INTRODUCTION

Cotton is one of the world's most important cash crops, contributing approximately 25% of all fiber used globally [1]. The United States is a leading cotton producer and exporter, with the Texas High Plains (THP) region producing about 25% of the nation's cotton and 65% of Texas's total production[1]. This region's cotton farming is economically vital, but it faces significant challenges due to water scarcity caused by the depletion of the Ogallala Aquifer, as a result, farmers are under pressure to adopt more efficient irrigation practices[1]. Accurate yield prediction models are crucial for optimizing water use, ensuring both economic viability and sustainability in cotton farming [2].

As noted by Haoyu Niu et al., earlier research on crop yield prediction using machine learning has primarily utilized satellite imagery as a remote sensing tool. While effective in some cases, satellite imagery faces limitations in accuracy due to challenges such as cloud cover and atmospheric disturbances. Moreover, the inflexibility in acquiring satellite images across different seasons further restricts its applicability. To mitigate these issues, **Unmanned Aerial Vehicles (UAVs)** have been identified as a promising alternative, offering enhanced precision and adaptability in agricultural monitoring and yield prediction.

UAVs can capture detailed imagery throughout the growing season, thereby allowing farmers to monitor crop conditions in near real-time and researchers to utilize those images for research. This improves the ability to detect subtle variations in crop health, irrigation needs, and growth patterns. [1]

The integration of UAV imagery with Computer Vision methods such as Convolutional Neural Networks (CNNs) has significantly advanced yield prediction[1, 2]. CNNs can automatically extract critical spatial features from images, such as vegetation indices and canopy cover, providing more accurate yield predictions[1, 2]. Furthermore, it is expected that advanced segmentation techniques like U-Net and Mask R-CNN enhance the precision of plant identification, improving the overall quality of data used in prediction models.

DATA SUMMARY

This project seeks to enhance cotton yield prediction by applying UAV-based image segmentation techniques and Computer Vision architectures. The cotton images were obtained from USDA - ARS Cropping Systems Research Laboratory (CSRL) in Lubbock, Texas, USA [1].

The data for this project was collected on November 9, 2022 using DJI Phantom 4, an Unmanned Aerial Vehicle (UAV) equipped with high-resolution RGB sensors to capture imagery of cotton fields. The UAV flight took place over a cotton field in the Texas High Plains (THP) region, a critical area for cotton production in the U.S. The UAV flew at a height of approximately 90 meters for approximately 18 minutes, capturing detailed images with a resolution of 4096 x 2160 pixels. [1, 2]

According to [2], the cotton field was divided into 12 drip zones (Figure. 1). The field was irrigated in 4 different ways, "rainfed", "full irrigation", "percent deficit of full irrigation", and "time delay of full irrigation". Each of these irrigation methods were used thrice in the field making it a total of 12 drip zones. Each drip zone was divided into 8 rows, and cotton was mechanically harvested from each row within the drip zone, resulting in 96(12 x 8) rows of cotton yield data. Every individual row was spanning 20 feet and it could fit approximately 150 cotton plants, and it had a 40 inch spacing between rows. These treatments are important for analyzing the impact of different water management strategies on cotton health and yield.

The cotton field images were split into a grid - scale which created 5376 images [2]. Out of these 1866 images were labeled using Anylabeling and the rest of the images were used to test the model performance.

Additionally, ground truth yield data was collected after the harvest. This data, labeled with the corresponding yield in pounds per acre for each image, is used to train and validate the yield prediction models. Each image corresponds to a specific area of the field, enabling precise mapping between UAV-captured imagery and actual yield outcomes.

