

GABI: Geometry-Aware Boundary Integration for Spacecraft Segmentation

Iason Georgios Velentzas Dhruv Ahuja Panagiotis Tsiotras

Georgia Institute of Technology

{ivelentzas3,dahuja8,tsiotras}@gatech.edu

Abstract

Accurate segmentation is crucial for autonomous spacecraft, as it directly affects downstream tasks related to 3D situational awareness. The harsh illumination conditions of space, however, produce images with high variability in appearance, hindering the generalization of segmentation approaches across different spacecraft and environments. In this work, we propose GABI, a lightweight boundary-aware multi-task segmentation architecture that augments a convolutional backbone with an auxiliary distance-field prediction head. The distance field provides dense geometric supervision around object boundaries, encouraging the network to learn spatially consistent representations of spacecraft structures while maintaining low model complexity suitable for onboard perception systems. We evaluated GABI against both an established convolutional baseline and a heavier transformer-based architecture. On the SPARK benchmark, distance-field supervision improves the baseline by up to 5% in Average Precision while achieving performance comparable to the transformer models. In generalization experiments, GABI improves Average Precision by more than 50% over the baseline. In cross-domain evaluation, the lightweight GABI variant performs within 5% in IoU and F1-score of the heavier transformer model while being approximately ten times smaller. At the same time, the heavier GABI variant surpasses the transformer architectures while remaining nearly three times lighter.

1. Introduction

Space exploration has long been at the forefront of technological advancement, with profound impacts on our everyday lives. From communication infrastructure and GPS, to weather forecasting, maintaining the sustainability of the space environment is critical to our society. Recent studies by the European Space Agency (ESA) [1, 2] have highlighted the urgent need for Active Debris Removal (ADR) and On-Orbit Servicing (OOS) due to increasing orbital congestion. Solving this problem requires that navigation

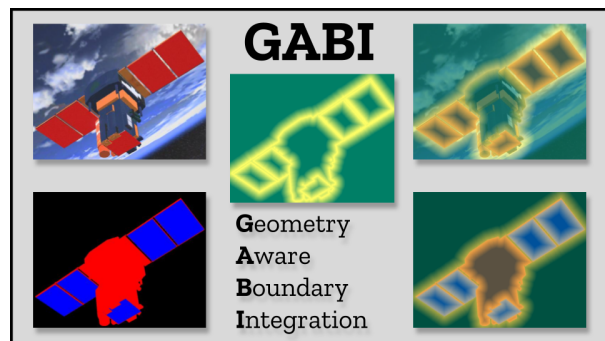


Figure 1. GABI is using the distance field transform to integrate geometric information in the segmentation architecture

in orbit be autonomous, efficient, robust, and at-scale [3].

Monocular cameras are an attractive modality for scalability [4, 5]. They provide rich information while meeting the weight, size, power, and cost requirements relative to other active sensing modalities. Moreover, advances in Deep Learning for perception have demonstrated remarkable robustness and efficiency. However, the scarcity of both synthetic and real-world images has limited the applicability of methods developed for space applications [6]. The SPARK 2026 Challenge provides a dataset of 10 spacecraft with a variety of pose configurations, accompanied by masks for part segmentation.

Spacecraft segmentation is a critical task for spacecraft autonomy. Accurate localization and identification of space objects are crucial for many downstream tasks, such as pose estimation and 3D situational awareness [7]. Target misidentification can lead to losing track of the target, wasting valuable fuel, and potentially failing the mission completely. Existing approaches to spacecraft segmentation address the challenge of developing efficient solutions, but robustness and generalization remain largely unexplored.

The space domain poses significant challenges for image processing algorithms [8]. Varying illumination conditions and the absence of atmospheric diffusion can produce images with markedly different photometric properties or lead to pronounced effects, such as sharp shadowing or lens

flares. Similarly, the variety of image backgrounds, ranging from Earth at daytime or nighttime to the vast emptiness of space, is an important factor affecting the overall appearance of the image. Specifically for the predominantly texture-based task of spacecraft segmentation, the developed algorithms must be flexible and adapt to significantly different appearances of the same texture.

Despite the harsh illumination conditions in orbit, the geometric information in spacecraft imagery can help achieve robustness and in-domain generalization of the solution [9]. Spacecraft geometry is typically characterized by rigid modular structures comprising planar or cylindrical surfaces with sharp edges. Despite the fact that structural elements in the images can appear overlapping and with fine detail, such as panel appendages or antennas, the observed spacecraft possess distinct geometric properties that can be leveraged as geometric priors or regularizers in perception to achieve robust target localization.

We propose **Geometry-Aware Boundary Integration (GABI)**, a solution to the spacecraft segmentation problem that involves a geometric regularizer to achieve robust performance across datasets. We construct distance fields from existing segmentation masks and integrate them into a baseline efficient segmentation network. We formulate our solution as a multi-task learning approach, in which the geometric distance field serves as a regularizer for the learned features. The goal is for the network to learn features that capture both texture and geometric details, such as sharp straight edges or corners. Although the idea of explicitly modeling object boundaries for segmentation has been explored in different forms in the literature, learning a distance field provides a richer geometric representation. Instead of encoding only the location of the boundary, the distance field describes the spatial relationship of every pixel with the object contour, enabling the network to better capture the global structure of the spacecraft components.

This paper investigates the effect of learning geometry and texture together in the spacecraft segmentation context with the goal of robustness and generalization. We compare our proposed architecture with the baseline approach that we selected across different spacecraft and datasets. Given the lack of masks for different parts in other datasets, we conduct experiments of instance segmentation, which treats the whole spacecraft as foreground. Our experiments demonstrate that our network, while only marginally improving performance in the SPARK 2026 validation set, it is capable of producing significantly more complete foreground proposals in other datasets and across vastly different illumination conditions.

