# Evaluation of Oriented Bounding Box (OBB) Models for UAV Video Analysis

Dhruv Patel, Devanshi Rathod, Garima Jain
Ahmedabad University, Ahmedabad, Gujarat 380009, India
dhruv.p12@ahduni.edu.in, devanshi.r@ahduni.edu.in, garima.j@ahduni.edu.in

***Abstract*—In many real-world scenarios, objects appear in arbitrary orientations, making traditional horizontal bounding box methods inadequate for precise object localization. This project aims to evaluate the performance of various Oriented Bounding Box (OBB) models on Unmanned Aerial Vehicle (UAV) video data. By leveraging publicly available OBB detection models trained on the DOTA v1.5 dataset, we develop a Python-based framework to benchmark these models in terms of accuracy, speed, and robustness. The evaluation involves preparing ground-truth annotations for UAV videos and comparing results using standard metrics like mean Average Precision (mAP) and Intersection over Union (IoU). The final outcome is an open-source tool that enables standardized performance evaluation of OBB models for oriented object detection in video surveillance and related applications.**

## I. Introduction

Object detection plays a crucial role in numerous computer vision applications, ranging from autonomous systems and video surveillance to aerial monitoring and remote sensing. While traditional object detection methods rely on horizontal bounding boxes (HBBs), they often fall short in scenarios where objects appear in arbitrary orientations. This limitation is addressed by Oriented Bounding Box (OBB) based detection methods, which offer better accuracy in representing rotated or tilted objects.

In this project, we focus on evaluating the performance of various state-of-the-art OBB detection models using Unmanned Aerial Vehicle (UAV) videos. The objective is to explore, benchmark, and compare the detection capability of these models under real-world conditions involving diverse object orientations. Using pre-trained models trained on the DOTA (Dataset for Object Detection in Aerial Images) dataset, the study aims to build a Python-based evaluation framework that allows for consistent and repeatable performance testing.

The project involves a detailed understanding of model architectures, identification of key modules responsible for orientation-specific detection, and the development of a ground-truth dataset for benchmarking. The final output is an open-source tool that provides a standardized evaluation pipeline for OBB-based object detectors in video-based environments.

## II. Literature Survey

Several state-of-the-art OBB detection models have been proposed:

| Paper | Key Idea | Advantages | Limitations |
|---|---|---|---|
| **TricubeNet** [arXiv:2104.11435v2] | Uses 2D Tricube kernel for anchor-free, heatmap-based OBB detection. | - No angle regression<br>- Efficient and robust<br>- Modular design | - Sensitive to kernel quality<br>- Less precise post-processing |
| **RIDet** [arXiv:2103.11636v3] | Introduces Representation Invariance Loss using Hungarian matching. | - Resolves OBB/QBB ambiguity<br>- Improves accuracy<br>- Easy integration | - Slower training<br>- Test-time ambiguity remains |
| **CFNet** [arXiv:2401.08174v3] | Segments occluded/dense objects using OBB prompts and BSMs. | - Handles occlusions<br>- Prompt-based, not box-dependent<br>- Efficient via distillation | - Needs accurate OBBs<br>- Heavy base models<br>- Post-processing needed |
| **BBAVectors** [WACV 2021] | Regresses directional vectors instead of (w, h, θ) for OBB detection. | - Fast and anchor-free<br>- No angle regression<br>- Unified vector space | - Ambiguity near axes<br>- HBB vs. RBB overhead<br>- Decoding may be unstable |

**Fig. 1:** Overview of recent oriented object detection models using heatmaps, vectors, and segmentation prompts.

## III. Dataset and Evaluation Criteria

We use the DOTAv1.5 dataset, which contains aerial images annotated with Oriented Bounding Boxes (OBBs). The dataset includes the following 16 object categories: Plane, Ship,Storage Tank,Baseball Diamond,Tennis Court,Basketball Court, Ground Track, Field, Harbor, Bridge, Large Vehicle, Small Vehicle, Helicopter, Roundabout, Soccer Ball Field, Swimming Pool, Container Crane

Evaluation metrics include:

- **Mean Average Precision (mAP)**
- **Intersection over Union (IoU)**
- **Angle Error**
- **Processing Speed (FPS)**

## IV. Methodology

The evaluation framework for oriented object detection using UAV-based aerial imagery involves the following stages:

1) **Model Selection:** We evaluate four state-of-the-art oriented object detection models: TricubeNet, RIDet, BBAVectors, and CFNet. These models are chosen for their distinct architectural strategies and proven performance in handling object rotation, occlusion, and dense object arrangements. All models are used in their pre-trained forms, originally trained on the DOTAv1.5 dataset.

