

Article

Improving Ground Cover Crop Fractional Vegetation Mapping via Causality-Based Deep Representation Learning

Atif Latif ¹, Masoumeh Hashemi ^{2,3,*} , Matt Yost ^{2,*} , Somayeh Esmaeili ⁴  and Xiaojun Qi ¹

¹ Department of Computer Science, Utah State University, Logan, UT 84322, USA

² Plants, Soils & Climate Department, Utah State University, Logan UT 84322, USA

³ Biology Department, Brigham Young University, Provo, UT 84602, USA

⁴ Horticultural Science Department, Shahid Chamran University of Ahvaz, Ahvaz 61357-8315, Iran; s.esmaeili@scu.ac.ir

* Correspondence: masoumeh.hashemi@usu.edu (M.H.); matt.yost@usu.edu (M.Y.); Tel.: +1-385-470-7593 (M.H.); +1-435-764-1456 (M.Y.)

Abstract

Semantic segmentation and deep learning methods have rarely been applied to fractional vegetation cover (FVC) segmentation tasks due to the lack of publicly available datasets for training deep learning models. FVC is a key indicator for assessing vegetation distribution, crop density, and crop responses to water availability and fertilizer application, yet conventional field-based measurement methods are time consuming, costly, labor intensive, and may lack the accuracy required for critical applications such as drought stress evaluation and water productivity. In this paper, we introduced causality-based deep learning techniques for FVC segmentation on a publicly available RGB dataset that consists of four ground cover crops: *Phyla nodiflora* L., *Cynodon dactylon*, *Frankenia thymifolia* Desf., and *Oxalis stricta* L. By separating causal from spurious correlations in pretrained features, using the stepwise intervention and reweighting (SIR) method at different encoder stages reduced confounding bias and enabled the models to learn more generalizable and task-relevant features. Extensive experiments on the FVC dataset, conducted with and without causality learning, showed that the proposed FCN + ResNet-50 model with causality learning and data augmentation achieved an accuracy of 94.80%, a precision of 94.97%, a recall of 94.35%, and an F1-score of 94.62%, which outperformed non-causal baselines and state-of-the-art transformer-based models including SegFormer and Mask2Former.

Keywords: causal learning; fractional vegetation cover; semantic segmentation; ground cover plants



Academic Editor: Javier Del Ser Lorente

Received: 30 September 2025

Revised: 6 February 2026

Accepted: 7 February 2026

Published: 11 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Fractional vegetation cover (FVC) is a key indicator for assessing ecosystem health and vegetation density, as well as monitoring climate change [1]. It is defined as the proportion of ground area covered by the vertical projection of vegetation [2]. Estimated FVC is essential for modeling vegetation productivity and supporting research in land surface dynamics, dynamics climatology, hydrology, ecology, and agricultural resource management [3]. Precise FVC measurements can contribute to monitoring ecosystem health and facilitate sustainable natural resource management, both of which are critical for environmental conservation [4].

Conventional techniques for measuring FVC typically involve ground-based surveys and manual sampling techniques, such as point intercept methods, line transects, or

quadrat sampling, which are labor intensive and impractical for large-scale monitoring due to limited spatial coverage and sampling frequency [5,6]. These traditional methods often suffer from inadequate sample representativeness and spatial scale discrepancies, especially when compared to pixel resolution in RGB and remote sensing images, which impair their ability to effectively capture spatial heterogeneity (e.g., in drylands) or temporal changes in vegetation cover [7,8]. Moreover, some image-based conventional methods rely on vegetation indices (VIs) and pixel dichotomy models, which may encounter saturation and nonlinear response issues, limiting their accuracy for moderate to high vegetation density areas [9]. To overcome these limitations, semantic segmentation has emerged as an effective technique which divides an image into multiple segments, simplifying its representation into a more meaningful and analyzable form [10]. By classifying each pixel according to the object it belongs to, semantic segmentation produces a segmented image where every pixel is assigned a category label, thereby transforming raw visual data into a structured and interpretable format [11,12]. Semantic segmentation models are more susceptible to overfitting and tend to be less stable when used in different settings. To tackle this challenge, our approach includes causality-guided learning, where the stepwise intervention and reweighting (SIR) framework mitigates these confounding effects and enables the models to focus on task-relevant causal attributes for more reliable vegetation cover prediction [13].

Accurately determining the vegetation of ground cover crops from RGB images is difficult due to various challenges and crop types. Challenges include variations or differences in lighting conditions, lightning angles, and crops colors during the time of day, all of which can affect image quality considerably [14,15]. The four ground cover crops differ in leaf shape, density, and color, which makes vegetation recognition difficult. Some have small or sparse leaves that blend with the soil, while others have larger, denser foliage that are easier to segment but may overlap. Despite these challenges, precise estimation of vegetation is important for plant science research and botanists in order to enable more reliable monitoring of plant growth, health, crop water use estimation, irrigation stress, and development across different species, as well as the estimation of drought and water availability in regions [16,17].

In this study, we employ a causality-based deep learning approach for enhancing FVC segmentation and estimation using a high-resolution RGB dataset of four ground cover crops: *Phyla nodiflora* L., *Cynodon dactylon*, *Frankenia thymifolia* Desf., and *Oxalis stricta* L. Our approach leverages the stepwise intervention and reweighting (SIR) framework to mitigate confounding biases in pretrained features, thereby improving segmentation performance [13]. Furthermore, we evaluate multiple deep learning architectures to assess the effectiveness of causality learning for robust vegetation cover estimation.

2. Related Work

Over the past decade, several approaches have been developed to estimate FVC, ranging from traditional field-based surveys to modern remote sensing and deep learning-based techniques [18,19]. This section reviewed both traditional methodologies and semantic segmentation-based techniques for FVC, highlighting their strengths, limitations, and evolving roles.

2.1. Traditional and Remote Sensing FVC Estimation Techniques

Traditional methods for FVC estimation have predominantly relied on manual ground-based surveys that have involved visual estimation or direct measurement of vegetation cover using quadrats, line transects, or photographic sampling [20,21]. While these approaches can provide detailed and accurate point-based data, they were often labor in-

tensive, time consuming, prone to observer bias, and restricted in spatial extent, limiting their applicability at larger scales [20,22]. Given the logistical challenges of ground-based monitoring, integrating remote sensing technologies has become integral to achieving spatially comprehensive and accurate FVC measurements [23–25].

Remote sensing approaches have used satellite imagery, aerial photographs, unmanned aerial vehicle (UAV)-based sensors, and multispectral or hyperspectral data to estimate FVC by analyzing vegetation spectral responses [26,27]. These methods can afford substantial spatial coverage and frequent data acquisition but face inherent challenges. One major difficulty can lie in reconciling spatial scale discrepancies between high-resolution ground measurements and coarser satellite pixels, which often can result in scale mismatch errors and uncertainties in fractional cover estimation [28,29]. Additionally, landscapes with sparse, heterogeneous, or desert vegetation may pose challenges due to spectral mixing and low signal-to-noise ratios. In such settings, subtle signals from vegetation amid bare soil or non-vegetated areas can complicate accurate FVC retrieval [30,31].

2.2. Semantic Segmentation Techniques for Fractional Vegetation Cover

The advent of deep learning technologies has significantly advanced automated FVC estimation through improved image analysis capabilities. Deep learning models, particularly convolutional neural networks (CNNs) [8,31], have been effectively employed for semantic segmentation tasks where the goal has been to classify each pixel in an image according to vegetation presence or type. Semantic segmentation techniques have enabled direct mapping of canopy cover over large areas with high delineation precision, outperforming traditional threshold-based or spectral index methods. This advancement can automate labor-intensive segmentation processes, extract richer spatial features, and accommodate complex background vegetation interactions [32,33].

Integrating deep learning with multispectral and RGB imagery can further enrich feature representation by leveraging spectral and spatial information simultaneously. Such fusion can improve discrimination between vegetation and non-vegetation classes, particularly under variable lighting or seasonal conditions, and may reduce saturation effects commonly encountered in traditional vegetation indices [34]. In remote sensing of crops like maize (*Zea mays*), deep learning models utilizing multispectral data have provided superior performance in estimating FVC and related biophysical parameters compared to classical approaches [35]. Additionally, urban vegetation monitoring has benefited from combining deep semantic segmentation with RGB imagery to generate novel indices like the multiview semantic vegetation index (MSVI) [36], demonstrating robust and accurate estimation of FVC across color variations, viewpoints, and seasonal changes. In UAV imagery analysis, CNNs have demonstrated success in identifying plant communities and estimating FVC with considerable accuracy by harnessing the fine spatial detail inherent in aerial images [37,38]. Beyond standard CNNs, advanced semantic segmentation architectures such as U-Net [38], Deeplabv3+ [39], and ResNet [40] have been explored for vegetation classification. U-Net's encoder–decoder architecture with skip connections can effectively preserve fine-grained spatial details crucial for segmenting vegetation patches and has demonstrated superior accuracy in multispectral land cover classification tasks [41]. Deeplabv3+ employs atrous spatial pyramid pooling and a strong encoder–decoder design to capture multi-scale contextual information, improving segmentation of complex vegetation structures. ResNet, when used as a backbone for segmentation models, can leverage deep residual connections to facilitate training of very deep networks, enhancing feature extraction and representation learning. Collectively, these developments have marked a paradigm shift toward more accurate, automated, and scalable FVC mapping supported by sophisticated deep representation learning frameworks.

