A Comparison of Incremental Semantic Segmentation Algorithms for the Recognition of Agricultural Threats from UAV Images

Lucas Oliveira Antonetti[1], Gabriel F. D. Pereira[1], Rafael Gomes Mantovani[3], Luiz F. S. Coletta, André Luís Debiaso Rossi[1]

[1]São Paulo State University (UNESP) – Bauru – SP – Brazil

[2]Federal University of Technology (UTFPR) – Apucarana – PR – Brazil

{lucas.antonetti,gabriel.fd.pereira}@unesp.br,

rafaelmantovani@utfpr.edu.br, luiz.coletta@alumni.usp.br,

andre.rossi@unesp.br

Abstract. *This study investigates two Incremental Semantic Segmentation (ISS) frameworks, ALIFE and SATS, using a data-scarce UAV imagery for agricultural threat recognition. Using a 5-fold cross-validation, the methods are evaluated through a three-stage protocol that incrementally introduces base categories, corn, and the weed Panicum. In this setup, both methods achieved comparable final performance (ALIFE 51.48% mIoU, SATS 51.38%), though SATS showed a clear advantage on Ground recognition, while differences on Corn and Panicum were comparable between methods, and ALIFE exhibited more stable training convergence. Notably, SATS experienced complete background suppression, a degradation absent in standard benchmarks.*

1. Introduction

Technological advancements have enabled continuous development in the agricultural sector, increasing its efficiency and sustainability. This progress is primarily driven by computer vision systems, which have paved the way for precision agriculture, also known as Agriculture 4.0 [Zhai et al. 2020]. However, agricultural production can still be optimized, as challenges related to the competition between crops and weeds [Rizzardi et al. 2001] present opportunities for the development of Machine Learning solutions.

Different techniques for recognizing weeds and other harmful agents in aerial crop images have been investigated in the literature [Chin et al. 2023, Hasan et al. 2021]. While these methods vary in their approaches, many rely on classification techniques [Jin et al. 2022], in which images, or sections thereof, are labeled based on the presence or absence of weeds. However, such techniques have limited precision, since large sections labeled as "containing weeds" may include weed-free areas. Semantic Segmentation techniques [Mo et al. 2022] mitigate this problem by assigning a label to each pixel in an image.

Furthermore, many Semantic Segmentation models can differentiate among multiple crop and weed species within a single scene [Gupta et al. 2023, Pereira 2023], which

is relevant for developing targeted responses against these harmful agents. The problem, however, is that traditional machine learning methods can only identify the classes they were originally trained on. If novel weed species or other plants appear in the field, they will be misidentified. Thus, in closed-set recognition, all classes within the domain are presumably represented in the training set.

Class-Incremental Learning (CIL) models [Zhou et al. 2024] are developed to address these limitations. These Continual Learning systems must be capable of learning new classes from a constant stream of data without forgetting the previous ones, maintaining a balance between stability and plasticity [Parisi et al. 2019]. One challenge in this process is recognizing when a novel class emerges. Some systems utilize Novelty Detection or Out-of-Distribution mechanisms to automatically flag the presence of an unknown plant in the field [Pimentel et al. 2014, Silva 2023].

Studies on incremental classification models are more common than their Semantic Segmentation equivalents [Wang et al. 2024], although the field has been receiving growing attention. While there has been significant progress in adjacent domains [Tasar et al. 2019, Shan et al. 2022], direct applications to UAV-based precision agriculture remain scarce.

Thus, to address this gap, this paper explores the adaptation of two distinct state-of-the-art Incremental Semantic Segmentation (ISS) approaches, ALIFE and SATS, applied to Unmanned Aerial Vehicles (UAV) imagery for agricultural threat recognition. The main contributions of this work are:

- The successful adaptation of Incremental Semantic Segmentation to UAV imagery through a three-stage incremental protocol, bridging a critical gap between continual learning and precision agriculture;
- The identification and correction of a critical instability in SATS's balanced initialization mechanism, which triggers complete classifier collapse when applied beyond the first incremental step;
- The optimization of ALIFE's loss components through empirical hyperparameter search, demonstrating that the Adaptive Logit Regularizer alone provides sufficient regularization to assimilate novel weed classes under severe data constraints.