2) **Data Preprocessing:** UAV video streams are converted into individual frames. Each frame is resized and normalized according to the input specifications of the selected models. Ground-truth annotations in the form of Oriented Bounding Boxes (OBBs) are prepared in a unified format compatible with each model (e.g., polygonal or angle-based representations).
3) **Model Inference:** The preprocessed images are passed through each model. Each architecture processes the input differently:
   - **TricubeNet:** Generates heatmaps for object centers, sizes, and orientations.
   - **RIDet:** Predicts OBBs using representation-invariant loss with Hungarian matching.
   - **BBAVectors:** Regresses directional vectors from object centers to define object boundaries.
   - **CFNet:** Uses detected OBBs as prompts for a segmentation backbone to extract masks.
4) **Post-Processing:** Models that output heatmaps or vectors (e.g., TricubeNet, BBAVectors) require post-processing using classical computer vision techniques such as non-maximum suppression (NMS), thresholding, and geometric decoding. CFNet employs additional shape priors to infer occluded object regions from partial detections.
5) **Performance Evaluation:** Predictions are compared against ground-truth annotations using the following metrics:
   - **Mean Average Precision (mAP)** at various IoU thresholds
   - **Intersection over Union (IoU)** to assess overlap accuracy
   - **Angle Error** to measure orientation prediction quality
   - **Inference Speed (FPS)** to evaluate real-time performance

   Evaluation is performed both quantitatively (e.g., metric scores) and qualitatively (e.g., visual inspection of predicted boxes in dense or occluded scenes).

## V. Qualitative Analysis

To complement the quantitative evaluation, we generated Precision-Recall (PR) curves for the four additional models: ORCNN, RFRCNN, RRN, and S2ANET. These curves visualize the trade-off between precision and recall across confidence thresholds, helping to assess category-wise detection consistency and reliability.

