EcoVision: AI-Powered Drone Imaging for Salt Marsh Vegetation Monitoring and Dominance Mapping

Innocent Onyenonachi, Peter J. Lawerance, Nadia Kanwal[a]

[a] *School of Computer Science and Mathematics, Keele University, ST5 3BG, United Kingdom*

[b] *School of Life Science, Keele University, ST5 3BG, United Kingdom*

Abstract

High-resolution RGB imagery acquired from low-altitude UAV surveys was processed through a modular pipeline incorporating transformer-based semantic segmentation, connected-component vegetation extraction, fine-grained species classification using a ConvNeXt architecture, and grid-based dominance scoring at 2×2m resolution. The framework targeted two ecologically significant halophytic grasses, Spartina maritima and Puccinellia maritima, and was trained using a curated dataset and manually annotated UAV imagery, along with biodiversity imagery sourced from publicly accessible datasets. In order to identify these plants from the imagery, our segmentation yielded reliable species masks (mean IoU ≈ 0.56; pixel-level accuracy ≈ 0.96), while object-level classification achieved very good discrimination (F1 ≈ 0.99). Dominance estimates closely matched quadrat-based field surveys, with mean absolute differences below 8%, preserving fine-scale spatial structure under realistic survey conditions. The developed system, named EcoVision, establishes a practical foundation for scalable, high-resolution salt marsh monitoring, demonstrating how AI-driven workflows can translate pixel-level predictions into ecologically interpretable metrics.

Keywords: perception, computer vision, segmentation, classification

1. Introduction

Perhaps more than landscapes, vegetation underpins and dominates the ecosystem service delivery and functioning in coastal areas, regulating biodiversity patterns, habitat structure, and key ecological processes [1]. Accurate, high-resolution monitoring is critical for assessing environmental change, guiding conservation strategies, and informing predictive ecological models [2, 3]. Despite their long-standing role in ecology, field surveys depend heavily on human judgement and manual effort, introducing variability and limiting scalability for both temporal and spatial coverage [4, 5].

These constraints are particularly pronounced in dynamic or inaccessible habitats, such as salt marshes, where interestingly species of quite different traits and morphology can co-exist in physical niches defined by mere tens of centimetres. The consequences of these very strong physical niches, yet also expansive, inaccessible, and heterogeneous settings, mean traditional surveys are both onerous, focused typically on transects, and often generalise a marsh rather than perhaps finding specific areas of concern that might be missed by linear walks [6]. Satellite imagery might counter this limitation, however, spatial coverage lacks the resolution to identify species (and thus their well-established ecosystem service values [7]), and it also lacks the temporal flexibility necessary for species-level analysis, further impeding timely ecological insight [8, 9]. Emerging UAV-based remote sensing offers high-resolution, repeatable imaging capabilities that capture fine-scale vegetation patterns. At the same time, machine learning methods enable the automated identification, segmentation, and dominance assessment of species [10]. By combining UAV imagery with computer vision and deep learning, the proposed framework addresses critical limitations of conventional monitoring [11], providing scalable, objective, and ecologically interpretable data. This integration also supports further advances such as near real-time ecosystem assessment, invasive species management, and evidence-based conservation planning, bridging the gap between high-resolution imagery and actionable ecological insight.

1.1. Research Objectives

This research aims to develop and evaluate an integrated UAV - AI framework for automated plant species identification and dominance assessment in salt marsh ecosystems. Existing approaches largely focus on species presence/absence mapping or pixel-level classification, offering limited insight into competitive dynamics and ecological dominance within defined spatial units [8]. EcoVision addresses this gap by combining high-resolution UAV imagery with advanced computer vision and deep learning to generate ecologically meaningful, patch-level outputs. The following are the primary objectives:

Figure 1: Caption

- to acquire high-resolution UAV imagery capable of supporting species-level analysis across heterogeneous coastal landscapes.
- to implement and assess transformer-based semantic segmentation and blob-level classification models [12] for accurate vegetation delineation and species identification.
- to design a spatial aggregation strategy that translates model predictions into quantitative dominance scores at a 2×2 m patch scale as per JNCC guidelines [13].
- evaluate the robustness and generalizability of the proposed pipeline across varying environmental conditions and datasets, ensuring suitability for real-world ecological monitoring.

1.2. Contributions of This Study

This study makes several substantive contributions to the field of AI-driven ecological monitoring. **First**, it presents an end-to-end UAV–AI pipeline that integrates high-resolution aerial data acquisition, species-level semantic segmentation, blob-based classification, and patch-level dominance scoring into a unified and reproducible workflow. To our knowledge, we currently know of no other pipeline of this type or utility. Unlike prior approaches that treat these components in isolation, EcoVision aims to bridge the gap between computer vision outputs and ecologically interpretable indicators within a single platform. **Second**, the study demonstrates the application of transformer-based segmentation (SegFormer) and modern convolutional architectures (ConvNeXt) for fine-grained vegetation analysis in complex salt marsh environments, addressing challenges such as overlapping canopies, background noise, and seasonal variability. This contributes empirical evidence for the suitability of state-of-the-art vision models in real-world ecological settings. **Third**, the introduction of a 2×2 m patch-based dominance framework provides a scalable method for quantifying species dominance based on ecologically and policy-driven guidance [13], thus aligned with traditional methods and may also offer insights to emerging priorities in the future, such as invasive species driven by these botanical experts' impact beyond simple presence–absence mapping. This spatially explicit approach thus supports similar conservation and management decisions to which onerous and costly manual surveys also deliver. **Finally**, the research emphasises modularity and reproducibility, offering a framework that can be adapted to other ecosystems, species groups, and monitoring objectives. Collectively, these contributions advance the practical integration of AI and UAV technologies into ecological monitoring, supporting more objective, scalable, and actionable environmental assessment.

2. Related Work

Traditional ecological surveys remain fundamental to biodiversity monitoring, yet they exhibit well-documented limitations that constrain their effectiveness, particularly for programmes requiring consistent spatial and temporal coverage. Field-based approaches rely heavily on manual species identification and visual estimation, introducing observer bias and uncertainty - especially for cryptic, morphologically similar, or rare species [2, 4]. These surveys are also labour-intensive and time-consuming, often requiring specialist teams to work in remote or physically challenging environments, which limits the frequency and scale at which monitoring can realistically occur [14, 5]. Scalability poses a further challenge. Extending surveys across large, remote, or environmentally dynamic landscapes, such as wetlands or coastal systems, requires resources far exceeding what most conservation or land-management programmes can sustain. Yet such repeated, fine-scale monitoring is essential for detecting seasonal or inter-annual ecological change and avoiding gaps in biodiversity records [15]. While current field methods may be optimised in terms of cost–benefit, a substantial spatial–temporal gap remains between what ecologists can practically survey and the resolution of data required for timely, actionable management. Another limitation lies in the delay between data collection and ecological interpretation. Traditional surveys require significant expert verification and taxonomic quality control, which extends the time before results can be incorporated into decision-making. This latency reduces the utility of field data for adaptive management, especially following rapid or transient disturbances such as storm events or short-lived but ecologically consequential anthropogenic impacts [16]. Taken together, these constraints - observer subjectivity, limited coverage, intensive labour requirements, and delays in data availability - highlight the need to augment (not replace) conventional survey methods with complementary, technology-driven tools. UAV imaging and machine learning offer a promising addition to the land-management "toolbox", providing rapid, consistent, and objective assessments that can be later supported or quality-checked by ecologists with far less time and cost[11, 17, 18]. Recent work demonstrates that convolutional neural networks can accurately segment and identify plant species from high-resolution imagery [10], enabling fast, good-faith estimates that ecologists can validate more efficiently than by conducting full field surveys.