2. Related Work

Efficient Semantic Segmentation Semantic segmentation aims to assign a semantic label to every pixel in an

image, enabling dense scene understanding in autonomous navigation and remote sensing. Although early architectures demonstrated strong performance on dense pixel-wise prediction tasks, these same models rely on heavy backbone networks and intensive encoder-decoder structures [10–12]. Mask R-CNN [13], for example, employs a two-stage approach that achieves strong instance segmentation performance but incurs a high computational cost, with approximately 44M parameters and a correspondingly high count of FLOPS in the ResNet-50-FPN backbone version [14].

To enable efficient inference while maintaining strong representational capacity, lightweight backbone networks such as MobileNet [15] and EfficientNet [16] were introduced. Specifically, MobileNet achieves efficiency through depthwise separable convolutions, which factorize a standard convolution into a depthwise convolution followed by a pointwise 1×1 convolution. This decomposition dramatically reduces the computation and parameter count while preserving spatial feature extraction. EfficientNet extends this line of work by introducing a compound scaling strategy that scales network depth, width, and image resolution.

These advances enabled the development of efficient segmentation architectures that leverage lightweight backbones with powerful segmentation heads [17–21]. Among these, DeepLabV3+ [17] has established itself as a widely adopted framework for its effective multi-scale context modeling via atrous spatial pyramid pooling (ASPP). The authors demonstrate that atrous convolution enables accurate multi-scale segmentation by expanding the receptive field through kernel dilation, without introducing additional learnable parameters. More recently, Segformer [18] introduced a hierarchical Mix Transformer (MiT) encoder with overlapping patch embeddings and spatial-reduction self-attention. The capacity to produce multi-scale feature maps, paired with a lightweight MLP decoder for feature fusion, enables efficient and accurate semantic segmentation.

Structure-Aware Image Segmentation A significant portion of the image segmentation literature has been devoted to coupling geometric and semantic information. The two main ways to use segmentation changes to infer geometry are either binary boundaries or continuous distance transforms.

The works of [22, 23] utilize binary boundary maps to improve local segmentation close to ambiguous boundary predictions. Gated-SCNN [22] proposed using the main and more powerful segmentation stream to guide a separate parallel stream that encodes shape information. The network’s performance is also boosted via a novel duality regularization loss between the two streams. On the other hand, SEMEDA [23] designed a semantic edge-aware loss in the feature space. These works achieve sharper predictions near the boundaries, boosting performance on thinner

structures, without completely eliminating boundary misdetections.

Boundary-aware segmentation typically focuses on improving local accuracy near the boundary and on capturing high-frequency details where masks change. Distance fields, on the other hand, are used to add global context from the image and to exploit the structure of the scene even beyond the boundaries. Distance fields are a rich representation that encodes dense per-pixel boundary information, offering robustness to occlusions and noise. The works of [24, 25] opt to learn a distance field for each label in the image. Learning signed distance fields [24] for each label and possibly the corresponding class in a multi-task approach [24] showcases the ability to capture spatial context in aerial images and a significant improvement in smoothness of predictions and lack of holes in the segmentation masks.

Similar to boundary approaches, many distance transform methods focus on enhancing predictions along the boundary. The work of [26] was motivated by the observation that the detect first and then segment approach of Mask R-CNN [27] leads to unrecoverable segmentation errors when the bounding boxes are detected erroneously. The authors aimed to enable masks to extend beyond the initial proposed bounding boxes and introduced a distance-based mask representation that predicts a binary mask per class. Investigating the distribution of mask prediction errors across baseline networks, the authors of [28] show that the mask predictions of the interior pixels are significantly more reliable. They propose a model-agnostic approach to segmentation refinement in which boundary pixels are localized by learning a boundary and direction map, and subsequently unstable predictions are replaced with the corresponding interior pixel.

In conclusion, the vast majority of works on structure-aware segmentation in terrestrial applications use distance fields and boundary masks to geometrically regularize the network prediction and enhance performance on fine-grained object boundaries. A commonly observed outcome is that pixel-wise losses often misconstrue the image topology, and adding extra geometric supervision helps to achieve structurally consistent mask predictions.

Spacecraft Segmentation In contrast to terrestrial applications, the space domain presents a substantially different set of challenges and characteristics. For spacecraft recognition, a typical categorization of interest is between the spacecraft body and panel material components, a convention also used in the SPARK 2026 challenge. Other datasets might involve more components, such as antennas [29], or even focus on instance segmentation, where the need is to distinguish the spacecraft from the background [30].

As the community addresses the problem of data scarcity

and more datasets are generated, baseline segmentation algorithms are being evaluated as performance benchmarks. The authors of [29] compared various state-of-the-art segmentation networks in both instance segmentation and parts recognition. The results of the latter experiment indicate that the baseline algorithms have trouble identifying the antennas, which are a complex and complex-shaped module, as well as facing problems when the antenna overlaps with the main body. Prior work in [31] proposes a semi-physical image collection system that focuses on the extreme effects of illumination conditions in space and on realistic, complex structures. The images have nine distinct annotated labels, and the results indicate that efficient segmentation architectures do not accurately capture small structures. Recently, [32] have offered a new synthetic dataset and provided results for finetuning different versions of YOLO segmentation models. The authors formalized the onboard constraint requirements based on NASA standards and deemed that YOLOv8n and YOLOv11n are the best performing variants that meet these requirements [33, 34].

