



The Future of Agriculture through Data Science: Challenges and Opportunities for Autonomous Field Systems

15th International Conference on Data Science, Technology and Applications

7th International Conference on Deep Learning Theory and Applications

João Paulo Papa

Department of Computing, São Paulo State University, Bauru, Brazil

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, USA

joao.papa@unesp.br

1. Introduction
2. Literature Review
3. Experiments
4. Conclusions

Introduction

USA Production Overview

Sugarcane is the most important row crop¹ grown in Louisiana. It has an economic impact of over **1.5 USD billion**.

- **Louisiana:** Highest economic dependence on sugarcane.
- **Florida:** Highest production volume.
- **U.S.A. industry:** Geographically concentrated and regionally specialized

¹Agricultural crop planted in distinct, spaced rows that are wide enough to allow for cultivation, weeding, and harvesting using machinery.

Why Sugarcane Matters for Brazil

Sugarcane is one of Brazil's most strategic agricultural commodities. It connects **agribusiness**, **exports**, **renewable fuels**, **bioelectricity**, and **technological innovation**.

- Brazil is a global leader in sugarcane production.
- The sector supports sugar, ethanol, electricity, and bio-based products.
- It is central to Brazil's low-carbon energy strategy.



Mechanized sugarcane harvesting in Brazil.

Scale of Production

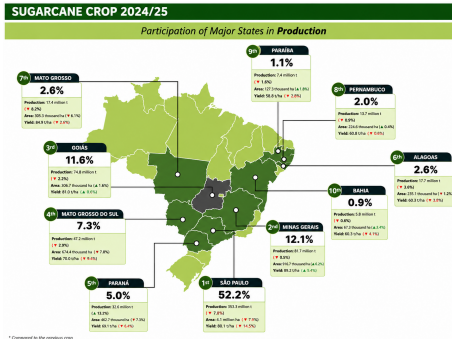
In recent harvests, Brazil has produced hundreds of millions of tonnes of sugarcane annually, making the crop one of the pillars of the national sugar-energy complex.

Indicator	Approx. value
Sugarcane	680 million t
Sugar	44 million t
Ethanol	37 billion L
Bioelectricity to grid	21 TWh



Indicative values from the Brazilian sugar-energy sector, 2024/2025 harvest.

Geographic Concentration



Sugarcane production is concentrated in the Center-South.

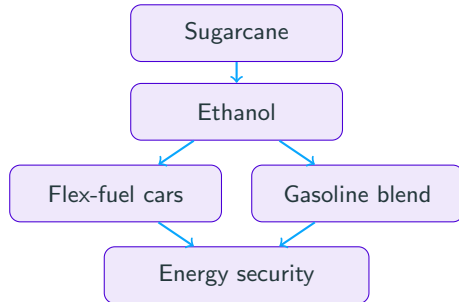
The main producing states include **São Paulo**, **Goiás**, **Minas Gerais**, **Mato Grosso do Sul**, and **Paraná**.

This reflects climate, logistics, industrial capacity, and high mechanization.

Ethanol and Energy Security

One of Brazil's major achievements is the integration of **sugarcane ethanol** into the national transportation fuel system.

- Hydrous ethanol is used directly in flex-fuel vehicles.
- Anhydrous ethanol is blended with gasoline.
- This reduces dependence on fossil fuels and strengthens domestic energy security.



Take-home Message: Reduce Costs

Sugarcane producers seek **greater** efficiency due to **rising** costs, with **weed control** being a major expense.



Source: Generated by ChatGPT (OpenAI), 2026.

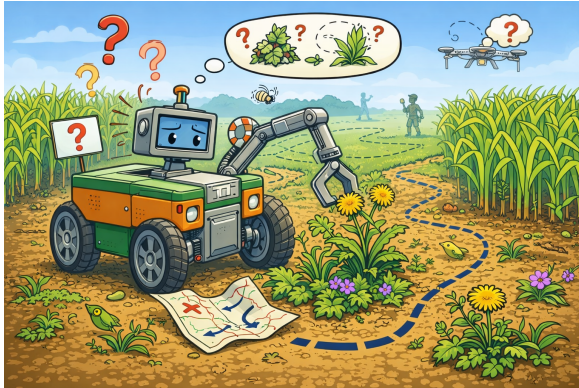
Common Practices

Uniform herbicide spraying is widely used but can be **costly**, **harm** non-target species, and pose **health** risks.