2.1. Role of AI and UAVs in Environmental Monitoring

The integration of unmanned aerial vehicles (UAVs) with artificial intelligence (AI) has greatly expanded the capabilities of ecological monitoring, offering novel ways to address long-standing limitations of traditional surveys. UAVs provide flexible, high-resolution imaging that captures fine-scale spatial patterns across heterogeneous, dynamic, or inaccessible regions such as coastal marshes, wetlands, and intertidal zones [19, 15]. Their capacity for repeated flights over fixed plots allows consistent temporal monitoring, facilitating the detection of short-term or seasonal changes that would otherwise be difficult to quantify [8, 20]. When paired with AI-driven analytical pipelines, UAV imagery can be processed for automated species detection, pixel-level segmentation, and patch-scale estimation of vegetation dominance[20, 12]. Machine-learning models, including transformer-based segmenters and advanced CNNs reduce subjectivity, increase repeatability, and allow scalable analysis across large datasets [10]. This enables the generation of ecologically meaningful outputs such as species distribution layers, dominance maps, and habitat-level descriptors that support evidence-based conservation planning [21, 22]. AI-enabled UAV monitoring thus provides a bridge between fine-resolution environmental imagery and the ecological indicators needed for land management, offering consistent, spatially explicit assessments that can complement and enhance traditional survey methods.

2.2. UAV-Based Ecological and Environmental Monitoring

UAVs have rapidly become essential tools for environmental monitoring due to their capacity to capture centimetre-scale imagery at flexible spatial and temporal resolutions. Operating at low altitudes, they offer detail beyond the reach of both field surveys and traditional satellite platforms [2, 4, 10]. Their rapid deployment makes them ideal for systems where vegetation undergoes rapid or seasonal shifts, such as salt marshes and wetlands

[5]. Early studies demonstrated the ability of UAV imagery to map vegetation at species-level resolution more efficiently than manual surveys [23, 11]. As their use expanded across agricultural, forestry, and coastal ecosystems, integration with machine learning further improved detection of invasive, structurally similar, or spatially intermixed species [10, 4]. Meta-analyses consistently show superior spatial detail and classification accuracy relative to traditional surveys, although performance remains sensitive to sensor selection, flight parameters, and species complexity [16]. Nevertheless, challenges remain in converting UAV imagery into ecologically interpretable outputs and standardising workflows across sites and species. These limitations underscore the need for integrated, end-to-end frameworks such as EcoVision, which aims to unify sensing, segmentation, classification, and ecological inference within a single operational pipeline.

2.3. Computer Vision for Plant and Vegetation Analysis

Computer vision has become central to vegetation analysis, allowing automated extraction of structure, composition and species-level information from imagery. Early feature-based approaches were constrained by lighting variability, background clutter, and species similarity [2]. Deep learning—especially CNNs—overcame many of these issues by learning hierarchical representations that capture texture, morphology, and contextual cues critical for differentiating plant species [23, 22]. Advanced semantic and instance segmentation models now allow pixel- or object-level classification with high spatial precision, improving the delineation of overlapping plants and complex canopy structures [4]. Transformer-based architectures further enhance performance by leveraging long-range spatial dependencies, which is particularly valuable in dense marsh environments [24]. Despite these improvements, many studies stop at the segmentation stage and do not translate outputs into ecological metrics such as species dominance, patch structure, or vegetation dynamics (Kattenborn et al., 2019). Addressing this gap requires coupling computer-vision predictions with ecological reasoning and spatial aggregation - an integration pursued in the EcoVision design.

Deep learning now dominates species classification and vegetation segmentation due to its capacity to learn discriminative features from complex imagery. CNN architectures such as ResNet, EfficientNet, and ConvNeXt have demonstrated strong performance when combined with high-resolution UAV data, enabling fine-grained species identification [2].

Semantic segmentation models, including U-Net variants, DeepLabV3+, and transformer-based models like SegFormer, provide robust pixel-level predictions even in challenging environments with complex illumination and vegetation structure [22]. Attention mechanisms and multi-scale feature extraction further enhance boundary accuracy and class separation. Persistent limitations, however, include generalisation across sites, the scarcity of annotated ecological datasets, and high computational demands for large-area imagery [10]. Hybrid pipelines that decouple segmentation and classification increasingly show promise for improving accuracy and interpretability, aligning with EcoVision's modular architecture [21].

2.4. Gaps in Existing Approaches

Despite significant advances, key gaps remain in UAV- and deep learning–based vegetation analysis. Many studies focus narrowly on either segmentation or classification rather than integrating these tasks into pipelines that reflect real ecological workflows. As a result, outputs often lack the ecological structure needed to inform habitat assessments, such as species dominance, patch organisation, or competitive dynamics [16]. Generalisation remains another major challenge. Most models are trained on limited datasets that fail to represent the variation introduced by differing sensors, flight conditions, seasons, or background vegetation, resulting in domain shift when methods are applied to new sites [10]. Ecological interpretability is further constrained by a lack of uncertainty quantification and by the absence of mechanisms to convert pixel-level data into meaningful ecological units [8]. Finally, computational overhead limits the scalability of existing methods, restricting their applicability for near–real-time ecological decision-making. These gaps motivate the need for end-to-end, ecologically grounded systems such as EcoVision, that integrate sensing, segmentation, classification, and ecological interpretation in a manner that balances accuracy, interpretability, and operational feasibility.

3. Methodology

EcoVision is positioned at the intersection of UAV-based ecological monitoring, modern computer vision, and ecologically interpretable machine learning. Unlike prior studies that treat segmentation, classification, and analysis as isolated tasks, EcoVision adopts an end-to-end pipeline aligned with real ecological workflows. High-resolution UAV imagery serves as the primary data source, while deep learning models are used not only for pixel-level accuracy but as building blocks for higher-level ecological inference (Gao et al. 2026) Within the current research landscape, EcoVision differentiates itself by prioritising structure over novelty. Semantic segmentation is used to delineate vegetation extents, followed by object-level extraction and species classification, enabling the transition from raw imagery to quantified species presence. Crucially, predictions are aggregated into fixed spatial units, allowing species dominance and distribution to be assessed in a manner consistent with field-based ecological methods. EcoVision also addresses key weaknesses in existing approaches by emphasising modularity, scalability, and interpretability. The system is designed to adapt to new species, datasets, and environments without retraining entire pipelines. By grounding model outputs in spatial and ecological context, EcoVision bridges the gap between computer vision research and applied environmental monitoring, positioning it as a practical and extensible framework rather than a task-specific experimental study.

