

Received 31 May 2025, accepted 23 June 2025, date of publication 30 June 2025, date of current version 18 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3584191

RESEARCH ARTICLE

Robust Multi-Input Multi-Output Analysis for Crop Row Segmentation and Furrow Line Detection in Diverse Agricultural Fields

MUHAMMAD IBRAHIM ZAIN UL ABIDEEN^{1,2}, DEWA MADE SRI ARSA^{1,2,3},
TALHA ILYAS^{1,2}, HYUNGGI JO^{1,4}, (Member, IEEE), SANG CHEOL KIM^{1,2},
AND HYONGSUK KIM^{1,2}, (Senior Member, IEEE)

¹Division of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea

²Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju 54896, Republic of Korea

³Department of Information Technology, Universitas Udayana, Bali 80361, Indonesia

⁴Division of Electronics Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea

Corresponding authors: Hyongsuk Kim (hskim@jbnu.ac.kr) and Hyunggi Jo (hygi@jbnu.ac.kr)

This work was partially supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) through the Innovative Human Resource Development for Local Intellectualization Program, funded by the Ministry of Science and ICT (MSIT), Republic of Korea (Grant No. IITP-2025-RS-2024-00439292); the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (Grant No. RS-2019-NR040079); and the Cooperative Research Program for Agricultural Science and Technology Development of the Rural Development Administration, Republic of Korea (Grant No. RS-2023-00232338).

ABSTRACT Agricultural robots have transformed traditional crop cultivation and harvesting by addressing labor shortages and enhancing precision and quality through artificial intelligence. Precise detection of furrow centerlines is critical for the seamless navigation of these robots. Previous methods often focused on specific furrow types, leading to issues with generalization and adaptability. This paper introduces a comprehensive deep-learning model designed to detect furrow centerlines across diverse types of furrows, thereby improving accuracy and robustness. Our algorithm utilizes RGB and depth images, processed through a dual encoder that combines their features. These features are refined through a channel-limiting network and then enhanced by Deep Multi-scale Feature Fusion (DFF), which maintains feature correlations across different scales. Finally, two interlinked decoders re-utilize the multiscale features to compute segmentations and lines. Our model significantly outperforms state-of-the-art methods, achieving a minimal lateral distance deviation of just 7.8 pixels, well within an acceptable range for agricultural robotics and achieves a detection line ratio (*mLR*) of 71.13%. Additionally, under a multi-task learning setup, our approach yields over a 10% improvement in *mIOU* for the furrow segmentation task. Our results demonstrate that the proposed model is robust and adaptable to various environments and conditions, ensuring reliable furrow navigation for agricultural robots.

INDEX TERMS Convolutional neural network, deep learning, precision agriculture, semantic segmentation.

I. INTRODUCTION

Precision agriculture is a farming management approach that relies on careful observation, accurate measurement, and quick responses to changes in crop conditions. The goal of precision agriculture research is to create a comprehensive

The associate editor coordinating the review of this manuscript and approving it for publication was Ayman El-Baz¹.

decision-making system for farm management. This system aims to improve agricultural yield and profitability by optimizing resource allocation, while also protecting and conserving valuable resources. Precision farming faces notable challenges, particularly in the identification of crucial knowledge gaps that demand immediate attention. Addressing these gaps is important for promoting the wider adoption and acceptance of precision agriculture technologies.

One critical aspect of precision agriculture is the navigation of agricultural robots, which requires accurate detection and tracking of furrow centerlines [1]. This capability is essential for the precise execution of tasks such as seeding [2], crop care [3], precision spraying [4], harvesting [5], selective harvesting [6], and data collection [7]. Navigating along furrows ensures that robots do not damage crops while performing their duties, thereby enhancing productivity and resource management.

Robotic navigation in agriculture involves several challenges. Furrows can vary in shape and size (see Figure 11), be infested with weeds, or have sharp turns. The ability to detect and follow these furrows accurately, regardless of their condition or environmental lighting, is crucial for the successful operation of agricultural robots. Inaccurate navigation can lead to crop damage and inefficient task execution, undermining the benefits of automation in agriculture.

Existing methods [8], [9] for furrow detection often focus on specific furrow types, which limits their generalizability and adaptability. This paper introduces a comprehensive deep-learning model designed to overcome these limitations by accurately detecting furrow centerlines across diverse types. Our approach leverages RGB and depth images to enhance the robustness and accuracy of furrow detection.

Our proposed framework consists of several key components that enhance the accuracy and reliability of furrow segmentation and center line detection.

- First, a dual encoder processes RGB and depth images, combining their features for comprehensive analysis. These combined features are then refined through a channel-limiting network, which prepares them for further enhancement.
- Next, the Deep Multi-scale Feature Fusion (DFF) component maintains feature correlations across different scales, ensuring robust feature integration for accurate segmentation.
- Finally, two interlinked decoders reutilize the multi-scale features to compute precise segmentations and centerlines.

Incorporating the concept of semantic graphics, our approach draws inspiration from [10], where convolutional encoder-decoder networks are used for learning semantic graphics in agricultural applications. The centerline detection in our framework (see Figure 3) can be considered a form of semantic graphics, simplifying the complex task of annotating furrow lines.

Additionally, we are publicly releasing a comprehensive dataset and the corresponding code for furrow segmentation and centerline detection, thereby promoting further research and development in this critical area of precision agriculture. These contributions work together to ensure our framework can accurately detect and follow furrow center lines, addressing a critical need for effective robotic navigation in precision agriculture.

A. LIMITATIONS OF EXISTING METHODS

Despite various advancements, existing furrow detection methods still face several limitations:

- **Limited Adaptability:** Many methods are restricted to specific furrow shapes, sizes, or soil types, thereby reducing their effectiveness across diverse agricultural environments [8].
- **Sensitivity to Environmental Changes:** Variations in lighting, occlusions from weeds, and field conditions frequently reduce the performance of traditional vision-based methods [9].
- **Single-Modality Dependency:** Many methods that depend on the RGB or depth images only reduce their robustness when only one modality is reduced or unpredictable [11].
- **Lack of Multi-Task Optimization:** Previous approaches used segmentation and centerline detection as separate tasks, which leads to inconsistent performance results [10].