3. Materials and Methods

Previous research has primarily employed U-Net architecture and image processing techniques for FVC segmentation [31]. While effective, these models have remained susceptible to confounding factors and spurious correlations can be present in image data. In this study, we extended beyond the U-Net baseline by adopting a causality-guided stepwise intervention and reweighting (SIR) framework for semantic segmentation of high-resolution RGB images [13,40]. The SIR approach systematically intervened in feature learning and reweighted representations to prioritize causal over spurious associations.

3.1. Dataset

A total of 432 images were captured from four different groundcover plants, namely, *Phyla nodiflora* L., *Frankenia thymifolia* Desf., *Oxalis stricta* L., and *Cynodon dactylon*, using an LG G5 phone camera (LG Electronics, South Korea) to build the dataset. The images have spatial resolution of 2988×5312 pixels, with 344 used for training and 88 for testing. Due to limited dataset size, a separate validation set was not created, while a training set was used for hyperparameter tuning and optimization. The data were randomly sampled to ensure mutual independence between training and testing sets to enhance the robustness of the models. Each pixel in these images was labeled into two categories: vegetation and background. Figure 1 shows example images of the four different types of crops of FVC.

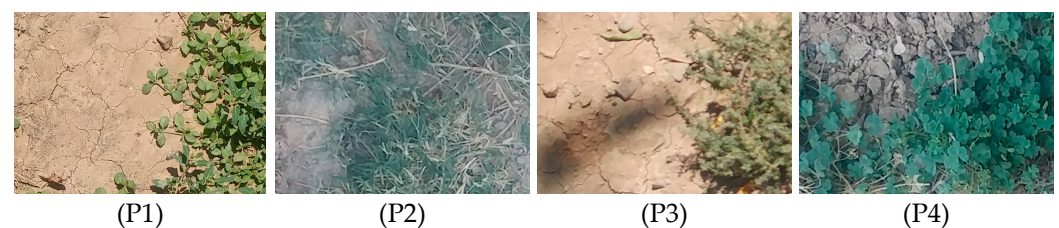


Figure 1. Example images from dataset. (P1) *Phyla nodiflora* L.; (P2) *Cynodon dactylon*; (P3) *Frankenia thymifolia* Desf.; and (P4) *Oxalis stricta* L.

The dataset was augmented using the Albumentations library [41], specifically using techniques such as horizontal and vertical flipping, rotation, occlusion, and color jitter, as well as random cropping, to generate additional variations in the original images as shown in Figure 2. This allowed each image to be transformed into three new versions, effectively tripling the dataset size and introducing meaningful variability that helped in model development and robustness.

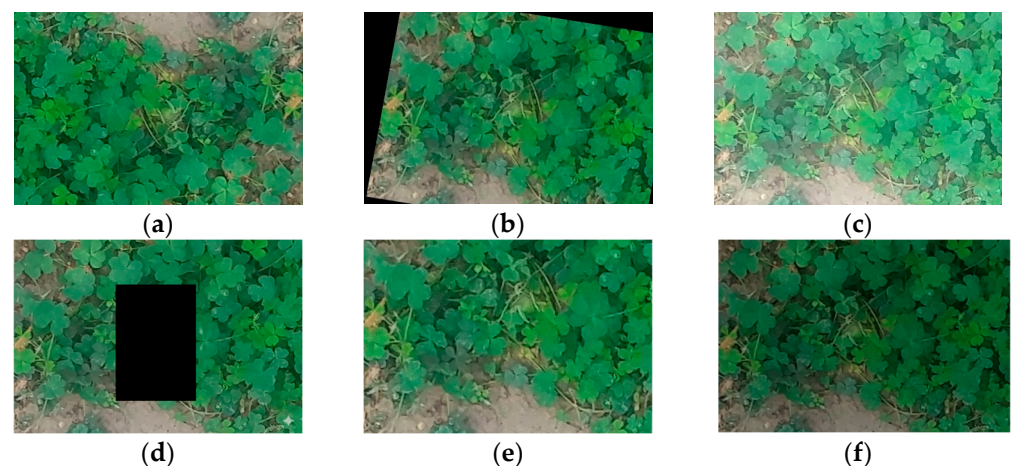


Figure 2. Example of images generated through data augmentation. (a) Flipping; (b) rotation; (c) brightness adjustment; (d) occluded; (e) cropping; (f) color jittering.

By applying these augmentations, the dataset captured more diverse spatial orientations and scale variations in objects, which was especially useful for segmentation tasks where small changes in perspective or framing can significantly affect feature recognition. The expanded dataset not only balanced class representation but also reduced overfitting, ensuring that the trained model became more robust and reliable in real-world applications.

To mitigate bias due to the limited dataset size, we employed repeated five-fold cross-validation with 10 repetitions. For each repetition, the dataset was reshuffled, and the model was trained and evaluated across all five folds. Fold-wise F1-scores were collected for statistical analysis.

3.2. Segmentation Models and Backbone Architectures

Each segmentation model can pair with different backbone architecture to achieve different segmentation results. Here we reviewed three widely used segmentation models (U-Net, FCN, and Deeplabv3+) and three common backbone architectures (ResNet50, VGG16, and EfficientNet).

3.2.1. Segmentation Models

U-Net

U-Net architecture [42] is a convolutional neural network designed specifically for semantic segmentation tasks. It follows an encoder–decoder structure, where the encoder progressively reduces spatial dimensions to capture contextual information, while the decoder restores spatial resolution through upsampling operations. A key feature of U-Net is the use of skip connections that transfer fine-grained spatial details from the encoder to the decoder, enabling precise boundary delineation even in complex images.

FCN

A fully convolutional network (FCN) [43] is a deep learning architecture designed for dense prediction tasks such as semantic segmentation, where each pixel of an image is classified into a category. Unlike traditional CNNs which end with fully connected layers for image-level classification, FCNs replace these layers with convolutional layers to preserve spatial information throughout the network. This design enables FCNs to process input images of varying sizes and produce output maps that align with the original image dimensions. By incorporating both upsampling layers and skip connections (as in FCN-8s and FCN-16s), FCNs can combine high-level semantic information with low-level spatial details, allowing for sharper segmentation boundaries.

Deeplabv3+

DeepLabv3+ model [44] is an advanced semantic segmentation architecture that builds upon DeepLabv3 by combining atrous (dilated) convolutions with an encoder–decoder structure. Its encoder leverages Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale contextual information, while the decoder refines object boundaries by recovering spatial details lost during downsampling. This design allows DeepLabv3+ to achieve strong performance in segmenting complex scenes with varied object sizes. It has been widely adopted in applications such as medical imaging, autonomous driving, and environmental monitoring, and is often used with backbones like ResNet, Xception, or EfficientNet for robust feature extraction.

SegFormer

SegFormer [45] is a transformer semantic segmentation model that combines a transformer encoder (hierarchical) with a lightweight multilayer perceptron (MLP) decoder, which results in multi-scale feature representations. SegFormer introduces a hierarchical en-

coder that does not rely on positional encoding, helping to eliminate the need for positional interpolation while avoiding performance degradation when the image resolutions in the test set differ from those in the training set. The multilayer perceptron (MLP) decoders fuse the features from multiple encoder stages, which effectively capture both local and global contextual information.

Mask2Former

Mask2Former [46] is a semantic segmentation model capable of addressing segmentation tasks such as pixel-level, instance-level, and semantic-level. It consists of a *Clude* masked attention block that extracts localized features by constraining cross-attention within predicted mask regions. This design enables a unified framework across different segmentation tasks, improving both efficiency and performance, reducing research effort by at least $3\times$, and outperforming the best segmentation architectures.

3.2.2. Backbone Architectures

ResNet50

Residual Network (ResNet) [47] is a deep convolutional neural network that introduces the concept of residual learning to address the problem of vanishing gradients in extremely deep models. Instead of learning direct mapping, ResNet uses skip connections that bypass one or more layers, allowing the network to learn residual functions. This design enables the training of networks with hundreds of layers while maintaining stable convergence and improving generalization.