2. Foundations and Related Work

Semantic Segmentation [Mo et al. 2022] is a task in which a Machine Learning (ML) model must classify each pixel of an image based on the class to which it belongs. It is applied in a variety of contexts, ranging from autonomous driving to precision agriculture. In the agricultural domain, some authors [Gupta et al. 2023, Pereira 2023] have successfully applied semantic segmentation models to differentiate among multiple crop and weed species within a single scene, offering finer spatial precision than classification-based approaches. Despite achieving high accuracy, these models assume a static, predefined environment, unlike real-world agricultural fields, which are highly dynamic.

The field of Continual Learning [Wang et al. 2024] seeks to develop models capable of learning from a continuous stream of data without forgetting previously acquired knowledge [Parisi et al. 2019]. Three possible scenarios for incremental learning are task-incremental, domain-incremental, and class-incremental learning [Van de Ven et al. 2022]. This work focuses on CIL [Zhou et al. 2024], in which an

algorithm must incrementally learn to distinguish among a growing number of classes using a single model, without access to task identity at inference time. One of the Continual Learning strategies relevant to this project is the use of a novelty detector [Pimentel et al. 2014], a system capable of discovering instances of new classes and separating them so that an external agent can label them, thereby updating the dataset on which the model is trained.

Applying CIL to semantic segmentation introduces challenges that do not arise in classification settings, the most significant one being the background shift problem. Formally, at step t, the background label absorbs any pixel whose true class has not yet been introduced (i.e., $y \notin C_t$), meaning the background distribution changes composition at every stage rather than corresponding to a single stable category [Cermelli et al. 2020]. For example, in this study's base phase, pixels of the weed *Panicum* are masked as Background, leading the model to inadvertently associate the weed's features with the Background and causing severe confusion when it is later introduced as a new target class. While the field has been receiving growing attention [Yuan and Zhao 2024, Cong et al. 2025], the application of ISS to an agricultural context remains largely unexplored.

While direct applications to UAV-based precision agriculture are scarce, relevant progress has been made in adjacent domains. For instance, [Tasar et al. 2019] proposes a methodology for dealing with incremental large-scale remote sensing data, mitigating catastrophic forgetting by maintaining a frozen copy of the previously trained network, which is then used to guide the updated network. Similarly, [Shan et al. 2022] tackles the ISS problem in aerial imagery, presenting a deep feature distillation (DFD) module and a label-mixed (LM) module that balance learning new classes while retaining past representations without storing old data. Bridging the gap toward practical drone development, [Cong et al. 2025] addresses the computational limitations of edge devices by proposing a lightweight ISS framework.

To address catastrophic forgetting and background shift, recent extensive reviews [Yuan and Zhao 2024] and comparative studies [Kalb et al. 2021] in Continual Semantic Segmentation (CSS) categorize architectures into two main branches: data-replay methods and data-free methods. Data-replay strategies mitigate forgetting by storing a small subset of historical data or compressed representations to refresh the model's memory during new training phases. Conversely, data-free methods do not store prior data; instead, they use alternative strategies to preserve acquired knowledge.

Based on this literature review and the availability of code, this work selects one high-performing representative from each of these major paradigms to evaluate in the agricultural domain. The chosen algorithms are ALIFE [Oh et al. 2022], which represents the data-replay paradigm via latent-feature replay, and SATS [Qiu et al. 2023], a data-free approach that transfers structural knowledge via self-attention maps.

ALIFE addresses catastrophic forgetting through the ALI, a loss term that combines the strengths of Calibrated Cross-Entropy (CCE) and Calibrated Knowledge Distillation (CKD) while mitigating their drawbacks. CCE extracts semantic information from unlabeled regions by reducing the logit values of new categories to prevent overfitting. However, it simultaneously increases the logits of previous categories without verifying

their accuracy, leading to flawed representations. CKD mitigates forgetting by distilling knowledge from the old model's outputs. While this successfully retains prior information, it blurs the distinction between the new target categories and the Background. ALI resolves this dilemma by adaptively adjusting logit values based on the discrepancy between the current model's predictions and those of the previous model, effectively neutralizing the drawbacks of both CCE and CKD. The stored feature vectors from ALIFE's latent feature replay are subsequently used to fine-tune the classifier at each incremental stage, preventing the overwriting of prior knowledge without modifying the feature extractor.