In order to focus on such challenges, our proposed method integrates RGB and depth features using a Multi-Input Multi-Output (MIMO) network with multi-task learning, as illustrated in Table 1.

TABLE 1. Summary of key contributions.

No.	Contribution
1	Development of a MIMO-based deep learning model integrating RGB and depth information for robust furrow segmentation and centerline detection.
2	Application of multi-task learning to jointly optimize segmentation and centerline prediction, improving spatial consistency.
3	Analysis of model behavior under different occlusion and lighting conditions, demonstrating improved generalization.

II. LITERATURE REVIEW

Furrow line detection is a critical technique used to identify the path or centerline of furrows, narrow and shallow trenches created in soil for planting crops (see Figure 1). Precision agriculture faces a challenge in accurately detecting the centerline of furrows, making furrow detection the backbone of modern agriculture. It plays a crucial role in improving crop management, enhancing farm productivity, and overcoming the limitations of manual farming. Autonomous agriculture robots contribute to labor productivity, promote crop health, reduce costs, and advance and optimize farming practices. The implementation of precision agriculture not only enhances productivity and product quality but also promotes resource preservation in an environmentally friendly manner, increasing overall efficiency.

According to a study by [12], using a vanishing point and line-scanning technique, researchers have developed a method to identify hybrid rice rows. To overcome the limitations of traditional image processing methods, they have employed image processing and deep learning to accurately predict rice rows. However, maintaining identification accuracy and generality under changing field

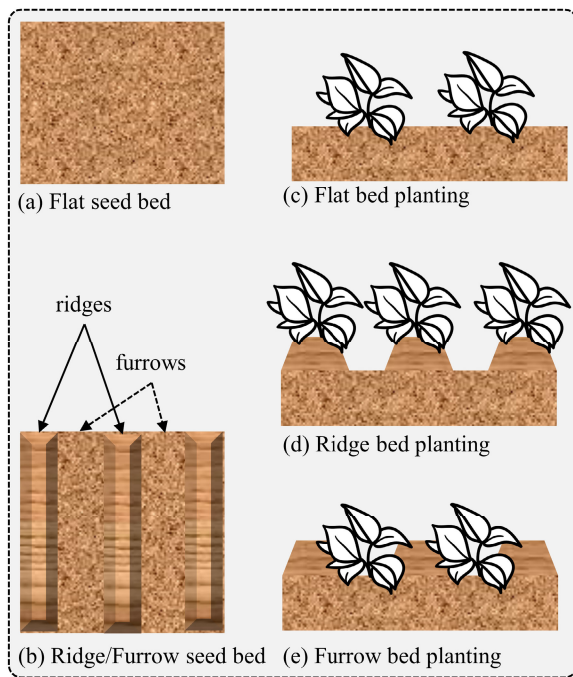


FIGURE 1. Representation of a seeding bed with ridges and furrows, illustrating the structured layout required for precision agricultural tasks such as seeding, crop care, furrow navigation and furrow line detection.

conditions remains a challenge. The study has shown an increasing interest in using Convolutional Neural Networks (CNN) for crop row identification, such as in AI-based Paddy Headland Detection using RGB-D. This approach aims to segment paddy headlands by fusing RGB images and depth information, and it introduces Long Short-Term Memory (LSTM) for temporal information between frames. Challenges involve the need to combine various techniques for segmentation results. Another related study focuses on AI-based furrow analysis [13], utilizing edge extraction algorithms for furrow identification. The natural distortion of furrows presents challenges as it can lead to the extraction of inaccurate edges. To address this issue, the study artificially modifies furrow slopes in order to obtain precise edge labels. However, it's worth noting that applying this method to real furrow images may introduce errors.

In addition, further research has introduced a method for identifying ridges and furrows in row-crop farms using Deep Lab, which is trained with RGB data. Some of the challenges associated with this approach include the spectral similarity between weeds and crops, inconsistencies in lighting, and variations in growth stages [9]. Key challenges identified in previous works involve the accurate detection of field rows in diverse environments and a lack of practical problem-addressing results in variation in real environments.

Further research introduces a method for identifying ridges and furrows in row-crop farms using Deep Lab, trained with RGB color data. Challenges include spectral similarity between weeds and crops, lighting inconsistencies,

and differences in growth stages [12]. While GNSS-based straight-line navigation technology is considered mature [14], related machine vision research offers numerous crop line recognition and tracking solutions [15], [16]. However, research on autonomous steering for agro-machinery at the headland is still in its infancy, necessitating manual intervention in most cases. Advancing autonomous steering technology at the headland is crucial for achieving fully autonomous navigation during field operations [17].

Machine vision systems excel in headland position perception due to their cost-effectiveness, flexibility, and rich data collection capabilities [18], [19], [20]. Previous studies have primarily used machine vision to detect headland boundaries, which are essential for planning subsequent turning paths. These boundaries were often identified based on differences in brightness or color between farmland and non-farmland areas [21], [22]. However, these methods are limited to scenarios where the headland boundary and the direction of machinery movement are approximately perpendicular. Using support vector machines (SVM) for detecting headland areas from superpixel blocks in field images has been explored, but performance remains unsatisfactory due to the reliance on manually created shallow features and varying outdoor lighting conditions [23], [24].

Integrating convolutional neural networks (CNN) into headland detection offers improvements in accuracy and robustness by leveraging vast data for self-learning [25]. The MobileV2-UNet network was applied to RGB images for headland segmentation and boundary extraction in a recent study [26]. DNN-based farmland boundary extraction methods have also been applied to remote sensing images from UAVs [27], [28]. However, distinguishing headlands from farmland can be challenging due to minimal color or texture differences, especially under uniform illumination. Employing height differences as supplementary information to improve headland segmentation is a promising approach. RGB and depth sensors, such as Kinect and RealSense, are increasingly used in agriculture for tasks like picking, weed identification, and crop row detection [11], weed detection [29], [30], and crop row identification [8]. Depth images provide valuable geometric information, enhancing RGB image representation and robustness to environmental disturbances. Effective multimodal feature fusion methods are key to leveraging complementary information from RGB and depth images, improving segmentation accuracy [31].