EfficientNet

EfficientNet [48] is a network developed with the goal of balancing model accuracy and efficiency. Its design is based on a compound scaling method, which uniformly scales network depth, width, and input resolution using a set of fixed coefficients. This systematic scaling strategy allows EfficientNet models to achieve state-of-the-art performance while using significantly fewer parameters and FLOPs compared to traditional backbones like VGG16 or ResNet.

Due to this efficiency–accuracy trade-off, EfficientNet has become a widely adopted backbone in computer vision tasks such as image classification, object detection, and semantic segmentation. Variants from EfficientNet-B0 through to B7 (and later EfficientNetV2) provide a spectrum of options depending on computational resources, making it suitable for both mobile devices and large-scale server applications. In segmentation pipelines, EfficientNet is often employed as an encoder, where its lightweight yet expressive features improve model robustness without excessive computational cost.

VGG16

VGG16 network [49] is a deep CNN with 16 layers that relies on stacking small $3 \times 3 \times 3$ convolutional filters to extract hierarchical image features. It has gained prominence for its simplicity and strong performance on large-scale datasets like ImageNet, but its high parameter counts, and computational costs limit its efficiency compared to newer models. Despite these drawbacks, VGG16 remains widely used as a backbone for transfer learning and as an encoder in other segmentation architectures such as FCN and U-Net, where its structured feature maps provide a solid foundation for pixel-level prediction.

3.3. Causality-Guided Stepwise Intervention and Reweighting Method

In this work, we introduced a causality-guided stepwise intervention and reweighting (SIR) method for FVC, recognizing that features learned from pretrained models consisted of both causal factors and spurious correlations. Only the causal factors provided

generalizable information that remained consistent across different tasks and datasets. The SIR framework [13] mitigated confounding bias inherent in pretrained knowledge and enhanced the model's capacity to learn robust domain-invariant features. This enabled semantic segmentation of FVC images to benefit more effectively from pretraining. Figures 3 and 4 illustrate the semantic causal model for semantic segmentation and the framework of the segmentation model with causality-based reweighting, respectively. In a typical semantic segmentation network, the encoder is initialized from a pretrained model, so all intermediate feature representations are influenced by pretrained knowledge. Consequently, X_{pre} acts as a confounder between pairs of intermediate representations R_i and R_j . On the causal path $X \rightarrow R_1: m \rightarrow Y$, this creates non-causal dependencies such as $X \rightarrow R_i \leftarrow X_{pre} \rightarrow R_j \rightarrow Y$. In this way, the transferred knowledge introduces confounding bias into the learning process, affecting the flow of information from one encoder layer to the next. To address this, a balancing weight $W_{j-1,j} = \frac{P(L = l_p)}{P(L = l_p | R_{j-1})}$ is applied to correct the confounding effect of X_{pre} between adjacent representations. By reweighting as such, the distribution of features is adjusted toward a post-intervention distribution, reducing spurious influences and ensuring that the model can capture causal, task-relevant signals for more robust segmentation.

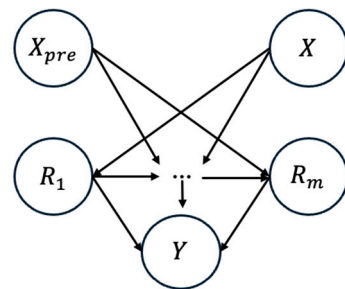


Figure 3. Semantic causal model for semantic segmentation.

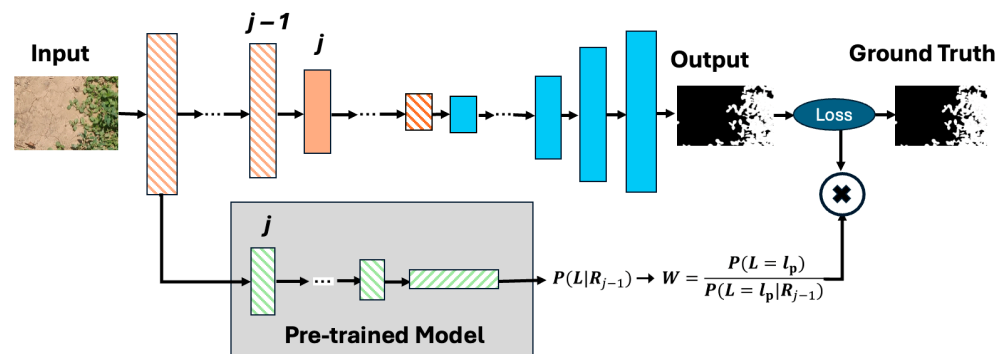


Figure 4. Framework of the segmentation model with causality-based reweighting, where intermediate features are adjusted using a pretrained model to reduce confounding effects.

Confounding bias can occur when an external factor, known as a confounder, influences both the input variables and the output, leading to misleading or spurious associations. In machine learning, particularly in deep learning for image segmentation, confounding bias can arise when pretrained features or irrelevant background information affect how the model learns the relationships between inputs and labels [49], as shown in Figure 5. These confounders can distort the true relationship between input data and output variables, and inadequate handling of these factors can often lead to misleading or biased associations. Traditional machine learning and statistical approaches can address this issue through techniques such as dataset matching, stratification, or residualizing imaging measurements [50].

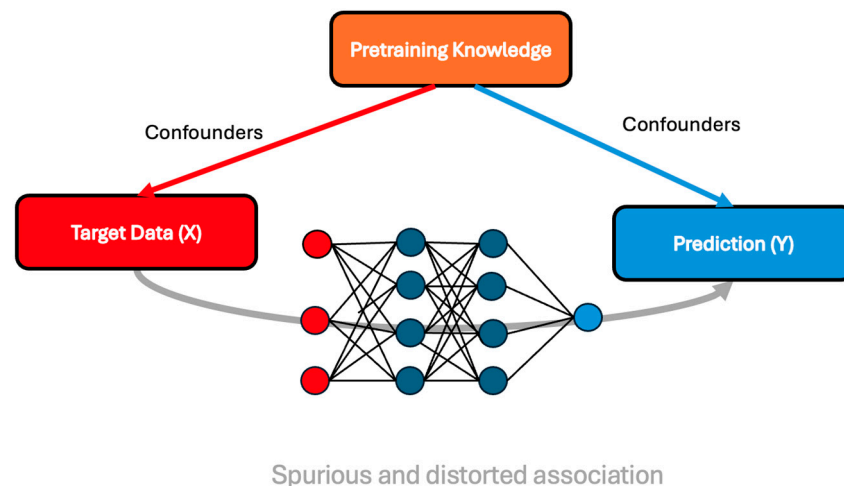


Figure 5. Illustration of confounding bias introduced by pretraining knowledge, leading to spurious and distorted associations between target data (X) and prediction (Y).

In our work, we addressed the issue of confounding bias by applying the SIR method [13], which explicitly reduced the influence of confounders that were introduced by pretraining or dataset specific artifacts. By encouraging the model to focus on causal and task-relevant features, SIR led to more reliable and generalizable segmentation performance, improving robustness to variations in unseen data and enhancing real-world applicability.

3.4. Stepwise Intervention and Reweighting (SIR) for Fractional Vegetation Cover

SIR is combined with semantic segmentation models which can work on dense pixel-level estimation rather than image-level prediction. The intervention and reweighting operations are applied to encoders of segmentation models where confounding bias introduced by pretrained features is most noticeable, which can help in the effective suppression of spurious representations.

Secondly, the reweighting strategy is applied to vegetation-background scenarios by focusing on task-specific feature distributions that are consistent across crop types and environmental conditions. This can encourage the model to focus on vegetation-related characteristics such as structural and spectral patterns, instead of background-specific textures.

Finally, SIR method is combined with segmented models trained on augmented data and evaluated on held-out test data, which can help to reduce overfitting and improve robustness across different conditions. These adaptations can enable a SIR-based framework to achieve improved generalization and reliability on the dataset of RGB images.

3.5. Evaluation Metrics for Performance Analysis

Evaluation metrics indicated how well the model performs across different classes, providing insights into its strengths and weaknesses for each class. In our study, we used accuracy, precision, recall, and F1-score to evaluate the effectiveness of a SIR-based causality method [13] and to assess the performance of state-of-the-art (SOTA) models on our dataset. Accuracy, in contrast, provided an overall measure of correctness by calculating the proportion of all correctly classified instances (both positives and negatives) out of the total number of predictions. Precision evaluated how accurate the model's positive predictions were by measuring the proportion of correctly predicted positives out of all predicted positives. It reflected the model's ability to minimize false positives. Recall, on the other hand, assessed the model's capability to detect all actual positive instances, showing how effectively it avoided missing positives (false negatives). The F1-score, calculated as the harmonic mean of precision and recall, provided a balanced metric that captured both

aspects. This measure was especially valuable in cases of imbalanced class distributions, as it highlighted the trade-off between precision and recall. Equations (1)–(4) represent accuracy, precision, recall, and F1-score, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

where TP indicates the total number of correct predictions; FP denotes the total number of false positives; TN represents the total number of true negatives; and FN indicates the total number of false negatives.

