Resolved Hyperspectral Imaging

Kimmy Chang[1], Lauren Fisher[2], J. Zachary Gazak[1], and Justin Fletcher[1]
[1]*Space Systems Command (A&AS)*
[2]*KBR*

ABSTRACT

We present a novel approach to advancing Space Domain Awareness (SDA) using resolved hyperspectral imaging. Traditional methods for extracting information from space assets rely on spatially resolved imagery, mainly applicable to low Earth orbit (LEO) objects. Despite high-resolution imagery, analysts often struggle to manually discern and accurately identify key structural components of space objects. Our solution addresses this issue by automatically segmenting and identifying structural information across spectral bands. Recent research indicates that valuable, distance-invariant information is embedded in the spectrum of an object's reflected sunlight, providing insights into the object's composition and state. We employ a custom convolutional neural network (CNN) model for image classification and segmentation. The primary 2D CNN model is trained on a diverse set of twenty simulated satellites, followed by fine-tuning on smaller, satellite-specific 2D CNNs for detailed analysis. In total, we simulated 27 satellites, each annotated with material (e.g., Germanium, Titanium Alloy) and component (e.g., Attitude Control System, Payload) semantic segmentation. Our method achieves satellite classification accuracies exceeding 98%, with macro average precision and recall exceeding 80%. Notably, we demonstrate the primary 2D CNN model's ability to be effectively fine-tuned on satellites not included in the initial training. Furthermore, our method requires only 80 samples from each satellite for both material and component semantic segmentation, and once the initial primary model is trained, fine-tuned model results can be obtained in less than five minutes.

1. INTRODUCTION

As the number of operational satellites continues to rise globally, the demand for autonomous solutions in Space Domain Awareness (SDA) has become increasingly critical for effectively monitoring and managing satellite systems. Traditional methods of satellite recognition and analysis, often reliant on visible light data, face limitations in distinguishing between materials and components, particularly in complex environments [11, 5]. Recent advancements in hyperspectral imaging technology offer a promising alternative by capturing detailed spectral information across a wide range of wavelengths, enabling enhanced material identification and discrimination [14].

In this study, we present a novel approach that integrates hyperspectral imaging with a custom convolutional neural network (CNN) for the segmentation of satellite materials and components (see Fig. 1 for our pipeline). We contribute a diverse synthetic dataset generated by the Computational Omni-Analysis Simulation Toolkit (COAST), which comprises 27 unique satellite geometries and includes semantic segmentations of various materials and components across both visible light and hyperspectral bands.

To tackle the challenges of effective SDA, we aim to create a real-time solution that not only delivers accurate satellite material and component segmentation but also generalizes well to satellites outside of the training dataset. Our hybrid Convolutional Neural Network (CNN) model is designed to adapt efficiently with minimal fine-tuning, allowing it to identify and analyze satellites not included in the initial training set. Additionally, we seek to quantify the minimum number of data points required for effective training, ensuring that our approach remains practical for various operational scenarios. Ambitiously, our model targets the segmentation of 10 and 11 distinct material and component classes, respectively, thereby enhancing the comprehensiveness of satellite recognition and analysis in real-world applications.

The remainder of this paper is organized as follows: Section 2 reviews related works in the fields of SDA and hyperspectral imaging. In Section 3, we detail our methodology, including the COAST dataset, hybrid CNN architecture, and metrics. Section 4 presents our experimental results and performance evaluations, while Section 5 discusses the limitations of our findings and potential future work. Finally, Section 6 concludes the paper, summarizing the key contributions of our research and their impact on advancing SDA.

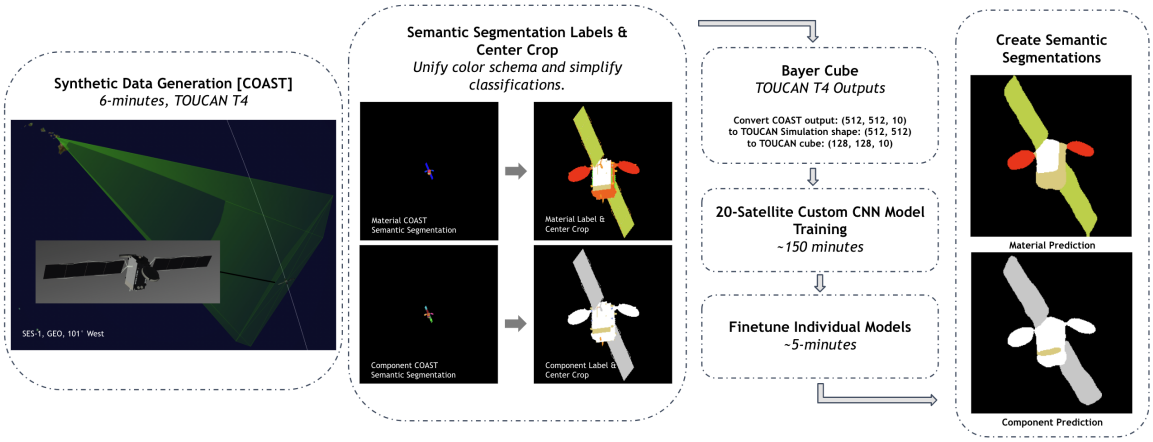

Fig. 1: The end-to-end pipeline of our approach for material and component semantic segmentation using hyperspectral images. We employ a hybrid custom CNN model trained on a diverse set of satellite data. The visualization includes steps from synthetic data generation in COAST, data preprocessing, model training, to the final semantic segmentation predictions for the satellite SES-1.

2. RELATED WORKS

Satellite recognition from ground-based observations is crucial for space situational awareness (SSA). Traditional methods, which rely on spatially resolved imagery, are limited to low Earth orbits (LEO). Recent advancements in spectroscopic techniques and deep learning have introduced new methods for identifying and characterizing artificial satellites in geostationary orbits (GEO). Spectroscopic imagery offers a distance-invariant method for satellite identification, overcoming the limitations of traditional imaging techniques. The SpectraNet approach, employing a modified residual CNN to map spectroscopic data to object identity, has shown promising results, achieving over 80% classification accuracy for a simulated 64-class satellite and 72% accuracy for a nine-class problem in an astronomical observing campaign [7]. However, spectroscopy has lagged behind imaging methods due to challenges such as the inaccuracy of material reflection models, the complexity of interpreting spectroscopic data, and difficulties in obtaining underlying truth data on material and geometry [7].

Recent work has contributed two large datasets of simulated spectrally enriched space imagery of 110 satellite classes, along with astronomical observations through multispectral cameras. The study analyzes the effectiveness of multispectral versus multi-aperture systems for these tasks, utilizing affordable, easy-to-deploy ground-based equipment [8]. Notably, hyperspectral imaging poses challenges due to the high dimensionality of data and limited training samples. Deep learning models, particularly CNNs, have proven effective in extracting highly discriminatory features from hyperspectral data. Various processing techniques, such as spectral unmixing [4], resolution enhancement [10], and anomaly detection [1], have been developed to efficiently analyze hyperspectral data cubes. Notable models include hybrid CNN2D-3D architectures and end-to-end CNN3D with recurrent feedback structures [2, 6, 18].

Recent advancements in spectroscopic techniques and deep learning have significantly improved the capabilities of satellite recognition from ground-based observations. Methods such as SpectraNet [7] and SpectraNet-SO(3) [9] demonstrate the potential of utilizing spectroscopic data for high-accuracy satellite identification and orientation mapping. Furthermore, hyperspectral data has shown versatility in various semantic segmentation applications beyond satellite imagery. For example, hyperspectral imaging is employed in agriculture for crop monitoring and disease detection, enabling the differentiation of plant species and the identification of stressed plants based on their spectral signatures [12]. Additionally, hyperspectral data proves effective in environmental monitoring, assisting in land cover classification and water quality assessment by detecting specific contaminants [15]. These applications underscore the potential of hyperspectral data to enhance semantic segmentation tasks. In this research, we aim to evaluate the applicability of hyperspectral data for the semantic segmentation of satellite components and materials.