3.1. Conceptual Architecture of EcoVision

The EcoVision framework is designed as a modular, end-to-end system that transforms raw UAV imagery into ecologically meaningful outputs. At the **perception** layer, semantic segmentation models identify vegetation and non-vegetation regions at the pixel level, establishing spatial structure within each scene. This is followed by object-level extraction, where contiguous vegetation regions are isolated to enable species-level reasoning. A dedicated classification module assigns species labels to these regions, allowing heterogeneous vegetation to be decomposed into ecologically relevant units (**interpretation**). The final layer **aggregates** predictions within fixed spatial grids, translating model outputs into metrics such as species presence and dominance (Kent, 2012). This design choice aligns computational outputs with conventional ecological sampling frameworks. By separating acquisition, perception, interpretation, and aggregation into distinct but connected stages, EcoVision emphasises clarity, adaptability, and ecological validity, enabling robust monitoring across sites and time without redesigning the entire system. A step-by-step description is given below:

- High-resolution UAV imagery is first acquired under standardised survey conditions to ensure consistency across flights and sites. These images form the sole input to the computational pipeline.

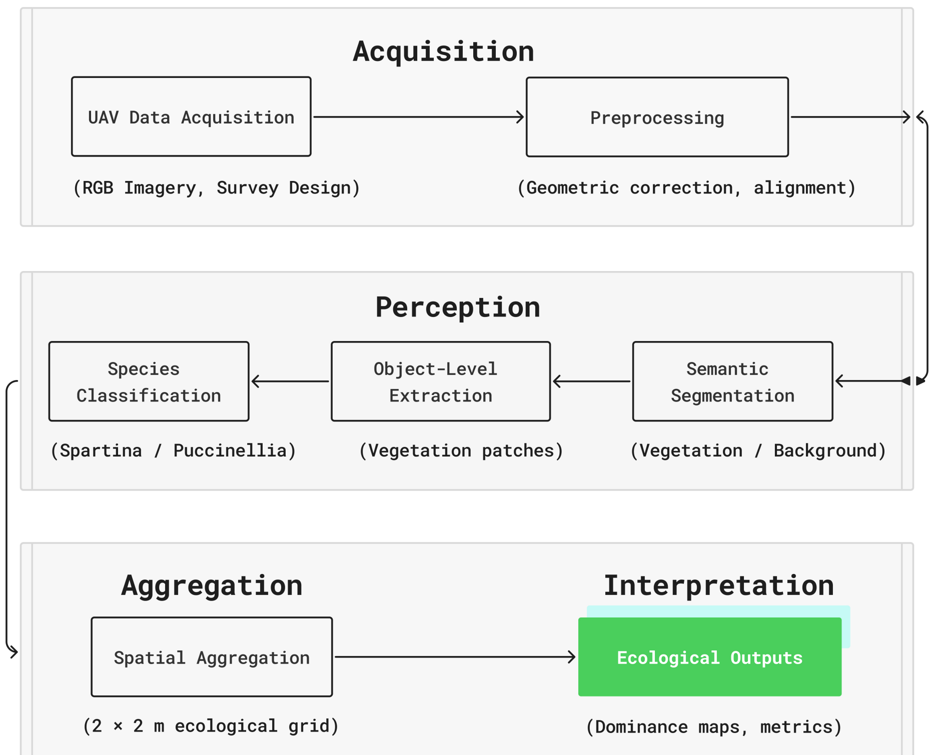

Figure 2: Conceptual architecture of the EcoVision framework. The system is organised as a modular, end-to-end pipeline. High-resolution UAV imagery is progressively transformed through perception, object-level interpretation, and spatial aggregation into ecologically meaningful outputs such as species presence and dominance.

- A transformer-based semantic segmentation model is then applied to isolate vegetation from background elements such as water, sediment, and exposed substrate, producing structured vegetation masks that capture spatial extent and patch boundaries (Xie et al., 2021).
- Connected component analysis is used to extract contiguous vegetation regions from these masks, transforming pixel-level predictions into discrete vegetation units (Haralick and Shapiro, 1992; Gonzalez and Woods, 2018).

- These regions are subsequently classified using a convolutional neural network to assign species-level labels, enabling discrimination between dominant marsh taxa and non-target classes. Confidence thresholds are applied to ensure only high-probability classifications are aggregated (Dash et al., 2017). To translate predictions into ecological indicators, classified vegetation units are aggregated within fixed 2 × 2 m spatial grids.
- Species dominance is computed as proportional areal coverage per grid cell, aligning model outputs with established ecological sampling scales.
- Final outputs are visualised as georeferenced maps and dominance heatmaps, supporting spatial analysis, monitoring, and management decisions.

The modular structure of the pipeline allows individual components to be upgraded without disrupting the overall framework.

3.1.1. Species of Interest (Spartina maritima, Puccinellia maritima)

(a) Spartina maritima (b) Puccinellia maritima

Figure 3: Spartina maritima and Puccinellia maritima with distinct morphological signatures

The study focused on dominant British salt marsh species, particularly Spartina maritima and Puccinellia maritima, selected due to their ecological importance in coastal zonation and distinct morphological signatures [25, 4]. Additional vegetation and non-vegetated classes were included to reduce false positives and improve discrimination in heterogeneous marsh mosaics.

3.2. Data Acquisition

UAV data were collected using a DJI Mavic 4 Pro equipped with a 20-megapixel, 1-inch CMOS RGB sensor and a 24 mm equivalent lens. This configuration provides a practical balance between spatial resolution, flight stability, and operational accessibility, without reliance on specialised multispectral payloads [11, 10]. The sensor provides sufficient spatial detail to resolve fine-scale vegetation morphology required for species-level segmentation and classification (reference). RGB imagery was captured in both RAW and JPEG formats, preserving spectral fidelity for analysis while maintaining efficient data handling for large survey volumes. Flights were executed at 10 m above ground level, yielding an ultra-high-resolution ground sampling distance of 0.5–1.0 cm per pixel[PL27.1]. Survey coverage targeted representative salt marsh zones, including pioneer, low-marsh, and mid-marsh areas, aligned with established ecological zonation frameworks [25, 4].

All UAV flights were conducted during low-tide windows to maximise visibility of vegetated patches and minimise tidal interference [16, 9]. Data collection was restricted to stable atmospheric conditions, with clear or uniform overcast skies preferred to reduce shadowing and illumination variability. Imagery was geotagged using onboard GPS and later refined during photogrammetric processing. Captured datasets were reserved exclusively for validation and analysis, ensuring independence from training data.

Model training data were sourced from global biodiversity repositories, including iNaturalist [26] and GBIF [27], using programmatic API access and species-specific filtering. Approximately 260 curated images per species were retained following manual quality control. UAV imagery was stored separately as independent validation data. All datasets followed consistent naming conventions and metadata standards to support traceability, reproducibility, and seamless integration with segmentation and classification pipelines. This two-tiered data strategy combined taxonomically diverse public imagery with site-specific UAV data used exclusively for validation [28].