A. CATEGORIZATION OF EXISTING FURROW DETECTION METHODS

Based on the reviewed literature, existing furrow detection methods can be broadly categorized into two main groups:

- **Traditional Image Processing Methods:** These include edge detection, thresholding, and vanishing point estimation techniques [12], [13]. **Advantages:** Low computational cost, easy to implement. **Disadvantages:** Highly sensitive to environmental conditions

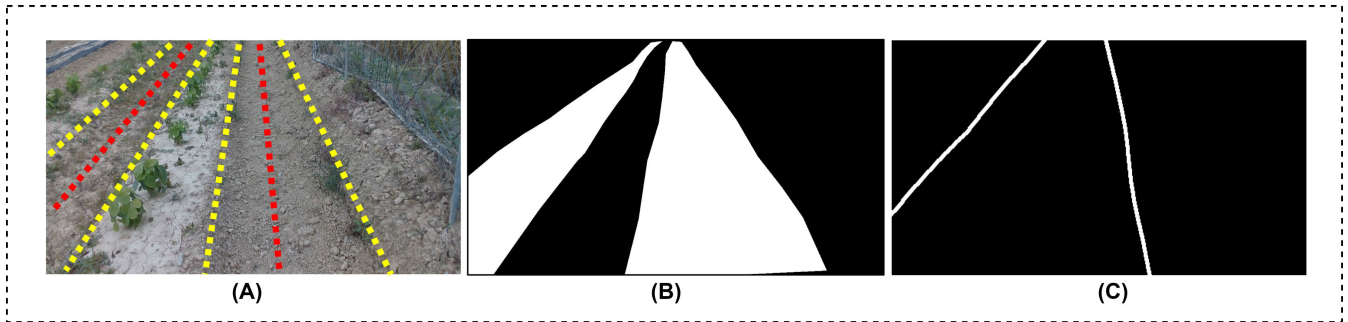


FIGURE 2. (A) Input RGB image sample, where yellow dotted lines indicate the furrow edges and red dotted lines represent the center lines of the furrows. (B) Segmentation mask of the sample image. (C) Extracted furrow midline from the sampled image.

such as lighting, soil texture, and weed occlusion; limited generalization to different field types.

- **Deep Learning-Based Methods:** These include convolutional neural network (CNN) architectures, such as UNet and DeepLab variants [9], [26], applied to RGB or depth images. **Advantages:** Improved adaptability, better feature learning, and robustness to varying field conditions. **Disadvantages:** Require large labeled datasets, may overfit to specific environments, and often depend on single-modality inputs (RGB or depth) without considering multi-modal fusion.

While both categories offer valuable contributions, their limitations highlight the need for more robust solutions. Our proposed method addresses these gaps by integrating RGB and depth information through a Multi-Input Multi-Output (MIMO) network combined with multi-task learning. This approach improves generalization, enhances segmentation accuracy, and ensures continuity in furrow centerline detection, as demonstrated in our experimental evaluations.

In recent years, MIMO networks have gained attention for their ability to handle multiple inputs and outputs simultaneously, enhancing performance and accuracy. For example, a study [32] showed the effectiveness of MIMO networks in high-precision cell segmentation using multiple input channels to capture diverse features. A study by [33] and [34] emphasized the effectiveness of MIMO networks in processing remote sensing data and improving the accuracy of satellite image segmentation.

This study proposes a sophisticated DL-based method for training a furrow centerline detection model using RGB and depth images. Through the incorporation of depth images, valuable insights are gained into object distances, enhancing the overall robustness of scene understanding. This innovative approach should be deployed on an autonomous robot as a future direction. Ultimately, the goal is to enable the robot to navigate autonomously in agricultural fields with exceptional precision, accurately tracing and following the detected furrow centerline.

B. ESTIMATION OF FURROW CENTER LINE USING SEMANTIC GRAPHICS

This study focuses on identifying furrows in agricultural fields to enable robots to autonomously navigate along

them, performing various farming tasks. Specifically, it aims to allow agricultural robots to follow the furrow center line, enhancing precision in their movement. Traditional methods, such as those cited by [35], often involve multiple steps: detecting identifiable areas or edges of furrows and then post-processing to find the center line. This multi-step approach can require complex algorithms and is not always efficient [36].

Our approach leverages deep learning (DL) to directly produce high-level, human-friendly solutions without needing extra processing steps. By utilizing graphical sketches, DL networks can recognize and learn the intent behind human annotations. This method simplifies complex agricultural tasks and has shown promise in related fields, as demonstrated by [10] and [37], where graphical sketches are used to encode target concepts for DL models [38].

Semantic graphics differ from traditional semantic segmentation by focusing on higher-order concepts rather than low-level image details. For instance, in our study, furrow center lines are annotated as thin lines in the image. These annotations are then used to train a convolutional encoder-decoder network, enabling the model to identify and follow furrow center lines effectively, as illustrated in Figure 2. This approach eliminates the need for detailed commands or multiple processing steps, allowing the network to utilize all available information to accurately produce furrow center lines.

C. SEMANTIC SEGMENTATION VERSUS SEMANTIC GRAPHICS

Semantic graphics offer a more concise representation than semantic segmentation. In furrow detection, semantic graphics represent the center line of the furrow, providing a reference for robotic control. In contrast, semantic segmentation involves labeling groups of pixels that belong to the same semantic region. Semantic graphics focus on higher-order concepts, using lines, points, or symbols to express human intentions, as supported by [39]. While semantic segmentation is typically used for structured object representation [40], [41], [42], semantic graphics can depict ambiguous and non-structured concepts [43]. Combining both approaches can enhance computer vision

applications, particularly in agricultural robotics [44], [45]. Semantic graphics can provide contextual information and prior knowledge to guide semantic segmentation, improving object recognition [46], reducing ambiguity [47], [48], and enhancing object tracking [49]. This integration leads to more robust and efficient systems, where the strengths of both methods complement each other, ultimately improving the understanding and analysis of complex scenes, such as predicting the center of furrows. This synergy not only boosts segmentation accuracy but also adapts models to specific scenarios, enhancing generalization and producing smoother, more intuitive predictions.