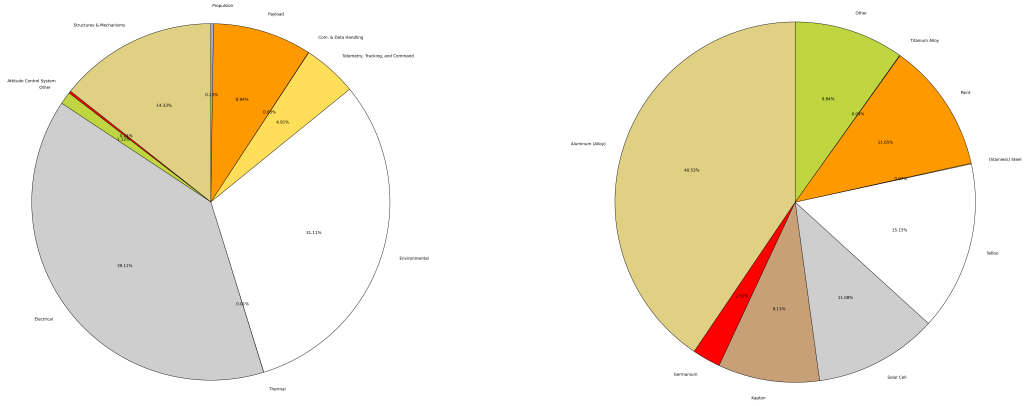

Fig. 2: Visualization of Component (left) and Material (right) Classes with their respective percentages. The background class, which accounts for 83.44% in the component dataset and 83.43% in the material dataset, is not shown. Even with the exclusion of background classes, the under-representation of certain classes, such as "Thermal" and "Titanium Alloy," is evident.

3. APPROACH

3.1 Dataset

COAST is a versatile software capable of creating satellite scenarios for a wide range of target satellites and ground-based sensors. Each scenario simulates Low Earth Orbit (LEO) and Geostationary Orbit (GEO) satellites orbiting the Earth. We generated 29,700 simulated examples, featuring a variety of satellite designs. These range from GEO communications satellites (e.g., Galaxy 4R, SES-1, and Spaceway-1) to weather satellites (e.g., Meteor-3 and DMSP F7) and defense/military satellites (e.g., DSCS-3B6 and Almaz). We included a family of six AMC satellites to test the robustness of our model's detail-oriented predictions. Within this family, design differences are minimal, ensuring consistent overall designs.

Our COAST scenarios begin at 3:00:00, with a time step of 6 seconds, centered at coordinates (20.709, -156.2535) and an altitude of 3048 meters. We use a detector size of (2048, 2048) and an aperture diameter of 3.6 meters. For the hyperspectral scenarios, the detector is configured to simulate the Silios Toucan, which utilizes 10 filters covering wavelengths from 400 to 900 nanometers in a 16-pixel Bayer pattern. This configuration enables the simultaneous detection, recognition, and localization of geosynchronous satellites from ground-based imagery, as demonstrated by [8]. Unlike visible light imaging, which only the 400-nanometer wavelength, hyperspectral imaging gathers information across a wide range of narrow spectral bands, offering greater detail and differentiation of materials. The Silios Toucan's design, along with the matrix size of its Bayer filters, can be customized for various industrial and scientific applications. Current commercial products from Ximea [19, 20, 21] and Silios [17] offer configurations with nine, sixteen, and twenty-five filters, all spanning approximately 400 to 900 nanometers.

In addition to generating hyperspectral data, we utilize COAST to procedurally create raw visible light as a baseline for our models and to annotate our renders with component and material semantic segmentations. Examples from the dataset are illustrated in Figure 3. The component and material classes are detailed in Figure 2. Notably, the "Other" class in the material pie chart encompasses a diverse range of materials, including brass, silica, gold plating, mylar, graphite-epoxy, beryllium, copper, and fiberglass. Similarly, the "Other" class in the component pie chart comprises data return systems, wideband systems, and ascent phase equipment. To streamline our analysis, we reclassified the components and materials into the top eleven and ten classes, respectively, reducing the previously large number of classes (numbered in the hundreds) that originated from COAST.

Given the large image output size (2048, 2048) of the renders, we center cropped the image outputs to (256, 256). This preprocessing step significantly improved the performance of our models by addressing the large class imbalance of background pixels. For each hyperspectral scenario, detections are extracted from simulated imagery to construct data cubes with a 10×10 spatial size and 10 spectral channels, matching the mode used in an active Space Domain

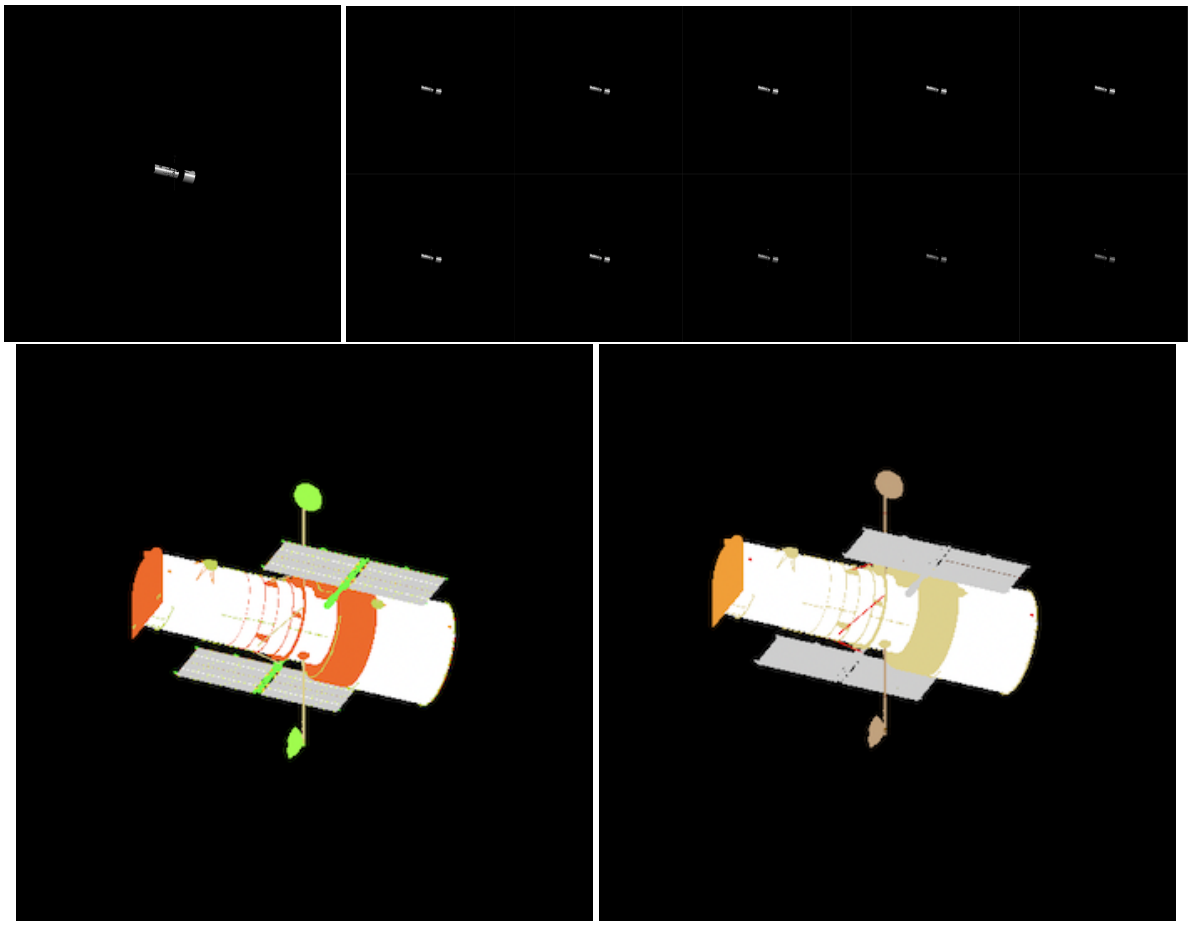

Fig. 3: Example of a Hubble Space Telescope render. From left to right and top to bottom: visible light image; ten-channel images; component semantic segmentation; material semantic segmentation. Note that the top images are not center-cropped, while the bottom images are center-cropped.