In contrast to ALIFE's methods, SATS addresses catastrophic forgetting through a relational knowledge distillation approach built on a SegFormer-B2 Transformer architecture. It preserves the structural relationships between image elements by transferring self-attention maps from the previous model to the current one. To avoid over-constraining the model, it applies CRP, which aggregates pixel-level features into class-specific region vectors. Because Background regions are too visually heterogeneous to form a cohesive representation, they are explicitly excluded from this process, allowing the model to calculate and distill attention scores solely between meaningful semantic regions rather than individual pixels.

3. Experiments

This section presents the experimental methodology performed to evaluate different Incremental Semantic Segmentation models for UAV imagery in agricultural threat recognition.

3.1. Dataset

The images used for training and evaluating the models were acquired by a DJI Phantom 4 Pro UAV in the state of Mato Grosso do Sul, Brazil, over corn plantations [Carnevalli 2020]. A portion of the dataset was annotated in [Carnevalli 2020], whereas additional images were manually annotated in this study using the open source platform Label Studio[1]. The final dataset comprised 90 images, divided via random assignment into 5 folds of 18 images each for cross-validation. In each validation step, the folds were distributed using a 3:1:1 split, allocating 3 folds for training, 1 for validation, and 1 for testing.

These data were pre-processed, including resizing, cropping, and pixel-level labeling, as illustrated in Figure 1. To evaluate the models in an Incremental Semantic Segmentation (ISS) scenario, the four semantic classes were distributed across a base phase and two sequential incremental learning steps. The base categories consisted of Background and Ground. The novel classes introduced incrementally were Corn (Step 1) and the invasive weed *Panicum* (Step 2). The dataset exhibits notable class imbalance, with Corn comprising 57.67% of annotated pixels, followed by Ground (23.59%), Background (10.41%), and *Panicum* (8.33%), making the weed class the rarest and the last to be introduced.

To simulate this sequential introduction of classes, a three-stage training protocol was employed. During the initial base training phase, all pixels corresponding to both

[1]https://github.com/HumanSignal/label-studio

Corn and *Panicum* were masked and relabeled as Background, hiding their true identity from the models and forcing them to form representations solely for Background and Ground. In the first incremental learning phase (Step 1), the true labels for Corn were unmasked and introduced to the ground truth, while *Panicum* pixels remained hidden under the Background mask. At this stage, the models were tasked with learning to recognize the crop class while retaining past representations. Finally, in the second incremental learning phase (Step 2), the true labels for *Panicum* were introduced. The frameworks were then challenged to incorporate this new weed class into their classifier without triggering catastrophic forgetting of the previously acquired categories.



Figure 1. Cropped image before and after segmentation. Black represents Background, Red represents Corn, Green represents Ground and Yellow represents *Panicum*.

3.2. Experimental Setup

To evaluate the frameworks under their officially optimized settings, different architectural backbones are utilized based on each method's design. The ALIFE framework employs a DeepLab-v3 model with a ResNet-101 feature extractor, capturing multi-scale context via an Atrous Spatial Pyramid Pooling (ASPP) module applied to the spatial feature maps. Conversely, SATS is built on a SegFormer-B2 Transformer architecture that natively generates hierarchical self-attention maps for structural knowledge transfer. The backbones of both models were initialized with ImageNet pre-trained weights. In contrast, the segmentation classifiers and all incremental learning modules were initialized from scratch and trained exclusively on the UAV dataset.

The experiments were conducted using the PyTorch deep learning framework. Both frameworks apply the same preprocessing pipeline: during training, a random-resized-crop augmentation samples a region of random size and aspect ratio, then resizes it via bilinear interpolation to 512×512 pixels; at validation and test time, images are resized and center-cropped to the same 512×512 resolution. During the base training phase, both models were trained for 60 epochs with a batch size of 9. Optimization was performed using Stochastic Gradient Descent (SGD) with Nesterov momentum, setting an initial learning rate of 0.01, a momentum of 0.9, and a weight decay of 1×10^{-4}. ALIFE was trained with a fixed seed of 28200, and SATS used its framework's default seed of 42. For the incremental training stages, the optimization strategies diverged to accommodate each method's specific pipeline. SATS was updated end-to-end for an additional 60 epochs per stage with a reduced learning rate of 0.001. Conversely, ALIFE used a multi-stage incremental pipeline: after updating the feature representations for 60 epochs, the model trained latent-feature rotation matrices for 10 epochs, followed by a final classifier fine-tuning stage of 2 epochs with a learning rate of 0.001. SATS training additionally employed mixed-precision optimization (Apex AMP O1) to reduce memory requirements during the SegFormer-B2 forward passes.