4. Results

This section reports the performance of the evaluated models on the FVC dataset. Detailed comparisons have been provided to highlight the effectiveness of different architectures and the improvements achieved through the proposed causality learning techniques.

4.1. Results of State-of-the-Art Segmentation Models

Table 1 summarizes the quantitative performance of representative CNN- and transformer-based segmentation models on the FVC dataset using accuracy, precision, recall, and F1-score. Among the baseline models without data augmentation, U-Net performed competitively by obtaining the highest recall of 93.04% and the highest F1-score of 91.98%, reflecting its strength in capturing fine-grained vegetation structures through skip connections. Mask2Former also achieved a strong overall balance, with the second-highest 91.56% F1-score and the highest precision score of 91.75%, indicating robust pixel-level discrimination. DeepLabv3+ and SegFormer showed comparable accuracy (above 91%), with DeepLabv3+ favoring higher precision and SegFormer slightly favoring recall, while FCN lagged behind across all metrics.

Table 1. Quantitative performance of CNN- and transformer-based segmentation models on the FVC dataset, with and without augmentation, using accuracy, precision, recall, and F1-score.

Model	Accuracy	Precision	Recall	F1
U-Net [42]	91.09%	90.95%	93.04%	91.98%
FCN [43]	90.10%	89.59%	90.66%	90.12%
DeepLabv3+ [44]	91.17%	91.63%	90.53%	91.08%
SegFormer [51]	91.42%	90.40%	91.50%	90.95%
Mask2Former [45]	91.37%	91.75%	91.37%	91.56%
U-Net [38] with Augmentation	91.07%	91.37%	92.48%	91.92%
FCN [39] with Augmentation	90.69%	89.32%	92.35%	90.81%
DeepLabv3+ [40] with Augmentation	91.58%	92.26%	90.71%	91.48%
SegFormer [46] with Augmentation	90.84%	90.48%	90.02%	90.25%
Mask2Former [47] with Augmentation	91.13%	91.19%	91.13%	91.11%

With data augmentation, most models exhibited modest but consistent improvements. DeepLabv3+ with augmentation achieved the highest accuracy (91.58%) and precision (92.26%), demonstrating improved robustness to appearance variations. U-Net with augmentation maintained strong recall (92.48%) and a high F1-score (91.92%), con-

firming its stability under enhanced data diversity. FCN benefited noticeably from augmentation in recall, though its overall performance remained lower than more advanced architectures. In contrast, SegFormer and Mask2Former showed marginal performance changes with augmentation, suggesting that their transformer-based designs already captured sufficient variability from the original data. In the following performance evaluation, we focused exclusively on CNN-based segmentation methods due to the higher computational cost and comparatively lower performance from these transformer-based approaches.

Overall, the results indicated that encoder–decoder CNNs such as U-Net and hybrid architectures like DeepLabv3+ remained highly effective for FVC segmentation, while transformer-based models achieved a competitive performance with a strong precision–recall balance. Data augmentation generally enhanced robustness, though its impact varied by architecture.

To assess model generalization across the four ground cover crops, namely, *Phyla nodiflora* L. (P1), *Cynodon dactylon* (P2), *Frankenia thymifolia* Desf. (P3), and *Oxalis stricta* L. (P4), we compared the performance of CNN-based segmentation models including U-Net, FCN, and DeepLabv3+ with data augmentation using accuracy, precision, recall, and F1-score, as shown in Figure 6. U-Net outperformed the other models, particularly for phyla and bermuda grass, achieving F1-scores of 96.63% for *phyla* and 96.40% for P2 (bermuda grass), demonstrating its strong ability to capture spatial features and preserve fine-grained vegetation boundaries. However, U-Net’s performance declined on *franken* and *oxalis*, with F1-scores dropping to 85.36% and 86.70%, respectively, indicating challenges in segmenting smaller or visually less distinct vegetation.

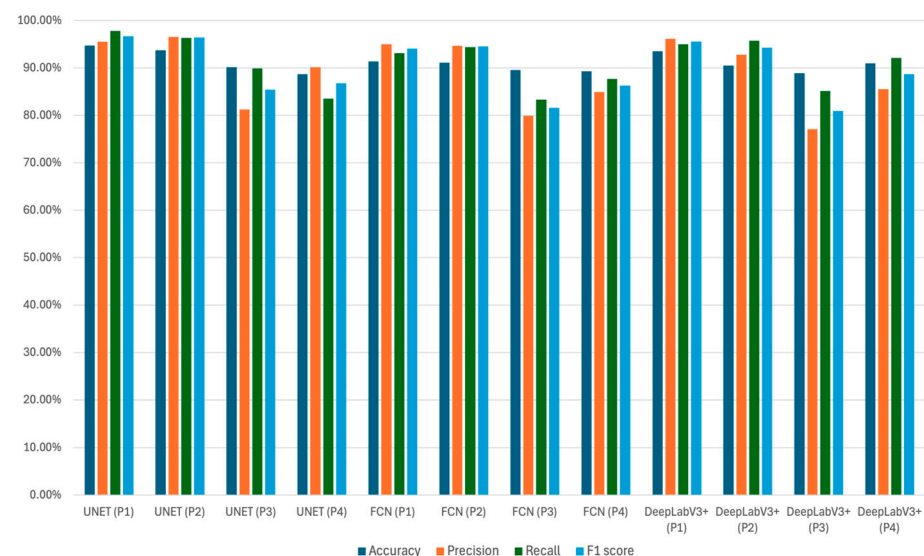


Figure 6. Performance comparison of SOTA CNN-based segmentation models (U-Net, FCN, DeepLabv3+) across different crop types (*Phyla nodiflora* L., *Cynodon dactylon*, *Frankenia thymifolia* Desf., and *Oxalis stricta* L.) with data augmentation.

FCN delivered competitive performance on the larger crops P1 (*phyla*) (94.02% F1), and P2 (bermuda grass) (94.52% F1), but struggled with P3 (*franken*) (81.54% F1) and P4 (*oxalis*) (86.29% F1), reflecting reduced adaptability to subtle crop textures. DeepLabv3+ showed balanced performance for P1 (*phyla*) (95.53% F1) and P2 (bermuda grass) (94.24% F1) and maintained relatively strong precision and recall, but like the other models, it experienced a notable performance drop for P3 (*franken*) (80.88% F1) and P4 (*oxalis*) (88.67% F1).

Overall, the results highlighted U-Net as the most robust architecture for multi-crop FVC segmentation, excelling especially for the dominant vegetation types of *phyla* and

bermuda grass. In contrast, FCN and DeepLabv3+ remained competitive for well-defined crops, but they showed reduced accuracy and F1-scores on smaller or visually ambiguous crops like frankenia and oxalis, underscoring the persistent challenge of segmenting less distinct vegetation.

4.2. Results of the Stepwise Intervention and Reweighting (SIR) Method

The application of the causality-guided SIR method yielded improvements across a range of backbone and segmentation architectures (Table 2). The results indicated that all tested combinations achieved strong performance with accuracy, precision, recall, and F1-scores consistently above 90%, except U-Net with EfficientNet and DeepLabV3+ with ResNet-50. Among these, the fully convolutional network with ResNet-50 achieved the best overall performance, reaching an F1-score of 94.62% and accuracy of 94.80%. These outcomes highlighted the effectiveness of the SIR framework in reducing confounding effects introduced by pretrained knowledge, enabling the models to learn more generalizable and task-relevant features.

Table 2. Performance comparison of CNN-based segmentation models (U-Net, FCN, and Deeplabv3+) with EfficientNet, ResNet-50, and VGG16 backbones using causality-based learning. Superscripts denote F1-score significance (*, ^{ns}) relative to FCN–ResNet-50.

Models	Backbone	Abbreviation	Accuracy	Precision	Recall	F1
U-Net	EfficientNet	U-EF	89.89%	91.80%	88.14%	89.21% ^{ns}
	ResNet-50	U-R50	92.86%	92.77%	92.54%	92.65% ^{ns}
	VGG16	U-VGG	91.12%	90.83%	90.97%	90.90% ^{ns}
FCN	EfficientNet	F-EF	91.90%	91.71%	91.60%	91.66% ^{ns}
	ResNet-50	F-R50	94.80%	94.97%	94.35%	94.62%
	VGG16	F-VGG	94.46%	94.27%	91.37%	94.32% ^{ns}
Deeplabv3+	EfficientNet	D-EF	91.20%	90.99%	90.88%	90.93% ^{ns}
	ResNet-50	D-R50	86.99%	87.83%	85.55%	86.28% [*]
	VGG16	D-VGG	91.68%	92.05%	90.88%	91.34% ^{ns}
SegFormer	MiT-B0	Seg-MiT	92.33%	92.48%	91.74%	92.05% ^{ns}
Mask2Former	Swin-tiny	Mask-Swin	93.58%	93.90%	93.77%	93.83% ^{ns}

The Friedman test showed statistically significant differences among the evaluated models ($\chi^2 = 67.684$, $p < 0.001$), indicating that at least one model exhibited different performance across the repeated cross-validation folds. To identify which models differed from the reference, pairwise comparisons were conducted using the Wilcoxon signed-rank test with Holm–Bonferroni correction ($\alpha = 0.05$), using FCN–ResNet-50 as the reference model. The results showed that only DeepLabV3+–ResNet-50 demonstrated a statistically significant difference compared to the reference model ($p_{\text{holm}} = 0.0195$), while all other comparisons were not statistically significant.