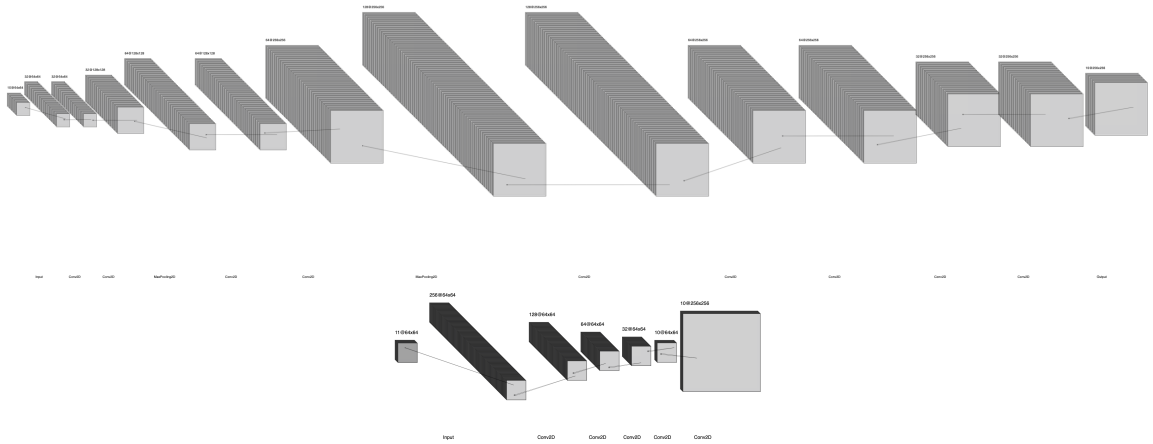

Fig. 4: The primary model architecture is depicted at the top, with the smaller fine-tuning models shown at the bottom. Notably, a Dropout layer with a probability of 0.2 is inserted after each Conv2D layer in both the primary and fine-tuning models. Additionally, concatenation layers are included in the downsampling layers of the primary model.

Awareness (SDA) scenario, where the preceding detection step allows for cube extraction. For the Silios Toucan systems, the frame is first resized from (2048, 2048) pixels to (512, 512, 10) and then center cropped to (128, 128, 10).

3.2 Model

In this research, we create a custom hybrid CNN model designed for combined and fine-tuned training on data from a collection of diverse satellites. The primary model is trained on data from twenty satellites and is based on a U-Net style architecture, which is well-suited for semantic segmentation tasks [3, 16, 13]. This model is then used as a feature extractor for fine-tuning individual models for each satellite. The details of the architecture for both the primary model and the fine-tuned satellite models are outlined in Fig. 4.

While our model is inspired by the U-Net architecture, it incorporates two key modifications. First, we introduce dropout layers after each convolutional block in both the encoder and decoder paths to regularize the model and mitigate overfitting. This is particularly crucial as our primary goal is for the model to generalize to satellite geometries not included in the training data during the fine-tuning stage for individual satellites. Second, our model is significantly simplified, featuring half the number of convolutional layers typical of U-Net and smaller filter sizes. As observed in [18], this reduction in computational complexity is essential for efficiently processing high-dimensional hyperspectral data. Additionally, the simplified architecture enhances the model's ability to generalize to unseen data and satellite-specific features addressed during the fine-tuning stage. We also resize our output to align with the unique characteristics of our data. Furthermore, the inclusion of concatenation layers, as implemented in [18], and max pooling layers, as utilized in [6], contributed positively to the model's performance.

For visible light data, the input height and width are set to $(256, 256)$, while for hyperspectral data, the input dimensions are reduced to $(64, 64)$ with 10 channels. The output shape of the primary model is $(256, 256)$ with the number of classes being 10 and 11 for the material and component models, respectively. The fine-tuned satellite models have an input shape of $(64, 64)$ with 10 channels.

We train our models using an NVIDIA DGX system equipped with four A100 GPUs. We use sparse categorical crossentropy for our loss as well as the Adam optimizer with a learning rate of 1e-4 and a clipnorm of 1.0. Each model (primary and fine-tuned) are trained for up to 1000 epochs. However, to prevent overfitting, we implemented early stopping with patience of 100 epochs. We use the aforementioned optimizer, loss function, and early stopping in our individual smaller models.

3.3 Metrics

Accuracy is defined as the ratio of correctly predicted pixels to the total number of pixels. While accuracy is a good general measure of model performance, it does not provide a complete picture for imbalanced datasets, such as ours

for satellite materials and components. Therefore, we also report the macro average precision and recall to better understand the performance of each class independently, giving equal importance to each class.

Macro average precision (MAP) is the unweighted mean of the precision values for all classes:

$$\text{Macro Average Precision} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i}$$

where N is the number of classes and TP_i and FP_i represent the true positive and false positive pixels for class i, respectively. This metric helps us understand the extent of over-prediction for certain classes and the prevalence of false positives.

Similarly, macro average recall (MAR) is the unweighted mean of the recall values for all classes:

$$\text{Macro Average Recall} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i}$$

where N is the number of classes and FN_i represents the false negative pixels for class i. This metric helps us assess how well each class is identified, ensuring that classes with less data are not ignored.

4. EXPERIMENTS

4.1 Satellite Component Semantic Segmentation

Overall, our satellite component results (see Table 2) indicate that the Hybrid Model consistently achieves the highest accuracy across almost all satellites, outperforming both the Visible Light Combined Model and the Hyperspectral Combined Model (see Figure 5. Notably, the Hyperspectral Combined Model generally achieves higher accuracy than the Visible Light Combined Model. On average, there is an increase in accuracy of 0.0722 when comparing the Hybrid Model to the Combined Visible Light Model and 0.0302 when contrasting it with the Combined Hyperspectral Model.

Specifically, the accuracy improvement can reach as high as 0.1431 for Almaz when transitioning from the Visible Light Combined to the Hybrid Model, and 0.1358 for AMC-23 when moving from the Hyperspectral Combined to the Hybrid Model. An interesting case is DMSP-F16, where the accuracy of the Visible Light Combined Model (0.9400) exceeds that of the Hyperspectral Combined Model (0.9320), but the Hybrid Model significantly enhances accuracy to 0.9705. Despite DMSP-F16's relatively low MAP and MAR values in the Combined Models, the Hybrid Model shows a notable improvement. This case illustrates the generalization capabilities of our approach, as DMSP-F16 was not included in the training of the primary combined hyperspectral model embedded in the Hybrid Model.

MAP metrics further demonstrate that the Hyperspectral Combined Model often surpasses the Visible Light Combined Model, indicating better performance in identifying true positives while minimizing false positives. The Hybrid Model consistently outperforms both the Combined Visible Light and Hyperspectral Models, with average increases of 0.2968 and 0.1593, respectively. For instance, Meteor-3 shows a significant increase in MAP from 0.1686 in the Visible Light Combined Model to 0.7539 in the Hyperspectral Combined Model, followed by further improvement to 0.8626 in the Hybrid Model. AMC-23 also exhibits substantial performance jumps, with MAP values of 0.4156 and 0.5237 when moving from the Visible Light and Hyperspectral Combined Models to the Hybrid Model, respectively. This highlights the powerful generalization capabilities of our method, especially since AMC-23 was not included in the training of the primary Hyperspectral Combined Model.

