Impact of Geometric Priors: Advanced Fine-grained Airplane Detection with Geometric Details in High-resolution Satellite Images

Tobias Traiser, Hai Huang and Helmut Mayer

Institute for Applied Computer Science, University of the Bundeswehr Munich, Germany

Keywords: Satellite imagery, Object detection, Bayesian inference, Airplane, Fine-grained classification

Abstract

Improved availability and quality of high-resolution satellite imagery allow for reliable airplane detection. Yet, fine-grained classification, especially of commercial airliners, remains a formidable challenge. Besides common difficulties, such as varying image artifacts and occlusions, the main challenge lies in the strong visual similarity between airliner families. This paper presents a geometry-aware classification that enhances oriented object detectors by integrating absolute measures and geometric features – fuselage length, wingspan, wing sweep angle, engine count, and fuselage width – in the form of priors into a Bayesian maximum a posteriori (MAP) estimation. The proposed pipeline is detector-agnostic by updating class posteriors without retraining the main detector. On the Gaofen Challenge dataset, it results in consistent improvements based on untuned baseline detectors, which outperform the top scores of the sophisticated fine-tuned models. An oracle experiment reveals the potential of the approach with an upper limit of the overall mean Average Precision of up to 0.96 and 0.98 for Gaofen and SuperView data, respectively. Furthermore, the impact of the employed geometric attributes is quantitatively evaluated.

1. Introduction

High-resolution optical satellites capture airports worldwide at sub-metre ground sampling distances (GSDs), enabling large-scale airplane monitoring. Yet, it also exposes the limits of purely appearance-based recognition in *fine-grained* settings where airplane families, e.g., A330/B787/B777, share highly similar planforms, i.e., shape and layout of fuselage and wing. Rotation-aware detectors localise airplanes robustly, yet family-level confusion persists (Ding et al., 2022, Xia et al., 2018). Domain knowledge suggests that physical geometry – fuselage length, wingspan, wing sweep angle, engine count, and fuselage width – carries discriminative information that is largely invariant to illumination, background, and (moderate) blur.

In this paper, we propose a refinement: Interpretable geometric attributes are estimated from the oriented bounding boxes and, where necessary, from auxiliary classifiers. These are subsequently fused with the detector's appearance-based posterior using a Bayesian maximum a posteriori (MAP) formulation (Berger, 1985). This probabilistic update refines class scores based on the consistency between the measured attributes and class-conditional geometric priors. By combining visual and geometric evidence, the method enhances fine-grained classification accuracy without the need for retraining or modification of the detector architecture.

Experiments on the Gaofen Challenge dataset (Sun et al., 2021) demonstrate that bivariate length–span priors and learned wing sweep angle priors reliably improve recognition of visually similar airplane families, while oracle analysis reveals substantial additional potential if geometric attributes could be measured perfectly. These results highlight that physically grounded priors remain a powerful complement to data-driven features in remote-sensing object detection.

The remainder of the paper is organized as follows. Section 2 reviews related work in object detection and recognition of specific airplanes in satellite data. The employed geometric attributes are introduced in Section 3 and the derivation of their priors is described in Section 4. Section 5 demonstrates experimental results with the proposed Bayesian inference framework. The paper ends with a conclusion in Section 6 outlining possible future work.

2. Related work

2.1 Object detection in optical remote sensing data

Object detection in aerial and satellite imagery must cope with arbitrary orientations, large-scale variation, cluttered backgrounds, and dense layouts of small objects. These challenges have driven a migration from horizontal bounding boxes (HBBs) to *oriented* bounding boxes (OBBs), where the box angle is a first-class variable. Large-scale benchmarks such as DOTA catalysed this shift and established standardised evaluation protocols for rotated detection (Xia et al., 2018, Ding et al., 2022).

In the beginning, two-stage CNN detectors were adapted to OBBs by transforming features or proposals. RoI Transformer learns geometry-aware pooling to align features with instance orientation (Ding et al., 2019), while Oriented R-CNN integrates rotation into proposal generation and refinement for end-to-end training (Xie et al., 2021). Gliding Vertex converts HBBs into accurate OBBs by regressing sliding vertex offsets (Xu et al., 2021). Rotation-equivariant backbones (ReDet) encode rotational symmetries directly in the network to improve generalisation (Han et al., 2021).

A complementary thread improves regression stability with geometry-aware losses: PIoU couples angle and overlap (Chen et al., 2020), RSDet mitigates rotation sensitivity via modulated loss (Qian et al., 2021), and Gaussian Wasserstein Distance and Kullback–Leibler formulations provide smooth surrogates for rotated IoU with favourable optimisation properties (Zhang and Zhang, 2024, Yang et al., 2023). Community toolkits such as MMRotate have further standardised datasets, models, and metrics, accelerating reproducible research (Zhou et al., 2022).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-289-2026 | © Author(s) 2026. CC BY 4.0 License.

Transformer-based detectors have also entered the remote-sensing domain. End-to-end formulations (DETR and derivatives) replace hand-crafted anchors with set prediction and global attention (Carion et al., 2020, Zhu et al., 2021, Yao et al., 2021), and vision transformers (ViT/CvT) or hybrid CNN–Transformer designs (Dosovitskiy et al., 2021, Wu et al., 2021) provide alternative backbones. Recent surveys summarise the state-of-the-art (SOTA) approaches in fine-grained detection in remote-sensing images, highlighting trends toward attention mechanisms, multi-scale fusion, and probabilistic modelling (Chu et al., 2024, Trigka and Dritsas, 2025, Gui et al., 2024). According to them, most methods optimise appearance features and bounding-box geometry jointly. Yet, explicit *attribute-level* priors (e.g., physical dimensions) are rarely incorporated, leaving a gap our approach addresses.