Figure 1. Cotton Field used for experimentation

PROJECT METHODOLOGY

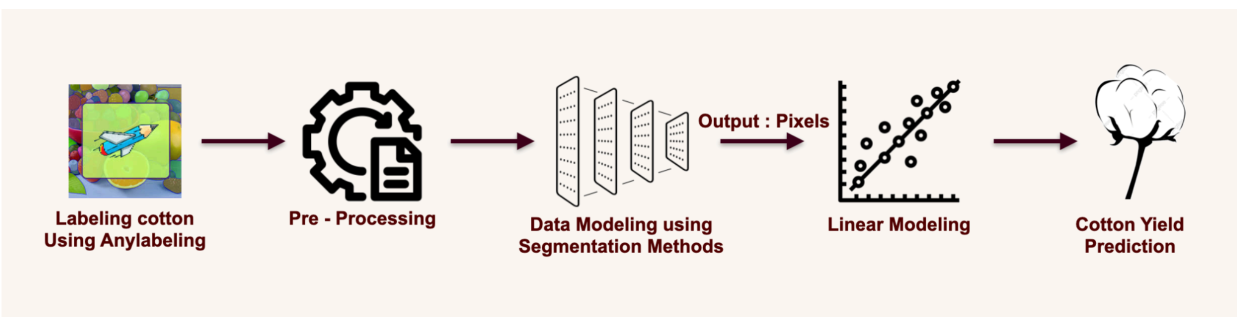

Figure 2. Project Flow

This project develops a pipeline for predicting cotton yield using machine learning and image segmentation. The process begins with **labeling cotton images** using AnyLabeling to create a dataset for training segmentation models. After **pre-processing**, where raw images are cleaned and resized, the data is passed through **segmentation models** to generate **pixel-wise masks** identifying cotton regions. The **pixel counts** from these masks are then used in **linear regression** modeling to predict yield data. Finally, the model predicts the **cotton yield**, providing valuable insights for precision agriculture and efficient resource management.

DATA PREPARATION

To process the raw images and predict cotton yield using computer vision techniques, we utilized a combination of tools and models designed for image labeling and segmentation. The process involved multiple stages, from labeling to generating masks, as outlined below:

For labeling cotton growth in raw images, we used **Anylabeling** tool that supports YOLOv5 and the Segment Anything Model (SAM). The **Segment Anything Model (SAM),** specifically the ViT-B(Base) variant of the Vision Transformer, was chosen given its effectiveness in the labeling process. While it may be slightly less precise than larger versions, its rapid processing capabilities were crucial for efficiently labeling our extensive dataset of cotton field images, significantly speeding up the data preparation phase without compromising overall performance.

To segment cotton objects, we used the **rectangle object tool** in Anylabeling's auto-segmentation marking feature. By drawing a rectangle around cotton areas in the image, the SAM model automatically segmented the object. The model then stored the coordinates of the segmented cotton objects in a JSON file for further processing. The below Figure 2. shows the raw UAV image and Figure 3. Shows the parts labeled as cotton using Anylabeling.

Figure 2. Raw UAV image

Figure 3. Labeled image

JSON Output generated for the above image - { "label": "cotton", "text": "", "points": [[59.0, 15.0], [58.0, 16.0], [58.0, 17.0], [60.0, 19.0], [60.0, 20.0], [61.0, 21.0], [61.0, 22.0], [62.0, 23.0], [64.0, 23.0], [64.0, 15.0]], "group_id": null, "shape_type": "polygon", "flags": {} } - The points represent the coordinates of cotton in the image.

Once labeling was completed for the dataset of **1866 non-overlapping raw cotton field images**, the next step was to generate segmentation masks. The annotated JSON files and raw images were used to create these masks. Figure 4 and Figure 5 show the generated masks for an image. A Python script using OpenCV's fillPoly was implemented to automate this process.

Figure 4. Raw UAV Image

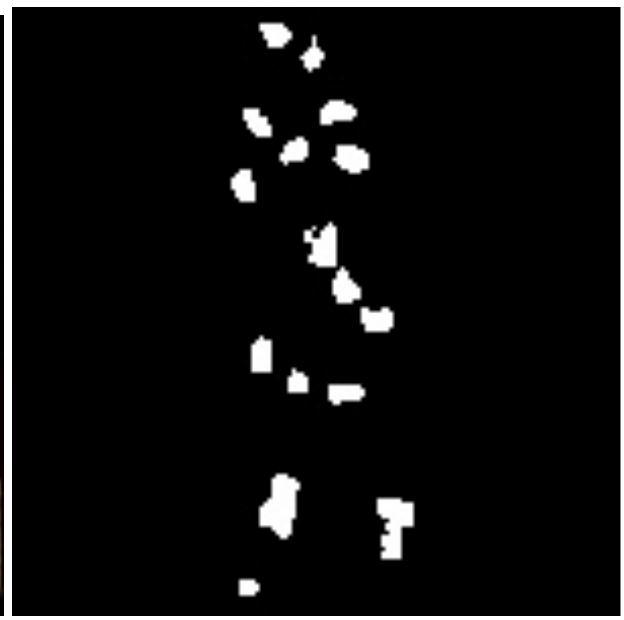

Figure 5. Corresponding mask

These masks were stored as binary images and were essential for further stages of the project, such as training and evaluating Computer Vision models on cotton field segmentation tasks.

Data Modeling using Segmentation Methods

To predict the cotton yield, after generating the masks above, the next step is to train various Computer Vision algorithms. For this project we plan to try out our images on various CNN architectures detailed below.