Similarly, MAR metrics reveal that both the Hyperspectral Combined Model and the Hybrid Model frequently achieve higher values than the Visible Light Combined Model. Some satellites, such as DMSP F7, exhibit considerable differences in MAR across models, with the Visible Light Combined Model at 0.1246 and the Hybrid Model reaching 0.4154. On average, MAR increases are 0.2861 and 0.1306 when comparing the Hybrid Model to the Visible Light and Hyperspectral Combined Models, respectively. An intriguing case is DirecTV-4s, which experiences a 0.027 decrease in MAR when transitioning from the Hybrid Model to the Hyperspectral Combined Model.

Satellites such as Anik F3, Argos, AMC-1, and ASTEX 1 exhibit high accuracy across all models, with the Hybrid Model performing the best. Conversely, lower-performing satellites like AMC-4, AMC-18, and AMC-23 show relatively lower metrics, particularly in the Visible Light and Hyperspectral Combined Models. However, significant

improvements are noted in the Hybrid Model, likely due to the absence of these satellites in the training data for the Visible Light and Combined Models.

In summary, the Hyperspectral Combined Model typically outperforms the Visible Light Combined Model, suggesting that hyperspectral data provides more valuable information for these metrics. The substantial improvements seen with the Hybrid Model underscore the effectiveness of combining data sources for satellite classification tasks. Additionally, the Hybrid Model demonstrates the ability to generalize to satellite designs both included and excluded from its primary combined model.

4.2 Satellite Material Semantic Segmentation

The semantic segmentation performance for the identification of satellite materials was evaluated across Combined (Visible Light and Hyperspectral) and Hybrid models, focusing on metrics including accuracy, MAP, and MAR for each satellite. As shown in Table 3, the Hybrid Model consistently outperforms the Combined Models in most cases, with under-represented classes improving most notably. Notably, the most significant improvements are observed in the MAP and MAR metrics when comparing the Hybrid Model to the Combined models (see Fig 6). Specifically, the average MAR increases by 0.4131 from Combined Visible Light to Hybrid, while the average MAP improves by 0.1416 from Combined Hyperspectral to Hybrid.

Across the satellites, the Hybrid Model demonstrates superior accuracy, with the highest recorded accuracy being 0.9886 for the "Argos" satellite. This performance is markedly higher than that achieved by the Combined Models, particularly when using visible light data alone. The improvement in accuracy within the Hybrid Model can be attributed to its ability to leverage the complementary strengths of both visible light and hyperspectral data, resulting in a more robust classification framework. For instance, the accuracy for "Almaz" increased from 0.6371 in the visible light model to 0.9004 in the Hybrid Model, highlighting a significant enhancement.

The MAP further highlights the advantages of the Hybrid Model. For most satellites, precision values significantly improve with the integration of multiple data sources. For example, the "Anik F3" satellite saw its precision increase from 0.2554 in the visible light model to 0.7714 in the Hybrid Model. This substantial gain suggests that the Hybrid Model is more effective at correctly identifying true positives, especially for materials with similar spectral characteristics when analyzed with a single data source.

In terms of MAR, the Hybrid Model again shows clear improvement over the Combined Models. For satellites such as "Ciel 2," recall improved from 0.1427 in the visible light model to 0.6343 in the Hybrid Model. This improvement is particularly notable for underrepresented materials in the dataset, indicating that the Hybrid Model offers a more balanced approach to material classification.

The marked improvements in accuracy, precision, and recall in the Hybrid Model parallel the observations made in the semantic segmentation of satellite components. The overall accuracy of the Hybrid Material Model (0.9554) exceeds that of the Hybrid Component Model (0.9513), which is expected given that the material model has 10 classes compared to the component model's 11 classes. However, the overall MAP and MAR of the material model are lower than those of the component model. This discrepancy is likely due to the fact that the component classes are slightly more balanced and exhibit less general class overlap. In contrast, the material classes include an ambiguous "Other" and several classes that are easily confused (e.g. metals like "(Stainless)" Steel and "Titanium Alloy," which have relatively high numbers of false positives for "Aluminum (Alloy)").

4.3 Ablation Studies

4.3.1 Holdout Satellites

In the primary model training of our Hybrid and Combined Models, we deliberately excluded the following satellites: AMC-4, AMC-8, AMC-18, AMC-23, DMSP-F16, and Echostar XIV. These are referred to as "holdout satellites" throughout the remainder of this paper. For comparison purposes, we included AMC-1 and DMSP-F7 in the training set to evaluate the models' generalization capabilities, given the presence of similar satellite designs in the training data. We aimed to test the robustness of the models in fine-tuning when detailed differences within satellite families are present, particularly when only one satellite from a family is included in the primary training set. Echostar XIV was specifically included to assess how the Hybrid model performs on satellites with "unseen" designs.

Material Analysis

Table 3 shows that the holdout satellites generally exhibit lower performance, with average differences of 0.045,