2.2 Airplane detection in optical satellite data

Airplanes are a canonical OBB target in very high-resolution imagery due to strong orientation variability. Dedicated datasets have emerged to benchmark their detection, including HRPlanes for high-resolution airplane detection (Bakirman and Sertel, 2023) and ADCOS as a broader airplane benchmark (Hu et al., 2024). FlightScope provides a comparative study of airplane detectors under realistic conditions (El Ghazouali et al., 2024). FineAir (Osswald et al., 2025) introduces multiple-level annotations, including a finer category for subtypes of airplane families. SOTA oriented detectors such as Oriented R-CNN and OrientedFormer report strong localisation and category recognition on aerial benchmarks (Xie et al., 2021, Zhao et al., 2024), and recent architectures continue to refine rotation modelling and receptive fields, e.g., Strip R-CNN (Yuan et al., 2025). One-stage designs, e.g., R2YOLOX, target efficiency for dense scenes with rotated targets (Liu et al., 2022), while task-specific improvements comprise attention mechanisms (Li et al., 2023), active/efficient annotation (Liang et al., 2023), and weak supervision (Xie et al., 2024). Work on loss functions for rotated regression has proven especially impactful for airplanes, where small angular errors translate to large IoU changes (Yang et al., 2023, Chen et al., 2020).

Despite these advances, *fine-grained* airplane recognition remains challenging: Airplane families such as A330/B777/B787 exhibit near-identical planforms at metre-level GSD, producing systematic confusion even for advanced detectors. Recent surveys on Fine-Grained Image Recognition (FGIR) in remote sensing emphasise the need for local, part-aware cues and structured priors (Chu et al., 2024). However, most detection pipelines still rely predominantly on appearance features. In contrast, our work introduces *explicit, physically interpretable* geometric attributes – fuselage length, wingspan, wing sweep angle, engine count, and fuselage width – modeled as class-conditional priors and fused via a Bayesian MAP update. This design complements appearance-based scores with geometry-informed evidence, improving fine-grained classification without retraining and aligning with the photogrammetric tradition of leveraging metric information for image understanding.

3. Measures and Geometric Attributes

Geometric measures and attributes provide physical and, therefore, stable cues to complement image interpretation solely based on appearance-based features in conventional detectors. In this case, high-resolution satellite imagery with an at least approximately known GSD has advantage compared to terrestrial and drone images.

In the proposed framework, each detected OBB is regarded as a geometric observation of a potential target airplane. A set of attributes – fuselage length, wingspan, wing sweep angle, engine count, and fuselage width – is derived and used to model class-conditional priors. These quantities are either directly measurable from the OBBs or estimated by auxiliary classifiers.

3.1 Airplane families

Although airplanes exhibit individual specifications, they – especially commercial airliners – share highly similar geometric characteristics, because they are concerned comparable with aerodynamic and structural constraints. This low inter-class variability is the main reason why fine-grained classification of commercial airplanes is challenging.

This work focuses on commercial airliners. We evaluate ten classes defined in the Gaofen Challenge dataset (Sun et al., 2021): Airbus A220, A321, A330, A350; Boeing B737, B747, B777, B787; COMAC ARJ21; and an *other* class that aggregates remaining types. The latter is rather heterogeneous: It may include additional Airbus/Boeing families and regional or legacy types that are not modeled as separate classes in the benchmark (e.g., A318/A319/A320 or A340; B757/B767).

Across all members (subtypes) within each airliner family, certain geometric characteristics remain consistent. In this work, this holds for fuselage width, wing sweep angle, and the number of engines. They define the overall aerodynamic configuration and structural layout of the whole family. In contrast, fuselage length and wingspan can vary between subtypes, reflecting differences in capacity, range, and performance requirements.

3.2 Fuselage length and wingspan

Length (fuselage length – L) and width (wingspan – S) are the main geometric measures of airplanes. They jointly characterize the overall size and proportions of an airplane. Here, they are estimated based on the OBBs proposed by the detector and scaled to metric units using the known GSD. L is derived from the bounding box side parallel to the fuselage axis, while the other side approximates S.

L and S exhibit systematic variability concerning both inter- and intra-family aspects. Inter-family differences primarily reflect design philosophies and operational roles – narrow-body families are optimized for short- to medium-haul missions, while wide-body types support higher payloads and longer ranges. Within the same airliner family, variability arises from distinct subtypes that share a common aerodynamic and structural architecture, but differ in overall size. Stretch and shrink variants modify fuselage length to adjust passenger capacity, while wingspan variations result from optional wingtip configurations or generation updates. Consequently, L and S are discriminative, yet correlated descriptors that capture both inter- and intra-class geometric diversity.

Modeling the fuselage length and the wingspan jointly as a bivariate Gaussian distribution (cf. Section 4.2.1) captures their strong correlation and enables a realistic probabilistic representation of each family's geometric footprint.