Structure-Aware Spacecraft Applications Despite the data scarcity in the space domain, there are some works that explore the intersection of geometric and semantic information of spacecraft. For the problem of line segment detection, [35] used an illumination augmentation scheme, similar to typical homographic augmentation, to detect line segments robustly under various illumination conditions. The authors showed that the global context that is injected into the problem by using attraction fields [36] is able to generate more complete line segments under harsh conditions, such as low brightness or low contrast. Moreover, IRoLSD [35] exhibited remarkable generalization performance on unseen spacecraft geometries and environmental conditions in both synthetic and real-world images.

In particular for the problem of spacecraft segmentation, [37] created a boundary and a distance field from the mask boundaries to augment the segmentation stream with two parallel heads: a boundary prediction head that learns the binary boundary mask and a head that uses the distance field contrastive context feature learning. This work focuses on improving segmentation locally across the boundary and, despite not learning a continuous distance field representation, it shows significant improvement in the overall boundary prediction.

A more influential work in the field was presented in [38]. Focused on instance segmentation, the authors explore the robustness of the segmentation pipeline using geometric primitives in a multi-task learning scenario. In addition to the segmentation head, the authors develop two heads, one for pose estimation and one for keypoint detection. The keypoints encode 2D geometric information, whereas the poses provide strong 3D projective-geometry

information to the network. The complete evaluation suite employs various space-specific image augmentations, such as sun flare and style augmentations, and, respecting mission-operational constraints, it develops an online domain refinement module to assess performance gains in real-world scenarios.

Overall, prior work suggests that incorporating geometric representations alongside semantic predictions can improve the structural consistency of segmentation models. In this work, we investigate whether learning a continuous distance field can introduce a geometric inductive bias in the network, encouraging representations that remain robust across varying imaging conditions and unseen spacecraft geometries.

3. Method

This section presents a boundary-aware semantic segmentation framework developed for the SPARK 2026 Spacecraft Perception Challenge, which involves multi-task spacecraft perception tasks including classification, detection, and semantic segmentation. In our approach, semantic segmentation serves as the primary prediction task, while detection outputs are derived directly from the predicted segmentation masks via bounding box extraction. As a result, improvements in segmentation quality also directly impact detection performance.

To improve segmentation accuracy, particularly along object boundaries, we augment a standard encoder–decoder segmentation architecture with an auxiliary distance field prediction task. We first describe the baseline segmentation network used for the challenge. We then introduce the proposed distance field prediction module and describe its integration as an auxiliary head in the architecture. Finally, we present the training objectives used to jointly optimize segmentation and distance field supervision.

3.1. Multi-Task Baseline Architecture

The baseline architecture, depicted in Figure 2, follows the encoder–decoder style architecture used in DeepLabV3+ and can be summarized by the components *Shared Feature Encoder* and *Baseline Heads*. Feature extraction is performed using the lightweight MobileNetV2 backbone [15], from which multiple resolution maps are extracted and fed as input to the *Bidirectional Feature Pyramid Network (BiFPN)* [39]. Projecting each level to a fixed channel dimension before fusion allows efficient cross-scale aggregation, combining detail-oriented features with semantic features. The backbone network produces a pyramid of features P2 – P5, with progressively lower spatial resolution. These features are fused using a *BiFPN* module, which combines information across multiple scales through top-down and bottom-up pathways. The segmentation decoder follows the standard DeepLabV3+ architecture, aggregat-

ing multi-scale context via ASPP and fusing high-level semantic features with lower-level spatial features to generate pixel-wise predictions for the classes *background*, *spacecraft body*, and *spacecraft panel*.

In our design, spacecraft classification is performed using a global pooling on the coarser output of *BiFPN* feature map, followed by dropout and a linear layer to predict the spacecraft class. The shared head structure enables efficient joint-learning of semantic features related to spacecraft recognition. The segmentation head and classification head are supervised simultaneously, with both tasks monitored using cross-entropy loss.

3.2. Boundary Integration

While binary segmentation masks define a discontinuous indicator function in the image domain, the distance transform defines a continuous scalar field that encodes the distance to the nearest object boundary [36]. The distance field is computed from mask boundary pixels where label transitions occur. For each pixel in the image domain, we compute the Euclidean distance to the nearest mask boundary and clip the value at a maximum distance $d_{\max}$.

Distance Field Implementation The network predicts a bounded distance field $\hat{D}$ by upscaling the encoder-generated features through a series of convolutional layers and bilinear interpolation, which are followed by ReLU activations and Batch normalization. Similarly to DeepLSD’s line neighborhood, we only use a region around the boundary as valid for distance field learning.

Boundary Gating for Segmentation To utilize the geometric information encoded from the boundary and improve segmentation performance along the boundary, we propose to use a boundary-gating block that modulates segmentation features using the predicted distance field. More specifically, let $\mathbf{F}$ denote the segmentation decoder feature map. Using the predicted distance field $\hat{D}$, we construct a gating map $B = w \exp\left(-\frac{\hat{D}^2}{2\sigma^2}\right)$, where w and σ are hyperparameters controlling the strength of the gating signal and the decay rate of the Gaussian weighting, respectively. The gated feature representation is then computed as

$$\mathbf{F}_{\text{gate}} = \mathbf{F} \odot (1 + \alpha \mathbf{B}), \quad (1)$$

where $\odot$ denotes element-wise multiplication and α controls the strength of the boundary modulation. This mechanism is similar to the Gated-SCNN attention map, such that it amplifies feature responses near predicted object boundaries, encouraging the segmentation decoder to focus on high-frequency structural cues while preserving global contextual information [22].