III. PROPOSED METHODOLOGY

Recent studies by [32] and [33] have shown that advancements in Multi-Input Multi-Output (MIMO) networks improve performance and accuracy by simultaneously handling multiple inputs and outputs. Expanding on these advances, we propose a Multi-Input Multi-Output (MIMO) network to further integrate deep learning methods. The introduction of the MIMO network methodology in our research is intended to revolutionize furrow detection distinctively. The significance of the MIMO network lies in its dual encoder-decoder structure, incorporating unique elements such as dual independent encoders and dual fused decoders. This meticulously engineered architecture aims to optimize both furrow segmentation and centerline detection concurrently, distinguishing itself from existing models. The design of the MIMO network specifically addresses deficiencies observed in other models. By simultaneously optimizing furrow segmentation and centerline detection, our proposed method seeks to achieve enhanced performance in terms of accuracy, smoothness, completeness, and other relevant metrics. A pivotal aspect of the MIMO network is its utilization of RGB+D information from Intel RealSense camera images/videos. The integration of dedicated encoders for processing both RGB and depth modalities separately enhances the model's capacity to capture rich information, thereby improving its ability to discern furrow characteristics and structure.

One drawback associated with a dual-encoder single-decoder architecture is the underutilization of features represented by the concatenated encoder feature map. Conversely, a larger single decoder, while contributing minimally to feature representations, exponentially increases the time complexity and potentially exposes the model to overfitting risks. This issue is addressed by incorporating an additional decoder responsible for learning the furrow segment. This supplementary decoder shares features with the line decoder, establishing a correspondence between furrow segment pixels and furrow line pixel positions. Before reaching the decoders, the concatenated feature map undergoes a series of sequential Deep Multi-scale Feature Fusion (DFF) modules. The introduction of the DFF module adaptively combines features of varying scales and resolutions, enhancing the network's robustness in furrow detection tasks. The inclusion

of the DFF module in the MIMO network distinguishes it from traditional methods, contributing to its innovative features. To validate the effectiveness and superiority of the MIMO network, a comparison with different pre-trained models is conducted, considering metrics such as mean smoothness, mean completeness score, mean length ratio, and mean average lateral distance. This comparative analysis serves to underscore the MIMO network's prowess over existing methods.

In addressing the challenges associated with a dual-encoder single-decoder architecture, we have implemented connected dual decoders sharing features at different scales. These interconnected decoders reinforce features learned by each other at various levels, establishing a multi-level correspondence between them. Both decoders are applied in two separate ways: firstly, culminating in a single loss function that calculates the line loss, and secondly, with one decoder learning the entire furrow segment while the other focuses on line features. The introduction of an additional decoder is instrumental in resolving connectivity issues, improving both connectivity and the accuracy of line detection. To further mitigate these challenges, a dedicated segmentation decoder is incorporated into the model, complemented by the integration of DFF modules. This sequential feature combination across different scales and resolutions allows the model to make informed decisions and accurately interpret furrow-related information, thereby enhancing its predictive ability for aligning furrow lines.

In summary, our proposed MIMO-network proposal aims to overcome specific challenges in furrow detection through its unique architecture, feature fusion techniques, and utilization of RGB+D information, positioning itself to outperform existing models and advance the field of furrow segmentation and centerline detection.

In the following sections, we will first describe the MIMO-Network architecture and multi-modality encoders. Then, we will explain deep multi-scale feature fusion. Finally, we will describe how the multi-decoders work.

A. NETWORK ARCHITECTURE (MIMO-NET)

Our research introduces a novel dual encoder-decoder structure, termed MIMO-Net, which integrates an enhanced Deep Multi-scale Feature Fusion (DFF) for simultaneous furrow segmentation and centerline detection. We have engineered several innovative modules, specifically dual independent encoders and dual fused decoders, to optimize furrow segmentation and centerline detection. In a departure from traditional methods, we process features through DFF modules before feeding them to the decoders, resulting in a network that exhibits robust performance in furrow detection tasks.

The architecture of MIMO-Net processes 384×768 RGB+D images/videos captured by an Intel RealSense camera, as illustrated in Figure 3. The network is adept at pinpointing furrows and their centerlines on a pixel-by-pixel basis. The detailed block-wise architecture of MIMO-Net