3.3 Fuselage width

Fuselage width represents the approximate cross-sectional diameter of the airplane body and complements the length–wingspan measurements. Here, this attribute is predicted by a dedicated auxiliary classifier trained to estimate the most probable width category from the oriented airplane crop. Table 1 shows the clear separation of so-called narrow-body and wide-body airliner families. Typical fuselage widths range from 3.2–3.8 m for narrow-body types to 5.6–6.2 m for wide-bodies. Please note that this attribute is particularly discriminative because it is the only geometric measure that is **consistently inter-class variable and intra-class identical**. As a measure with a lower and narrower range of values (in comparison with L and S), it is often hard to determine it exactly from noisy images with limited resolution. In practice, it can reasonably be expected that the fuselage width differentiates at least between narrow- and wide-body airliners and contributes synergistically with other attributes, such as wing sweep angle.

3.4 Wing sweep angle

The wing sweep angle (ϕ) is defined as the angle between the wing's leading edge (the foremost edge encountering the airflow during forward flight) and the longitudinal axis of the fuselage. Here, ϕ is not measured geometrically, but predicted by a dedicated CNN classifier that estimates **discrete** sweep angle values as predefined classes. The classifier is trained based on airplane crops (corresponding to the annotated bounding boxes) derived from the ground truth (GT) data. It captures subtle planform variations that are often imperceptible to conventional appearance-based detectors or even to human annotators. Significant as well as systematic differences in ϕ across airliner families make it a powerful discriminative feature for fine-grained classification.

Table 1 lists the representative sweep angles of the airplane families investigated in this study. All narrow-body airplanes including A220, A321, B737, and ARJ21, share the same sweep angle of 25°, while most wide-body families occupy a range between 30° and 32°. The Boeing 747 with its distinctive four-engine design has a larger angle of 37.5°. Incorporating this attribute into the MAP framework (see Section 5.3) particularly contributes to the separation of wide-body families that are otherwise nearly indistinguishable in their appearance.

Table 1. Representative geometric attributes of airplane families.

Classes	Fuselage	Wing sweep angle (°)	Engine count	Fuselage width (m)
A220	Narrow-body	25.0	2	3.28
A321	Narrow-body	25.0	2	3.70
B737	Narrow-body	25.0	2	3.76
ARJ21	Narrow-body	25.0	2	3.14
A330	Wide-body	30.0	2	5.64
B777	Wide-body	31.6	2	6.20
A350	Wide-body	31.9	2	5.96
B787	Wide-body	32.2	2	5.77
B747	Wide-body	37.5	4	6.08

3.5 Engine count

The number of engines provides an additional discrete cue with a clear geometric as well as visual representation. Within the Gaofen dataset, all families except the Boeing 747 are twin-engine airplanes, while only the B747 has four engines mounted on the wings. A lightweight CNN classifier is trained to infer the binary engine configuration, i.e., two or four engines, from individual airplane crops. Despite its simplicity, this categorical feature yields high reliability when engines are visible and implausible hypotheses in ambiguous visual conditions can be reliably suppressed in the MAP estimation.

Each of the above five attributes contains complementary information about the aircraft's geometry. Their combination creates a high-dimensional feature space and the joint class-conditional priors are able to enhance the fine-grained classification not only with high accuracy, but also interpretability in terms of physical meaning.

4. Data and priors

4.1 Gaofen dataset

The Gaofen challenge airplane dataset provides high-resolution optical satellite imagery along with annotations, including OBBs and fine-grained airplane labels (Sun et al., 2021). It comprises 3,000 images and 7,591 annotated airplanes across ten categories (B737, B747, B777, B787, A220, A321, A330, A350, ARJ21, and *other*). The given satellite data consist of 3,000 image patches with a size of 1,024 by 1,024 pixels. The dataset predominantly originates from Gaofen-2 satellites, which account for 84.8% of the imagery and provide a GSD of 0.8m. It is complemented by SuperView acquisitions at 0.5m GSD. Additionally, these sources introduce substantial variation in viewing geometry, and scene context (Sun et al., 2021). Besides, the challenges also include: (1) inconsistent source/quality of images (dominance of Gaofen-2 imagery), (2) instance imbalance between classes, (3) a heterogeneous *other* category with high intra-class variance concerning basic geometry, and (4) occasional occlusion by smoke or clouds.

Please note, that to fairly show the performance of the proposed method, we use the 1,000 final test images of the challenge as our test data.

4.2 Derivation of priors of geometric attributes

We treat the annotations in the training partition of the Gaofen dataset as observations of the geometric attributes $A = \{L, S, \phi, E, W\}$. Class-conditional priors for individual attributes are derived from the airplane instances, i.e., image crops defined by the OBBs.

4.2.1 Prior distributions from instances *Fuselage length* L and *wingspan* S are computed directly from the OBB axes together with the known GSD. The values for all instances are collected for statistical analysis. Figure 1 shows the distribution of fuselage-length values with the colors indicating different airliner families. The value ranges, overlaps as well as data imbalance can be observed.

We employ Gaussian distributions, as shown in Figure 2, to approximate the distributions of the instance values and model the priors. Without occlusion of data points (cf. Figure 1), the distributions of individual classes can be more clearly observed. The distribution of the “other” class demonstrates the high diversity of the instances.