3.3. Dataset Preparation and Annotation

The dataset comprised two complementary data sources: ultra-high-resolution UAV imagery and curated public biodiversity images. UAV data captured fine-scale salt marsh vegetation structure at 0.5–1.0 cm spatial resolution, preserving morphological detail critical for species-level analysis. Public images sourced from iNaturalist [26] and GBIF [27] provided taxonomic breadth and visual variability across growth stages, lighting conditions, and viewing angles, which is essential for ensuring model generalisability [2, 29]. Together, these sources supported model training while preserving ecological realism during validation.

All images were resized to 512 × 512 pixels to standardise spatial input for deep learning models [12]. Preprocessing prioritised preservation of vegetation structure over aggressive radiometric normalisation. To mitigate class imbalance and improve generalisation, probabilistic data augmentation was applied, including geometric transformations, rotations, brightness and contrast variation, and controlled noise injection [28].

3.3.1. Annotation Strategy and Labelling Protocol

Species-level annotations were produced manually using polygonal delineation in LabelMe [30]. Masks captured precise vegetation boundaries, as pixel-accurate segmentation is required to resolve overlapping canopies in salt marsh environments [10, 12]. Annotations were stored in JSON format and converted to binary and multi-class PNG masks via custom scripts. Strict naming conventions ensured traceability between images, masks, and metadata throughout the pipeline.

3.3.2. Data Quality Control and Validation

specify the total number of images in the dataset, which are then divided into three different sets

All images underwent manual quality screening to exclude blurred captures, degraded specimens, and ecologically irrelevant samples. The final dataset was split into training (70%), validation (15%), and testing (15%) subsets. UAV imagery was reserved exclusively for independent validation to ensure the model could generalise from taxonomic training data to aerial monitoring contexts [28, 10].

3.4. Data processing and classification

EcoVision addresses vegetation mapping as a hierarchical perception problem. At the pixel level, the task is formulated as multi-class semantic segmentation, separating background, Spartina maritima, and Puccinellia maritima [12, 31], Cunha and Baptista, 2025). At the object level, contiguous vegetation patches are treated as individual units for species classification. Finally, dominance estimation is framed as a spatial aggregation problem, where species-specific cover is quantified within ecologically meaningful grid cells. This formulation ensures consistency between computer vision outputs and traditional ecological survey metrics.

Given a patch P, the dominance score of species s was calculated as:

$$D_2P = \frac{A_s(P)}{\sum_{i=1}^{N} A_i(P)} \times 100 \quad (1)$$

Where:
$A_s(P)$ = total area (pixels or $\mathrm{m}^{\mathbf{2}}$) covered by species s in patch P,
N = total number of species considered (here, N=2: Spartina maritima and Puccinellia maritima),
$D_2(P)$ = dominance percentage of species s in patch P.

This formulation yields species-specific proportional cover within each patch, ensuring dominance is expressed as a standardised percentage. Such an approach is consistent with ecological survey conventions, where dominance is commonly derived from relative cover or biomass measures []

3.4.1. Segmentation Models

Semantic segmentation provides the spatial foundation of the pipeline by delineating vegetation boundaries and suppressing non-vegetated background elements. Transformer-based architectures were selected to capture both local texture and global spatial context inherent in heterogeneous salt marsh environments, effectively overcoming the receptive field limitations of traditional convolutional layers [24]. SegFormer-B5 uses a hierarchical transformer encoder with a lightweight MLP decoder, allowing high-resolution segmentation without explicit positional encoding, which is suitable for fine vegetation boundaries. Initialised with ImageNet-pretrained weights, the model was fine-tuned on 512 × 512 UAV imagery using cross-entropy loss with class weighting. This architecture was selected for its ability to preserve fine vegetation boundaries while maintaining robustness under variable illumination and canopy structure.

3.4.2. Classification Model

Species-level discrimination was performed using ConvNeXt, a convolutional network that incorporates design elements inspired by transformer

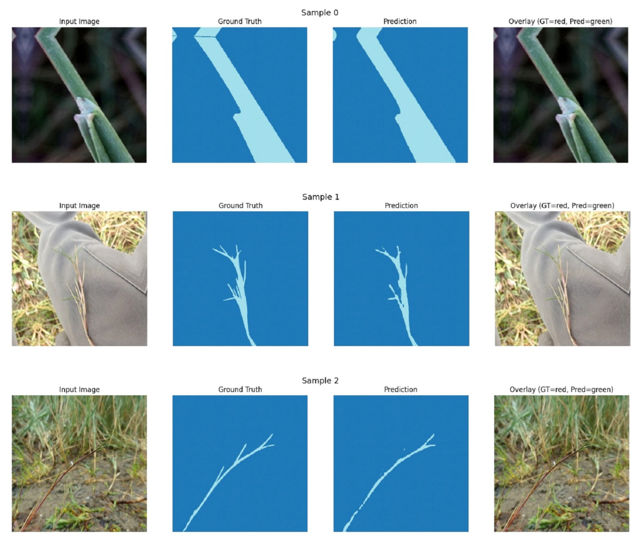

Figure 4: Caption

models, maintaining CNN computational efficiency while improving feature representation for classification [32]. Cropped vegetation blobs were classified into Spartina and Puccinellia classes. Pretrained ImageNet weights were fine-tuned to adapt feature representations to salt marsh morphology, improving performance under limited labelled data conditions.

Both segmenter and classifier models were trained using transfer learning. SegFormer employed a small batch size constrained by GPU memory, while ConvNeXt used larger batches for stable optimisation. To enhance convergence and mitigate overfitting, we used AdamW optimisation with linear learning-rate warm-up and mixed-precision training. Best-performing checkpoints were selected based on validation metrics.

3.4.3. Loss Functions and Optimisation

CrossEntropyLoss was used for both segmentation and classification tasks. For segmentation, class weighting mitigated the imbalance between dominant vegetation and background pixels [33]. AdamW optimisation ensured stable weight decay, improving generalisation across heterogeneous marsh conditions compared to standard stochastic gradient descent [34].

$$\mathcal{L}_{\text{WCE}} = -\sum_{c=1}^{C} w_c \, y_c \log \hat{p}_c \tag{2}$$

where: $C = 3$ is the number of classes (*Spartina maritima*, *Puccinellia maritima*, background), $y_c \in \{0, 1\}$ is the ground-truth one-hot label for class c, $\hat{p}_c$ is the predicted softmax probability for class c, and $w_c = \frac{1}{f_c} \Big/ \sum_{c'} \frac{1}{f_{c'}}$ is the inverse-frequency class weight, with f_c the pixel frequency of class c in the training set.

3.4.4. Inference and Post-Processing

During inference, vegetation blobs were extracted from segmentation masks using connected component analysis [35]. Small artefacts were filtered by area thresholding. Classified blobs were spatially aggregated within $2 \times 2m$ grid cells, and species dominance was computed as proportional areal coverage. Final outputs consisted of georeferenced dominance maps and species heatmaps suitable for ecological interpretation and long-term monitoring.