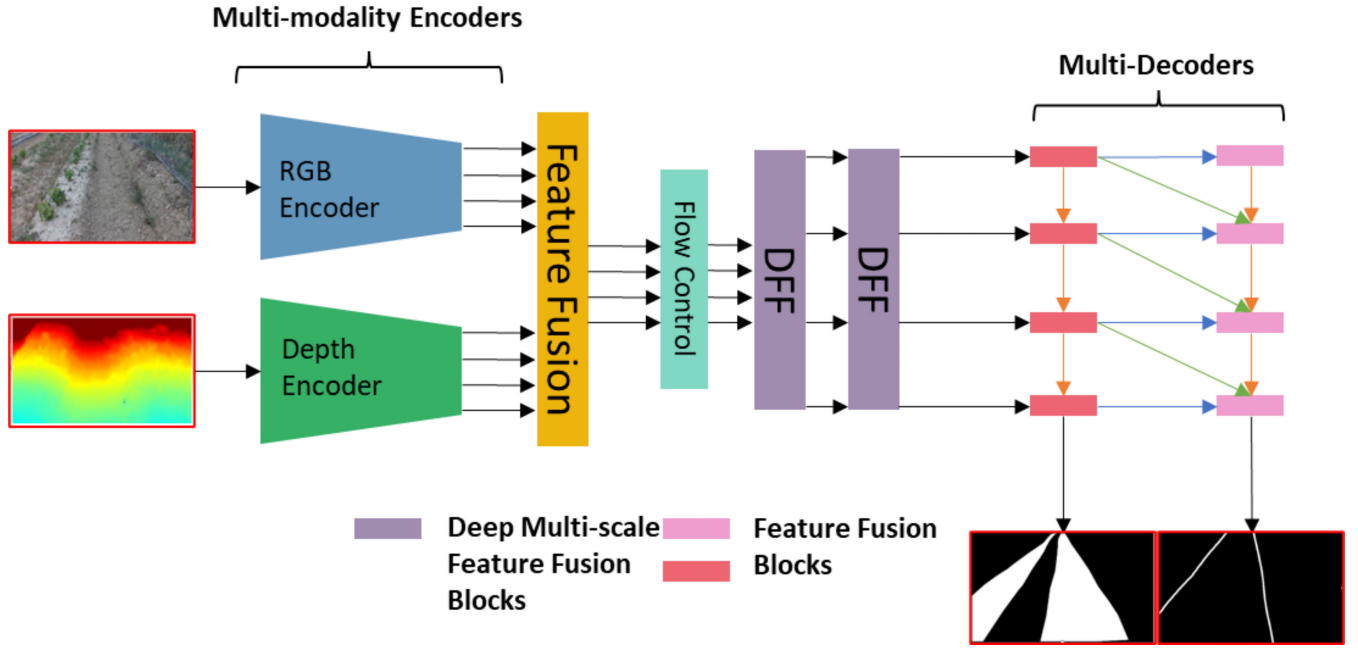


FIGURE 3. Overview of the proposed architecture: RGB and depth images are processed through separate encoders, followed by feature fusion. The fused features are then refined via flow control and Deep Multi-scale Feature Fusion (DFF) blocks. Finally, multi-decoder modules produce segmentation and furrow centerline masks.

is illustrated in Figure 4. It comprises 3 parts: (a) Multi-Modality Encoders (MME): one for RGB and one for Depth modalities, (b) Deep Multi-scale Feature Fusion (DFF), and (c) Multi Decoders (MD): one for segmentation and the other for line detection. Each encoder processes its respective inputs; the resulting features are fused and directed into the Deep Multi-scale Feature Fusion (DFF) module. The DFF module adaptively combines features of various scales and resolutions based on their relative importance, culminating in two deeply connected decoders tasked with furrow segmentation prediction and centerline detection.

B. MULTI-MODALITY ENCODERS (MME)

To elevate feature extraction and acquisition, our MME module incorporates elements from hybrid neural network architectures that synergize the spatial understanding of Convolutional Neural Networks (CNNs) with the contextual reasoning capabilities of Transformers. This fusion adeptly captures intricate field nuances, including variable weather and lighting conditions, while maintaining a broad perspective.

The encoders in the MME are grounded in the MaxViT architecture proposed by [50]. This architecture is chosen for its ability to capture both global and local spatial interactions at all network stages, including early high-resolution stages Figure 4a.

The MaxViT generates five feature maps at different resolutions for RGB and Depth, denoted as $\{P_1^r, P_2^r, \dots, P_5^r\}$ for RGB and $\{P_1^d, P_2^d, \dots, P_5^d\}$ for Depth. Here, the ‘r’ and ‘d’ in the superscript denote the features from the backbone for

RGB and Depth, and the digit in the subscript indicates the level.

In the Feature Fusion block, the feature maps from both the RGB and Depth pathways are combined element-wise addition to integrate the complementary information from the two modalities. Given feature map at i^{th} level from both encoders, the output of fusion process FF_i^o can be formalized as follows.

$$FF_i^o = P_i^r + P_i^d \quad (1)$$

After the feature fusion, the fused features go through a series of Flow Control (FC) blocks. The FC block is used to control the channel size for each level. The process inside the FC block can be formalized as shown in Eq. 2.

$$FC_i^o = f_{drop}(f_{bn}(f_{SiLU}(\Psi(FF_i^o)))) \quad (2)$$

where Ψ , f_{SiLU} , f_{bn} , f_{drop} are a 3×3 convolution layer function, SiLU activation function, batch normalization layer, and dropout layer, respectively. i and o indicating the level of feature map and output from flow control block, respectively.

This feature fusion and control process is crucial in this stage. By integrating features from depth and rgb encoders, the model becomes more robust to variations and noise in individual modalities. For instance, in scenarios where the RGB data may be insufficient due to poor lighting conditions, depth information can compensate, leading to more reliable predictions. Moreover, the use of a flow control block to align channel sizes at each level ensures dimensional consistency, which is crucial for efficient feature fusion. This alignment simplifies subsequent processing