Affordable automated targeted weed spraying is **scarce**, but **autonomous** precision agriculture can **boost** yields, **reduce** inputs and chemicals, and **minimize** human error.

What do we have today?

Emerging agricultural robots face challenges in **navigation**, **efficiency**, **tool control**, and **reliable** weed detection.



Source: Generated by ChatGPT (OpenAI), 2026.

Primary Objective

Our main goal here is to **answer** the question: "Are we close"? How did we do that?

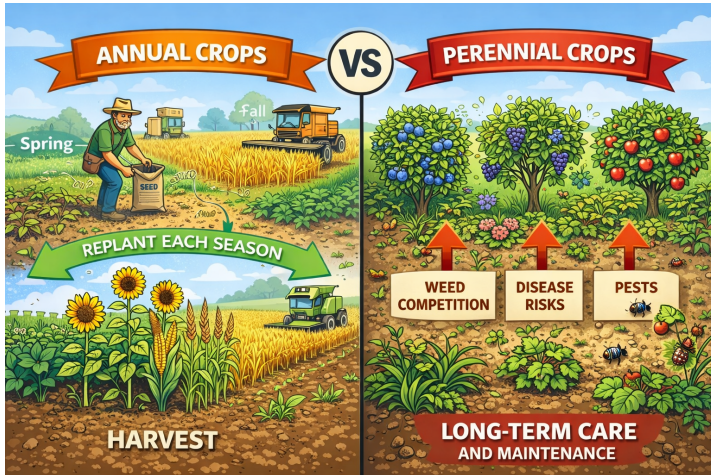
- A literature review on the topic.
- The curation of a new dataset with more **realistic** conditions, e.g., weeds with varying sizes, colors, and degrees of overlap with sugarcane crops².
- The evaluation of several approaches for weed **detection**, **classification**, and **segmentation**.

²<https://github.com/MIT-The77Lab/SugarcaneWeedDataset>

Literature Review

Annual × Perennial Crops

Most works deal with **annual** and not **perennial** crops.



Source: Generated by ChatGPT (OpenAI), 2026.

Summary and Comparison

Study & Citation	Crop(s)	Primary Task(s)	Year	Data Acquisition	Operational Context
Singh and Singh ⁹	Sugarcane	Classification	2022	UAV/Drone Imagery	Aerial (Early Weed Classification Patches)
Singh et al. ¹⁰	Sugarcane	Classification	2024	UAV/Drone Imagery	Aerial (Green-on-Green Camouflage)
Sun et al. ¹¹	Mixed crops	Segmentation	2024	RGB Images	Aerial (Green-on-Brown/Bare Soil)
Romero and Heenkenda ¹²	Sugarcane	Prescription mapping	2024	UAV + LiDAR	Aerial (UAV + LiDAR Multi-modal Mapping)
Ajayi and Ashi ¹⁴	Mixed crops	Classification	2023	UAV Imagery	Aerial (Multi-Crop Classification)
Verçosa et al. ¹⁵	Sugarcane	Classification	2022	Remote Sensing	Remote Sensing (Cross-Device Modality)
Santiago et al. ¹⁶	Sugarcane	Classification	2019	Field Images	Ground-Level (Static Images/High Weed Density)
Andrade and Ramires ¹⁷	Sugarcane	Classification	2022	Color Spectral Data	Non-Image (Color Spectral Data)
Souza et al. ¹⁸	Sugarcane	Classification	2020	Spectral Information	Non-Image (Laboratory Spectral Analysis)
Verçosa et al. ¹⁵	Sugarcane	Classification	2022	CNN-based	Aerial (Crop, Weed, and Forest Classification)
Thammoonlest et al. ¹⁹	Sugarcane	Gap forecasting	2025	Low-resolution UAV	Aerial (Low-Resolution Gap Forecasting)
Our Study	Sugarcane	Detection, Classification, Segmentation		Ground-level Sensors	Ground-Level (Simulated On-Machinery, Real-Time)