3.5. Experimental Setup

3.5.1. Hardware and Computational Resources

All experiments were conducted on a CUDA-enabled Nvidia GeForce RTX 4060 8GB GPU, which supported both transformer-based segmentation and high-throughput classification workloads. GPU memory constraints directly informed batch size selection for the segmentation model, while classification training leveraged larger batches for optimisation stability. CPU resources were used for preprocessing, blob extraction, and fallback inference where GPU access was unavailable. This mixed-resource setup reflects realistic deployment conditions for ecological research environments.

3.5.2. Software Stack and Frameworks

The pipeline was implemented in Python using PyTorch as the primary deep learning framework [36], with HuggingFace Transformers providing pretrained SegFormer [37, 24]. Torchvision was used for image transformations, normalization, and tensor handling. Training, validation, and checkpointing adhered to established PyTorch practices to maintain transparency and facilitate extensibility. All dependencies were selected for long-term support and reproducibility within the computer vision research community.

3.5.3. Training and Evaluation Environment

add all the mathematical equations that can help explaining the processing or evaluation metrics

Segmentation models were trained on UAV images resized to 512 × 512 pixels, employing mixed-precision (FP16) to conserve GPU memory and speed up training [38]. Classification models were trained on 224 × 224 vegetation patches extracted from segmentation outputs. Training employed AdamW optimisation with linear learning-rate scheduling and warm-up. Model selection was based on validation Intersection over union (IoU) for segmentation and accuracy–F1 convergence for classification. Evaluation was conducted under consistent preprocessing and metric definitions to avoid distributional bias.

$$\mathrm{IoU_c} = \frac{TP_c}{TP_c + FP_c + FN_c} \tag{3}$$

where
TP_c = true positive pixels,
FP_c = false positive pixels, and
FP_c‘= false negative pixels for class c.

3.5.4. Reproducibility Considerations

To support reproducibility, input resolutions were fixed, hyperparameters documented, and model checkpoints, including weights, class labels, and input dimensions, were serialized [39]. Pretrained initialised sources were documented, and deterministic preprocessing pipelines were enforced. These measures ensure that the EcoVision framework can be independently replicated, audited, and extended across ecological monitoring studies.

4. Results

Segmentation and classification modules are evaluated independently before analysing the integrated workflow. Quantitative performance metrics are complemented by visual inspection of model outputs to verify both computational accuracy and ecological plausibility.

4.1. Segmentation Performance Metrics

The SegFormer-B5 model demonstrated strong performance in delineating salt marsh vegetation from UAV imagery. Across the held-out test set, the model achieved a mean IoU(mIoU) of 0.557, pixel accuracy of 0.962, and a mean F1-score of 0.606 when evaluated on vegetation classes only. At the