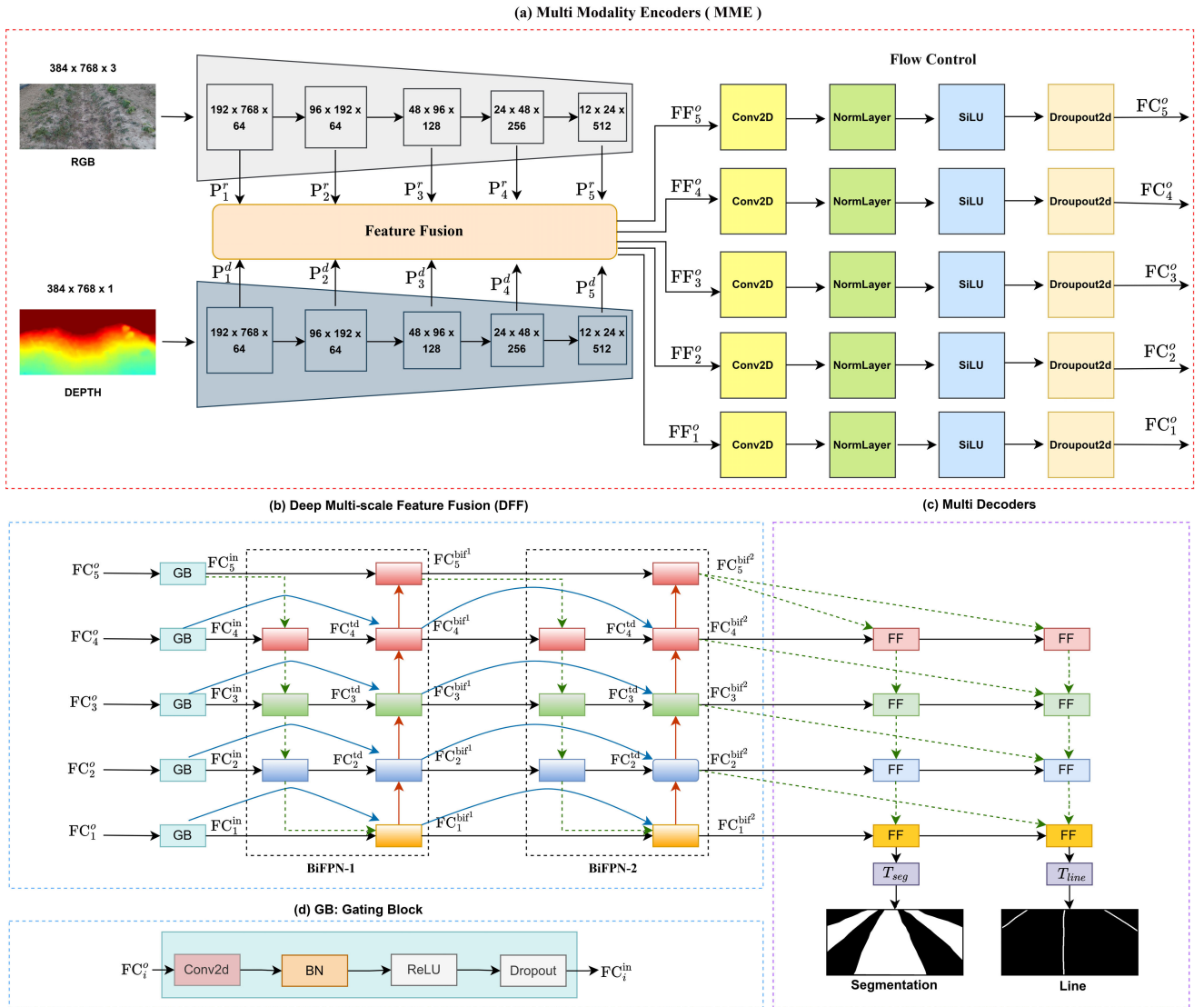


FIGURE 4. The details diagram of a) Multi Modality Encoders (MME) including feature fusion and flow control modules, b) Deep Multi-scale Feature Fusion (DFF), c) multi decoders, and d) gating block (GB).

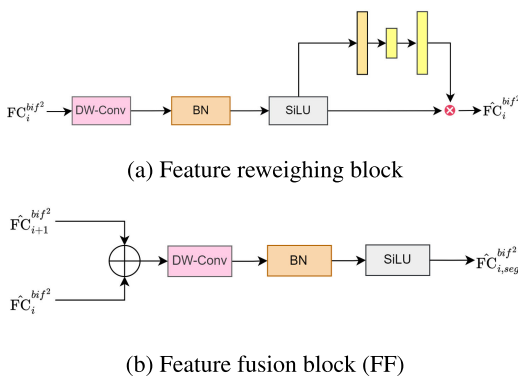


FIGURE 5. The proposed (a) feature reweighing block and (b) feature fusion block (FF) used in our multi-decoders.

steps, such as concatenation or addition of features, and prevents issues related to mismatched dimensions. It also

reduces computational overhead, as operations on feature maps of consistent size are more optimized.

C. DEEP MULTI-SCALE FEATURE FUSION (DFF)

In our approach, the combined MME features are processed through Deep Multi-scale Feature Fusion (DFF) blocks. The initial backbone features first go through a Gating Block (Figure 4d), which restricts the number of channels of all incoming features to a constant value. The Gating block manages the computational complexity and memory usage of the network. Contrary to previous methods that indiscriminately sum features of different resolutions, we assert that features from different spatial resolutions contribute variably to the output feature map. The DFF block, consisting of multiple top-down (green arrows) and bottom-up paths (orange arrows), is built upon insights from [51]. This block, referred to as Deep Multi-scale Feature Fusion (DFF),

is further subdivided into a feature reweighing block and a feature fusion block.

The feature reweighing block introduces a per-channel weighting vector $\tilde{V}^{1 \times 1 \times C_l}$ (per channel) for each input feature $R^{W_l \times H_l \times C_l}$. This enables the network to adjust the importance of each channel in the input features.

A feature fusion block takes the output from the feature reweighing block and combines the improved features from different input resolutions. Before combining the features at each level (l), the feature maps are resized to the required size using bilinear interpolation ($\hat{B}$).

The feature maps at the 2nd level ($l = 2$) of the first BiFPN ($i = 1$) can be illustrated as:

$$FC_2^{td} = \text{Conv} \left[(\tilde{V}_{in}^{1 \times 1 \times C_2}) * FC_2^{in} + (\tilde{V}_{td}^{1 \times 1 \times C_3}) \cdot \hat{B}(FC_3^{td}) \right] \quad (3)$$

$$FC_2^{bif^l} = \text{Conv} \left[(\tilde{V}_{in}^{1 \times 1 \times C_2}) * FC_2^{in} + (\tilde{V}_{td}^{1 \times 1 \times C_2}) * FC_2^{td} + (\tilde{V}_{bif^l}^{1 \times 1 \times C_1}) \hat{B}(FC_1^{bif^l}) \right] \quad (4)$$

where, $FC_1^{bif^l}$, FC_2^{in} , and FC_2^{td} represent the feature maps after the gating block (input to DFF), the top-down pathway (shown with green arrows), and the output of the first BiFPN, respectively. The term $\tilde{V}_j^{1 \times 1 \times C_l}$ denotes the reweighing vector for the feature map at level 2, where j is either in or td. Similarly, $\tilde{V}_{bif^l}^{1 \times 1 \times C_1}$ denotes the reweighing vector for the feature map at level 1 ($l = 1$). The proposed architecture connects two BiFPNs in series, with $i = 2$.

D. MULTI DECODERS (MD)

The dual decoder architecture comprises two interlinked decoders that synchronously fuse multi-scale features from the DFF, generating precise segmentation and furrow centerline masks. Figure 4 illustrates the architecture of our dual decoder structure. Each output feature map from the DFF undergoes processing through a series of FF blocks, resized appropriately via bilinear interpolation indicated by the dashed arrow. This iterative process is repeated until the final prediction maps at the final pyramid level ($l - 1$) are obtained. The primary purpose of the dual-decoder structure is to emphasize local-level feature information by mutually decoding each output, focusing on the visual characteristics of furrow segments and centerlines. The MD serves as a guiding mechanism, directing the network to utilize segmentation predictions in generating the more complex centerline predictions. This approach significantly enhances overall performance in centerline detection, as validated by our experiments section.