1) **Attention U-Net** : To enhance U-Net's performance, **soft attention mechanisms** are introduced to refine segmentation. Unlike hard attention, which focuses on one region at

a time, soft attention assigns weights to multiple regions, emphasizing relevant areas and suppressing irrelevant ones. In **Attention U-Net**, soft attention is applied in skip connections to filter out redundant features, ensuring only meaningful information is passed from the encoder to the decoder. This approach reduces noise and improves segmentation precision.

2) **MANet** : **MANet** enhances U-Net by integrating **channel attention** and **spatial attention** mechanisms to emphasize important features and suppress irrelevant ones. It incorporates **residual blocks** in the encoder for efficient feature extraction, while attention mechanisms are applied throughout the encoder, decoder, and skip connections. This architecture improves focus on critical regions, achieving high segmentation accuracy with reduced computational cost.

3) **FPN** : **Feature Pyramid Network** employs a hierarchical feature pyramid structure for effective multi-scale feature extraction. It combines a **bottom-up pathway**, extracting low-resolution, semantically rich features, with a **top-down pathway**, which upsamples coarser features and integrates them via lateral connections. This design preserves both high-level semantics and spatial resolution, making FPN efficient for tasks like object detection and semantic segmentation.

4) **PAN : Pyramid Attention Network** incorporates a **Feature Pyramid Attention (FPA)** module to fuse multi-scale contextual information using convolutions with different kernel sizes. In the decoder, it employs a **Global Attention Upsample (GAU)** mechanism to restore the original resolution while preserving spatial details. This architecture is well-suited for semantic segmentation, excelling at handling objects of varying sizes and maintaining pixel-level accuracy.

5) **Linknet : LinkNet** simplifies the encoder-decoder structure by directly passing encoded features from each encoder layer to its corresponding decoder layer through skip connections. This approach minimizes information loss during upsampling and ensures

efficient recovery of spatial details. LinkNet is computationally lightweight while maintaining good segmentation accuracy, making it suitable for real-time applications

Results -

The IOU Score (Jaccard Similarity) obtained from each of these models can be seen in Table 1. IOU score is used to evaluate the performance of segmentation models, it measures how well the predicted mask aligns with the ground truth. The Attention Unet model yielded a high IOU score of 70 compared to other segmentation models. The cotton pixels predicted by the Attention Unet model can be seen in Figure 6 which is almost accurate in predicting the cotton pixels.

Model	IOU Score
MAnet(Multi-scale Attention Net)	0.68
Linknet	0.677
FPN	0.676
PAN(Pyramid Attention Network)	0.66
Attention Unet	**0.70**

Table 1. IOU Scores for various models

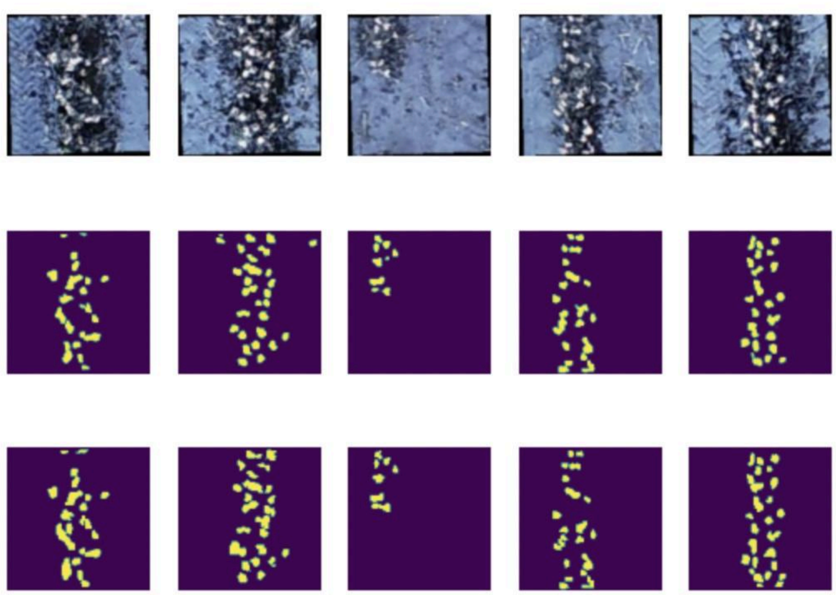

Figure 6. Actual Image, Predicted Image, Expected Image in corresponding rows

Linear Modeling Method and Results

The outputs of the segmentation model are processed to calculate the total number of white pixels in each image, representing cotton regions. As mentioned earlier, 1866 images were labeled using anylabeling tool, the total images were 5376, so the segmentation model was used to obtain white pixels for all those images. These individual images are aggregated to construct 96 composite images, each corresponding to a specific cotton yield measured in pounds. A **linear regression model** is employed to predict cotton yield based on the number of white pixels, effectively capturing the relationship between segmented cotton areas and yield. The linear regression model was particularly employed because as observed in Figure there is a linear correlation between the cotton region and cotton yield. and The model demonstrates its efficacy by explaining **88% of the variance** in the data, as evaluated by the R^2 score metric. The MSE value was **23.1** and the MAE value was **3.9**. The Figure shows the linear relationship between the white pixel area and the actual yield on the test set.

