed true centers and spurious peaks.

Typically, the ground truth uses a small Gaussian radius around the center so that near-misses are penalized less. The radius can be fixed or proportional to object size (if known). Some methods use rotated Gaussians to reflect orientation[?], but symmetric Gaussians are common for simplicity.

If outputs for object size or shape are required, additional regression heads can be added (e.g., width/height of a box, offset, etc., alongside the heatmap)[?]. Classification can be embedded by using one heatmap per class so that the peak resides in the channel corresponding to the class[?] (e.g., ship, buoy, land). This enables simultaneous localization and classification. When classes may overlap spatially (e.g., near shorelines), static land is often handled separately.

During training, a heatmap loss is computed (e.g., pixel-wise logistic focal loss as in CenterNet[?]). During inference, local maxima above a threshold are treated as detected centers with associated class and confidence. Variants of this approach are also used in multi-object tracking (e.g., CenterTrack, CenterFusion) because they naturally produce a variable number of outputs per frame[?].

Overall, center-point annotation is an efficient strategy for localization and classification. It has been applied in camera-radar fusion by detecting centers in the image domain[?]. Point labels are converted into heatmaps and trained with focal-style objectives, yielding anchor-free detectors that output object centers.

References

[1] F. Ma, Z. Kang, C. Chen, J. Sun, X.-B. Xu, and J. Wang, "Identifying ships from radar blips like humans using a customized neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 7, pp. 7187–7205, 2024. doi: 10.1109/TITS.2023.3347761.

[2] W. Huang, H. Feng, H. Xu, X. Liu, J. He, L. Gan, X. Wang, and S. Wang, "Surface vessels detection and tracking method and datasets with multi-source data fusion in real-world complex scenarios," *Sensors*, vol. 25, no. 7, 2179, 2025. doi: 10.3390/s25072179.

[3] X. Chen, J. Guan, X. Mu, Z. Wang, N. Liu, and G. Wang, "Multi-dimensional automatic detection of scanning radar images of marine targets based on Radar-PPInet," *Remote Sensing*, vol. 13, no. 19, 3856, 2021. doi: 10.3390/rs13193856.

[4] X. Yang, Q. Zhang, Q. Dong, Z. Han, X. Luo, and D. Wei, "Ship instance segmentation based on rotated bounding boxes for SAR images," *Remote Sensing*, vol. 15, no. 5, 1324, 2023. doi: 10.3390/rs15051324.

[5] J. Blackman *et al.*, "FCN for extended target segmentation and size estimation in radar," in *Proc. IEEE Radar Conf.*, 2025. Online preprint link.

[6] F. Ma, X. Xu, J. Sun, Z. Kang, and J. Wang, "MrisNet: Robust ship instance segmentation in challenging marine radar environments," *J. Mar. Sci. Eng.*, vol. 12, no. 1, 72, 2024. doi: 10.3390/jmse12010072.

[7] R. Nabati and H. Qi, "CenterFusion: Center-based radar and camera fusion for 3D object detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, 2021, pp. 1527–1536. doi: 10.1109/WACV48630.2021.00157.

[8] J. Choi, D. Cho, G. Lee, H. Kim, G. Yang, J. Kim, and Y. Cho, "PoLaRIS dataset: A maritime object detection and tracking dataset in Pohang canal," 2024. arXiv:2412.06192.

[9] Z. Shao, Z. Dai, W. He, W. Lin, and L. Wang, "An anchor-free rotation ship detector based on Gaussian-mask in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3518–3531, 2021. doi: 10.1109/TGRS.2020.3018106.

[10] D. P. Papadopoulos, J. R. R. Uijlings, F. Keller, and V. Ferrari, "Extreme clicking for efficient object annotation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 4930–4939. doi: 10.1109/ICCV.2017.528.

[11] C. Agnew, E. M. Grua, P. van de Ven, P. Denny, C. Eising, and A. G. Scanlan, "Pre-training instance segmentation models with bounding box annotations," *Patterns*, vol. 5, no. 12, 100984, 2024. doi: 10.1016/j.iswa.2024.200454.

[12] B. Cheng, O. Parkhi, and A. Kirillov, "Pointly-supervised instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 2617–2626. doi: 10.1109/CVPR52688.2022.00264.

[13] Z. Tian, C. Shen, X. Wang, and H. Chen, "BoxInst: High-performance instance segmentation with box annotations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 5443–5452. Open access PDF.