During the multi-stage experiments, a critical instability was identified in SATS's balanced initialization strategy. This mechanism is designed to mitigate the background shift by suppressing the background channel's classification bias when introducing a new category. While effective for the initial incremental step, applying it in the subsequent stage led to a complete classifier collapse, resulting in 0% validation accuracy for both the background and the newly introduced class. To fix this, balanced initialization was explicitly turned off after the first step, allowing the network to learn the subsequent class naturally.

To optimize performance on the agricultural dataset, both models underwent a targeted hyperparameter ablation study over 9 distinct combinations. For ALIFE, the grid search varied the Adaptive Logit Regularizer weight (λ_{ALI}) and Knowledge Distillation weight (λ_{KD}), leading to a final configuration that eliminated the KD penalty ($\lambda_{KD} = 0$) and halved the regularizer weight ($\lambda_{ALI} = 0.5$). For SATS, the grid search cross-examined the self-attention transfer weight ($\lambda_a \in \{0, 10, 20\}$) and the output knowledge distillation weight ($\lambda_d \in \{0, 10, 20\}$). Validation performance remained largely stable across most configurations; however, accuracy degraded noticeably when output distillation was removed entirely ($\lambda_d = 0$), confirming its necessity. Ultimately, the original default values of $\lambda_a = 20$ and $\lambda_d = 20$ were retained as they provided the most stable validation performance.

3.3. Model Evaluation

Performance was evaluated using mean Intersection over Union (mIoU), Mean Accuracy (MA), and Harmonic mean IoU (hIoU). In precision agriculture, mIoU serves as the primary metric because it rigorously penalizes false positives and negatives, ensuring exact boundary predictions for localized herbicide application (Equation 1). To account for the dataset's severe class imbalance, MA provides a measure of pure sensitivity by computing the mean proportion of correctly classified pixels per class (Equation 2).

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \frac{|A_i \cap B_i|}{|A_i \cup B_i|} \tag{1}$$

$$\text{MA} = \frac{1}{N} \sum_{i=1}^{N} \frac{|A_i \cap B_i|}{|A_i|} \tag{2}$$

where N is the total number of classes; A_i and B_i are the ground truth and predicted masks for class i, respectively.

Finally, hIoU is adopted to explicitly quantify the stability-plasticity trade-off during incremental stages. While standard mIoU provides a global average, hIoU prevents high performance on previously seen classes ($mIoU_{old}$) from obscuring failures on newly introduced classes ($mIoU_{new}$), or the catastrophic forgetting of older knowledge (Equation 3).

$$\text{hIoU} = \frac{2 \times mIoU_{old} \times mIoU_{new}}{mIoU_{old} + mIoU_{new}} \tag{3}$$

4. Results

To understand the models' learning behavior before final performance benchmarking, Figure 2 presents the training loss curves concatenated across the three learning stages (Base, Step 1, and Step 2), averaged across all five folds, with shaded regions representing the standard deviation. ALIFE exhibits a smooth convergence trajectory, consistently remaining at the bottom of the loss scale with near-zero variance across all three stages. Conversely, SATS exhibits an escalating pattern of instability. During the base stage, its variance and loss magnitude are relatively low; however, both the training loss value and its volatility across the data splits increase significantly with each subsequent incremental step.

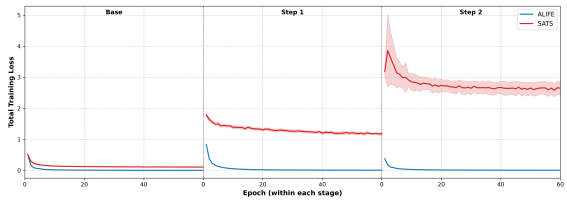


Figure 2. Training loss curves concatenated across the Base (60 epochs), Step 1 (60 epochs), and Step 2 (60 epochs) training phases. Solid lines represent the mean across 5 folds, and shaded regions indicate the standard deviation.