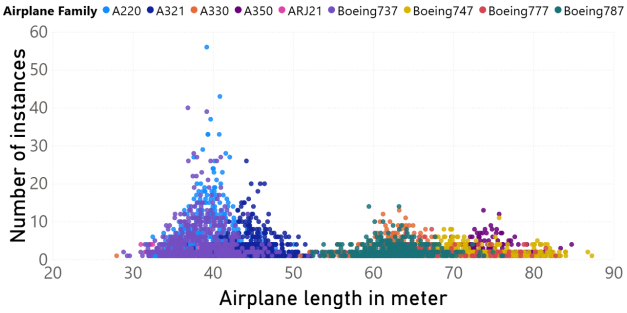

Figure 1. Instance distributions: Fuselage length

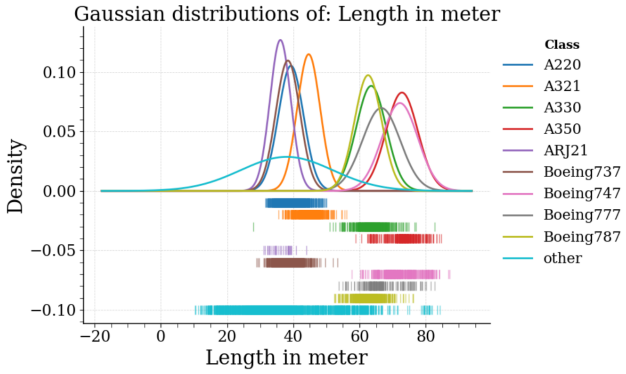

Figure 2. Fitting Gaussian distributions: Fuselage length (with additional class “other”)

Considering the strong correlation of fuselage length and wingspan, we further model (L, S) jointly with a bivariate Gaussian $p(L, S \mid c)$ for each class c. As demonstrated in Figure 3, the joint prior in two dimensions is consistently more discriminative than separate univariate distributions of length and wingspan.

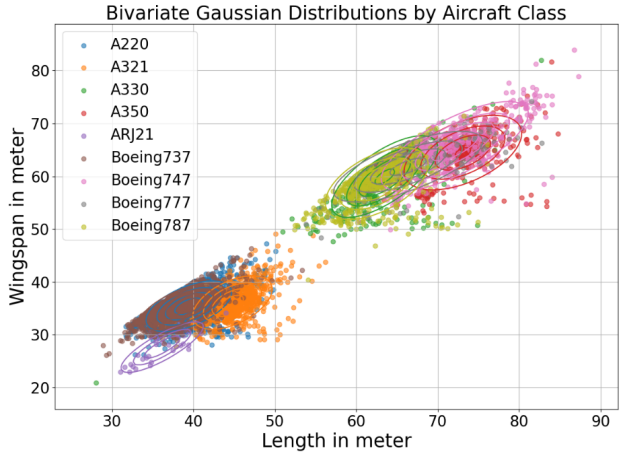

Figure 3. Bivariate priors of fuselage length and wingspan

4.2.2 Priors embedded in classifiers The remaining attributes, *wing sweep angle* ϕ and *fuselage width* W, are intra-class invariant features. I.e., a prior distribution can be easily modeled with the known specification value as the mean along with a given standard deviation. Such “perfect” priors, however, do not represent the real data characteristics or scene complexity. Instead of explicit distributions, we employ individual auxiliary classifiers, in which the collective prior information is implicitly embedded by training from the ground truth instances.

The *engine count* E is also intra-class invariant, but rather an appearance-based feature than a measurement. A specific CNN classifier is employed as the carrier of prior information collected from instances.

Just like the conventional prior distributions, the auxiliary classifiers provide an estimation of plausibility in the form of probabilities and fit well within the Bayesian maximum a posteriori (MAP) framework (van de Schoot et al., 2021, Van Dongen, 2006) (cf. Section 5.3).

5. Fine-grained classification of airplanes

5.1 Baseline models

We select four classic and SOTA oriented detectors: Strip R-CNN, RoI Transformer, Oriented R-CNN, and OrientedFormer as our baseline models for conventional appearance-only detection. In the basic evaluation, OrientedFormer attains the highest mean Average Precision (mAP), followed by Oriented R-CNN, Strip R-CNN, and RoI Transformer. Table 2 reports the performance of the appearance-only detectors as reference for further geometry-aware refinement.

Table 2. Performance of the appearance-only baseline models for all airplane families (Average Precision).

Classes	Strip R-CNN	RoI Transformer	Oriented R-CNN	Oriented Former
B737	0.279	0.296	0.317	**0.416**
B747	**0.900**	0.884	0.889	0.885
B777	0.041	0.087	**0.109**	**0.109**
B787	0.725	0.678	**0.731**	0.699
A220	0.461	0.423	0.449	**0.529**
A321	0.720	0.720	**0.737**	**0.737**
A330	0.783	0.771	**0.790**	0.722
A350	0.850	0.847	0.855	**0.893**
ARJ21	0.156	0.182	0.216	**0.222**
Other	0.755	0.765	0.737	**0.798**
mAP	0.567	0.565	0.578	**0.601**

We employ OrientedFormer, which is one of the SOTA transformer-based detectors and reaches the top performance among the previously reported sophisticated models by the competition (about 0.60), as the basis for further improvement in the following Sections.

5.2 Determination of geometric attributes

The attribute set $A = \{L, S, \phi, E, W\}$ is estimated for each detection proposal. The metric attributes fuselage length L and wingspan S are computed from the axes of the OBB using the known GSD. The remaining attributes – wing sweep angle ϕ, count of engines $E \in \{2, 4\}$, as well as fuselage width W – are determined by specific classifiers (cf. Section 4.2.2) applied on oriented airplane crops. All values will then be used to update the appearance scores in a MAP framework.