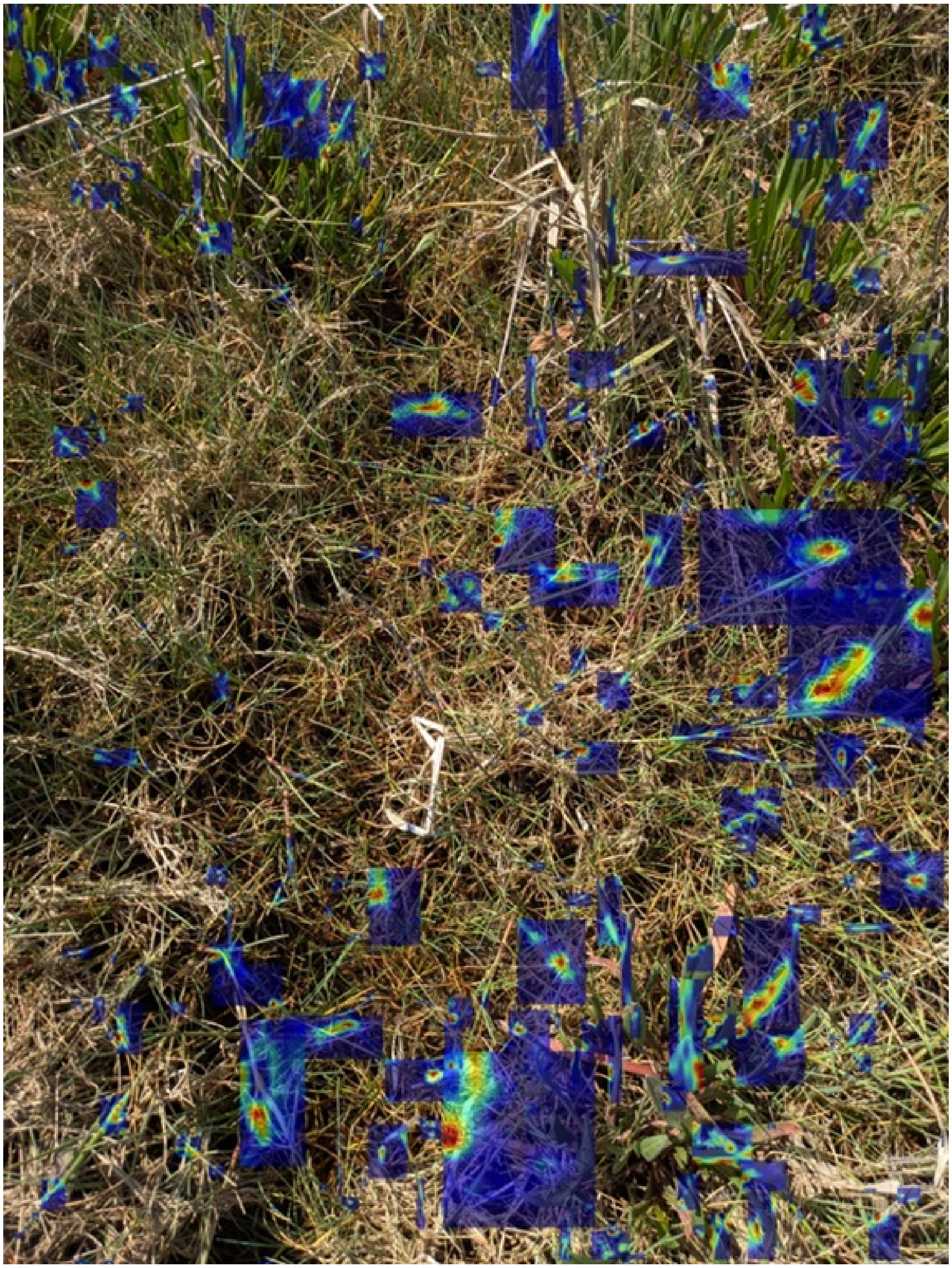

Figure 5: Caption

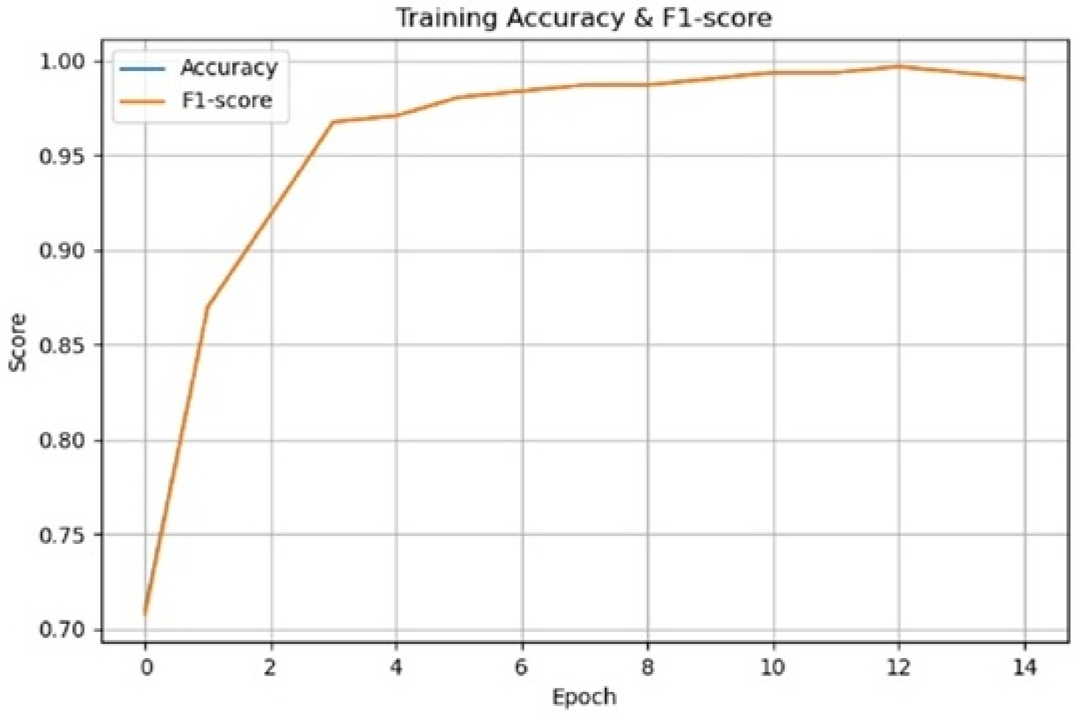

Figure 6: Caption

species level, Spartina maritima showed consistently higher segmentation accuracy, achieving an IoU of 0.904 and an F1-score of 0.950. Puccinellia maritima was segmented with an IoU of 0.767 and an F1-score of 0.868. These results indicate reliable discrimination between morphologically similar halophytic grasses, despite heterogeneous canopy structures and varying illumination conditions. Background pixels were excluded from performance reporting to prevent metric inflation due to class imbalance and to maintain ecological focus on vegetation detection, consistent with standard practices in UAV-based vegetation segmentation [10]. This approach aligns with established UAV-based ecological segmentation practices where non-vegetative classes are treated as ancillary.

4.2. Classification Accuracy and Confidence Analysis

The ConvNeXt-based classifier exhibited rapid convergence and high confidence in species discrimination. Final test accuracy reached 99.02%, with a corresponding F1-score of 0.990. Per-class accuracy was balanced, with

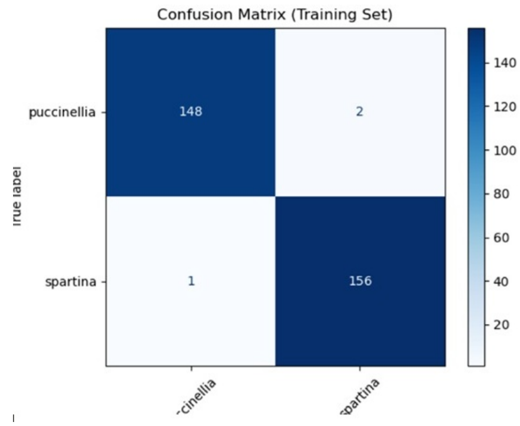

Figure 7: Caption

Spartina maritima classified at 99.4% accuracy and Puccinellia maritima at 98.7%. The confusion matrix was dominated by correct class predictions, with minimal off-diagonal errors, with only three misclassifications across 307 test samples. Prediction confidence remained high across illumination variations and patch-level heterogeneity, indicating that the classifier successfully captured discriminative morphological and textural features rather than overfitting to superficial cues. Such reliability is critical, as classification errors propagate directly into dominance estimation and downstream ecological interpretation [10, 12].

4.3. Comparative Model Evaluation

While this study did not conduct an exhaustive benchmark across multiple architectures, performance trends align with recent literature demonstrating the superiority of transformer-based segmentation and modern convolutional backbones in fine-grained vegetation analysis. SegFormer-B5 provided robust spatial generalisation compared to traditional CNN-based segmenters, particularly in handling complex canopy boundaries [24]. Similarly, ConvNeXt outperformed classical ResNet-style architectures reported in comparable ecological studies by maintaining high accuracy under limited training data [32]. Overall performance indicates that the chosen architectures provide a practical compromise between predictive accuracy, inference cost, and ecological interpretability.

4.4. Visual Results and Qualitative Analysis

Qualitative inspection of the segmentation outputs revealed strong alignment between the predicted masks and the ground truth annotations. Vegetation patches were consistently captured with correct species attribution, and large-scale spatial patterns were preserved across entire marsh sections. Boundary inaccuracies were primarily confined to thin vegetation edges, overlapping canopies, and transitional zones. However, these boundary errors were spatially limited and had a negligible impact on object extraction or downstream patch-level aggregation. Dominance maps generated from the integrated pipeline revealed ecologically plausible zonation patterns consistent with established salt marsh vegetation structure reported in ecological studies [40, 41], including contiguous Spartina-dominated stands and fragmented Puccinellia maritima distributions shaped by micro-topography [4, 25]. Overlay visualisations further confirmed that predicted dominance aligned with observed field patterns.

4.5. Error Analysis

Most segmentation errors occurred in areas with dense, interwoven vegetation or partial occlusion by shadows, resulting in boundary ambiguity. Over-segmentation occasionally occurred in regions with fine leaf structures, while under-segmentation was observed in low-contrast patches under diffuse lighting. Classification errors were rare and predominantly associated with small, fragmented blobs extracted from ambiguous segmentation regions. These errors were most common in mixed-species transition zones, where ecological boundaries are inherently diffuse [2, 12]. Importantly, spatial aggregation within the $2 \times 2m$ dominance grid reduced the impact of pixel and blob-level errors, preserving robustness at the ecological scale of interest (Yang et al. 2025). As a result, system-level dominance predictions remained stable despite localised inaccuracies.

5. Discussion

EcoVision demonstrates how UAV imagery combined with modern deep learning can shift vegetation monitoring toward quantitatively grounded and repeatable ecological assessment [](Dandois & Ellis, 2013; Ma et al., 2019). Quadrat-based field surveys remain methodologically robust, but their application is limited by labour demands, sparse spatial coverage, and observer-dependent variability [25](Elzinga et al., 2001). EcoVision offers a reproducible alternative, delivering comparable dominance estimates with substantially greater spatial coverage and consistency. UAV-derived assessments captured fine-scale heterogeneity often missed in manual sampling [42, 43], particularly in transitional or mixed-species zones. EcoVision is not intended as a replacement for field surveys, but as a complementary system that extends spatial coverage and analytical consistency, thereby extending observational reach and enabling repeatable landscape-scale analysis. Field data remain essential for calibration, validation, and ecological context, while automated pipelines reduce workload and increase temporal frequency. The discussion below situates the system's performance, limitations, and implications within current ecological and remote sensing practice.