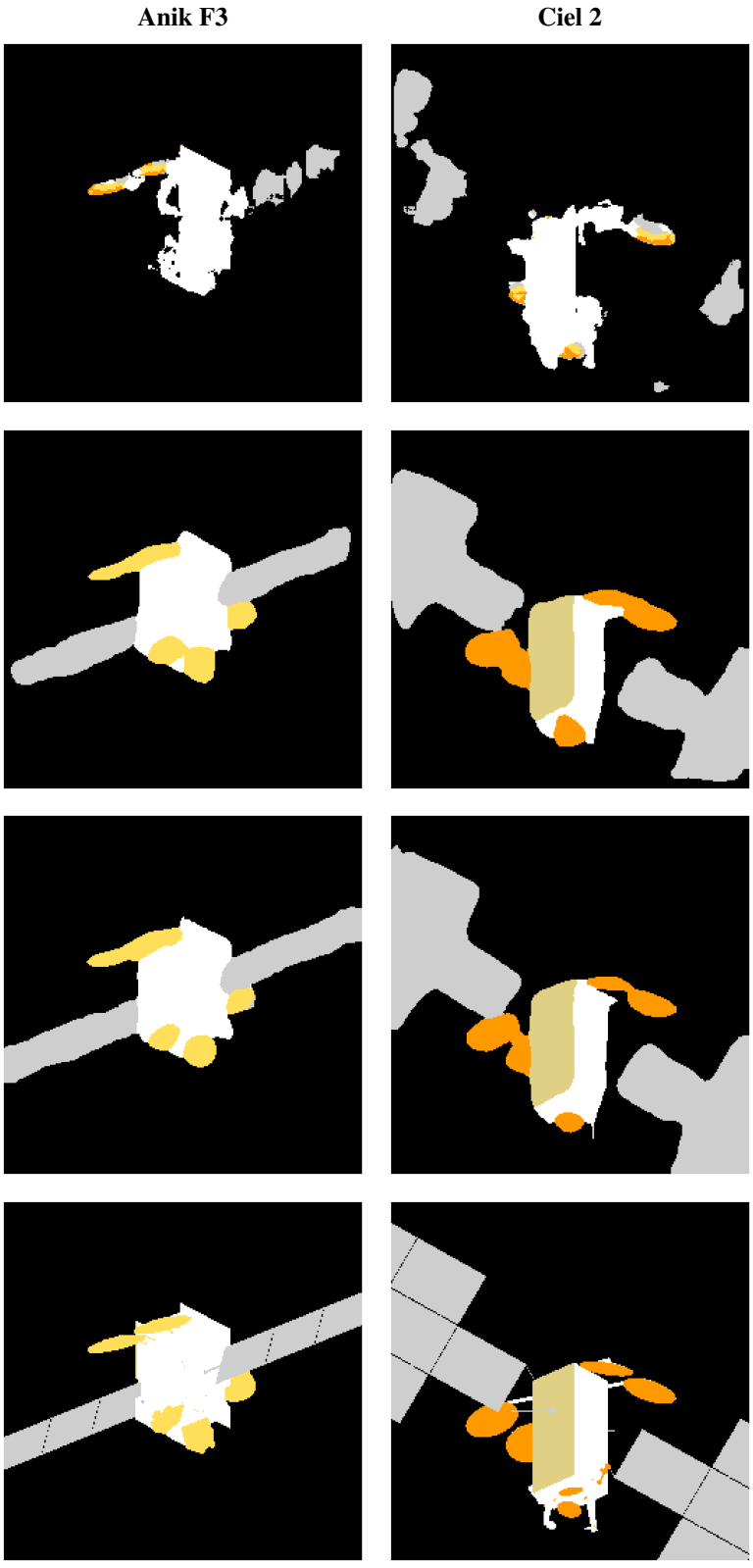

Fig. 5: Satellite component semantic segmentation predictions are presented in rows 1, 2, and 3 for the Combined Visible Light, Combined Hyperspectral, and Hybrid Hyperspectral Models, respectively, with the ground truth in row 4. The results demonstrate a notable improvement in accuracy when transitioning from visible light to hyperspectral data in the combined models. The Hybrid Hyperspectral model, in contrast to the Combined Hyperspectral model, accurately captures the entire span of solar panels on Anik F3 as the "Electrical" class, while the predictions for Ciel 2 show more clearly defined circular Antenna Reflectors, correctly identified as part of the "Payload" class. This provides a visual example of how fine-tuning from the Combined Hyperspectral to the Hybrid Hyperspectral model enhances both Macro Average Precision and Macro Average Recall.

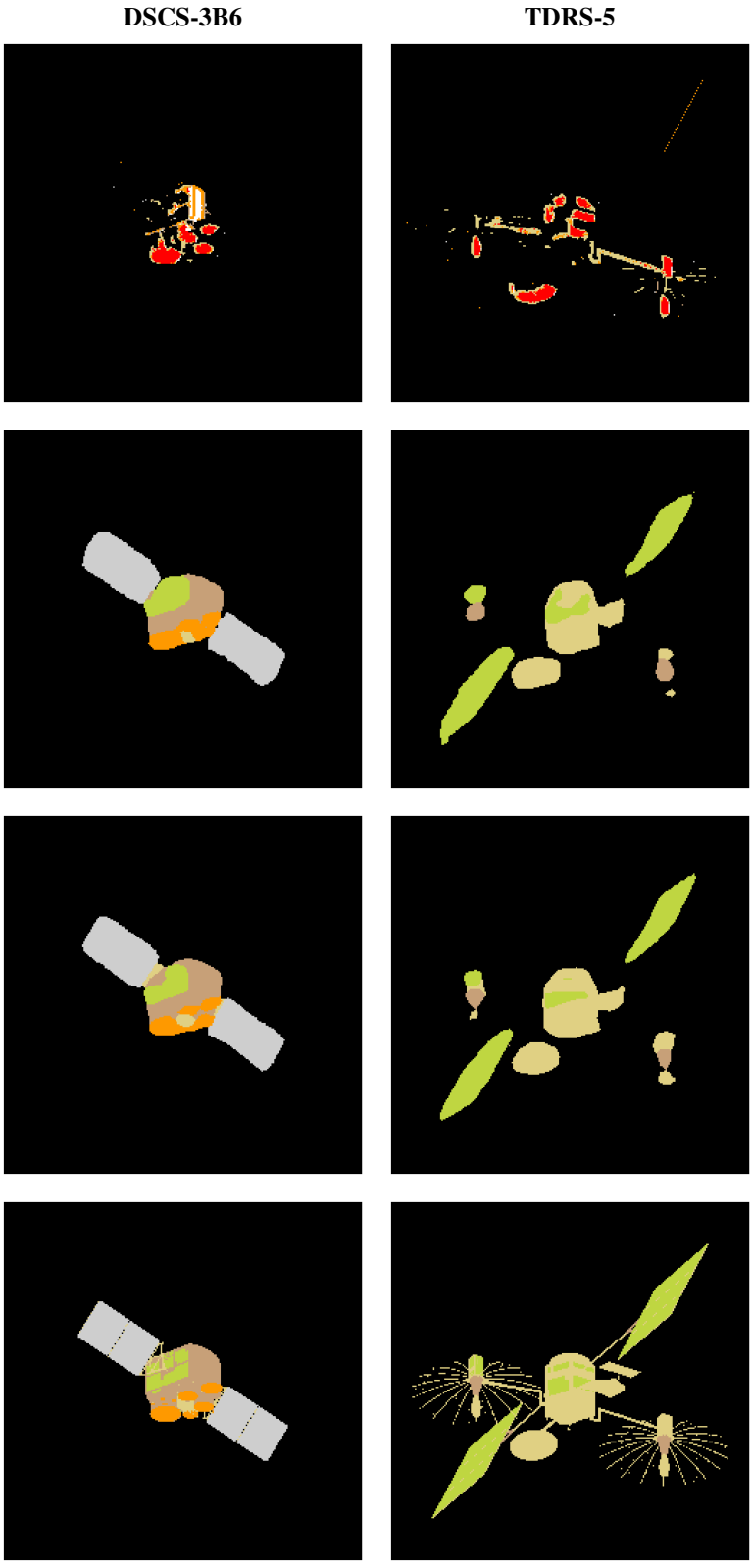

Fig. 6: Satellite material semantic segmentation predictions are presented in rows 1, 2, and 3 for the Combined Visible Light, Combined Hyperspectral, and Hybrid Hyperspectral Models, respectively, with the ground truth in row 4. While the Combined Visible Light model captures the general shapes of the two satellites, it incorrectly predicts 'Paint' for other materials. Material predictions significantly improve in the Combined Hyperspectral Model, but detailed predictions, particularly for DSCS-3B6's 'Aluminum (Alloy)' regions and the trapezoidal shape of TDRS-5's satellite wings, are most accurately captured in the Hybrid Hyperspectral Model.

0.1179, and 0.1743 in accuracy, MAP, and MAR, respectively. This trend is consistent across all models, with the Combined Hyperspectral Model showing the largest discrepancies in average accuracy, MAP, and MAR (0.1101, 0.2407, 0.2529, respectively). Although the Combined Visible Light Model demonstrates similar differences between training and holdout satellites, it produces equally poor segmentation predictions, indicating that the fine-tuning stage may help the model generalize to holdout satellites to some extent.

Among the holdout satellites, the AMC satellites tend to generalize best in our Hybrid model results. This could be due to the relatively consistent materials used across the AMC family, especially those shared with the non-holdout satellite AMC-1. Conversely, DMSP-F16 shows notably low MAP and MAR despite its high accuracy. This discrepancy can likely be attributed to the significantly different material compositions between DMSP-F16 and DMSP-F7, which may have confused the Hybrid model due to the relatively similar satellite designs. Although Echostar XIV performs poorly compared to other holdout satellites, it still achieves higher accuracy, MAP, and MAR than the Combined Visible Light and Hyperspectral models. This suggests that while the primary model generalizes reasonably well, there may be room for improvement in the fine-tuning process.

Component Analysis

In contrast to the material results, the differences in error metrics between the training and holdout satellites are significantly smaller for component identification (see Table 2). Specifically, the training satellites outperform the holdout satellites by only 0.0319 in average accuracy and 0.0545 in MAR. Interestingly, the holdout satellites achieve an average MAP that is 0.0007 higher than that of the training satellites. Similar trends are observed in the Combined Visible Light and Hyperspectral models, as well as in the performance of the AMC, DMSP-F16, and Echostar XIV holdout satellites.