Performance varied across crop types. Models achieved the highest accuracy and F1-scores on phyla and bermudagrass, both of which presented clearer structural and spectral characteristics, while their performances were lower for P3 (*Frankenia thymifolia* Desf.) and P4 (*Oxalis stricta* L.), which exhibited finer textures and less distinct visual boundaries. For instance, DeepLabv3+ with VGG16 achieved F1-scores of 93.48% for P1 (*phyla*) and 92.29% for P2 (*bermudagrass*), but only 87.48% for P3 (*frankenian*) and 88.41% for P4 (*Oxalis stricta* L.). These results indicated that while the SIR method improved robustness across architectures, segmentation of visually subtle crops remains challenging (Figure 7). Overall, the experiments confirmed that incorporating SIR into pretrained models enhanced their capacity to extract causal features, improving both precision and recall while ensuring stronger generalization across diverse vegetation types.

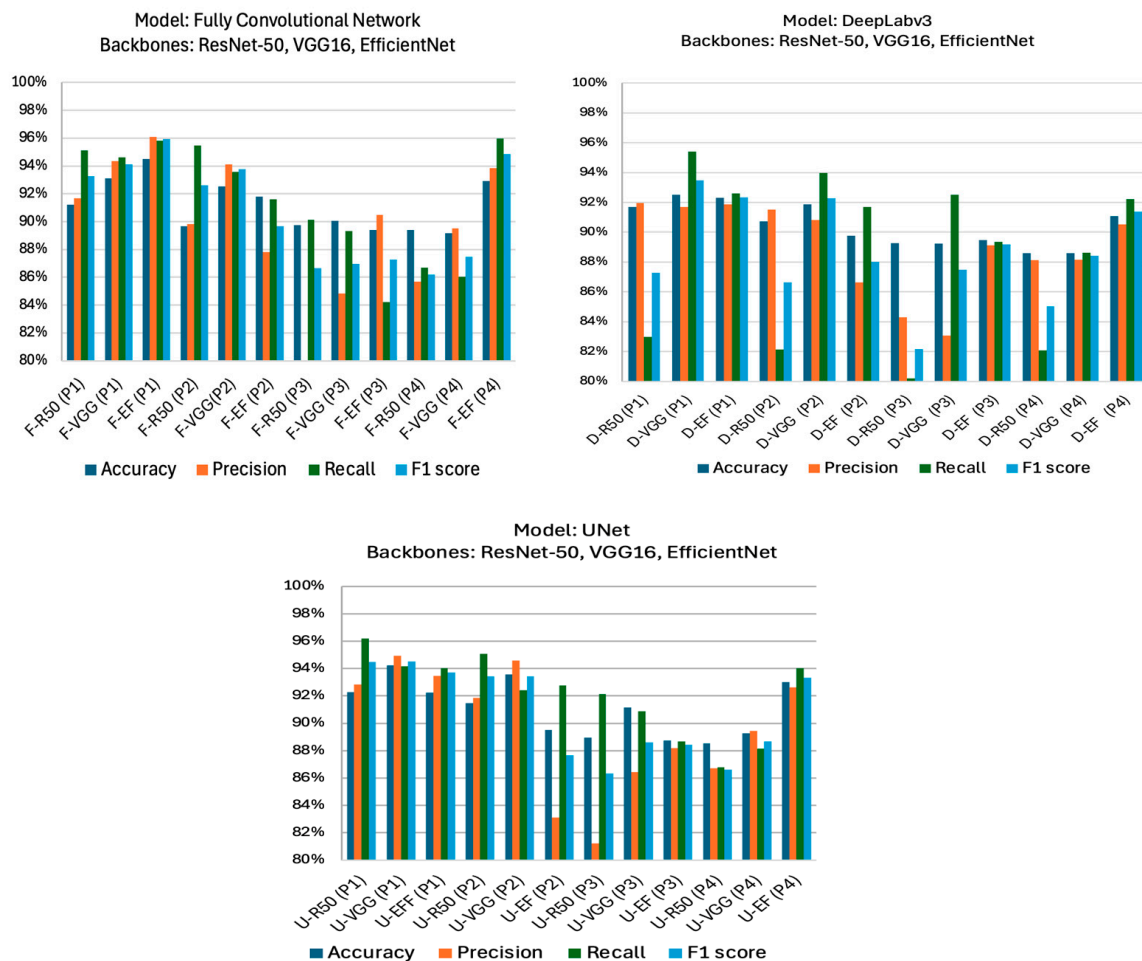


Figure 7. Performance comparison of FCN, U-Net, and DeepLabv3+ models with backbones (ResNet-50, VGG16, EfficientNet) across plant types (*Phyla nodiflora* L., *Cynodon dactylon*, *Frankenia thymifolia* Desf., and *Oxalis stricta* L.).

A comparison of the state-of-the-art (SOTA) models with those enhanced by the SIR framework showed that the proposed method improved results across all models. In the baseline models, U-Net achieved an F1-score of 91.98%, FCN reached 90.12%, and DeepLabv3+ obtained 91.08%, with modest gains observed when augmentation was applied. Under the SIR framework, FCN with ResNet-50 maintained an F1-score of 94.62% but showed improved stability across accuracy, precision, and recall. FCN with VGG16 achieved an F1-score of 94.32%, surpassing its baseline result of 90.12%. DeepLabv3+ also benefited from SIR, with EfficientNet and VGG16 backbones producing F1-scores around 91%, compared to 91.48% and 91.08% in the baseline models. Overall, SIR reduced variation between backbones, enhanced the balance between precision and recall, and demonstrated more robust performance than the baseline SOTA models, confirming its effectiveness in mitigating confounding effects from pretrained knowledge. Figure 8 shows the comparison of ground truth data to the segmentation outputs from U-Net and U-Net + VGG16.

Table 3 shows a comparison of semantic segmentation models in terms of computational complexity and runtime efficiency. The evaluation considered four key parameters: the number of trainable parameters (#Params), floating-point operations (FLOPs), memory usage (MB), and inference time (ms). The comparison included transformer-based architectures such as SegFormer and Mask2former and CNN-based models including U-Net, FCN, and DeepLabv3+ with different backbone networks.

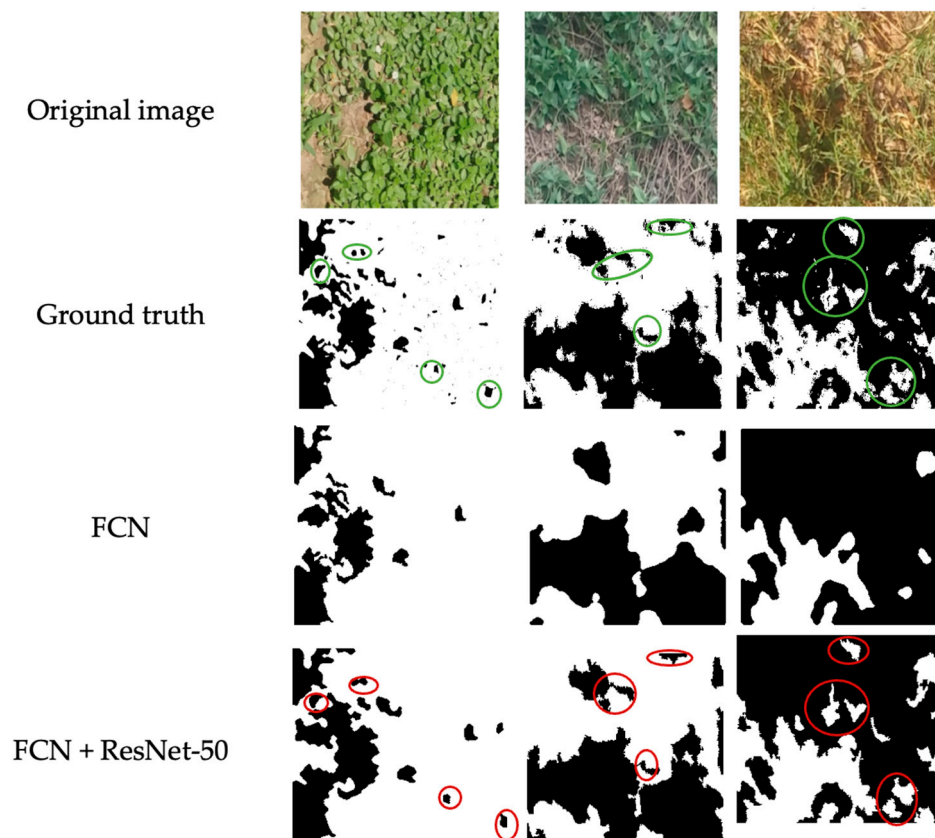


Figure 8. Comparison of segmentation results for baseline and SIR method.