Experiments

Proposed Work

Sugarcane-Weed Dataset

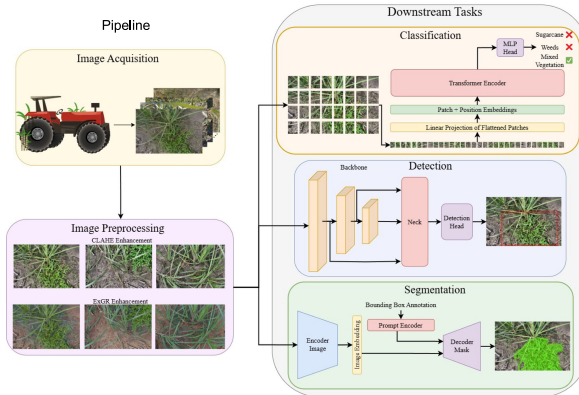


Sugarcane

Sugarcane and
Weeds

Weeds

Pipeline



The dataset comprises **2,139** high-resolution ($4,608 \times 3,072$) RGB images captured from sugarcane fields at the USDA/ARS, Sugarcane Research Unit's Ardoyne Research Farm in Schriever, Louisiana, USA in April 2020.

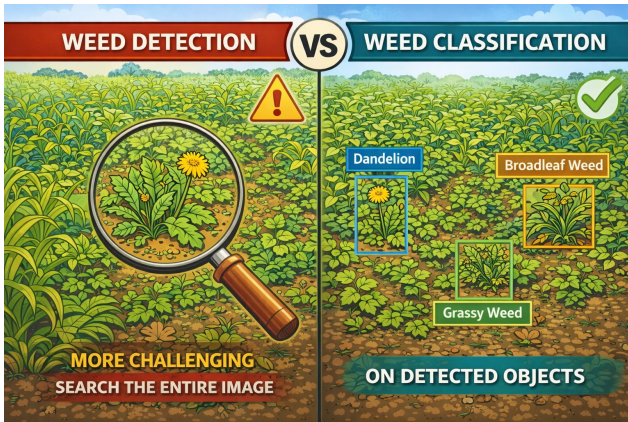
It is structured into three categories:

- **Sugarcane-only:** 869 images.
- **Weed-only:** 855 images.
- **Mixed:** 415 images.

A challenging task due to:

- **Subtle** visual differences between crops and invasive plants.
- **Overlapping** foliage.
- **Inconsistent** scaling.
- **Imperfect** annotations.

Detection × Classification



Source: Generated by ChatGPT (OpenAI), 2026.

Detection Methodology

- **Evaluation Measure:** AP50 (Intersection over Union (IoU) threshold of 50%).
- Adoption of the **Complete Intersection over Union (CloU)** rather than IoU. Why? It aims to obtain **geometrically** plausible predictions even when ground truth boxes are **oversized** or inconsistently **grouped**.



CloU considers:

- **Overlap:** Same as IoU.
- **Center distance penalty:** Penalizes how far apart the box centers are.
- **Aspect ratio consistency:** Penalizes differences in width/height.

Tuning Strategies

- All architecture backbones used ImageNet32 **pre-trained weights**, while the downstream components, including the feature pyramid necks and detection heads, were **randomly** initialized using Kaiming initialization³.
- The downstream components were trained from scratch using Adam optimizer with an initial learning rate of $4e-5$ and a cosine learning rate schedule over 120 epochs.
- The dataset is stratified into 60% training, 20% validation, and 20% testing sets. The primary input resolution is set to 640×640 pixels.

³It sets the weight variance based on the number of input connections.

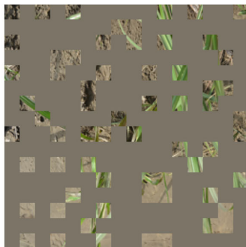
Tuning Strategies

In addition to pretrained ImageNet weights, we investigated other tuning strategies for architectures with the SwinViT backbone.

- ImageNet-only pretraining.
- ImageNet with UM-MAE⁴ pretraining.
- ImageNet with UM-MAE pretraining followed by classification fine-tuning.