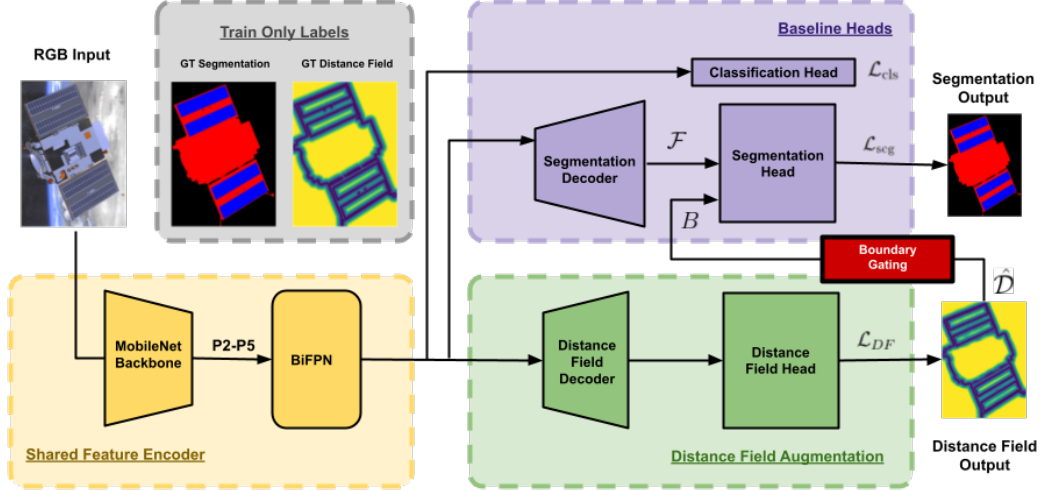


Figure 2. Overview of the proposed multi-task architecture. The predicted Distance Field is further used to generate a boundary-weighting map that modulates segmentation prediction, encouraging boundary-aware feature learning.

3.3. Loss Functions

The multi-task perception objective function can be expressed as a combination of the distance field, segmentation, and classification losses, which are described below:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{DF}} \mathcal{L}_{\text{DF}} + \lambda_{\text{seg}} \mathcal{L}_{\text{seg}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} \quad (2)$$

Classification Loss For $\mathcal{C}$ spacecraft models with y_c ground-truth labels, the spacecraft classification loss is defined as the cross-entropy loss on p_c :

$$\mathcal{L}_{\text{cls}} = - \sum_{c=1}^{\mathcal{C}} y_c \log p_c, \quad (3)$$

Distance Field Loss The distance field prediction is trained using a regression loss between the normalized distance field prediction $\hat{D}_n$ and the normalized ground-truth distance field D_n , around a valid neighborhood region of size $r \in \mathbb{R}_+$.

$$\mathcal{L}_{\text{DF}} = \|\hat{D}_n - D_n\|_1, \quad D_n = -\log\left(\frac{D}{r}\right) \quad (4)$$

Segmentation Loss The segmentation objective is defined as a standard pixel-wise cross-entropy loss applied over the image domain Ω and the predicted probability of the ground-truth segmentation class $p_{S_{\text{gt}}}$:

$$\mathcal{L}_{\text{seg}} = -\frac{1}{|\Omega|} \sum_{\Omega} \log p_{S_{\text{gt}}}, \quad (5)$$

4. Experiments

In this section, we evaluate the performance of the proposed architecture against different levels of generalizability. First, we evaluated the segmentation performance for every mask within the SPARK dataset. Second, we withhold part of the dataset and test the ability of the network to generalize in unseen spacecraft geometries within the same dataset domain. Finally, we evaluate the generalization of GABI across other synthetic domains. With these experiments, we show that distance-field supervision improves structural consistency beyond standard pixel-wise segmentation performance.

4.1. Experimental Setup

Datasets We used the SPARK 2026 Challenge dataset for the primary segmentation task. It is generated with Unreal Engine with ten distinct spacecraft models of varying scales, viewpoints, backgrounds, and lighting conditions. Each model contains 6,000 images for training and 2,000 images for evaluation. A subset of the dataset includes challenging visual effects, such as intense lens flares that introduce additional variability in the appearance of the models. For the domain-generalization task, we used the SPE3R dataset [30], which contains 100k synthetic images of 100 spacecraft models. SPE3R exhibits greater variation in brightness and contrast, whereas in SPARK images, the illumination profile is more homogeneous. Furthermore, images of SPE3R have prominent self-shadowing effects, rendering portions of the spacecraft nearly indistinguishable from the background. The two datasets differ both in spacecraft geometries and in illumination properties, allowing a meaningful cross-dataset generalization experiment.

Evaluation metrics Segmentation performance is evaluated using standard region-based metrics, for instance, and semantic segmentation. We report Average Precision (AP) at multiple Intersection-over-Union (IoU) thresholds. The evaluation metric of the SPARK 2026 Challenge for the segmentation task is the mean Average Precision, calculated by averaging AP scores across IoU thresholds from 0.50 to 0.95 with a step size of 0.05. This metric provides a comprehensive evaluation across multiple levels of spatial overlap. In addition, we report the standard evaluation metrics AP_{50} and AP_{75} , and when meaningful, the even stricter threshold of AP_{90} . For experiments conducted on the SPARK dataset, we evaluated the segmentation performance separately for the spacecraft body and solar panel components. For cross-domain experiments, the evaluation is performed in the foreground extraction task, which combines body and panel predictions. Since the spacecraft occupies a large portion of the image, it is deemed appropriate to also report the F1-score, predicated on precision and recall metrics to avoid misinterpreting high IoU that stem from network’s over-prediction.