Table 3. Computational and runtime comparison of semantic segmentation models.

Models	#Params (M)	FLOPs (G)	Inference Time (ms)	Memory (MB)
U-Net + EfficientNet-B0	6.25	2.54	20.2	23.85
U-Net + ResNet-50	32.52	10.72	30.9	124.06
U-Net + VGG16	23.75	24.76	34.7	90.59
FCN + EfficientNet-B0	5.76	2.03	20.4	21.97
FCN + ResNet-50	26.12	7.85	26.1	99.63
FCN + VGG16	16.80	22.44	31.6	64.08
DeepLabv3+ + EfficientNet-B0	4.91	2.7	26.3	18.72
DeepLabv3+ + ResNet-50	26.68	9.23	32.6	101.77
DeepLabv3+ + VGG16	19.5	24.0	32.50	74
SegFormer	27.37	62	85.6	106.12
Mask2former	43.98	190	220.4	167.78

5. Ablation Study of SIR Method

In the ablation study, we conducted a series of experiments to evaluate the performance of SIR framework at different stages of encoder block. Our main objective was to systematically evaluate the impact of SIR framework on performance of segmentation models we used in our study. The models were also evaluated to assess the performance of the segmentation models without the SIR framework; therefore, we experimented solely with the segmentation models as shown in Table 4. The results of SIR framework at different stages of encoder are presented in Table 5.

Table 4. Performance comparison of segmentation models (FCN, U-Net, Deeplabv3+) with EfficientNet, ResNet-50, and VGG16 backbones without the SIR framework.

Models	Backbone	Accuracy	Precision	Recall	F1
U-Net	EfficientNet	89.23%	87.63%	90.91%	88.7%
	ResNet-50	88.56%	85.27%	91.82%	88.47%
	VGG16	90.87%	90.74%	90.07%	90.31%
FCN	EfficientNet	90.12%	89.37%	91.04%	90.21%
	ResNet-50	88.67%	85.42%	91.53%	88.91%
	VGG16	89.79%	88.93%	90.13%	89.56%
Deeplabv3+	EfficientNet	89.94%	88.42%	90.52%	89.41%
	ResNet-50	88.72%	87.18%	81.82%	87.44%
	VGG16	89.21%	87.07%	90.31%	88.61%

Table 5. Performance of segmentation models with SIR at different encoder stages.

Model	Backbone	Encoder Stage	Accuracy	Precision	Recall	F1-Score
U-Net	EfficientNet	−4 only	90.02%	87.94%	92.41%	90.12%
		−3 only	90.44%	88.35%	93.02%	90.63%
		−4, −3	90.63%	88.62%	93.19%	90.88%
		−4, −3, −2	90.86%	88.87%	93.33%	91.05%
	ResNet-50	−4 only	89.67%	85.91%	94.01%	89.79%
		−3 only	89.94%	86.33%	94.45%	90.21%
		−4, −3	90.11%	86.61%	94.71%	90.43%
		−4, −3, −2	90.32%	86.89%	94.89%	90.71%
	VGG16	−4 only	91.26%	91.05%	90.87%	90.96%
		−3 only	91.71%	91.88%	91.23%	91.55%
		−4, −3	91.88%	92.07%	91.42%	91.74%
		−4, −3, −2	92.04%	92.31%	91.65%	91.98%
FCN	EfficientNet	−3 only	91.42%	90.88%	92.01%	91.44%
		−2 only	91.10%	90.35%	91.72%	91.03%
		−3, −2	91.68%	91.02%	92.24%	91.63%
		−4, −3, −2	91.99%	91.46%	92.57%	92.01%
	ResNet-50	−3, −2	89.74%	86.31%	93.89%	90.01%
		−4, −3, −2	90.01%	86.77%	94.34%	90.40%
		−3 only	90.62%	89.94%	91.05%	90.49%
		−2 only	90.31%	89.12%	90.77%	90.22%
	VGG16	−3, −2	90.79%	90.21%	91.33%	90.77%
		−4, −3, −2	91.13%	90.68%	91.61%	91.14%
		−3 only	90.61%	88.93%	92.12%	90.49%
		−2 only	90.78%	89.12%	92.33%	90.70%
DeepLabv3+	EfficientNet	−3, −2	90.94%	89.41%	92.68%	91.02%
		−4, −3, −2	91.15%	89.71%	92.91%	91.28%
		−3 only	89.82%	88.94%	81.02%	89.16%
		−2 only	90.12%	89.12%	81.45%	89.97%
	ResNet-50	−3, −2	90.36%	89.31%	81.71%	90.22%
		−4, −3, −2	90.57%	89.49%	81.86%	90.66%
		−3 only	90.24%	87.92%	93.11%	90.44%
		−2 only	90.39%	88.11%	93.29%	90.63%
	VGG16	−3, −2	90.61%	88.34%	93.55%	90.85%
		−4, −3, −2	90.85%	88.57%	93.87%	91.09%

We applied the SIR framework at different encoder stages, but our experiments indicated that the best performance resulted from combinations that were jointly applied at early (−4), intermediate (−3), and deep (−2) encoder stages. From these experiments, we also observed that SIR was effective when it was used in deeper encoder stages rather than

at initial stages. These observations suggested that using SIR at different stages performed better than applying it at a single encoder stage, although using SIR at only one stage still improved model performance when compared to not using SIR.

The experiments were conducted using different hyperparameters, such as batch sizes, learning rates, and numbers of epochs. The best performance was achieved with a batch size of eight, a learning rate of 1×10^{-4} , and 20 epochs. We also experimented with different augmentation rates, such as $2\times$, $3\times$, and $5\times$. Although the augmentation did not significantly affect performance, all models were trained using $3\times$ augmentation.

Challenges

One of the key challenges with this dataset was the high variability in lighting and shadows, which created inconsistency in plant appearance across images. The presence of strong shadows altered the texture and tone of vegetation, making segmentation more difficult for models that relied on color cues. This problem was further amplified in outdoor field conditions, where light intensity and direction can change throughout the day producing heterogeneous data.

Another major challenge was the color similarity between plants and soil, particularly for crops such as oxalis, where the plant exhibited a brownish or soil-like appearance under certain lighting conditions. This visual overlap caused confusion between background and vegetation classes, leading to misprediction. Additionally, small plant patches and dense overlapping vegetation increased intra-class variability, while some species shared similar leaf structures and colors, all of which made distinguishing them even more complex. Figure 9 illustrates these challenges along with the segmentation performance of different models. The regions highlighted in the U-Net + VGG16 results show areas where the model demonstrated better segmentation performance compared to the U-Net.

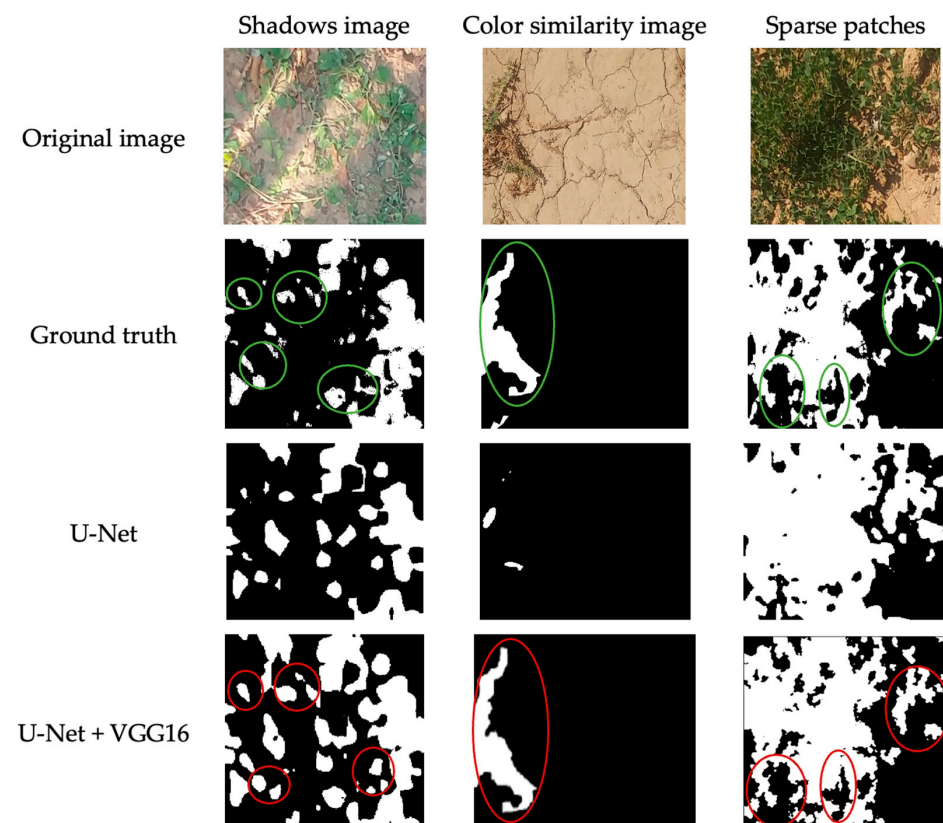


Figure 9. Example segmentation results under variable lighting, color similarity, and vegetation sparsity.