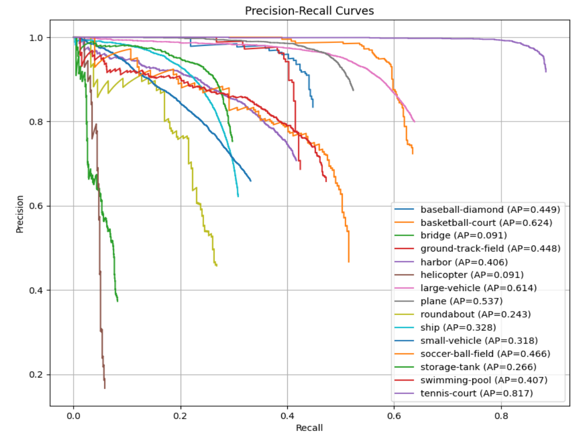


**Fig. 2:** Precision-Recall curves for ORCNN model on UAV video frames.

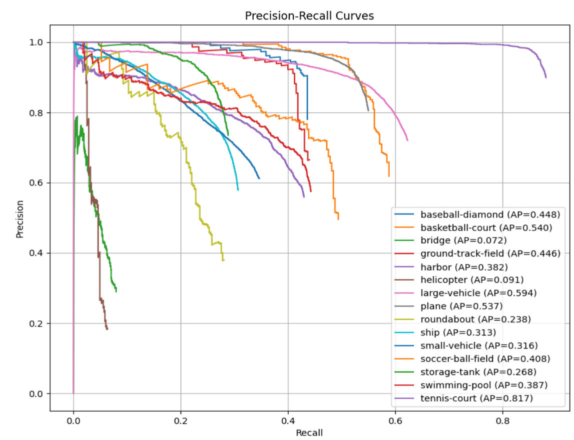


**Fig. 3:** Precision-Recall curves for RFCNN model on UAV video frames.

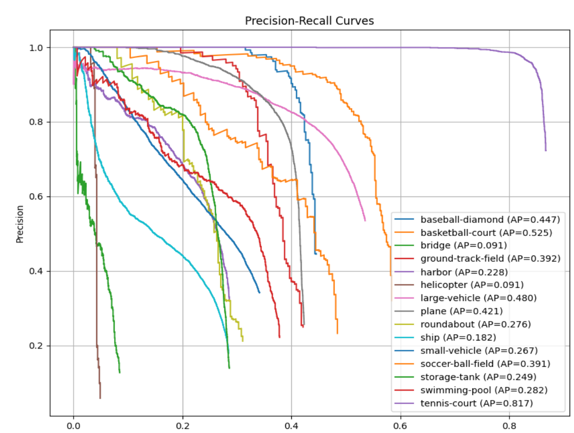


**Fig. 4:** Precision-Recall curves for RRN model on UAV video frames.

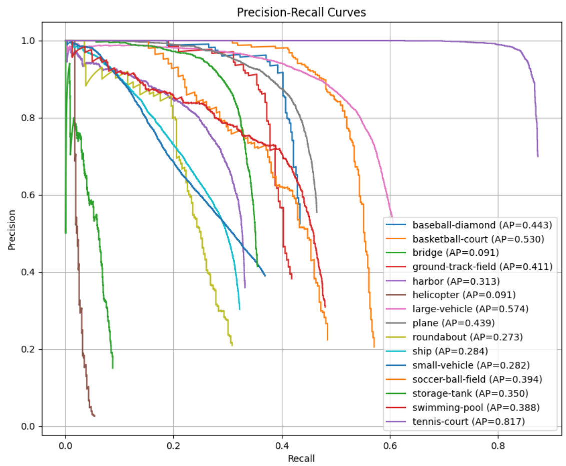


**Fig. 5:** Precision-Recall curves for SZANET model on UAV video frames.

## VI. Quantitative Evaluation

The mean Average Precision (mAP) at an IoU threshold of 0.5 is **0.4070**. Table I presents per-class detection performance metrics, including Average Precision (AP), Precision, Recall, F1-score, True Positives (TP), False Positives (FP), and False Negatives (FN).

TABLE I: Per-class evaluation results on UAV dataset at IoU=0.5

| Class | AP | Precision | Recall | F1-score | TP | FP | FN |
|---|---|---|---|---|---|---|---|
| Baseball Diamond | 0.4495 | 0.8341 | 0.4482 | 0.5831 | 186 | 37 | 229 |
| Basketball Court | 0.6236 | 0.7235 | 0.6350 | 0.6763 | 327 | 125 | 188 |
| Bridge | 0.0909 | 0.3731 | 0.0826 | 0.1352 | 169 | 284 | 1878 |
| Ground Track Field | 0.4484 | 0.6866 | 0.4246 | 0.5247 | 138 | 63 | 187 |
| Harbor | 0.4059 | 0.7073 | 0.4172 | 0.5248 | 2496 | 1033 | 3487 |
| Helicopter | 0.0909 | 0.1667 | 0.0587 | 0.0869 | 37 | 185 | 593 |
| Large Vehicle | 0.6143 | 0.7995 | 0.6385 | 0.7100 | 10835 | 2718 | 6134 |
| Plane | 0.5369 | 0.8741 | 0.5239 | 0.6551 | 4220 | 608 | 3835 |
| Roundabout | 0.2434 | 0.4573 | 0.2682 | 0.3381 | 107 | 127 | 292 |
| Ship | 0.3280 | 0.6219 | 0.3083 | 0.4122 | 8653 | 5261 | 19415 |
| Small Vehicle | 0.3176 | 0.6587 | 0.3319 | 0.4414 | 8672 | 4494 | 17454 |
| Soccer Ball Field | 0.4661 | 0.4667 | 0.5153 | 0.4898 | 168 | 192 | 158 |
| Storage Tank | 0.2665 | 0.7526 | 0.2977 | 0.4266 | 1497 | 492 | 3532 |
| Swimming Pool | 0.4069 | 0.6579 | 0.4729 | 0.5503 | 821 | 427 | 915 |
| Tennis Court | 0.8169 | 0.9180 | 0.8842 | 0.9008 | 2093 | 187 | 274 |

## VII. Quantitative Evaluation - RFRCNN

The mean Average Precision (mAP) at an IoU threshold of 0.5 for the RFRCNN model is **0.3904**. Table II provides per-class performance including Average Precision (AP), Precision, Recall, F1-score, and the counts of True Positives (TP), False Positives (FP), and False Negatives (FN).