Implementation details As mentioned above, we compare the performance of GABI against its baseline (BL), which uses the same architecture but without the distance-field augmentation. To ensure a fair comparison, the baseline model and the proposed architecture share the same backbone, training procedures, and optimization settings. Only the distance-field prediction head and its associated loss function are unique to GABI. We compare two variants of each model, *v2* and *v3s*, which are named after the mobilenet backbone that they use. In addition to these models, we include the two lightest versions of SegFormer [18], *b0* and *b1*. SegFormer is a transformer-based segmentation network with a relatively lightweight architecture that has demonstrated strong performance and robustness across multiple segmentation benchmarks. Including this model allows us to compare GABI not only against its direct convolutional baseline, but also against a competitive architecture with strong generalization capabilities.

To preserve detail, we train and test with full size images (1024x1024). We use the AdamW optimizer with the initial learning rate set to 3×10^{-4} and the weight decay set to 1×10^{-2} . Training is performed by finetuning pretrained models for 20 epochs using a batch size of 10 images on RTX-5090 GPUs.

4.2. Multi-Task Experiment

Tab. 1 presents the segmentation results for the two classes averaged on all spacecraft of the SPARK dataset. Overall, GABI consistently outperforms the baseline configurations with minimal parameter overhead. For the panel class, which consists of smaller, finely detailed structures, the per-

formance increase is more than 5% and is mainly observed at higher IoU thresholds, which denotes better segmentation accuracy. The fact that performance on the body class is satisfactory in both convolutional approaches can be explained by the homogeneity of the SPARK dataset in terms of appearance of texture, with limited variability in contrast and shadowing. The marginal improvement we observe in GABI corroborates the conclusion that the performance enhancement is due to the distance field’s ability to encode clear geometry and better represent boundaries, as qualitatively observed also in Fig. 3a.

Regarding SegFormer architectures, we observe that performance on the body class is significantly lower at higher IoU thresholds than with CNN-based architectures. This behavior is consistent with previous studies showing that CNNs benefit from strong spatial inductive biases such as locality and translation equivariance, allowing them to learn geometric structures efficiently from limited data, whereas transformer architectures require larger datasets to learn these priors implicitly [40, 41]. This is also visible in Fig. 3a, in which the SegFormer prediction fully captures the semantic meaning of panels in the image, but fails to add details in the spacecraft body prediction.

4.3. Generalization in SPARK

For this experiment, we exclude the *Soho* and *Proba3ocs* spacecraft models from the training. These models were specifically selected due to their strong geometric distinctions from other spacecraft classes. The panels in *Proba3ocs* resemble a combination of the structure of different models, whereas the *Soho* spacecraft is labeled with panels only on one side of the panel frames. Furthermore, the *Soho* panels possess a distinct glow that is not observed to that extent in other spacecraft models.

Tab. 2 reports the segmentation performance on the two hidden spacecraft models. All methods achieve relatively strong performance on the body class, while performance degrades at higher IoU thresholds for *Soho*, mainly due to panel misclassification. In this challenging scenario, GABI achieves a substantial improvement in mAP 30 - 45% compared to the competing models.

These results suggest that although segmentation performance remains stable at lower IoU thresholds due to the homogeneous texture of the dataset, the distance-field supervision enables more precise boundary localization. Similar to the observations in Sec. 4.2, SegFormer architectures exhibit limited performance in this setting.

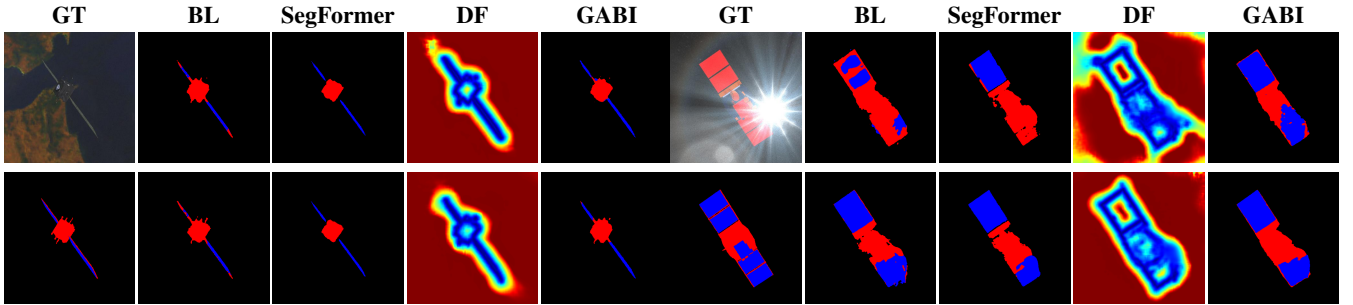
The qualitative comparison in Fig. 3b offers the results of a case in which intense lens glare affects a portion of the panel. We can see that all heavier models have similar performance in panel identification, with GABI better predicting the shape and, hence, the mask of the spacecraft body. Notably, in the top row, where the lighter networks

Table 1. Segmentation performance comparison on SPARK dataset. For each class (body, panel), we report AP at IoU thresholds of 0.50, 0.75, and 0.90, along with mAP averaged over IoU thresholds from 0.50 to 0.95 with a step size of 0.05.