4.3.2 Other Architectures

The results of our ablation study demonstrate the effectiveness of the custom CNN architecture, particularly in handling hyperspectral imaging for satellite component and material classification. Notably, the custom CNN outperforms the other popular semantic segmentation model architectures, achieving an overall accuracy of 0.939 for component classification and 0.951 for material classification in the hyperspectral domain. These results highlight the value of our architectural modifications, including reduced convolutional layers and the strategic use of dropout, which appear to enhance the primary model's ability to capture and generalize the complex spectral information inherent in hyperspectral data.

When comparing hyperspectral data to visible light data, the results show a consistent performance advantage for hyperspectral data across all models. The custom CNN achieved an overall accuracy of 0.844 in material classification with visible light, which is notably lower than its performance with hyperspectral data. This difference emphasizes the richer information content in hyperspectral data, which provides the model with a broader range of features to distinguish between different classes. However, the custom CNN still managed to perform robustly with visible light data, demonstrating its versatility and generalization capabilities. This robustness is critical for space domain awareness applications, where operational scenarios vary widely, and the ability to generalize from training data to unseen satellite configurations is essential.

Overall, the results underscore the importance of tailoring model architectures to the specific challenges of the task at hand. The modifications we introduced to the U-Net-inspired architecture have proven effective, particularly in improving the model's performance on high-dimensional hyperspectral data, making it a strong candidate for space domain awareness applications.

4.3.3 Dataset Size

This experiment explores the impact of dataset size on the performance of a Hybrid model trained on Hyperspectral data. The study is divided into two key parts.

The first part examines how varying the number of satellites in the primary training set affects the Hyperspectral model's performance. This includes identifying the minimum number of satellites required to achieve reliable performance and evaluating how changes in the training dataset influence the model's ability to generalize to unseen data.

Classification Task	Spectral Range	Model Architecture	Macro Average Precision	Macro Average Recall	Accuracy
Component	Visible Light	UNet	0.076	0.091	0.833
		SegNet	0.137	0.164*	0.808
		Fcn-8	0.118	0.151	0.666
		Custom CNN	0.142*	0.091	0.836*
Component	Hyperspectral	UNet	0.137	0.144	0.677
		SegNet	0.105	0.126	0.350
		Fcn-8	0.144	0.130	0.796
		Custom CNN	0.436*	0.412*	0.939*
Material	Visible Light	UNet	0.152	0.103	0.836
		SegNet	0.170	0.184*	0.789
		Fcn-8	0.135	0.158	0.666
		Custom CNN	0.202*	0.156	0.844*
Material	Hyperspectral	UNet	0.111	0.102	0.833
		SegNet	0.131	0.135	0.733
		Fcn-8	0.152	0.146	0.788
		Custom CNN	0.667*	0.650*	0.951*

Table 1: Performance Comparison of Semantic Segmentation Model Architectures for Satellite Component and Material Classification. This table presents the accuracy, Macro Average Precision, and Macro Average Recall for various model architectures (UNet, SegNet, Fcn-8, and our Custom CNN) across both hyperspectral and visible light data. The Custom CNN model consistently outperforms other architectures, particularly in the hyperspectral domain, achieving the highest overall accuracy of 0.939 for component classification and 0.951 for material classification. These results demonstrate the effectiveness of the architectural modifications made to the custom CNN, which enhance its ability to capture and generalize complex spectral information.

The analysis provides insights into the relationship between dataset size and model accuracy, precision, and recall.

The second part of the study investigates how the model's performance is affected when the number of satellites included during the fine-tuning stage is increased. This part contrasts the model's ability to generalize to new, unseen satellites, particularly those not included in the initial training phase, offering a deeper understanding of the model's robustness and adaptability.

Component Analysis

As the number of satellites in the training set increases, the Component Macro Average Precision (MAP) shows a gradual improvement, particularly on the training data, where it increases from 0.203 with 2 satellites to 0.436 with 80 satellites as in the main study. However, the holdout data exhibits a less pronounced improvement, moving from 0.308 to 0.401.

Component Macro Average Recall (MAR) and Overall Accuracy follow similar trends. The training data shows a steady increase in MAR from 0.305 to 0.412 and in accuracy from 0.836 to 0.939 as the dataset size grows. The holdout data, however, improves at a slower rate, indicating that the model becomes more accurate and precise with a larger dataset but still struggles somewhat to generalize these improvements to unseen data.

Material Analysis

The Hyperspectral Hybrid Model generally performs better on the material task than the component task across all metrics, particularly as the dataset size increases. Material MAP on the training set improves from 0.432 with 2 satellites to 0.667 with 80 satellites, while the holdout MAP also improves significantly, albeit with a slight lag compared to the training data.

Material MAR and Overall Accuracy show a strong upward trend with increasing dataset size, especially for the training data. MAR improves from 0.454 to 0.650 and accuracy from 0.779 to 0.951. Similar with the component task, the Hyperspectral Hybrid Model shows slower improvement on the holdout data, but the gap between training and holdout performance narrows as the dataset size increases.

Generalization and Holdout Performance

Both models exhibit a common pattern where performance on the holdout data improves as the dataset size increases, though not as sharply as on the training data. This suggests that while larger datasets improve the model's ability to fit

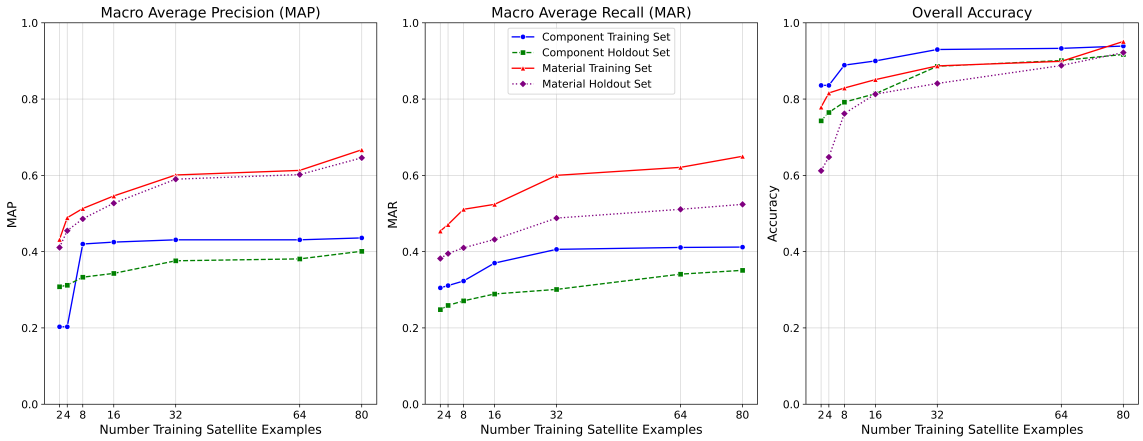

Fig. 7: Impact of Dataset Size on Model Performance for Component and Material Classification. This figure shows the relationship between the number of training satellite examples and model performance across three key metrics: Macro Average Precision (MAP), Macro Average Recall (MAR), and Overall Accuracy. The plots compare the performance on both the training and holdout sets for Component and Material classification tasks. The results highlight the importance of a sufficient number of training examples for achieving robust model performance, particularly in hyperspectral data contexts.

the training data (higher MAP, MAR, and accuracy), they also contribute to better generalization, as evidenced by the gradual improvement in holdout performance.

Despite the improvements, there is a consistent gap between training and holdout performance for both models, which suggests some degree of overfitting, particularly in the earlier stages with smaller datasets. However, this gap decreases as the number of satellites increases, implying that larger datasets help mitigate overfitting by providing more diverse examples for the model to learn from.

Increasing the dataset size has a positive effect on Hyperspectral Hybrid model on both the material and component semantic segmentation. The larger dataset sizes contribute to better generalization, as seen in the improved holdout performance, though there remains a persistent gap between training and holdout metrics, indicating room for further improvement in model generalization.