In response to these challenges, U-Net with VGG16 and FCN with EfficientNet demonstrated strong capacity to handle variability in lighting, shadows, and color similarity. The encoder–decoder design of U-Net with skip connections allowed it to preserve fine spatial details, improving discrimination between plants and soil even when color cues overlapped. FCN with EfficientNet benefited from efficient multi-scale feature extraction, which enhanced robustness against inconsistent lighting conditions and shadow effects.

6. Conclusions

In this work, we employed a causality-based deep learning approach to predict FVC with an RGB dataset of four ground cover crops: *Phyla nodiflora* L., *Cynodon dactylon*, *Frankenia thymifolia* Desf., and *Oxalis stricta* L. The technique employed in this study, which leveraged the SIR framework [13], effectively mitigated confounding biases in pretrained features and enhanced model generalization and segmentation performance. Among the evaluated architectures, the fully convolutional network model with ResNet-50 and causality learning achieved superior results with an accuracy of 94.80%, precision of 94.97%, recall of 94.35%, and an F1-score of 94.62%, outperforming other non-causal baseline models.

Overall, this study demonstrated the effectiveness of incorporating causality-guided learning into semantic segmentation for vegetation analysis. By applying the SIR framework [13], the models were able to focus on causal and task-relevant features, reducing the influence of spurious correlations introduced by pretrained knowledge. This resulted in more robust and generalizable segmentation outcomes across multiple architectures. The findings highlighted the potential of causality-based approaches to improve environmental monitoring applications such as crop health assessment, ecosystem management, and precision agriculture. Future work may extend this framework to larger and more diverse datasets, as well as integrate additional modalities such as multispectral or hyperspectral imagery to further enhance vegetation cover estimation.

Beyond the application to estimating vegetation for individual crop types in RGB images, the SIR method can be applied for broader ecological and environmental monitoring tasks. For instance, this approach could be applied to aerial or satellite imagery to assess regional vegetation cover, detect changes in forest density, and monitor the effects of deforestation. By applying automated and high-precision analysis of data over large areas, such methods could support timely identification of environmental risks and facilitate decision-makers about conservation and land management. Our current study was limited and only focused on small-scale RGB dataset, but the results could suggest that the SIR framework can be applied to larger-scale applications.

Our study has the limitation of using a relatively small dataset collected from a specific location and device, which may restrict generalization across different environmental conditions and locations. In future work, we plan to capture images using multiple cameras under diverse conditions to increase data variability. This will help optimize the models for improved robustness and more consistent performance across varying scenarios.

Author Contributions: Conceptualization, A.L., M.H., M.Y., S.E., and X.Q.; methodology, A.L., M.H., M.Y., and X.Q.; software, A.L.; validation, A.L., M.H., M.Y., and X.Q.; formal analysis, A.L., M.H., M.Y., and X.Q.; investigation, M.H., M.Y., and X.Q.; data curation, S.E. and M.H.; writing—original draft preparation, A.L.; writing—review and editing, A.L., M.H., M.Y., S.E., and X.Q.; visualization, A.L., and M.H.; supervision, M.H., M.Y., and X.Q.; funding acquisition, M.H., and M.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Tehran Municipality, Iran, in year 2020 (TMZ2020193).

Data Availability Statement: The original data presented in the study are openly available in Kaggle (<https://www.kaggle.com/datasets/masoumehhashemi/images-of-cover-crops-to-estimate-vegetation> (accessed on 29 September 2025)).

Acknowledgments: During the preparation of this manuscript/study, the authors used Grammarly (accessed September 2025), research rabbit (accessed July 2025), connecting papers (accessed July 2025), for the purposes of segmentation models selection and related articles, rephrasing and removing grammatical mistakes. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Andreatta, D.; Gianelle, D.; Scotton, M.; Dalponte, M. Estimating Grassland Vegetation Cover with Remote Sensing: A Comparison between Landsat-8, Sentinel-2 and PlanetScope Imagery. *Ecol. Indic.* **2022**, *141*, 109102. [CrossRef]
2. Liu, Y.; Mu, X.; Wang, H.; Yan, G. A Novel Method for Extracting Green Fractional Vegetation Cover from Digital Images. *J. Veg. Sci.* **2012**, *23*, 406–418. [CrossRef]
3. Chen, X.; Lv, X.; Ma, L.; Chen, A.; Zhang, Q.; Zhang, Z. Optimization and Validation of Hyperspectral Estimation Capability of Cotton Leaf Nitrogen Based on SPA and RF. *Remote Sens.* **2022**, *14*, 5201. [CrossRef]
4. Li, Y.; Sun, J.; Wang, M.; Guo, J.; Wei, X.; Shukla, M.K.; Qi, Y. Spatiotemporal Variation of Fractional Vegetation Cover and Its Response to Climate Change and Topography Characteristics in Shaanxi Province, China. *Appl. Sci.* **2023**, *13*, 11532. [CrossRef]
5. Karl, J.W.; McCord, S.E.; Hadley, B.C. A Comparison of Cover Calculation Techniques for Relating Point-Intercept Vegetation Sampling to Remote Sensing Imagery. *Ecol. Indic.* **2017**, *73*, 156–165. [CrossRef]
6. Song, W.; Mu, X.; Ruan, G.; Gao, Z.; Li, L.; Yan, G. Estimating Fractional Vegetation Cover and the Vegetation Index of Bare Soil and Highly Dense Vegetation with a Physically Based Method. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *58*, 168–176. [CrossRef]
7. Chen, X.; Sun, Y.; Qin, X.; Cai, J.; Cai, M.; Hou, X.; Yang, K.; Zhang, H. Assessing the Potential of UAV for Large-Scale Fractional Vegetation Cover Mapping with Satellite Data and Machine Learning. *Remote Sens.* **2024**, *16*, 3587. [CrossRef]
8. Wang, H.; Gui, D.; Liu, Q.; Feng, X.; Qu, J.; Zhao, J.; Wang, G.; Wei, G. Vegetation Coverage Precisely Extracting and Driving Factors Analysis in Drylands. *Ecol. Inform.* **2024**, *79*, 102409. [CrossRef]
9. Yao, Y.; Shen, J.; Yue, J.; Liu, Y.; Feng, H.; Shu, M.; Fu, Y.; Qiao, H.; Sun, T.; Zheng, G. Remotely Estimate the Cropland Fractional Vegetation Cover Using Linear Spectral Mixture Analysis and Improved Band Operations. *Int. J. Remote Sens.* **2025**, *46*, 3466–3486. [CrossRef]
10. Khan, S.; Jan, A.; Seo, K. Accurate Structural Crack Detection Using NestedUNet from Drone and Handheld Camera Images. *KSCE J. Civ. Eng.* **2025**, *29*, 100204. [CrossRef]
11. Csurka, G.; Volpi, R.; Chidlovskii, B. Semantic Image Segmentation: Two Decades of Research. *arXiv* **2023**. [CrossRef]
12. Khan, S.; Jan, A.; Seo, S. Structural Crack Detection Using Deep Learning: An in-Depth Review. *Korean J. Remote Sens.* **2023**, *39*, 371–393.
13. Shi, S.; Li, B.; Zhang, L.; Kuang, K.; Wu, S.; Feng, T.; Yan, Y.; Du, Z. Causality-Guided Stepwise Intervention and Reweighting for Remote Sensing Image Semantic Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5633517. [CrossRef]
14. Milioto, A.; Lottes, P.; Stachniss, C. Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 2229–2235.
15. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
16. Haddadi, S.R.; Hashemi, M.; Peralta, R.C.; Soltani, M. Visible Image-Based Machine Learning for Identifying Abiotic Stress in Sugar Beet Crops. *Algorithms* **2025**, *18*, 680. [CrossRef]
17. Li, W.; Niu, Z.; Chen, H.; Li, D.; Wu, M.; Zhao, W. Remote Estimation of Canopy Height and Aboveground Biomass of Maize Using High-Resolution Stereo Images from a Low-Cost Unmanned Aerial Vehicle System. *Ecol. Indic.* **2016**, *67*, 637–648. [CrossRef]
18. Wang, Z.; Song, D.-X.; He, T.; Lu, J.; Wang, C.; Zhong, D. Developing Spatial and Temporal Continuous Fractional Vegetation Cover Based on Landsat and Sentinel-2 Data with a Deep Learning Approach. *Remote Sens.* **2023**, *15*, 2948. [CrossRef]
19. Yue, J.; Tian, Q. Estimating Fractional Cover of Crop, Crop Residue, and Soil in Cropland Using Broadband Remote Sensing Data and Machine Learning. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102089. [CrossRef]
20. Chen, J.; Zhao, X.; Zhang, H.; Qin, Y.; Yi, S. Evaluation of the Accuracy of the Field Quadrat Survey of Alpine Grassland Fractional Vegetation Cover Based on the Satellite Remote Sensing Pixel Scale. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 497. [CrossRef]
21. Chen, J.; Yi, S.; Qin, Y.; Wang, X. Improving Estimates of Fractional Vegetation Cover Based on UAV in Alpine Grassland on the Qinghai–Tibetan Plateau. *Int. J. Remote Sens.* **2016**, *37*, 1922–1936. [CrossRef]