Model	Params	FLOPs	Body				Panel			
			AP ₅₀	AP ₇₅	AP ₉₀	mAP	AP ₅₀	AP ₇₅	AP ₉₀	mAP
BL-v3s	2.7 M	4.8 G	99.8	94.1	74.2	88.5	80.2	73.6	53.1	68.6
GABI-v3s	2.8 M	8.6 G	99.8	94.3	74.5	89.2	84.3	76.2	56.9	73.1
BL-v2	3.9 M	20.5 G	99.8	96.1	81.8	91.4	82.3	77.0	61.3	71.8
GABI-v2	4.2 M	34.5 G	99.8	96.5	82.3	91.9	84.5	79.8	66.4	76.1
SegFormer-b0	3.7 M	44.0 G	99.6	89.9	62.4	74.1	84.4	78.7	65.9	74.8
SegFormer-b1	13.7 M	86.6 G	99.7	90.1	64.1	85.2	86.1	80.4	68.9	77.2

Table 2. Segmentation performance comparison on hidden spacecraft of SPARK dataset. For each spacecraft (Soho, Proba3ocs) and for each class (body, panel), we report AP at IoU thresholds of 0.50 and 0.75, together with mAP averaged over IoU thresholds from 0.50 to 0.95 with a step size of 0.05.

Model	Soho - Body			Soho - Panel			Proba3ocs - Body			Proba3ocs - Panel		
	AP ₅₀	AP ₇₅	mAP	AP ₅₀	AP ₇₅	mAP	AP ₅₀	AP ₇₅	mAP	AP ₅₀	AP ₇₅	mAP
BL-v3s	71.5	20.8	29.3	05.9	00.2	01.4	96.4	89.4	82.0	74.1	70.1	60.8
GABI-v3s	<u>86.1</u>	<u>35.9</u>	<u>42.3</u>	25.9	10.4	12.3	<u>98.9</u>	91.1	83.9	<u>81.3</u>	<u>78.1</u>	<u>69.1</u>
BL-v2	82.7	35.8	36.4	24.7	12.0	11.7	97.3	91.9	84.2	75.0	72.4	64.4
GABI-v2	89.2	39.5	45.8	38.7	26.2	24.3	99.2	93.8	88.2	82.1	79.1	74.4
SegFormer-b0	67.7	17.5	26.9	19.6	07.0	08.8	98.4	89.7	82.0	71.7	68.0	58.7
SegFormer-b1	68.8	25.0	31.2	<u>31.9</u>	<u>18.4</u>	<u>17.5</u>	98.7	90.6	83.2	71.4	68.5	58.9



(a) Multi-task experiment.

(b) Generalization in SPARK experiment on the Soho spacecraft.

Figure 3. Qualitative comparison of BL, SegFormer, and GABI. In both subfigures, the first column contains the input image and the ground-truth mask. Columns 2–3 show the segmentation predictions of BL and SegFormer [18], while columns 4–5 show the distance field and segmentation prediction of GABI, respectively. The top row corresponds to the light models (BL-v3s, GABI-v3s and SegFormer-b0), while the bottom row corresponds to the heavier models (BL-v2, GABI-v2 and SegFormer-b1).

are depicted, we observe that GABI leverages the geometry learned from the distance-field prediction to strongly outperform the competitors and match the performance of the heavier networks. It is evident that the distance field enables smoother, spatially connected prediction masks.

4.4. Cross-domain Generalization in SPE3R

To test generalization across domains, we employ SPE3R, a dataset significantly different in texture and geometry. The evaluation results on three previously unseen spacecraft models: Apollo–Soyuz, Grace, and Iro are summarized in Tab. 3. Across all spacecraft models, GABI consistently outperforms its convolutional counterparts and

achieves performance competitive with the much heavier transformer-based architectures. The most notable improvements occur at higher IoU thresholds. For example, on the Apollo spacecraft, GABI-v2 improves AP₇₅ from 15–20% to 65–70%, while the transformer architectures achieve between 54–67%. Similar improvements are observed in Grace, where the Average Precision for high IoU is almost doubled and it is still 20% higher than the heavier and best-performing SegFormer. These gains also translate into strong overall segmentation quality, as indicated by the consistent improvement in the F1-score.

Fig. 4 shows that GABI achieves performance comparable to, and in some cases better than, the heavier Seg-

Table 3. Segmentation performance comparison on three spacecraft models of SPE3R [30] dataset. For each spacecraft (Apollo, Grace, Iro), we report AP at IoU thresholds of 0.50 and 0.75, together with mAP averaged over IoU thresholds from 0.50 to 0.95 with a step size of 0.05. We provide the F1-scores and mean IoU over all images of the spacecraft.

Model	Apollo					Grace					Iro				
	AP ₅₀	AP ₇₅	mAP	F ₁	IoU	AP ₅₀	AP ₇₅	mAP	F ₁	IoU	AP ₅₀	AP ₇₅	mAP	F ₁	IoU
BL-v3s	60.4	16.2	25.1	67.7	54.0	69.5	52.3	49.1	72.1	78.7	73.1	35.4	38.5	74.0	62.1
GABI-v3s	<u>92.1</u>	65.2	54.4	84.6	75.2	<u>90.0</u>	78.7	68.7	<u>88.2</u>	<u>80.9</u>	89.7	65.3	56.9	84.5	75.0
BL-v2	<u>67.0</u>	19.5	29.2	70.9	57.6	<u>75.8</u>	62.2	58.9	78.7	71.1	75.4	47.7	46.3	76.1	64.7
GABI-v2	94.9	71.9	58.6	86.8	76.8	95.9	86.8	74.2	89.4	83.9	96.2	87.7	69.1	89.4	81.2
SegFormer-b0	86.5	54.2	<u>51.9</u>	82.5	72.2	84.2	72.9	68.3	84.6	78.2	90.6	62.7	60.9	86.4	77.5
SegFormer-b1	89.4	<u>67.4</u>	60.9	85.5	76.8	84.5	73.2	<u>69.1</u>	84.7	78.5	<u>92.7</u>	<u>74.9</u>	<u>66.1</u>	<u>87.2</u>	<u>79.4</u>