Motivating the design of the proposed modules is the need for robust and precise feature extraction and processing to improve segmentation and centerline detection accuracy. For each input feature map to our FF block, we perform additional operations by calibrating the feature map, filtering important features before they are mixed in the FF block. This operation is illustrated in Figure 5a. First, a depth-wise convolution block is utilized, followed by batch normalization and SiLU

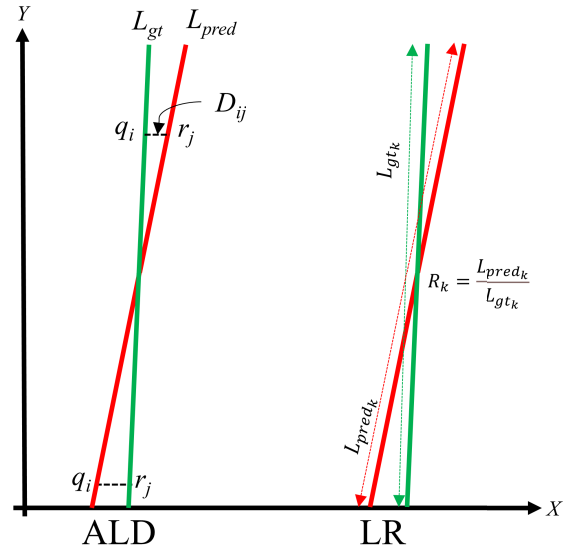


FIGURE 6. Diagram illustrating the accuracy evaluation metrics for detected centerlines, including Average Lateral Distance (ALD) and Length Ratio (LR).

activation function. Subsequently, we compute the mean value of the feature map across the height and width dimensions. Two sequential fully connected layers then learn these features, and a sigmoid function scales the values between 0 and 1, representing a weight to the feature map. The calculated weight is then multiplied by the output feature map, producing a calibrated feature map $\hat{FC}_i^{bif^2}$.

The operation inside the proposed FF block is illustrated in Figure 5b. In this figure, we illustrate the process for two inputs. Initially, we mix the weighted feature maps through pixel-wise addition. This is followed by depth-wise convolution, batch normalization, and SiLU activation. The motivation behind this design is to ensure effective feature calibration and integration, enhancing the network's ability to capture and utilize critical feature information, thereby improving the precision of the segmentation and centerline detection tasks.

E. FURROW DETECTION VIA RGB AND DEPTH MAPS COMPARISON

Objective: To enhance model interpretability and understand the attention behavior during furrow centerline detection, we conducted a comprehensive visualization analysis. Inspired by interpretability tools such as Grad-CAM, we examined attention patterns at multiple levels of the Feature Pyramid Network (FPN) and performed occlusion-based experiments. This analysis allows us to reveal which input modalities and image regions the model relies on most during prediction.

Methodology:

1) Model Training:

- **Separate Models:** Two models were trained independently—one using only RGB images and the other using only depth images.

- Combined Model: A third model was trained using both RGB and depth images as inputs.

2) Occlusion Experiments:

- Occlusions were applied to both RGB and depth images at identical locations to assess their impact on furrow line detection.
- Attention maps were computed at five different levels of the Feature Pyramid Network (FPN) to visualize which parts of the inputs were attended to by deep and shallow layers.

Results:

1) Separate Models (RGB vs Depth):

- Both models exhibited similar changes in output attention maps when the same parts of the furrow were occluded.
- The depth-only model required approximately twice the training time compared to the RGB-only model. The results are illustrated in Figure 7 (A and B).

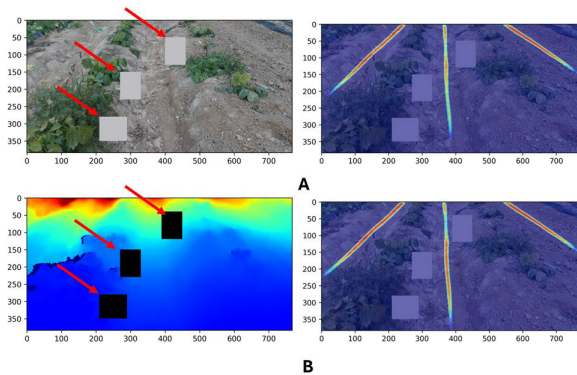


FIGURE 7. Comparison of RGB and Depth modality models for furrow detection. (A) Results from the RGB-trained model showing furrow centerline prediction overlaid on RGB images. (B) Results from the Depth-trained model showing furrow centerline prediction overlaid on depth images. Both models demonstrate similar attention to occluded regions, but the RGB model shows higher visual clarity and smoother centerline alignment.

2) Combined Model (RGB and Depth):

- Occluding the RGB image significantly affected furrow line detection.
- Occluding the depth image had a lesser impact.
- Interpretation: The model relies more heavily on RGB information for furrow detection. The results are illustrated in Figure 8 (A and B).
- Consistency Across Locations: Similar patterns were observed in Figure 9 (A and B), reinforcing the conclusion about the dominance of RGB input.

3) Attention Maps Analysis:

- **Deep Layers (Layers 4 and 5):** These layers capture high-level, abstract features that are not easily interpretable by the human eye. The attention maps at these levels are low-resolution and do not directly correspond to specific image regions.

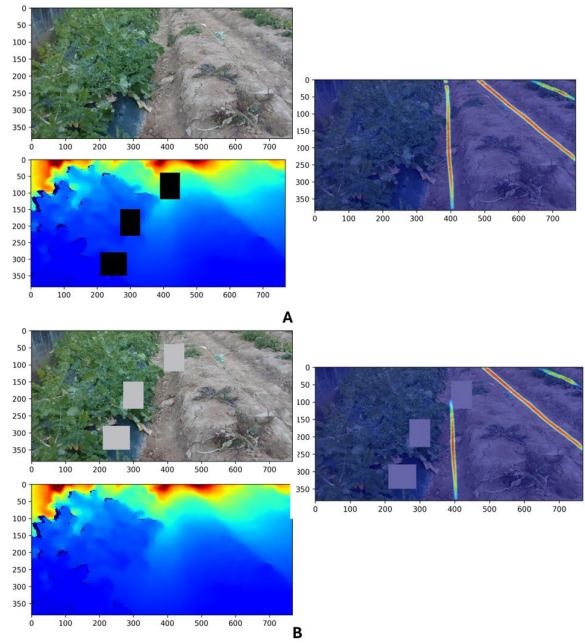


FIGURE 8. Effect of occlusions on furrow detection using RGB and depth modalities. (A) Results when occlusions are applied to the RGB images, showing a significant disruption in the predicted furrow centerline. (B) Results when occlusions are applied to the depth images, showing a smaller impact on the predicted furrow centerline.