22. Henrys, P.A.; Jarvis, S.G. Integration of Ground Survey and Remote Sensing Derived Data: Producing Robust Indicators of Habitat Extent and Condition. *Ecol. Evol.* **2019**, *9*, 8104–8112. [[CrossRef](#)] [[PubMed](#)]
23. Fernández-Guisuraga, J.M.; Verrelst, J.; Calvo, L.; Suárez-Seoane, S. Hybrid Inversion of Radiative Transfer Models Based on High Spatial Resolution Satellite Reflectance Data Improves Fractional Vegetation Cover Retrieval in Heterogeneous Ecological Systems after Fire. *Remote Sens. Environ.* **2021**, *255*, 112304. [[CrossRef](#)]
24. Li, L.; Mu, X.; Jiang, H.; Chianucci, F.; Hu, R.; Song, W.; Qi, J.; Liu, S.; Zhou, J.; Chen, L.; et al. Review of Ground and Aerial Methods for Vegetation Cover Fraction (fCover) and Related Quantities Estimation: Definitions, Advances, Challenges, and Future Perspectives. *ISPRS J. Photogramm. Remote Sens.* **2023**, *199*, 133–156. [[CrossRef](#)]
25. Lin, X.; Chen, J.; Lou, P.; Yi, S.; Qin, Y.; You, H.; Han, X. Improving the Estimation of Alpine Grassland Fractional Vegetation Cover Using Optimized Algorithms and Multi-Dimensional Features. *Plant Methods* **2021**, *17*, 96. [[CrossRef](#)]
26. Xie, Y.; Sha, Z.; Yu, M. Remote Sensing Imagery in Vegetation Mapping: A Review. *J. Plant Ecol.* **2008**, *1*, 9–23. [[CrossRef](#)]
27. Yan, K.; Gao, S.; Yan, G.; Ma, X.; Chen, X.; Zhu, P.; Li, J.; Gao, S.; Gastellu-Etchegorry, J.-P.; Myneni, R.B.; et al. A Global Systematic Review of the Remote Sensing Vegetation Indices. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *139*, 104560. [[CrossRef](#)]
28. Kim, J.; Kang, S.; Seo, B.; Narantsetseg, A.; Han, Y. Estimating Fractional Green Vegetation Cover of Mongolian Grasslands Using Digital Camera Images and MODIS Satellite Vegetation Indices. *GIScience Remote Sens.* **2020**, *57*, 49–59. [[CrossRef](#)]
29. Gao, L.; Wang, X.; Johnson, B.A.; Tian, Q.; Wang, Y.; Verrelst, J.; Mu, X.; Gu, X. Remote Sensing Algorithms for Estimation of Fractional Vegetation Cover Using Pure Vegetation Index Values: A Review. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 364–377. [[CrossRef](#)] [[PubMed](#)]
30. Han, J.; Zhu, J.; Cao, X.; Xi, L.; Qi, Z.; Li, Y.; Wang, X.; Zou, J. Extraction of Sparse Vegetation Cover in Deserts Based on UAV Remote Sensing. *Remote Sens.* **2025**, *17*, 2665. [[CrossRef](#)]
31. Sotomayor, L.N.; Lucieer, A.; Turner, D.; Lewis, M.; Kattenborn, T. Mapping Fractional Vegetation Cover in UAS RGB and Multispectral Imagery in Semi-Arid Australian Ecosystems Using CNN-Based Semantic Segmentation. *Landsc. Ecol.* **2025**, *40*, 169. [[CrossRef](#)]
32. Zhang, Y.; Yang, R.; Dai, Q.; Zhao, Y.; Xu, W.; Wang, J.; Wang, L. Boosting Semantic Segmentation of Remote Sensing Images by Introducing Edge Extraction Network and Spectral Indices. *Remote Sens.* **2023**, *15*, 5148. [[CrossRef](#)]
33. Said, Y.; Saidani, O.; Algarni, A.D.; Algarni, M.H.; Flah, A. Lightweight Multiscale Information Aggregation Network for Land Cover Land Use Semantic Segmentation from Remote Sensing Images. *Sci. Rep.* **2025**, *15*, 30265. [[CrossRef](#)]
34. Hu, J.; Feng, H.; Wang, Q.; Shen, J.; Wang, J.; Liu, Y.; Feng, H.; Yang, H.; Guo, W.; Qiao, H.; et al. Pretrained Deep Learning Networks and Multispectral Imagery Enhance Maize LCC, FVC, and Maturity Estimation. *Remote Sens.* **2024**, *16*, 784. [[CrossRef](#)]
35. Waqar, R.; Grbović, Ž.; Khan, M.; Pajević, N.; Stefanović, D.; Filipović, V.; Panić, M.; Djuric, N. End-to-End Deep Learning Models for Gap Identification in Maize Fields. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 17 June 2024; pp. 5403–5412.
36. Khan, A.; Asim, W.; Ulhaq, A.; Robinson, R.W. A Multiview Semantic Vegetation Index for Robust Estimation of Urban Vegetation Cover. *Remote Sens.* **2022**, *14*, 228. [[CrossRef](#)]
37. Sa, I.; Popović, M.; Khanna, R.; Chen, Z.; Lottes, P.; Liebisch, F.; Nieto, J.; Stachniss, C.; Walter, A.; Siegwart, R. WeedMap: A Large-Scale Semantic Weed Mapping Framework Using Aerial Multispectral Imaging and Deep Neural Network for Precision Farming. *Remote Sens.* **2018**, *10*, 1423. [[CrossRef](#)]
38. Celikkan, E.; Kunzmann, T.; Yeskaliyev, Y.; Itzerott, S.; Klein, N.; Herold, M. WeedsGalore: A Multispectral and Multitemporal UAV-Based Dataset for Crop and Weed Segmentation in Agricultural Maize Fields. In Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Tucson, AZ, USA, 26 February–6 March 2025.
39. Han, H.; Kim, S.; Joo, H.; Hong, S.; Lee, J. Mitigating Spurious Correlations via Disagreement Probability. In Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS 2024), Vancouver, BC, Canada, 10–15 December 2024.
40. Albumentations: Fast and Flexible Image Augmentations. Available online: <https://albumentations.ai/> (accessed on 20 September 2025).
41. Yu, H.; Jiang, D.; Peng, X.; Zhang, Y. A Vegetation Classification Method Based on Improved Dual-Way Branch Feature Fusion U-Net. *Front. Plant Sci.* **2022**, *13*, 1047091. [[CrossRef](#)] [[PubMed](#)]
42. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015.
43. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
44. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

45. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In Proceedings of the Advances in Neural Information Processing Systems 34 (NeurIPS 2021), Online, 6–14 December 2021.
46. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-Attention Mask Transformer for Universal Image Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022.
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**. [[CrossRef](#)]
48. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.
49. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015.
50. Pearl, J. *Causality: Models, Reasoning, and Inference*; Cambridge University Press: Cambridge, UK, 2009.
51. Satterthwaite, T.D.; Wolf, D.H.; Loughhead, J.; Ruparel, K.; Elliott, M.A.; Hakonarson, H.; Gur, R.C.; Gur, R.E. Impact of In-Scanner Head Motion on Multiple Measures of Functional Connectivity: Relevance for Studies of Neurodevelopment in Youth. *Neuroimage* **2012**, *60*, 623–632. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.