The wing sweep angle classifier predicts discrete angle classes (25°, 30°, 31.6°, 31.9°, 32.2°, 37.5°) and reaches an overall accuracy of 0.9463. The performance is strongly angle-dependent: The extremes are robust (recall 25°: 0.9989; 37.5°: 0.9886), whereas the narrow mid-bin at 31.6° is frequently confused with neighbouring classes (recall 0.0374), reflecting small inter-class distances of 1–2° and measurement noise at

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-289-2026 | © Author(s) 2026. CC BY 4.0 License.

Table 3. Wing sweep angle classifier: Selected per-class recalls.

Sweep class	Recall	Comments
25°	0.9989	Dominant, relatively stable
31.6°	0.0374	Confused with 30° and 32.2°
37.5°	0.9886	Distinct extreme (B747)
Overall		Accuracy = 0.9463

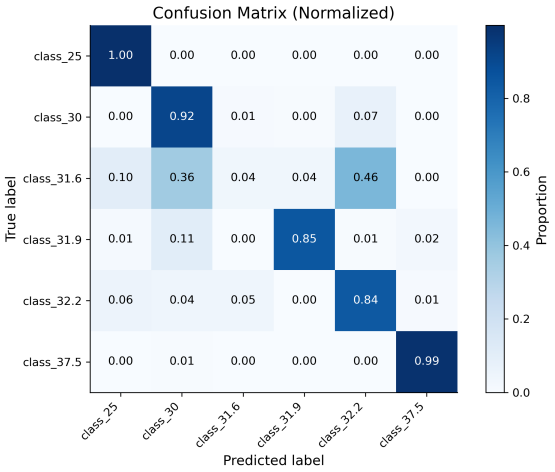

Figure 4. Wing sweep angle classifier: confusion matrix

metre-level GSD. Key results are listed in Table 3, while Figure 4 shows the confusion matrix.

The engine count classifier provides a highly reliable categorical cue. Only the B747 exhibits four engines, whereas all other families are twin-engine. Its overall accuracy reaches 0.9961 on the holdout data. The four-engine minority class is found to be more difficult, but still achieves a precision of 0.9698 and a recall of 0.9735; the twin-engine class is recognized almost perfectly (a recall of 0.9978). Absolute error counts are small, but relevant for post-processing (seven four-engine samples predicted as two; eight twin-engine samples predicted as four), as summarised in Table 4.

Table 4. Engine-count classifier: summary metrics on holdout data.

Class	Precision	Recall	Comments
2-engine	**0.9981**	0.9978	Few FP to four-engine
4-engine	0.9698	0.9735	Rare class; 7 FN; 8 FP
Overall		Accuracy = 0.9961	

The overall accuracy of the fuselage width classifier is low (0.6657), with pronounced confusion between adjacent narrow-body bins (3.76 m → 3.28 m: 638 cases; 3.28 m → 3.76 m: 311 cases). Extreme bins are more reliable; the only widest class at 6.20 m suffers low recall due to limited support. Table 5 summarises the main results. More details can be found in the confusion matrix in Figure 5.

In summary, the three classifiers provide complementary, physically interpretable geometric prior information. The engine-count classifier achieves near-perfect discrimination when engines are visible, while the wing sweep angle classifier reliably separates specific classes and assists within wide-body families. Fuselage width classification contributes additional evidence for narrow versus wide bodies, but cannot really deal with sub-metre margins and localization noise. In the following

Table 5. Fuselage-width classifier: accuracy

Measure	Instances	Comments
3.76 m → 3.28 m	638	Heavy confusion
3.28 m → 3.76 m	311	Heavy confusion
6.20 m (B777)	(Low recall)	Few samples
Overall	Accuracy = 0.6657	

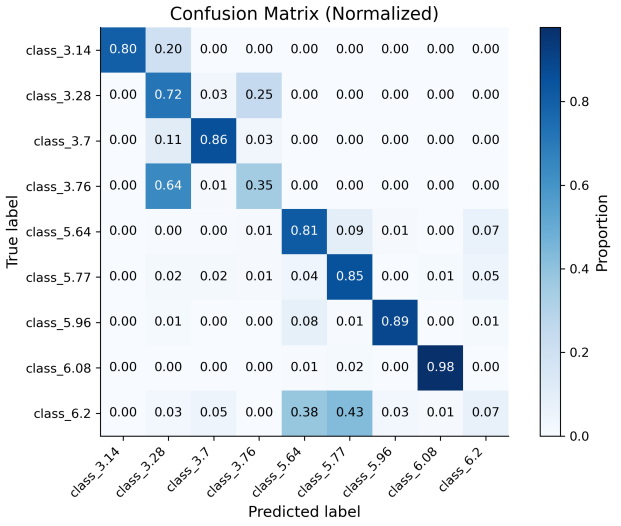

Figure 5. Fuselage width classifier: confusion matrix

MAP optimization, the geometric prior information is used to “re-weight” the class probabilities stemming directly from the baseline detector, i.e., solely based on the visual appearance in images.