Following the multi-stage training phase, the models' final segmentation performance was evaluated. Performance is reported in terms of mIoU and hIoU, computed per class and averaged across the five cross-validation folds. As detailed in Table 1, the evaluation presents individual scores for each class, aggregated metrics across all classes, and an aggregation excluding the Background class to better reflect the models' agronomic performance.

Table 1. Evolution of the segmentation metrics across the three learning stages (mean IOU % ± standard deviation over 5 folds). The Harmonic mIoU (hIoU) evaluates the stability-plasticity trade-off at each incremental step.

Class / Metric	Base Phase		Step 1 (+Corn)		Step 2 (+Panicum)	
	ALIFE	SATS	ALIFE	SATS	ALIFE	SATS
Background	8.44 ± 0.97	**93.07** ± 1.20	16.06 ± 3.66	**20.45** ± 6.63	**11.76** ± 7.33	0.00 ± 0.00
Ground	70.07 ± 2.24	**79.40** ± 1.49	70.77 ± 1.04	**79.14** ± 1.85	70.43 ± 1.23	**79.17** ± 1.64
Corn (Step 1)	Masked		72.42 ± 8.94	**78.12** ± 4.07	78.58 ± 4.76	**78.73** ± 3.82
Panicum (Step 2)	Masked		Masked		45.15 ± 7.33	**47.63** ± 10.99
Global mIoU (Active Classes)	39.26 ± 1.32	**86.23** ± 1.33	53.09 ± 3.86	**59.24** ± 3.50	**51.48** ± 3.88	51.38 ± 3.89
mIoU (Without Background)	70.07 ± 2.24	**79.40** ± 1.49	71.60 ± 4.57	**78.63** ± 2.83	64.72 ± 3.71	**68.51** ± 5.19
Harmonic mIoU (hIoU)	N/A		54.14 ± 3.75	**60.77** ± 3.66	48.80 ± 5.09	**49.46** ± 6.94

Both methods achieved comparable final performance after the complete three-stage protocol, with ALIFE reaching a Global mIoU of 51.48% and SATS 51.38% at Step 2. However, this near-equivalence stems primarily from SATS's complete degradation of the Background class; when excluded from the mIoU, SATS numerically outperforms ALIFE (68.51% vs. 64.72%), though the standard deviations for this metric overlap (±3.71 and ±5.19, respectively). This dynamic evolved across the training stages. During the Base Phase, SATS had a considerably higher Background mIoU (93.07% vs. 8.44%)

and consequently a higher overall mIoU (86.23% vs. 39.26%). As the incremental steps were introduced, this gap narrowed. By Step 1, SATS's Background mIoU decreased to 20.45%, and at Step 2, it collapsed to 0%.

To compare the incremental performance against an upper bound, both frameworks were additionally trained offline on all four classes jointly, bypassing the incremental protocol entirely (Table 2). This joint training achieved higher performance for both methods (ALIFE: 61.06% mIoU, 72.74% without Background; SATS: 67.73% mIoU, 77.73% without Background), indicating that the performance degradation observed during the sequential steps is driven by the incremental learning protocol itself, rather than an inability of the models to learn agricultural features. Notably, even under joint training, Background remained the weakest class for both methods (26.00% and 37.73% IoU for ALIFE and SATS, respectively), far from the collapse observed at incremental Step 2, but still comparatively low. This suggests that Background is an intrinsically harder class in this dataset, independent of the incremental mechanism.

Table 2. Baseline joint training metrics (mean IOU % ± SD). Offline results bypassing the incremental protocol to serve as an upper bound.