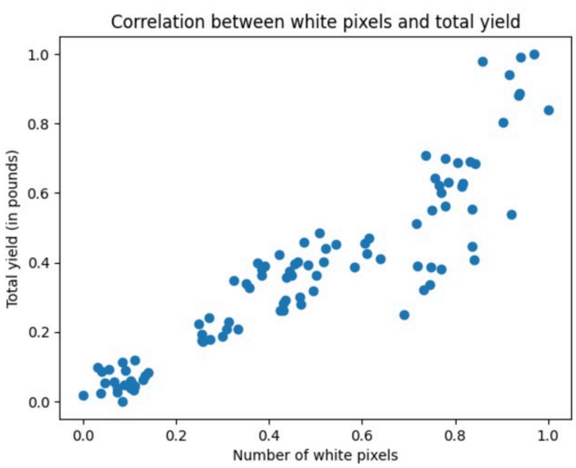

Figure 7. Correlation between cotton area and yield

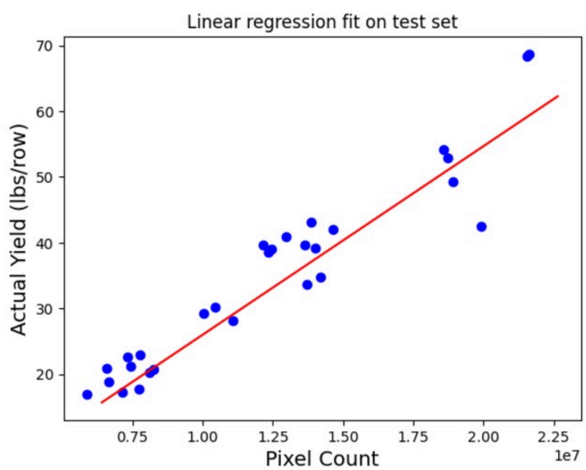

Figure 8. Linear Regression fit on test set

CONCLUSION

This project successfully integrates computer vision and machine learning for cotton yield prediction. Using advanced segmentation models such as Attention U-Net we obtained an IOU Score of 70 and we accurately identified cotton regions and established a strong linear relationship between segmented areas and yield, achieving an R^2 score of 88% with low error metrics (MSE: 23.1, MAE: 3.9). This approach demonstrates the potential of UAVs and deep learning in precision agriculture, providing a scalable solution for yield monitoring and resource optimization.

REFERENCES

1) Niu H, Landivar J, Duffield N. Classification of cotton water stress using convolutional neural networks and UAV-based RGB imagery. Advances in Modern Agriculture. 2024; 5(1): 2457. https://doi.org/10.54517/ama.v5i1.2457
2) Niu, Haoyu & Peddagudreddygari, Janvita & Bhandari, Mahendra & Landivar, Juan & Bednarz, Craig & Duffield, Nick. (2024). In-Season Cotton Yield Prediction with Scale-Aware Convolutional Neural Network Models and Unmanned Aerial Vehicle RGB Imagery. Sensors. 24. 2432. 10.3390/s24082432.
3) Risal, Avay & Niu, Haoyu & Landivar-Scott, Jose & Maeda, Murilo & Bednarz, Craig & Landivar-Bowles, Juan & Duffield, Nick & Pal, Pankaj & Lascano, Robert. (2024). Improving Irrigation Management of Cotton with Small Unmanned Aerial Vehicle (UAV) in Texas High Plains. Water. 16. 1-16. 10.3390/w16091300.
4) https://anylabeling.nrl.ai/docs
5) https://towardsdatascience.com/a-detailed-explanation-of-the-attention-u-net-b371a5590831
6) https://segmentation-models-pytorch.readthedocs.io/en/latest/models.html#
7) https://wiki.cloudfactory.com/docs/mp-wiki/model-architectures/linknet
8) https://wiki.cloudfactory.com/docs/mp-wiki/model-architectures/pan
9) https://wiki.cloudfactory.com/docs/mp-wiki/model-architectures/fpn
10) arXiv:1612.03144
11) Hettihewa, K., Kobchaisawat, T., Tanpowpong, N. *et al.* MANet: a multi-attention network for automatic liver tumor segmentation in computed tomography (CT) imaging. *Sci Rep* 13, 20098 (2023). https://doi.org/10.1038/s41598-023-46580-4