Original



Masked Image



UM-MAE Reconstruction

⁴Uniform Masking Masked Autoencoder. A reconstruction mask was used to mask 75% of the original image, and the model was trained to reconstruct these gaps.

Detection Effectiveness

Architecture	Backbone	Type	Pretrained On	Detector Head	Feature Fusion	Preprocessing Technique	AP50
Mask R-CNN	ResNet50	CNN	ImageNet	FPN	-	Histogram Equalization	29.67
Mask R-CNN	ResNet101	CNN	ImageNet	FPN	-	Histogram Equalization	31.04
Mask R-CNN	ResNeXt101	CNN	ImageNet	FPN	-	Histogram Equalization	31.66
Mask R-CNN	ResNet50	CNN	ImageNet	FPN	-	CLAHE	32.04
Mask R-CNN	ResNet101	CNN	ImageNet	FPN	-	CLAHE	33.21
Mask R-CNN	ResNeXt101	CNN	ImageNet	FPN	-	CLAHE	35.76
DINO	SwinViT-L	ViT	ImageNet	FFN	-	CLAHE	28.82
RetinaNet	ResNext101	CNN	ImageNet	FPN	-	CLAHE	35.6
YOLOv11N	-	CNN	ImageNet	-	-	-	30.5
YOLOv11N	-	CNN	ImageNet	-	-	CLAHE	28.6
YOLOv11S	-	CNN	ImageNet	-	-	-	31.0
YOLOv11S	-	CNN	ImageNet	-	-	CLAHE	31.0
YOLOv11M	-	CNN	ImageNet	-	-	-	35.5
YOLOv11M	-	CNN	ImageNet	-	-	CLAHE	32.3
YOLOv11L	-	CNN	ImageNet	-	-	-	34.5
YOLOv11L	-	CNN	ImageNet	-	-	CLAHE	35.1
YOLOv11X	-	CNN	ImageNet	-	-	-	34.6
YOLOv11X	-	CNN	ImageNet	-	-	CLAHE	33.3
RTMDeT	SwinViT-B	ViT	ImageNet	CSP-PAFPN	-	CLAHE	37.96
RTMDeT	ConvNeXt	CNN	ImageNet	CSP-PAFPN	-	CLAHE	<u>42.3</u>
RTMDeT	SwinViT-B	ViT	MAE	CSP-PAFPN	-	CLAHE	11.93
RTMDeT	SwinViT-B	ViT	MAE + Classification	CSP-PAFPN	-	CLAHE	33.31
TransConv RTMDeT	SwinViT + ConvNext	Hybrid	ImageNet + Classification	CSP-PAFPN	Add.	CLAHE	36.94
TransConv RTMDeT	SwinViT + ConvNext	Hybrid	ImageNet + Classification	CSP-PAFPN	Concat.	CLAHE	39.3
TransConv RTMDeT	SwinViT + ConvNext	Hybrid	ImageNet	CSP-PAFPN	Concat.	CLAHE	42.1
TransConv RTMDeT	SwinViT + ConvNext	CNN	ImageNet	CSP-PAFPN	Concat.	CLAHE + ExGR	40.9
RTMDeT	ConvNeXt	CNN	ImageNet	CSP-PAFPN	-	CLAHE + ExGR	41.3
RTMDeT + CIoU Loss	ConvNeXt	CNN	ImageNet	CSP-PAFPN	-	CLAHE + ExGR	44.2

AP50 scores are reported on the test set, where the most accurate approach is highlighted in bold.

Detection Time and Image Resolution

Architecture	Backbone	Type	Inference time (s)
Mask R-CNN	ResNet50	CNN	0.1545
Mask R-CNN	ResNet101	CNN	0.1885
Mask R-CNN	ResNeXt101	CNN	0.1554
DiNO	SwinViT-L	ViT	0.2685
RetinaNet	ResNext101	CNN	0.1010
YOLOv11N	-	CNN	0.0073
YOLOv11S	-	CNN	0.0075
YOLOv11M	-	CNN	0.0094
YOLOv11L	-	CNN	0.1470
YOLOv11X	-	CNN	0.1512
RTMDeT	SwinViT-B	ViT	0.1031
RTMDeT	ConvNeXt	CNN	0.0990

Inference times (seconds per image) for detection.