Method	Class mIoU				Global mIoU	
	Background	**Ground**	**Corn**	**Panicum**	**All Classes**	**No Background**
ALIFE	26.00 ± 4.40	71.57 ± 1.52	84.36 ± 1.60	62.29 ± 3.44	61.06 ± 1.43	72.74 ± 1.92
SATS	**37.73** ± 5.59	**79.17** ± 0.94	**87.36** ± 1.66	**66.66** ± 2.15	**67.73** ± 1.36	**77.73** ± 1.07

This inherent difficulty of the Background class, coupled with the incremental learning protocol, led to its complete collapse in SATS at Step 2. This failure is visually evident in the qualitative segmentation comparisons presented in Figure 3, where SATS completely fails to assign the Background label to any region of the target scene. This total suppression, unobserved in original SATS benchmarks [Qiu et al. 2023], likely stems from two compounding factors. First, SATS's Class-specific Region Pooling (CRP) inherently excludes the heterogeneous Background from its distillation process, meaning no structural knowledge is transferred across stages for this class. Second, the three-stage protocol, coupled with the small 90-image dataset, amplifies the background shift problem, as *Panicum* pixels are masked as Background in Step 1, substantially disrupting the Background distribution when their true labels are finally introduced in Step 2. In contrast, ALIFE retained partial Background recognition, achieving an mIoU of 11.76%.

For the agronomically relevant classes, both methods demonstrated strong performance on Corn at Step 2, achieving near-identical mIoU values of 78.58% (ALIFE) and 78.73% (SATS), likely reflecting its high visual regularity in the UAV imagery. The Ground class also yielded robust mIoU scores, although ALIFE's performance was comparatively lower than that of SATS (70.43% vs. 79.17%). Conversely, performance on the newly introduced *Panicum* class was noticeably lower overall, though the difference between the two frameworks remained minimal, with SATS achieving 47.63% compared to ALIFE's 45.15%. The distribution and cross-validation variance of these final metrics across all five data splits are explicitly illustrated in the Step 2 class-wise IoU boxplot (Figure 4), revealing that while the Ground class remains highly consistent across folds for both methods, the remaining classes exhibit more volatility, most notably *Panicum*, where SATS's wider spread (≈30%–64%) indicates greater sensitivity to fold composi-

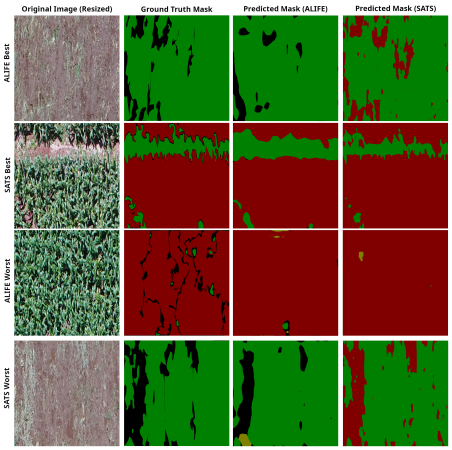


Figure 3. Best and worst segmentation outcomes. Columns (left to right): UAV image, ground truth, ALIFE, and SATS predictions. Colors: Black (Background), Red (Corn), Green (Ground), and Yellow (*Panicum*).

tion than ALIFE's tighter distribution. The Mean Accuracy metric further reveals that ALIFE detects *Panicum* pixels with high sensitivity (MA of 89.74%) but at the cost of boundary precision, as its considerably lower IoU (45.15%) indicates significant over-segmentation into neighboring regions. SATS, by contrast, presents a more balanced profile (MA 51.80%, IoU 47.63%), suggesting tighter but less complete detections.

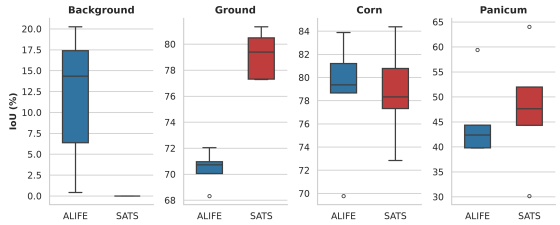


Figure 4. Step 2 Per-Class IoU Comparison (5-Fold Mean).