5. LIMITATIONS

A significant limitation in our current research is the pronounced class imbalance, which heavily skews the performance metrics of our Hybrid Component and Material models. As observed in the confusion matrices (Fig. 8), despite relatively strong overall metrics, there remain a considerable number of false positives across all classes. For instance, the Hybrid Material model fails to correctly identify any "(Stainless) Steel" instances and only classifies two "Titanium Alloy" instances correctly. This discrepancy highlights the limitation of accuracy as a reliable metric, as the model predominantly excels in predicting the more common "Background" and "Aluminum (Alloy)" classes, which constitute the majority of pixels. In contrast, "(Stainless) Steel" and "Titanium Alloy" account for a mere 0.07% and 0.04% of the non-background pixels, respectively.

Similarly, the Hybrid Component model struggles with less represented classes such as "Altitude Control System", "Communications & Data Handling", "Thermal", and "Propulsion", failing to correctly predict any pixels in these categories. These classes together represent less than 0.5% of the non-background pixels, while the more common classes like background, electrical, and environmental components dominate the predictions.

The inherent class imbalance leads to models that are disproportionately good at predicting well-represented classes while underperforming on minority classes, which is critical for comprehensive satellite component and material analysis. This issue is further underscored by the relatively better performance in the "Other" class of the Hybrid

Fig. 8: These confusion matrices highlight the impact of class imbalance in our Hybrid Component (left) and Material (right) models, with a significant number of false positives for underrepresented classes like Propulsion and Titanium Alloy, while more common classes like Background, Electrical, and Aluminum Alloy dominate predictions.

Component model, which, despite its low representation (approximately 1% of non-background pixels), achieves a fair number of correct predictions. To mitigate this limitation, we propose that future work prioritize methods that address the significant data imbalance, such as data augmentation, class-balanced sampling techniques, and ensemble methods.

6. CONCLUSION

In this study, we introduced a novel approach to advancing Space Domain Awareness (SDA) through hyperspectral imaging and a custom convolutional neural network (CNN) for satellite material and component segmentation. Our results demonstrate that the proposed hybrid CNN model, which trains and generalizes on hyperspectral data from various satellites, significantly outperforms models trained solely on visible light. The hybrid model consistently achieved higher accuracy, macro average precision, and macro average recall across different satellite types, including those not included in the initial training set. Notably, the hybrid approach proved particularly effective in generalizing to new satellites, requiring only minimal fine-tuning with a small number of samples. This efficiency, combined with the ability to produce detailed segmentation results in under five minutes, underscores the potential of this method for real-time space object characterization and tracking. Additionally, we contribute a large and diverse synthetic dataset comprising 27 unique satellite geometries, designed to simulate a variety of satellite materials and components, which enhanced the model's ability to generalize to real-world data.

Overall, our research highlights the promise of integrating hyperspectral data with deep learning models and leveraging synthetic datasets like COAST to improve the accuracy and efficiency of satellite recognition and analysis in SDA applications. Future work could focus on refining the hybrid model architecture, addressing the large class imbalances, and expanding the scope to include a broader range of satellite types and orbits, ultimately contributing to more comprehensive and reliable space situational awareness systems.

Satellite	Accuracy			Macro Average Precision			Macro Average Recall		
	Combined		Hybrid	Combined		Hybrid	Combined		Hybrid
	Visible Light	Hyperspectral		Visible Light	Hyperspectral		Visible Light	Hyperspectral	
Almaz	0.6804	0.7967	0.8235*	0.3613	0.4864	0.5078*	0.3083	0.4335	0.4592*
Argos	0.8728	0.9454	0.9652*	0.5171	0.7199	0.7694*	0.2617	0.7142	0.7489*
AMC-1	0.8411	0.9091	0.9543*	0.4436	0.6392	0.7440*	0.3322	0.5410	0.7227*
Anik F3	0.9594	0.9735	0.9814*	0.1925	0.3605	0.5432*	0.2222	0.2905	0.3703*
ASTEX 1	0.9374	0.9794	0.9854*	0.3280	0.5519	0.6419*	0.3312	0.5056	0.6491*
Ciel 2	0.7806	0.9183	0.9224*	0.4593	0.7235	0.8663*	0.3037	0.6854	0.8031*
DirecTV-4s	0.8732	0.9432	0.9645*	0.3842	0.6272	0.6275*	0.2712	0.5365*	0.5095
DMSP F7	0.9428	0.9737	0.9844*	0.1806	0.3656	0.4255*	0.1246	0.3759	0.4154*
DSCS-3B6	0.9627	0.9776	0.9820*	0.4895	0.6350	0.6557*	0.3633	0.5031	0.5388*
Galaxy 4R	0.8806	0.9357	0.9636*	0.3993	0.6002	0.6163*	0.2796	0.5291	0.5536*
GOES 14	0.9512	0.9728	0.9820*	0.2806	0.4126	0.5436*	0.2781	0.3951	0.4441*
Horizons-2	0.9389	0.969	0.9850*	0.3746	0.4619	0.7198*	0.2952	0.3668	0.4754*
Hubble	0.8334	0.9065	0.9258*	0.3757	0.6195	0.7874*	0.2877	0.5480	0.6906*
Landsat D	0.9427	0.9767	0.9794*	0.2181	0.3827	0.3997*	0.1856	0.3759	0.3771*
Meteor-3	0.9017	0.9771	0.9845*	0.1686	0.7539	0.8626*	0.2865	0.6330	0.6917*
Mexsat-3	0.9175	0.9645	0.9752*	0.4242	0.5658	0.6104*	0.2699	0.4900	0.5548*
SES-1	0.9406	0.9702	0.9829*	0.4332	0.6009	0.6979*	0.3465	0.5235	0.6761*
Spaceway-1	0.8036	0.9081	0.9388*	0.3307	0.5317	0.6049*	0.2370	0.4299	0.5608*
TDRS-5	0.8794	0.9332	0.9463*	0.4449	0.7761	0.8219*	0.3830	0.6182	0.7142*
Telstar 401	0.8742	0.9408	0.9636*	0.4494	0.5976	0.6795*	0.3235	0.5645	0.6115*
AMC-4	0.8422	0.8389	0.9443*	0.2588	0.1817	0.6876*	0.2672	0.2356	0.5764*
AMC-8	0.9099	0.9246	0.9461*	0.4315	0.4125	0.4905*	0.2448	0.3274	0.4509*
AMC-15	0.8799	0.8865	0.9349*	0.5740	0.3979	0.6941*	0.3537	0.3116	0.6845*
AMC-18	0.8911	0.8871	0.9287*	0.3378	0.2467	0.7073*	0.1650	0.1326	0.3879*
AMC-23	0.8174	0.7651	0.9009*	0.3423	0.2342	0.7579*	0.3512	0.1985	0.6623*
DMSP-F16	0.9400	0.9320	0.9705*	0.2265	0.2011	0.5364*	0.1487	0.1654	0.3700*
Echostar XIV	0.7389	0.7624	0.8682*	0.2835	0.3362	0.7250*	0.2877	0.2775	0.5346*

Table 2: Semantic Segmentation of Satellite Component Metrics across Combined (Visible Light and Hyperspectral) and Hybrid Models. This table presents accuracy, macro precision, and macro recall values for both model types, categorized by data type. Notably, the Hybrid Model consistently outperforms the individual models across most satellites, particularly in accuracy and macro precision, highlighting the effectiveness of combining data sources for enhanced satellite classification. The highest metrics for each satellite are indicated with asterisks. Satellites listed below and including AMC-4 were not used in training the Combined Models.