5.1. Model's Strengths and Limitations

EcoVision's primary strength lies in its modular integration of transformer-based segmentation and high-capacity classification. SegFormer-B5 reliably captured spatial structure across heterogeneous salt marsh canopies, while ConvNeXt delivered near-perfect species discrimination, minimizing taxonomic error propagation into dominance metrics. The object-based aggregation strategy further reduced sensitivity to pixel-level noise, preserving robustness at ecologically relevant scales. However, limitations remain. RGB-only imagery constrains spectral separability under variable illumination and moisture conditions [](Adam et al., 2010; Aasen et al., 2018). Boundary precision errors persist in dense or interwoven vegetation, occasionally affecting fine-scale patch delineation. Model performance is also contingent on high-quality manual annotations, creating a scalability bottleneck as species diversity increases. Environmental variability, tidal inundation, shadowing, and phenological change introduce residual uncertainty that cannot be fully mitigated by architecture alone.

5.2. Ecological Implications

EcoVision enables consistent, high-resolution assessment of species dominance across salt marsh landscapes. The ability to quantify vegetation composition at 2×2 m resolution supports detection of subtle zonation shifts, invasive expansion fronts, and early-stage habitat fragmentation. Such granularity is rarely achievable through manual surveys at scale (Kent, 2012; Legendre & Legendre, 2012).

5.3. Scalability and Generalization

EcoVision's architecture is inherently scalable. The modular pipeline supports substitution of models, expansion to additional species, and deployment across broader geographic regions with minimal structural modification. Inference efficiency enables batch processing of large UAV datasets, making regional-scale monitoring feasible. Generalization beyond salt marsh ecosystems will require retraining on habitat-specific data and, potentially, integration of multispectral sensors (Aasen et al., 2018; Ma et al., 2019), to improve spectral discrimination. Nevertheless, the underlying framework, semantic segmentation, object-level classification, and dominance aggregation, remains transferable across vegetation-dominated landscapes. With expanded datasets and infrastructure, EcoVision provides a viable foundation for scalable, automated ecological monitoring.

6. Conclusion

This research presented EcoVision, a fully integrated UAV-based framework for species-level vegetation mapping and dominance estimation in salt marsh ecosystems. The system integrates transformer-based semantic segmentation, object-level species classification, and spatial dominance scoring into a single, coherent pipeline. By combining SegFormer-B5 for pixel-level delineation with ConvNeXt for fine-grained species discrimination, EcoVision delivers accurate, repeatable, and ecologically interpretable outputs. The introduction of grid-based dominance aggregation bridges computational predictions with established field survey practices, enabling quantitative assessment at management-relevant scales. Collectively, this work demonstrates that modern deep learning can be translated beyond visual interpretation to support quantitatively robust ecological measurement.

References

[1] E. B. Barbier, S. D. Hacker, C. Kennedy, E. W. Koch, A. C. Stier, B. R. Silliman, The value of estuarine and coastal ecosystem services, Ecological monographs 81 (2) (2011) 169–193.

[2] J. Wäldchen, P. Mäder, Plant species identification using computer vision techniques: A systematic literature review, Archives of Computational Methods in Engineering 25 (2) (2017) 507–543.

doi:10.1007/s11831-016-9206-z.
URL `https://doi.org/10.1007/s11831-016-9206-z`

[3] L. Lei, Q. Yang, L. Yang, T. Shen, R. Wang, C. Fu, Deep learning implementation of image segmentation in agricultural applications: a comprehensive review, Artificial Intelligence Review 57 (6) (2024). doi:10.1007/s10462-024-10775-6.
URL `https://doi.org/10.1007/s10462-024-10775-6`

[4] A. Bertacchi, V. Giannini, C. Di Franco, N. Silvestri, Using unmanned aerial vehicles for vegetation mapping and identification of botanical species in wetlands, Landscape and Ecological Engineering 15 (2) (2019) 231–240. doi:10.1007/s11355-018-00368-1.
URL `https://doi.org/10.1007/s11355-018-00368-1`

[5] C. Cruz, K. McGuinness, P. M. Perrin, J. O'Connell, J. R. Martin, J. Connolly, Improving the mapping of coastal invasive species using uav imagery and deep learning, International Journal of Remote Sensing 44 (18) (2023) 5713–5735. doi:10.1080/01431161.2023.2251186.
URL `https://doi.org/10.1080/01431161.2023.2251186`

[6] R. A. Stevens, S. Shull, J. Carter, E. Bishop, N. Herold, C. A. Riley, et al., Marsh migration and beyond: A scalable framework to assess tidal wetland resilience and support strategic management, PLOS ONE 18 (11) (2023) e0293177. doi:10.1371/journal.pone.0293177.

[7] C. S. Reddy, Remote sensing of biodiversity: what to measure and monitor from space to species?, Biodiversity and Conservation 30 (2021) 2617–2631. doi:10.1007/s10531-021-02216-5.

[8] Z. Yang, T. Blount, C. Lester, N. Blackford, A. D'Alpaos, M. Marani, B. Murray, S. Silvestri, Unmixing marsh vegetation species across multiple sensors and spatial scales, GIScience & Remote Sensing 62 (1) (2025). doi:10.1080/15481603.2025.2481689.
URL `https://doi.org/10.1080/15481603.2025.2481689`

[9] M. Chang, P. Li, Z. Li, H. Wang, Time series satellite remote sensing reveals annual distribution, composition, and trajectory of tidal wetlands in the yellow river delta, Estuarine Coastal and Shelf Science 319 (2025)

109264. doi:10.1016/j.ecss.2025.109264.
URL https://doi.org/10.1016/j.ecss.2025.109264

[10] T. Kattenborn, J. Eichel, F. E. Fassnacht, Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution uav imagery, Scientific Reports 9 (1) (2019) 17656. doi:10.1038/s41598-019-53797-9.
URL https://doi.org/10.1038/s41598-019-53797-9

[11] J. P. Dash, M. S. Watt, G. D. Pearse, M. Heaphy, H. S. Dungey, Assessing very high resolution uav imagery for monitoring forest health during a simulated disease outbreak, ISPRS Journal of Photogrammetry and Remote Sensing 131 (2017) 1–14. doi:10.1016/j.isprsjprs.2017.07.007.
URL https://doi.org/10.1016/j.isprsjprs.2017.07.007

[12] N. Gao, X. Du, P. Xu, E. Gao, Y. Yang, High-resolution mapping coastal wetland vegetation using frequency-augmented deep learning method, Remote Sensing 18 (2) (2026) 247. doi:10.3390/rs18020247.
URL https://doi.org/10.3390/rs18020247

[13] Joint Nature Conservation Committee, National Vegetation Classification: Users' Handbook, JNCC, accessed: 2024-01-01 (2006).
URL https://data.jncc.gov.uk/data/a407ebfc-2859-49cf-9710-1bde9c8e28c7/JNCC-NVC-UsersHandbook-2006.pdf