To ensure a fair comparison, basic training hyperparameters such as learning rate, epochs, and batch size were kept identical, isolating the effect of each method's incremental mechanism from the optimization setup. However, the architectural differences between backbones should be noted when interpreting absolute performance differences. A preliminary ablation study tested different weights for ALIFE's specific loss components, Adaptive Logit Regularizer (ALI) and Knowledge Distillation (KD). The analysis revealed that high KD values led to widespread degradation of the network's discriminative ability, deteriorating performance across both the base and incremental classes, often resulting in excessive Background predictions. Furthermore, it was observed that high ALI weights also led to a gradual decline in overall segmentation accuracy, progressively penalizing the new *Panicum* class. Ultimately, eliminating the KD penalty and setting the ALI weight to $\lambda_{ALI} = 0.5$ provided the optimal balance. Similarly, an ablation study was conducted for SATS to test different configurations for its self-attention transfer weight (λ_a) and output knowledge distillation weight (λ_d). As the validation performance remained mostly robust across the tested variations, the original default values ($\lambda_a = 20$ and $\lambda_d = 20$) were retained.

While a direct comparison with the original SATS and ALIFE benchmarks is limited by differences in dataset scale and domain complexity, contrasting the results highlights the unique dynamics of agricultural imagery. The original ALIFE study reported mIoU scores up to 55.27% for new classes and 77.66% for old classes, whereas SATS achieved overall mIoUs of up to 78.72%. In this study, both methods converged to considerably lower overall mIoU (ALIFE's 51.48% and SATS's 51.38%), reflecting the challenges posed by the small 90-image dataset, class imbalance, and background shift. The newly introduced *Panicum* class was learned at 45.15% mIoU (ALIFE) and 47.63% mIoU (SATS), substantially below the new-class scores reported on standard benchmarks, a shortfall likely compounded by its minority status in the dataset (8.33% of pixels), which limits the discriminative signal available during the incremental stage. Conversely, both methods achieved strong performance on Corn (ALIFE's 78.58% and SATS's 78.73%), a result consistent with its visual regularity and its dominance in the dataset (57.67% of pixels), which provides abundant training signal across all learning stages.

5. Conclusion

This work investigated and adapted ALIFE and SATS, two state-of-the-art incremental semantic segmentation methods for the recognition of agricultural threats in UAV imagery. Both methods were applied to a dataset of 90 aerial images of corn plantations captured in Mato Grosso do Sul, Brazil, across a three-stage incremental protocol introducing Background and Ground at the base phase, Corn at Step 1, and the invasive weed *Panicum* at Step 2. Experiments were conducted using 5-fold cross-validation, and performance was measured by mean Intersection over Union (mIoU) and Harmonic mIoU (hIoU).

The results demonstrate that both methods can be adapted to the agricultural domain and converged to comparable final performance, with ALIFE reaching 51.48% mIoU and SATS 51.38% at Step 2. However, this near-equivalence conceals substantially different profiles. SATS, built on a SegFormer-B2 backbone, showed a consistent advantage on Ground recognition throughout all stages, while differences on Corn, *Panicum*, and hIoU (49.46% vs. 48.80%) fell within overlapping standard deviations and should not be read as a definitive advantage without formal significance testing. ALIFE, configured with a reduced ALI weight ($\lambda_{ALI} = 0.5$) and the KD term disabled ($\lambda_{KD} = 0$) following empirical hyperparameter search, demonstrated markedly more stable training convergence across all folds and stages, suggesting that the ALI term alone provides sufficient regularization under small-dataset agricultural constraints.

Regarding SATS, a severe instability was identified in its balanced initialization mechanism, which triggered complete classifier collapse when applied beyond the first incremental step and required an explicit correction. Additionally, the most critical finding was SATS's complete failure to predict the Background class, yielding zero IoU across all five folds at Step 2, a behavior not observed in the method's original benchmarks [Qiu et al. 2023]. This is consistent with background shift being a well-documented driver of catastrophic forgetting in class-incremental segmentation more broadly [Cermelli et al. 2020], rather than a phenomenon unique to this domain. Without any form of data replay, models can fail to retain discriminating features between classes learned in different steps that share similar visual characteristics [Kalb and Beyerer 2022], a limitation that likely applies to SATS's data-free design and its reduced ability to retain a discriminative representation for Background against the classes introduced in later steps.

This total suppression, rather than partial degradation, is more plausibly explained by this architectural vulnerability, compounded by the small dataset size and the extended three-stage protocol, than by any agriculture-specific factor, a conclusion supported by the joint-training baseline, where Background remained the weakest class but did not collapse entirely.