Satellite	Accuracy			Macro Average Precision			Macro Average Recall		
	Combined		Hybrid	Combined		Hybrid	Combined		Hybrid
	Visible Light	Hyperspectral		Visible Light	Hyperspectral		Visible Light	Hyperspectral	
Almaz	0.6371	0.8573	0.9004*	0.1948	0.3316	0.4187*	0.1132	0.2551	0.3990*
AMC-1	0.8040	0.9433	0.9718*	0.2228	0.5810	0.6172*	0.1208	0.5706	0.6015*
Anik F3	0.7587	0.9516	0.9634*	0.2554	0.6329	0.7714*	0.1172	0.6501	0.7642*
Argos	0.9632	0.9862	0.9886*	0.2041	0.3422	0.4330*	0.1234	0.3069	0.3486*
ASTEX 1	0.9129	0.9820	0.9852*	0.1840	0.5141	0.6985*	0.1250	0.4822	0.6735*
Ciel 2	0.7529	0.9160	0.9534*	0.2657	0.6530	0.7381*	0.1427	0.6208	0.6343*
DirecTV-4s	0.8232	0.9601	0.9697*	0.1941	0.4550	0.5306*	0.1158	0.4616	0.5352*
DMSP F7	0.9586	0.9794	0.9819*	0.1992	0.4005	0.4584*	0.1240	0.3949	0.4028*
DSCS-3B6	0.9391	0.9724	0.9832*	0.2585	0.6599	0.7892*	0.1493	0.6172	0.7586*
Galaxy 4R	0.7970	0.9492	0.9724*	0.2261	0.4442	0.5280*	0.1155	0.4555	0.5375*
GOES 14	0.9249	0.9711	0.9785*	0.2184	0.5146	0.5626*	0.1277	0.4842	0.5106*
Horizons-2	0.9061	0.9823	0.9831*	0.1822	0.5976	0.6797*	0.1651	0.5910	0.6687*
Hubble	0.6491	0.9001	0.9402*	0.2552	0.5026	0.6040*	0.1459	0.4628	0.5162*
Landsat D	0.9226	0.9692	0.9739*	0.1981	0.3859	0.5816*	0.1131	0.3570	0.3656*
Meteor-3	0.8956	0.9677	0.9747*	0.2313	0.5052	0.7646*	0.1284	0.4635	0.6042*
Mexsat-3	0.8882	0.9696	0.9775*	0.1864	0.6629	0.7092*	0.1896	0.6190	0.6507*
SES-1	0.8792	0.9735	0.9828*	0.1670	0.5838	0.6996*	0.1708	0.5724	0.6710*
Spaceway-1	0.7674	0.9185	0.9458*	0.2686	0.5993	0.7608*	0.1319	0.6050	0.7417*
TDRS-5	0.8701	0.9386	0.9498*	0.2729	0.7901	0.8404*	0.1474	0.6825	0.7196*
Telstar 401	0.8241	0.9317	0.9657*	0.1874	0.5989	0.7667*	0.1558	0.5863	0.7535*
AMC-4	0.8311	0.8299	0.9667*	0.1594	0.3311	0.7142*	0.1461	0.3443	0.5380*
AMC-8	0.8841	0.9110	0.9251*	0.2126	0.3422	0.4700*	0.1128	0.2342	0.3615*
AMC-15	0.8203	0.8772	0.9137*	0.1864	0.4539	0.6818*	0.1531	0.3851	0.5755*
AMC-18	0.8857	0.8893	0.9148*	0.2221	0.2854	0.4351*	0.1170	0.1889	0.3053*
AMC-23	0.7626	0.7366	0.8964*	0.1366	0.1978	0.6322*	0.1440	0.1992	0.4836*
DMSP-F16	0.9484	0.9422	0.9677*	0.1945	0.2155	0.3399*	0.1323	0.2472*	0.2369
Echostar XIV	0.6699	0.7001	0.8703*	0.1546	0.2536	0.4345*	0.1057	0.2145	0.4288*

Table 3: Semantic Segmentation of Satellite Material Metrics across Combined (Visible Light and Hyperspectral) and Hybrid Models. The table displays accuracy, macro average precision, and macro average recall values for both the Combined and Hybrid Models, differentiated by data type (Visible Light and Hyperspectral). The Hybrid Model demonstrates superior performance across most satellites, particularly in accuracy and macro average precision, underscoring the advantages of integrating data sources for enhanced satellite material classification. The highest metrics for each satellite are marked with asterisks. Satellites listed below and including AMC-4 were excluded from the training of the Combined Models.

7. REFERENCES

[1] Stein D.W.J.; Beaven S.G.; Hoff L.E.; Winter E.M.; Schaum A.P.; Stocker A.D. Anomaly detection from hyperspectral imagery. *IEEE*, 2002.

[2] Yanan Luo; Jie Zou; Chengfei Yao; Xiaosong Zhao; Tao Li; Gang Bai. Hsi-cnn: A novel convolution neural network for hyperspectral image. *IEEE*, 2018.

[3] Olaf Ronneberger; Philipp Fischer; Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CVPR*, 2015.

[4] José M. Bioucas-Dias; Antonio Plaza; Nicolas Dobigeon; Mario Parente; Qian Du; Paul Gader; Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE*, 2012.

[5] Ujjal; Bhattacharyya Dhruba Choudhury, Purbarag; Dutta. Multispectral satellite data classification using soft computing approach. *CVPR*, 2022.

[6] Jiabing Leng; Tao Li; Gang Bai; Qiankun Dong; Han Dong. Cube-cnn-svm: A novel hyperspectral image classification method. *IEEE*, 2016.

[7] J. Zachary Gazak; Ian McQuaid; Ryan Swindle; Matthew Phelps; Justin Fletcher. Spectranet: Learned recognition of artificial satellites from high contrast spectroscopic imagery. *WACV*, January 2022.

[8] J. Zachary Gazak; Ryan Swindle; Matthew Phelps; Justin Fletcher. Simultaneous detection, recognition, and localization of geosynchronous satellites from ground based imagery. *AMOS*, September 2023.

[9] Matthew Phelps; Thomas Swindle; J. Zachary Gazak; Andrew Vandenberg; Justin Fletcher. Spectranet-so(3): Learning satellite orientation from optical spectra by implicitly modeling mutually exclusive probability distributions on the rotation manifold. *IEEE*, 2023.

[10] M.T. Eismann; R.C. Hardie. Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions. *IEEE*, 2005.

[11] European Space Imaging. Short-wave infared imagery (swir).

[12] Marion F. Baumgardner; Larry L. Biehl; David A. Landgrebe. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3. *Purdue University*, 2015.

[13] Zongwei Zhou; Md Mahfuzur Rahman Siddiquee; Nima Tajbakhsh; Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *CVPR*, 2018.

[14] Lewis; Campbell Andrew; Marto Simão; Murray Paul; Marshall Stephen; Savitski Vasili Massimiliano, Vasile; Walker. Space object identification and classification from hyperspectral material analysis. *Scientific Reports*, August 2023.

[15] R. Richter. Hyperspectral sensors for military applications. *Emerging EO Phenomenology*, 2005.

[16] Ozan Oktay; Jo Schlemper; Loic Le Folgoc; Matthew Lee; Mattias Heinrich; Kazunari Misawa; Kensaku Mori; Steven McDonagh; Nils Y Hammerla; Bernhard Kainz;Ben Glocker; Daniel Rueckert. U-net: Convolutional networks for biomedical image segmentation. *CVPR*, 2018.

[17] CMS series. Multispectral cameras.

[18] Wenju Wang; Shuguang Dou; Sen Wang. Alternately updated spectral–spatial convolution network for the classification of hyperspectral images. *MDPI*, 2019.

[19] XIMEA. Hyperspectral snapshot usb3 camera 15 bands 600-860nm.

[20] XIMEA. Hyperspectral snapshot usb3 camera 16 bands 460-600nm.

[21] XIMEA. Ximea - hyperspectral snapshot usb3 camera 24 bands 665-960nm.