[14] S. Xie, Z. Tu, Holistically-nested edge detection, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1395–1403. doi:10.1109/iccv.2015.164.
URL https://doi.org/10.1109/iccv.2015.164

[15] M. Moreno, et al., Multi-camera machine learning for salt marsh species classification and mapping, Remote Sensing 17 (12) (2025) 1964. doi:10.3390/rs17121964.
URL https://doi.org/10.3390/rs17121964

[16] T. Deng, et al., Comparison of multi-class and fusion of multiple single-class segnet model for mapping karst wetland vegetation using uav images, Scientific Reports 12 (1) (2022) 13270. doi:10.1038/s41598-022-17620-2.
URL https://doi.org/10.1038/s41598-022-17620-2

[17] J. Iglhaut, et al., Structure from motion photogrammetry in forestry: a review, Current Forestry Reports 5 (3) (2019) 155–168. doi:10.1007/s40725-019-00094-3.
URL https://doi.org/10.1007/s40725-019-00094-3

[18] P. C. Gray, G. D. Larsen, D. W. Johnston, Drones address an observational blind spot for biological oceanography, Frontiers in Ecology and the Environment 20 (7) (2022) 413–421. doi:10.1002/fee.2472.
URL https://doi.org/10.1002/fee.2472

[19] R. Van Alphen, et al., Uav-based wetland monitoring: Multispectral and lidar fusion with random forest classification, Drones 8 (3) (2024) 113. doi:10.3390/drones8030113.
URL https://doi.org/10.3390/drones8030113

[20] M. Pöttker, K. Kiehl, T. Jarmer, D. Trautz, Convolutional neural network maps plant communities in semi-natural grasslands using multispectral unmanned aerial vehicle imagery, Remote Sensing 15 (7) (2023) 1945.
URL https://www.mdpi.com/2072-4292/15/7/1945

[21] D. Gautam, et al., Detection of invasive species (siam weed) using drone-based imaging and yolo deep learning model, Remote Sensing 17 (1) (2025) 120. doi:10.3390/rs17010120.
URL https://doi.org/10.3390/rs17010120

[22] G. R. Morgan, M. E. Hodgson, C. Wang, S. R. Schill, Deep learning of high-resolution aerial imagery for coastal marsh change detection: A comparative study, ISPRS International Journal of Geo-Information 11 (2) (2022) 100. doi:10.3390/ijgi11020100.
URL https://doi.org/10.3390/ijgi11020100

[23] A. Milioto, P. Lottes, C. Stachniss, Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns, in: IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 2229–2235. doi:10.1109/icra.2018.8460962.
URL https://doi.org/10.1109/icra.2018.8460962

[24] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo, Segformer: Simple and efficient design for semantic segmentation with transformers, arXiv preprint arXiv:2105.15203 (2021).
URL https://arxiv.org/abs/2105.15203

[25] M. Kent, Vegetation Description and Data Analysis: A Practical approach, 2nd Edition, John Wiley & Sons, Ltd, 2012.
URL https://download.e-bookshelf.de/download/0000/5957/22/L-G-0000595722-0002

[26] iNaturalist contributors, iNaturalist, iNaturalist Research-grade Observations, Available at: https://www.inaturalist.org, occurrence dataset (2023). doi:10.15468/ab3s5x.
URL https://www.inaturalist.org

[27] GBIF: The Global Biodiversity Information Facility, What is GBIF?, Available at: https://www.gbif.org/what-is-gbif, accessed: 17 May 2025 (2023).
URL https://www.gbif.org/what-is-gbif

[28] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6 (1) (2019). doi:10.1186/s40537-019-0197-0.
URL https://doi.org/10.1186/s40537-019-0197-0

[29] J. Carranza-Rojas, et al., Going deeper in the automated identification of herbarium specimens, BMC Evolutionary Biology 17 (1) (2017) 181. doi:10.1186/s12862-017-1014-z.
URL https://doi.org/10.1186/s12862-017-1014-z

[30] B. C. Russell, et al., Labelme: a database and web-based tool for image annotation, International Journal of Computer Vision 77 (1–3) (2007) 157–173. doi:10.1007/s11263-007-0090-8.
URL https://doi.org/10.1007/s11263-007-0090-8

[31] F. Santos, T. R. Cunha, P. Baptista, Deep learning-based semantic segmentation for automatic shoreline extraction in coastal video monitoring systems, Remote Sensing 17 (23) (2025) 3865. doi:10.3390/rs17233865.
URL https://doi.org/10.3390/rs17233865

[32] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference

on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 11976–11986.
URL https://openaccess.thecvf.com/content/CVPR2022/papers/Liu$_{AC}onvNet_{f}or_{t}he_{2}02$

[33] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440. doi:10.1109/cvpr.2015.7298965.
URL https://doi.org/10.1109/cvpr.2015.7298965

[34] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, arXiv preprint arXiv:1711.05101 (2017). doi:10.48550/arxiv.1711.05101.
URL https://doi.org/10.48550/arxiv.1711.05101

[35] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 4th Edition, Pearson Education Limited, 2018.
URL https://api.pageplace.de/preview/DT0400.9781292223070$_{A}37747583/preview-$ 9781292223070$_{A}$37747583.*pdf*

[36] A. Paszke, et al., Improving reproducibility in machine learning research, Journal of Machine Learning Research 22 (2021) 1–20.
URL https://jmlr2020.csail.mit.edu/papers/volume22/20-303/20-303.pdf

[37] Hugging Face, Transformers documentation, https://huggingface.co/docs/transformers/en/index, accessed: 2026-03-09 (2023).

[38] P. Micikevicius, et al., Mixed precision training, arXiv preprint arXiv:1710.03740 (2017). doi:10.48550/arxiv.1710.03740.
URL https://doi.org/10.48550/arxiv.1710.03740

[39] J. Pineau, P. Vincent-Lamarre, K. Sinha, V. Larivière, A. Beygelzimer, F. d'Alché Buc, E. Fox, H. Larochelle, Improving reproducibility in machine learning research: A report from the neurips 2019 reproducibility program, Journal of Machine Learning Research 22 (2021) 1–20.
URL https://dl.acm.org/doi/abs/10.5555/3546258.3546422

[40] P. Adam, Saltmarsh Ecology, Cambridge University Press, 1990. doi:10.1017/cbo9780511565328.
URL https://doi.org/10.1017/cbo9780511565328

[41] M. D. Bertness, Zonation of spartina patens and spartina alterniflora in new england salt marsh, Ecology 72 (1) (1991) 138–148. doi:10.2307/1938909.
URL https://doi.org/10.2307/1938909

[42] I. L. Turner, M. D. Harley, C. D. Drummond, Uavs for coastal surveying, Coastal Engineering 114 (2016) 19–24. doi:10.1016/j.coastaleng.2016.03.011.
URL https://doi.org/10.1016/j.coastaleng.2016.03.011

[43] K. Anderson, K. J. Gaston, Lightweight unmanned aerial vehicles will revolutionize spatial ecology, Frontiers in Ecology and the Environment 11 (3) (2013) 138–146. doi:10.1890/120150.
URL https://doi.org/10.1890/120150