5.3 Maximum posteriori estimation in classification

Let $L(c \mid x)$ denote the appearance-based likelihood of the baseline detector for class c given image evidence x, and $a_k \in A$ the estimated k-th geometric attribute of the target airplane. Then, the posterior of class assignment conditional on both x and A can be expressed as

$$P(c \mid x, A) \propto L(x \mid c) \prod_k p_k(a_k, c), \tag{1}$$

assuming conditional independence of the attributes given c. I.e., the class-conditional likelihood based on the visual appearance in the images is further “updated/corrected” with geometric plausibility.

In practice, the outputs of the baseline model as well as individual auxiliary classifiers (in the form of logits) are first converted into probabilities for the computation in Equation 1 together with probabilities derived from the explicit prior of L-S.

This update re-ranks the class scores with geometric consistency and is implemented for the OrientedFormer baseline, which has been selected in Section 5.1 as the reference model for this study. This training-free refinement regularises fine-grained decisions by favouring classes whose geometric profiles agree with the measured attributes, thereby reducing family-level confusions without modifying the underlying detector.

5.4 Experimental results

This section presents the quantitative evaluation of the proposed geometry-aware MAP refinement on the Gaofen chal-

lenge dataset, analysing its theoretical potential via oracle experiments, i.e., with perfect attribute information, its practical performance on the dataset, and the specific impact of individual geometric attributes.

5.4.1 Oracle experiments for theoretical potential The theoretical potential of the proposed approach is evaluated by means of oracle experiments, in which the estimated geometric attributes are replaced by their ground-truth values. This experimental setting simulates the maximally achievable improvement and, thereby, quantifies the upper bound of the method if geometric information can be perfectly derived. The results (cf. Figure 6) show that geometric priors can contribute significantly to resolving fine-grained classification ambiguities that remain in solely appearance-based detection.

As explained in Section 3.3, the fuselage width is expected to be the most powerful discriminator among all the attributes, as it is the only measure showing clear inter-class differences. Its potential is clearly demonstrated in the oracle experiments: When the correct values (in the form of classes derived from GT-labels) of fuselage width are applied, the overall mean Average Precision (mAP) increases from 0.601 (OrientedFormer) to 0.965, corresponding to a gain of +0.364.

The wing sweep angle provides the second-largest “solo” improvement, raising the overall mAP from 0.601 to 0.787 (a gain of +0.186). As expected, the impact is particularly strong for wide-body families with highly similar visual appearance: the B777 improved by +0.887 AP, the A330 by +0.278 AP, and the B787 by +0.215 AP. These results show that the wing sweep angle is a decisive indicator of the aerodynamic configuration and demonstrate its substantial potential for the fine-grained discrimination of aircraft families with similar planforms.

The number of engines, while just differentiating the only four-engine family of B747 in the given dataset, still positively affects the overall mAP and leads to a measurable gain from 0.601 to 0.611 (+0.010). Please note that, along with the improved detection of B747, which directly benefits with an AP gain of +0.024, false positives (FPs) in other classes have been simultaneously reduced.

The improvements presented above are summarized in Table 6. Furthermore, being aware of the difference between data sources, we give the evaluations also for the separate subdatasets from Gaofen and SuperView. With the same combination of oracle geometric attributes, the overall mAP increased from 0.625 to 0.965 (+0.340) for the images from the Gaofen Satellite and from 0.597 to 0.988 (+0.391) for the SuperView data.

The complementary nature of the attributes – size, aerodynamic configuration, and structural features – implies that each cue could reinforce a different aspect of class separability. The oracle results, therefore, confirm the capability of the proposed approach to suppress most of the inter-class confusion for an ideal estimation of geometric details.

5.4.2 Practical improvement In the practical setting, the geometric attributes A are determined as described in Section 5.2 and employed in the same MAP pipeline as above. As summarized in Table 7, the joint size prior over fuselage length and wingspan raises overall mAP from 0.625 to 0.637 (+0.012). The gain concentrates on families where size regularises otherwise similar textures: A330 (+0.066 AP), B777 (+0.041), B787 (+0.039), with smaller improvements for A321 (+0.029), A220 (+0.015) and B737 (+0.011), while ARJ21 (–0.073) and B747 (–0.008) slightly decline due to noise sensitivity for small sizes and residual model bias. Importantly, recall remains essentially unchanged. The benefit stems from better ranking (fewer spurious duplicates, fewer implausible assignments) rather than from creating additional true positives.

Table 6. Oracle experiments: mAP improvements and impact of individual attributes

Attribute (oracle)	Appearance only	With MAP	Δ mAP
Engine count	0.601	0.611	+0.010
Wing sweep	0.601	0.787	+0.186
Fuselage width	0.601	0.965	+0.364
All combined (Gaofen, 0.8 m)	0.625	**0.965**	+0.340
All combined (SuperView, 0.5 m)	0.597	**0.988**	+0.391

Figure 6. Example results of oracle experiment: Classification results of the baseline model (center) have been updated (right) with given priors and proved by GT labels (left). The misclassification of B737 (red), A220 (purple), and “other” (blue) has been corrected.

The estimation of wing sweep angle contributes a further +0.009 mAP (0.601 → 0.610). Improvements are strongest for wide-bodies with subtle but systematic sweep differences: A330 (+0.046 AP), B787 (+0.027), and B777 (+0.016). The classifier itself attains 94.63% overall accuracy on holdout data, with near-perfect recalls at the extremes (25° and 37.5°), but pronounced spillover among adjacent mid-bins (e.g., 31.6° vs. 30°/32.2°), which explains minor recall losses for some classes after employing this prior.