TABLE II: Per-class evaluation results for RFRCNN model at IoU = 0.5

| Class | AP | Precision | Recall | F1-score | TP | FP | FN |
|---|---|---|---|---|---|---|---|
| Baseball Diamond | 0.4481 | 0.7802 | 0.4361 | 0.5595 | 181 | 51 | 234 |
| Basketball Court | 0.5397 | 0.6184 | 0.5883 | 0.6030 | 303 | 187 | 212 |
| Bridge | 0.0718 | 0.2895 | 0.0796 | 0.1249 | 163 | 400 | 1884 |
| Ground Track Field | 0.4461 | 0.6651 | 0.4400 | 0.5296 | 143 | 72 | 182 |
| Harbor | 0.3819 | 0.5595 | 0.4301 | 0.4863 | 2573 | 2026 | 3410 |
| Helicopter | 0.0909 | 0.1826 | 0.0635 | 0.0942 | 40 | 179 | 590 |
| Large Vehicle | 0.5940 | 0.7201 | 0.6231 | 0.6681 | 10574 | 4111 | 6395 |
| Plane | 0.5367 | 0.8055 | 0.5500 | 0.6536 | 4430 | 1070 | 3625 |
| Roundabout | 0.2381 | 0.3797 | 0.2807 | 0.3228 | 112 | 183 | 287 |
| Ship | 0.3131 | 0.5787 | 0.3069 | 0.4011 | 8615 | 6272 | 19453 |
| Small Vehicle | 0.3161 | 0.6118 | 0.3466 | 0.4425 | 9055 | 5745 | 17071 |
| Soccer Ball Field | 0.4075 | 0.4954 | 0.4939 | 0.4946 | 161 | 164 | 165 |
| Storage Tank | 0.2677 | 0.7349 | 0.2883 | 0.4142 | 1450 | 523 | 3579 |
| Swimming Pool | 0.3870 | 0.5756 | 0.4430 | 0.5007 | 769 | 567 | 967 |
| Tennis Court | 0.8170 | 0.8991 | 0.8809 | 0.8899 | 2085 | 234 | 282 |

## VIII. Quantitative Evaluation - RRN

The mean Average Precision (mAP) at an IoU threshold of 0.5 for the RRN model is **0.3425**. Table III summarizes the performance metrics per class, including Average Precision (AP), Precision, Recall, F1-score, and the counts of True Positives (TP), False Positives (FP), and False Negatives (FN).

TABLE III: Per-class evaluation results for RRN model at IoU = 0.5

| Class | AP | Precision | Recall | F1-score | TP | FP | FN |
|---|---|---|---|---|---|---|---|
| Baseball Diamond | 0.4470 | 0.4447 | 0.4458 | 0.4452 | 185 | 231 | 230 |
| Basketball Court | 0.5250 | 0.3199 | 0.5845 | 0.4135 | 301 | 640 | 214 |
| Bridge | 0.0909 | 0.1265 | 0.0850 | 0.1017 | 174 | 1201 | 1873 |
| Ground Track Field | 0.3919 | 0.2491 | 0.4215 | 0.3131 | 137 | 413 | 188 |
| Harbor | 0.2276 | 0.3230 | 0.2863 | 0.3036 | 1713 | 3590 | 4270 |
| Helicopter | 0.0909 | 0.0579 | 0.0492 | 0.0532 | 31 | 504 | 599 |
| Large Vehicle | 0.4797 | 0.5345 | 0.5357 | 0.5351 | 9090 | 7917 | 7879 |
| Plane | 0.4205 | 0.2562 | 0.4236 | 0.3193 | 3412 | 9904 | 4643 |
| Roundabout | 0.2760 | 0.2112 | 0.3108 | 0.2515 | 124 | 463 | 275 |
| Ship | 0.1820 | 0.1793 | 0.2849 | 0.2201 | 7997 | 36599 | 20071 |
| Small Vehicle | 0.2674 | 0.3413 | 0.3417 | 0.3415 | 8927 | 17229 | 17199 |
| Soccer Ball Field | 0.3907 | 0.2320 | 0.4847 | 0.3138 | 158 | 523 | 168 |
| Storage Tank | 0.2487 | 0.1385 | 0.2863 | 0.1867 | 1440 | 8956 | 3589 |
| Swimming Pool | 0.2825 | 0.2213 | 0.3785 | 0.2793 | 657 | 2312 | 1079 |
| Tennis Court | 0.8167 | 0.7226 | 0.8673 | 0.7884 | 2053 | 788 | 314 |

## IX. Quantitative Evaluation - S2ANET

The mean Average Precision (mAP) at an IoU threshold of 0.5 for the S2ANET model is **0.3786**. Table IV presents the detailed per-class detection results including Average Precision (AP), Precision, Recall, F1-score, True Positives (TP), False Positives (FP), and False Negatives (FN).