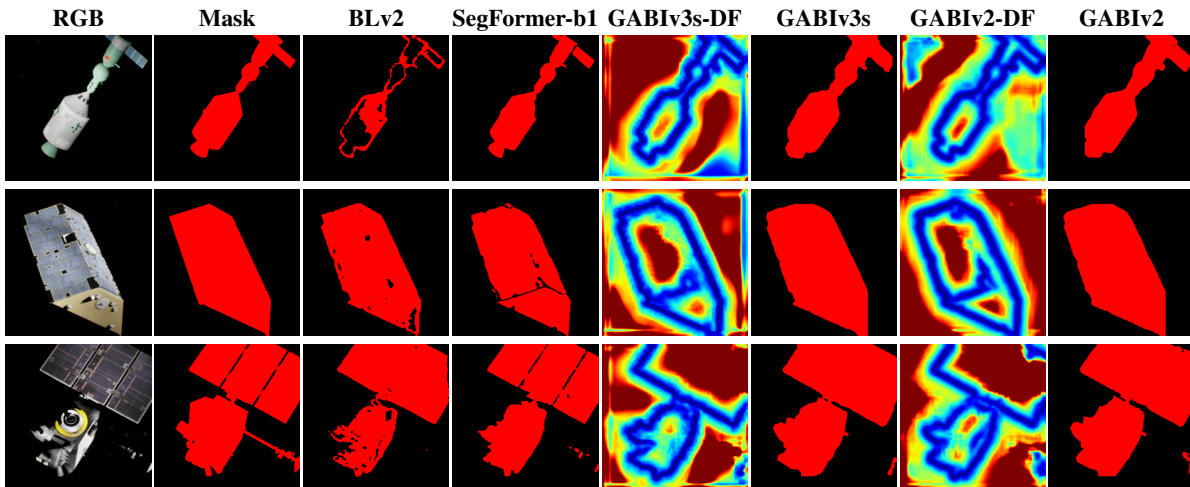


Figure 4. Qualitative comparison across three spacecraft from SPE3R. Columns 1–2 show the RGB inputs and ground truth masks. Columns 3–4 represent the predictions of BL-v2 and SegFormer-b1. Columns 5–6 and 7–8 contain the predictions for distance field and segmentation mask of GABI-v3s and GABI-v2, respectively. Note that the proposed GABI models have substantially fewer parameters and FLOPs than the SegFormer counterparts(Tab. 1).

Former architecture while using only 10%–30% of its complexity. Geometric supervision allows GABI to produce less fragmented segmentation masks while learning complete boundaries, even in dim regions. The findings of the previous experiment about smooth and connected predictions translate across domains as well. Overall, the learned distance field appears sufficiently accurate for downstream applications.

5. Conclusion

In this work, we propose GABI, an efficient segmentation network augmented with explicit geometric supervision. We show that the distance field prediction head can successfully localize the boundaries of the segmentation masks and the total spacecraft by integrating structural information. The multi-task segmentation setup with boundary gating helps the network learn structural regularities across different spacecraft models. This geometric regularization improves generalization across datasets, often leading to per-

formance increase comparable with much larger networks. The results also indicate that GABI-v2 and GABI-v3 consistently outperform their convolutional counterparts, with an improvement of up to 50 – 100% AP in some cases.

The strong performance of the proposed approach opens several promising directions for future research. We plan to investigate the robustness of the predictions under varying illumination conditions and in the presence of severe visual artifacts. Finally, it is essential to assess the network’s performance in a complete pipeline, in which GABI’s outputs are used for downstream tasks such as pose estimation or tracking.

6. Acknowledgments

Iason Georgios Velentzas was supported by the AFOSR SURI project (FA9550-23-1-0723). Iason would like to thank George Rapakoulias for their insightful discussions and feedback that improved the quality of this work.