By contrast, the learned engine-count cue – despite 99.61% test accuracy – does not improve the overall mAP (0.601 → 0.600, –0.001) because rare misclassifications in the four-engine minority class propagate into the re-weighting and reduce the B747 recall (–0.060). This sensitivity underlines that categorical constraints on rare attributes should be integrated softly. The learned fuselage width classifier has not been included in the experiments, because of its low accuracy (66.57%). Narrow-body mid-bins are frequently confused, and small rotation/localisation errors (at metre-level GSD) lead to substantial confusion with adjacent bins. Nonetheless, fuselage width has a large potential to contribute to the fine-grained classification if improved classifiers or, more importantly, data with higher resolution are available, which has been proven by the oracle experiments.

Overall, the practical test yields modest, but consistent gains when the geometric estimation provides reliable results (cf. Figure 7). The size prior systematically reduces family-level confusions through physically plausible sizes, the prior of wing sweep angle adds discriminative power exactly where sizes

Table 7. Practical improvement in classification

Attribute (estimated)	Appearance-only	With MAP	Δ mAP
Length–span (Gaofen)	0.625	0.637	+0.012
Length–span (Superview)	0.597	0.607	+0.010
Wing sweep ϕ	0.601	0.610	+0.009
Engine count E	0.601	0.600	−0.001

overlap, and the prior count of engines highlights the trade-off between strong categorical constraints and lack of rare-class robustness. Fuselage width, while theoretically dominant, needs means to be more reliably estimated before a meaningful deployment.

Figure 7. Example results of practical experiments: Classification results of the baseline model (center) have been updated (right) with given priors and proved by GT labels (left). The misclassification from A350 (red) to B747 (blue) has been corrected. One Helicopter (top left of the scene) has been additionally detected and labeled as “other” (purple).

5.4.3 Impact and balance of the selected attributes A detailed analysis of class-wise results reveals that the different attributes act on complementary confusion patterns. The joint fuselage length–wingspan prior is most effective for separating narrow-body types with similar appearance, but different sizes, notably A220 and B737, and in refining the boundaries between mid-size wide-body airliners such as A330, B777, and B787. The wing sweep angle prior proves essential in distinguishing wide-body airplanes that have a nearly identical size, but characteristic sweep angle differences of only 1–2°, for example, A330 (30°), B777 (31.6°), and B787 (32.2°). Despite the limited accuracy of the wing sweep angle classifier, this attribute improves global mAP and reduces misclassifications among these families. The engine count prior adds a categorical constraint that uniquely identifies the B747. Due to the rarity of four-engine aircraft in this dataset, however, it is highly sensitive to mispredictions and thus requires “soft” probabilistic integration rather than binary filtering. The fuselage width attribute is the strongest single discriminator in theory, but its limited estimation accuracy currently prevents effective use. Together, these findings underline that fine-grained recognition can benefit from a balanced combination of robust geometric measurements and probabilistic reasoning. The proposed geometry-aware MAP refinement transforms the physically interpretable cues into quantitative improvements in classification performance, providing a promising approach towards further gains with improved attribute estimation.

6. Conclusion

We have presented a geometry-aware MAP refinement framework for fine-grained airplane classification in high-resolution satellite imagery. The contributions of the proposed approach can be summarized as follows:

- Integration of interpretable geometric attributes, including measures of fuselage length, wingspan, wing sweep angle, fuselage width, as well as one configuration feature – the count of engines
- Derivation of priors for the proposed attributes and update of the class prediction within a Bayesian maximum a posteriori (MAP) framework
- Investigation of the impact of individual geometric priors on the fine-grained classification

In future work, we will first focus on improving attribute estimation. The determination of fuselage length and wingspan will not be limited to bounding boxes, but supported by finer segmentation. The latter provides an approximate mask as well as the overall area of the airplane, which can also contribute to the classification. The auxiliary classifiers, for which lightweight models with simple backbones have been used, could be replaced by fine-tuned sophisticated SOTA models. The proposed framework can be extended to new commercial and non-commercial (special-purpose) aircraft classes with reduced demand for training data, as long as reliable prior information is available. Furthermore, the inherent advantage of interpretable geometric information can be utilized to tackle another tricky problem, namely, partially occluded objects under clouds, hangars, or across the boundaries of image patches. Based on geometric constraints, the proposed attributes can, at least theoretically, be derived from one-half or even one-fourth of an airplane.

References

Bakirman, T., Sertel, E., 2023. A Benchmark Dataset for Deep Learning-Based Airplane Detection: HRPlanes. *International Journal of Engineering and Geosciences*, 8(3), 212–223.

Berger, J. O., 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-End Object Detection with Transformers. *Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I*, Springer-Verlag, Berlin, Heidelberg, 213–229.

Chen, Z., Chen, K., Lin, W., See, J., Yu, H., Ke, Y., Yang, C., 2020. PIoU Loss: Towards Accurate Oriented Object Detection in Complex Environments. A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (eds), *Computer Vision – ECCV 2020*, Springer International Publishing, Cham, 195–211.

Chu, Y., Ye, M., Qian, Y., 2024. Fine-Grained Image Recognition Methods and Their Applications in Remote Sensing Images: A Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 19640-19667.

Ding, J., Xue, N., Long, Y., Xia, G.-S., Lu, Q., 2019. Learning RoI Transformer for Oriented Object Detection in Aerial Images. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2844–2853.