TABLE IV: Per-class evaluation results for S2ANET model at IoU = 0.5

| Class | AP | Precision | Recall | F1-score | TP | FP | FN |
|---|---|---|---|---|---|---|---|
| Baseball Diamond | 0.4430 | 0.5248 | 0.4337 | 0.4749 | 180 | 163 | 235 |
| Basketball Court | 0.5298 | 0.2052 | 0.5709 | 0.3018 | 294 | 1139 | 221 |
| Bridge | 0.0909 | 0.1502 | 0.0874 | 0.1105 | 179 | 1013 | 1868 |
| Ground Track Field | 0.4113 | 0.3820 | 0.4185 | 0.3994 | 136 | 220 | 189 |
| Harbor | 0.3127 | 0.3591 | 0.3321 | 0.3451 | 1987 | 3546 | 3996 |
| Helicopter | 0.0909 | 0.0255 | 0.0540 | 0.0347 | 34 | 1298 | 596 |
| Large Vehicle | 0.5741 | 0.5208 | 0.6070 | 0.5606 | 10300 | 9479 | 6669 |
| Plane | 0.4385 | 0.5548 | 0.4651 | 0.5060 | 3746 | 3006 | 4309 |
| Roundabout | 0.2725 | 0.2092 | 0.3083 | 0.2492 | 123 | 465 | 276 |
| Ship | 0.2841 | 0.3026 | 0.3224 | 0.3122 | 9049 | 20855 | 19019 |
| Small Vehicle | 0.2821 | 0.3902 | 0.3692 | 0.3794 | 9646 | 15077 | 16480 |
| Soccer Ball Field | 0.3936 | 0.2235 | 0.4847 | 0.3059 | 158 | 549 | 168 |
| Storage Tank | 0.3499 | 0.4140 | 0.3555 | 0.3825 | 1788 | 2531 | 3241 |
| Swimming Pool | 0.3879 | 0.3088 | 0.4804 | 0.3759 | 834 | 1867 | 902 |
| Tennis Court | 0.8172 | 0.6994 | 0.8737 | 0.7769 | 2068 | 889 | 299 |

## X. Future Scope

The current framework has successfully implemented oriented object detection and instance segmentation using CFNet in conjunction with the Segment Anything Model (SAM). While the results show promising performance in handling occlusion and densely packed objects, several enhancements and extensions can be pursued in the future:

- **Integration of Alternative Segmentation Models:** In future work, we aim to experiment with other prompt-based segmentation backbones such as SEEM, HQ-SAM, and Grounded SAM to assess their effectiveness in scenarios involving partial occlusions and low-contrast objects.
- **End-to-End Model Unification:** Combining detection and segmentation into a single unified architecture can reduce inference time and avoid cumulative errors introduced by multi-stage processing.

These improvements will enhance the robustness, speed, and adaptability of the proposed oriented object detection and segmentation pipeline across real-world aerial applications.

## References

[1] Beomyoung Kim, Hyunseok Kim, Jeonghee Kim, Beomki Lee, and Youngjoon Yoo. TricubeNet: 2D Kernel-Based Object Representation for Weakly-Occluded Oriented Object Detection. arXiv preprint arXiv:2104.11435, 2021.

[2] Qi Ming, Xuepeng Shi, Pei Xu, Jianqiang Huang, and Xiaolin Hu. Optimization for Arbitrary-Oriented Object Detection via Representation Invariance Loss. arXiv preprint arXiv:2103.11636, 2021.

[3] Zhen Zhou, Qiyu Kang, Zizhao Shen, Fan Jin, and Song Bai. CFNet: An Efficient Instance Segmentation Framework Based on Oriented Bounding Boxes. arXiv preprint arXiv:2401.08174, 2024.

[4] Jingru Yi, Wenjing Jia, and Dacheng Tao. Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pages 1406–1415, 2021.