References

- [1] European Space Agency. ESA space environment report 2025. Technical report, European Space Agency, March 2025. [1](#)
- [2] Camilla Colombo, Martina Rusconi, Juan Luis Gonzalo, Wiebke Retagne, Andrea Muciaccia, FJ Simarro Mecinas, Diego Ramirez, Francesca Letizia, and Emma Stevenson. Extension of the esa tool for tracking the health of the environment and missions in space. In *9th European Conference on Space Debris, Bonn, Germany*, pages 1–4, 2025. [1](#)
- [3] Simone D’Amico, Mathias Benn, and John L Jørgensen. Pose estimation of an uncooperative spacecraft from actual space imagery. *International Journal of Space Science and Engineering* 5, 2(2):171–189, 2014. [1](#)
- [4] Guangcong Zhang, Michail Kontitsis, Nuno Filipe, Panagiotis Tsiotras, and Patricio A Vela. Cooperative relative navigation for space rendezvous and proximity operations using controlled active vision. *Journal of field robotics*, 33(2):205–228, 2016. [1](#)
- [5] Tae Ha Park, Sumant Sharma, and Simone D’Amico. Towards robust learning-based pose estimation of noncooperative spacecraft. *arXiv preprint arXiv:1909.00392*, 2019. [1](#)
- [6] Tae Ha Park, Sumant Sharma, and Simone D’Amico. Towards robust learning-based pose estimation of noncooperative spacecraft. *arXiv preprint arXiv:1909.00392*, 2019. [1](#)
- [7] Niccolò Faraco, Michele Maestrini, and Pierluigi Di Lizia. Instance segmentation for feature recognition on noncooperative resident space objects. *Journal of Spacecraft and Rockets*, 59(6):2160–2174, 2022. [1](#)
- [8] Leo Pauly, Wassim Rharbaoui, Carl Shneider, Arunkumar Rathinam, Vincent Gaudilliere, and Djamila Aouada. A survey on deep learning-based monocular spacecraft pose estimation: Current state, limitations and prospects. *Acta Astronautica*, 212:339–360, 2023. [1](#)
- [9] Antoine Legrand, Renaud Detry, and Christophe De Vleeschouwer. Nerf-based visualization of 3d cues supporting data-driven spacecraft pose estimation. *arXiv preprint arXiv:2509.14890*, 2025. [2](#)
- [10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. [2](#)
- [11] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. [2](#)
- [13] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. [2](#)
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. [2](#)
- [15] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. [2, 4](#)
- [16] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. [2](#)
- [17] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. [2](#)
- [18] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021. [2, 6, 7](#)
- [19] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint arXiv:1606.02147*, 2016.
- [20] Rudra PK Poudel, Stephan Liwicki, and Roberto Cipolla. Fast-scnn: Fast semantic segmentation network. *arXiv preprint arXiv:1902.04502*, 2019.
- [21] Sachin Mehta, Mohammad Rastegari, Anat Caspi, Linda Shapiro, and Hannaneh Hajishirzi. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In *Proceedings of the european conference on computer vision (ECCV)*, pages 552–568, 2018. [2](#)
- [22] Towaki Takikawa, David Acuna, Varun Jampani, and Sanja Fidler. Gated-scnn: Gated shape cnns for semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5229–5238, 2019. [2, 4](#)
- [23] Yifu Chen, Arnaud Dapogny, and Matthieu Cord. Semeda: Enhancing segmentation precision with semantic edge aware loss. *Pattern Recognition*, 108:107557, 2020. [2](#)
- [24] Dengfeng Chai, Shawn Newsam, and Jingfeng Huang. Aerial image semantic segmentation using dcnn predicted distance maps. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161:309–322, 2020. [3](#)
- [25] Nicolas Audebert, Alexandre Boulch, Bertrand Le Saux, and Sébastien Lefèvre. Distance transform regression for spatially-aware deep semantic segmentation. *Computer Vision and Image Understanding*, 189:102809, 2019. [3](#)
- [26] Zeeshan Hayder, Xuming He, and Mathieu Salzmann. Boundary-aware instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5696–5704, 2017. [3](#)
- [27] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. [3](#)
- [28] Yuhui Yuan, Jingyi Xie, Xilin Chen, and Jingdong Wang. Segfix: Model-agnostic boundary refinement for segmen-

- tation. In *European conference on computer vision*, pages 489–506. Springer, 2020. [3](#)
- [29] Hoang Anh Dung, Bo Chen, and Tat-Jun Chin. A spacecraft dataset for detection, segmentation and parts recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2012–2019, 2021. [3](#)
- [30] Taehyun H. Park and Simone D’Amico. Spe3r: Synthetic dataset for satellite pose estimation and 3d reconstruction, 2024. [3](#), [5](#), [8](#)
- [31] Yi Cao, Jinzhen Mu, Xianghong Cheng, and Fengyu Liu. Spacecraft-ds: A spacecraft dataset for key components detection and segmentation via hardware-in-the-loop capture. *IEEE Sensors Journal*, 24(4):5347–5358, 2024. [3](#)
- [32] Jeffrey Joan Sam, Janhavi Sathe, Nikhil Chigali, Naman Gupta, Radhey Ruparel, Yicheng Jiang, Janmajay Singh, James W Berck, and Arko Barman. A new dataset and performance benchmark for real-time spacecraft segmentation in onboard flight computers. *arXiv preprint arXiv:2507.10775*, 2025. [3](#)
- [33] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. [3](#)
- [34] Glenn Jocher and Jing Qiu. Ultralytics yolov11, 2024. [3](#)
- [35] Iason Georgios Velentzas and Panagiotis Tsiotras. Irolsd: Illumination robust line segment detection for spacecraft relative navigation. *Acta Astronautica*, 2025. [3](#)
- [36] Nan Xue, Song Bai, Fudong Wang, Gui-Song Xia, Tianfu Wu, and Liangpei Zhang. Learning attraction field representation for robust line segment detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. [3](#), [4](#)
- [37] Kun Huang, Yan Zhang, Feifan Ma, Jintao Chen, Zhuangbin Tan, and Yuanjie Qi. Wdicd: A novel simulated dataset and structure-aware framework for semantic segmentation of spacecraft component. *Acta Astronautica*, 225:1–15, 2024. [3](#)
- [38] Tae Ha Park and Simone D’Amico. Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap. *Advances in Space Research*, 73(11):5726–5740, 2024. [3](#)
- [39] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020. [4](#)
- [40] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. [6](#)
- [41] Stéphane d’Ascoli, Hugo Touvron, Matthew L Leavitt, Ari S Morcos, Giulio Biroli, and Levent Sagun. Convit: Improving vision transformers with soft convolutional inductive biases. In *International conference on machine learning*, pages 2286–2296. PMLR, 2021. [6](#)