Future work should expand the dataset by including additional weed species and plantation conditions to enable a more comprehensive evaluation of the incremental learning protocols. The integration of either framework with a novelty detection mechanism represents a particularly promising direction, allowing the system to identify previously unseen weed species autonomously rather than relying on supervised incremental stages, moving toward a fully open-world agricultural segmentation system deployable on UAV platforms.

References

Carnevalli, Silva e Souza, S. (2020). Estudo de deep learning para o reconhecimento de ervas invasoras na cultura do milho a partir de imagens de vants.

Cermelli, F., Mancini, M., Bulò, S. R., Ricci, E., and Caputo, B. (2020). Modeling the background for incremental learning in semantic segmentation.

Chin, R., Catal, C., and Kassahun, A. (2023). Plant disease detection using drones in precision agriculture. *Precision Agriculture*, 24(5):1663–1682.

Cong, W., Cong, Y., and Ren, Y. (2025). Lightweight class incremental semantic segmentation without catastrophic forgetting. *IEEE Transactions on Image Processing*, 34:4566–4579.

Gupta, S. K., Yadav, S. K., Soni, S. K., Shanker, U., and Singh, P. K. (2023). Multiclass weed identification using semantic segmentation: An automated approach for precision agriculture. *Ecological Informatics*, 78:102366.

Hasan, A. M., Sohel, F., Diepeveen, D., Laga, H., and Jones, M. G. K. (2021). A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture*, 184:106067.

Jin, X., Bagavathiannan, M., McCullough, P. E., Chen, Y., and Yu, J. (2022). A deep learning-based method for classification, detection, and localization of weeds in turfgrass. *Pest Management Science*, 78(11):4809–4821.

Kalb, T. and Beyerer, J. (2022). Causes of catastrophic forgetting in class-incremental semantic segmentation.

Kalb, T., Roschani, M., Ruf, M., and Beyerer, J. (2021). Continual learning for class- and domain-incremental semantic segmentation. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 1345–1351.

Mo, Y., Wu, Y., Yang, X., Liu, F., and Liao, Y. (2022). Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493:626–646.

Oh, Y., Baek, D., and Ham, B. (2022). ALIFE: Adaptive logit regularizer and feature replay for incremental semantic segmentation. In *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71.

Pereira, G. F. D. (2023). Investigação de algoritmos de aprendizado profundo para a segmentação semântica de plantas daninhas e culturas agrícolas usando imagens.

Pimentel, M. A. F., Clifton, D. A., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99:215–249.

Qiu, Y., Shen, Y., Sun, Z., Zheng, Y., Chang, X., Zheng, W., and Wang, R. (2023). Sats: Self-attention transfer for continual semantic segmentation. *Pattern Recognition*, 138:109383.

Rizzardi, M. A., Fleck, N. G., Vidal, R. A., Merotto Jr., A., and Agostinetto, D. (2001). Competição por recursos do solo entre ervas daninhas e culturas. *Ciência Rural*, 31(4):707–714.

Shan, L., Wang, W., Lv, K., and Luo, B. (2022). Class-incremental learning for semantic segmentation in aerial imagery via distillation in all aspects. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12.

Silva, M. G. (2023). Classificador open-world para o reconhecimento de ameaças em imagens aéreas de culturas agrícolas. Dissertação de mestrado, Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Engenharia, São João da Boa Vista.

Tasar, O., Tarabalka, Y., and Alliez, P. (2019). Incremental learning for semantic segmentation of large-scale remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9):3524–3537.

Van de Ven, G. M., Tuytelaars, T., and Tolias, A. S. (2022). Three types of incremental learning. *Nature Machine Intelligence*, 4(12):1185–1197.

Wang, L., Zhang, X., Su, H., and Zhu, J. (2024). A comprehensive survey of continual learning: Theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5362–5383.

Yuan, B. and Zhao, D. (2024). A survey on continual semantic segmentation: Theory, challenge, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10891–10910.

Zhai, Z., Martínez, J. F., Beltran, V., and Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, 170:105256.

Zhou, D.-W., Wang, Q.-W., Qi, Z.-H., Ye, H.-J., Zhan, D.-C., and Liu, Z. (2024). Class-incremental learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20.