Ding, J., Xue, N., Xia, G.-S., Bai, X., Yang, W., Yang, M. Y., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2022. Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11), 7778-7796.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations*.

El Ghazouali, S., Gucciardi, A., Venturini, F., Venturi, N., Rueegsegger, M., Michelucci, U., 2024. FlightScope: An Experimental Comparative Review of Aircraft Detection Algorithms in Satellite Imagery. *Remote Sensing*, 16(24).

Gui, S., Song, S., Qin, R., Tang, Y., 2024. Remote Sensing Object Detection in the Deep Learning Era—A Review. *Remote Sensing*, 16(2).

Han, J., Ding, J., Xue, N., Xia, G., 2021. ReDet: A Rotation-equivariant Detector for Aerial Object Detection. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2785-2794.

Hu, J., Zhi, X., Zhang, B., Shi, T., Cui, Q., Sun, X., 2024. A Benchmark Dataset for Aircraft Detection in Optical Remote Sensing Imagery. *Remote Sensing*, 16(24).

Li, Y., Wang, H., Dang, L. M., Song, H.-K., Moon, H., 2023. ORCNN-X: Attention-Driven Multiscale Network for Detecting Small Objects in Complex Aerial Scenes. *Remote Sensing*, 15(14).

Liang, D., Zhang, J.-W., Tang, Y.-P., Huang, S.-J., 2023. MUS-CDB: Mixed Uncertainty Sampling With Class Distribution Balancing for Active Annotation in Aerial Object Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1-13.

Liu, F., Chen, R., Zhang, J., Xing, K., Liu, H., Qin, J., 2022. R2YOLOX: A Lightweight Refined Anchor-Free Rotated Detector for Object Detection in Aerial Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-15.

Osswald, M., Niederloehner, L., Koejer, S., Ziedorn, T., Gulli, V., Mommert, M., Mayer, H., 2025. FineAir: Finest-grained Airplanes in High-resolution Satellite Images. *Proceedings of the Winter Conference on Applications of Computer Vision (WACV) Workshops*, 1191–1199.

Qian, W., Yang, X., Peng, S., Yan, J., Guo, Y., 2021. Learning Modulated Loss for Rotated Object Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(3), 2458-2466.

Sun, X., Wang, P., Yan, Z., Diao, W., Lu, X., Yang, Z., Zhang, Y., Xiang, D., Yan, C., Guo, J., Dang, B., Wei, W., Xu, F., Wang, C., Hänsch, R., Weinmann, M., Yokoya, N., Fu, K., 2021. Automated High-Resolution Earth Observation Image Interpretation: Outcome of the 2020 Gaofen Challenge. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 8922-8940.

Trigka, M., Dritsas, E., 2025. A Comprehensive Survey of Machine Learning Techniques and Models for Object Detection. *Sensors*, 25(1).

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., Yau, C., 2021. Bayesian Statistics and Modelling. *Nature Reviews Methods Primers*, 1(1), 1.

Van Dongen, S., 2006. Prior Specification in Bayesian Statistics: Three Cautionary Tales. *Journal of Theoretical Biology*, 242(1), 90-100.

Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., Zhang, L., 2021. CvT: Introducing Convolutions to Vision Transformers. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 22–31.

Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3974–3983.

Xie, X., Cheng, G., Feng, X., Yao, X., Qian, X., Han, J., 2024. Attention Erasing and Instance Sampling for Weakly Supervised Object Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-10.

Xie, X., Cheng, G., Wang, J., Yao, X., Han, J., 2021. Oriented R-CNN for Object Detection. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 3500–3509.

Xu, Y., Fu, M., Wang, Q., Wang, Y., Chen, K., Xia, G.-S., Bai, X., 2021. Gliding Vertex on the Horizontal Bounding Box for Multi-Oriented Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4), 1452-1459.

Yang, X., Zhou, Y., Zhang, G., Yang, J., Wang, W., Yan, J., Zhang, X., Tian, Q., 2023. The KFIoU Loss for Rotated Object Detection. *arXiv e-prints*, arXiv:2201.12558.

Yao, Z., Ai, J., Li, B., Zhang, C., 2021. Efficient DETR: Improving End-to-End Object Detector with Dense Prior. *ArXiv*, abs/2104.01318.

Yuan, X., Zheng, Z., Li, Y., Liu, X., Liu, L., Li, X., Hou, Q., Cheng, M.-M., 2025. Strip R-CNN: Large Strip Convolution for Remote Sensing Object Detection. *arXiv preprint arXiv:2501.03775*.

Zhang, H., Zhang, S., 2024. Focaler-IoU: More Focused Intersection over Union Loss. *ArXiv*, abs/2401.10525.

Zhao, J., Ding, Z., Zhou, Y., Zhu, H., Du, W.-L., Yao, R., El Saddik, A., 2024. OrientedFormer: An End-to-End Transformer-Based Oriented Object Detector in Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-16.

Zhou, Y., Yang, X., Zhang, G., Wang, J., Liu, Y., Hou, L., Jiang, X., Liu, X., Yan, J., Lyu, C., Zhang, W., Chen, K., 2022. MMRotate: A Rotated Object Detection Benchmark using PyTorch. *Proceedings of the 30th ACM International Conference on Multimedia*.

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J., 2021. Deformable DETR: Deformable Transformers for End-to-End Object Detection. *International Conference on Learning Representations*.