ImageSize	Model	Architecture	Detector Head	AP50
640 × 640	RTMDeT	ConvNeXt	CSP-PAFPN	44.2
1280 × 1280	RTMDeT	ConvNeXt	CSP-PAFPN	41.3
1920 × 1920	RTMDeT	ConvNeXt	CSP-PAFPN	23.3
2560 × 2560	RTMDeT	ConvNeXt	CSP-PAFPN	13.4

Detection performance (AP50) of our best architecture (RTMDeT-ConvNeXt) across different input resolutions.

Classification

We worked with a 3-class classification problem: (i) sugarcane, (ii) weed, and (iii) a mix between them.

Architecture	Type	Accuracy	Precision	Recall	F1-Score
ViT-B	Transformer	98.12	98.45	98.03	98.23
ViT-B w/ MAE	Transformer	99.04	93.74	93.18	93.45
SwinViT-B	Transformer	88.96	91.27	86.26	87.93
SwinViT-B w/ MAE	Transformer	99.05	89.98	89.44	89.71
ResNet-18	CNN	95.33	95.51	95.27	95.37
ResNet-34	CNN	93.72	93.12	93.45	93.09
ResNet-50	CNN	91.31	92.79	92.23	92.34
ResNet-101	CNN	92.48	94.21	91.73	92.85
ResNeXt-101	CNN	92.25	93.45	92.53	92.82
ConvNeXt-v2	CNN	68.13	79.36	45.05	52.35
EfficientNet	CNN	94.36	94.88	93.29	94.06
MobileViT	Transformer	81.22	85.93	79.63	81.19
DeiT	Transformer	85.21	88.25	73.81	79.25

Segmentation - Visual

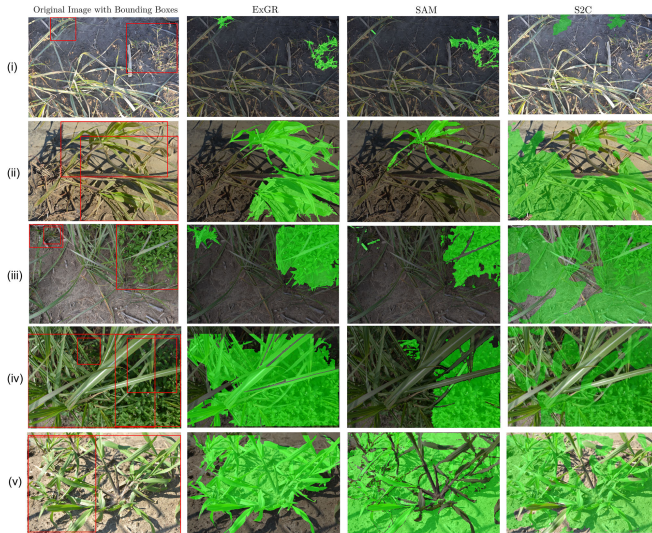
We compared an ExGR-based method⁵, SAM, and S2C⁶. Besides, we manually annotated all test images containing detection bounding boxes with pixel-level ground truth.



⁵ExGR - Excess Green Ratio.

⁶Saliency-to-Contour segmentation that combines region-level attention (saliency) with boundary refinement (contours).

Segmentation - Visual



Segmentation - Quantitative

Method	IoU	Dice	Precision	Recall	F1-Score	Accuracy
ExGR	0.3185	0.4384	0.3623	0.6308	0.4384	0.7458
SAM	0.2997	0.4003	0.5178	0.3805	0.4003	0.8380
S2C	0.2454	0.3540	0.2595	0.8227	0.3540	0.5199

Conclusions

Lessons Learned

- Although significant progress was made, precisely detecting weeds in perennial crops remains unsolved under **real-world** conditions.
- Each downstream task requires a **particular** architecture and tuning strategies.
- There is much to be done concerning unlabelled images. We could explore **semi-supervised**, **self-supervised**, and further **foundation models** to address the problem.
- **Different** crop sizes, shapes, and colors, combined with **harsh** conditions (dust and water in the acquisition sensor), make the problem worse.

Thank you!