- **Shallow Layers (Layers 1 to 3):** These layers focus on more concrete, detailed features. The attention maps become progressively clearer and more interpretable, highlighting specific areas relevant to furrow detection (see Figure 10), reinforcing the conclusion about the dominance of RGB input.

Technical Explanation: The Feature Pyramid Network (FPN) architecture leverages features at multiple scales to improve detection performance. Deep layers capture global context, while shallow layers preserve fine-grained details. Our results show that the model integrates both to achieve accurate furrow detection, with a stronger dependency on RGB features.

Conclusion: Both RGB and depth contribute to furrow detection, but the model relies more heavily on RGB information. Attention map analysis confirms the importance of shallow layers in detecting furrow edges. While depth-only models are less efficient, depth remains valuable when combined with RGB data, enhancing the model's robustness and generalization capabilities.

F. EVALUATION METRICS

In this section, we detail the metrics used to evaluate the performance and accuracy of our proposed furrow line detection model. These metrics provide a comprehensive assessment of the model's ability to accurately identify and delineate furrow centerlines within agricultural images. The chosen metrics include Median Average Lateral Distance

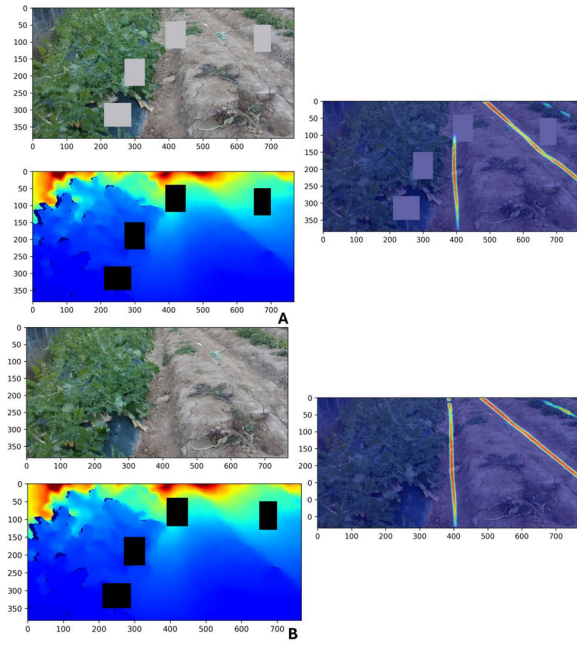


FIGURE 9. Consistency analysis demonstrating the model's dependency on RGB input. (A) Results with occlusions applied at different locations on the RGB images, consistently showing significant degradation in furrow detection accuracy. (B) Results with corresponding occlusions on the depth images, showing minimal impact on detection performance.

(*mALD*), Mean Length Ratio (*mLR*), and Mean Intersection Over Union (*mIOU*) for both line detection and segmentation tasks. Additionally, for furrow segmentation, Precision and Recall are used to provide a reliable measure of segmentation performance.

To evaluate the accuracy of the centerline extraction algorithm, we use the Average Lateral Distance (ALD), which measures the mean deviation between the manually labeled ground truth centerline (L_{gt}) and the detected centerline (L_{pred}). Specifically, we calculate the lateral distance D_{ij} between corresponding points (q_i, r_j) on the ground truth and detected lines, where q_i is a point on L_{gt} and r_j is a point on L_{pred} , selected at identical horizontal positions. The ALD is computed as follows:

$$mALD = \frac{1}{m} \sum_{i=1}^m D_{ij} \quad (5)$$

here, m represents the total number of point pairs, and D_{ij} is the lateral distance between the i th point on the ground truth line q_i and the corresponding j th point on the detected line r_j . The *mALD* metric provides a reliable measure of how accurately the algorithm can detect the centerline in comparison to the manually annotated ground truth [12].

To further assess the accuracy of the predicted line lengths, we introduce the Mean Length Ratio (MLR). This metric compares the length of each predicted line segment (L_{pred_k}) against the corresponding ground truth line segment (L_{gt_k}), providing an indication of how closely the algorithm's predictions match the actual line lengths. The Length Ratio

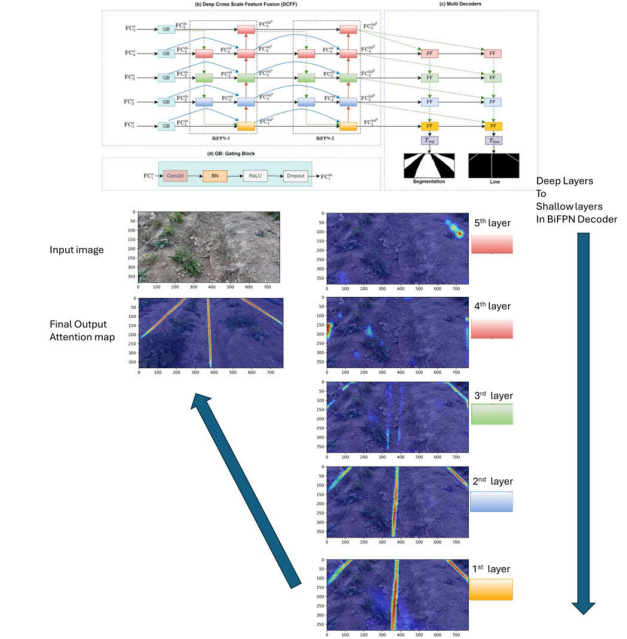


FIGURE 10. Visualization of attention maps across different layers of the Feature Pyramid Network (FPN). The figure highlights how shallow layers (1st to 3rd) focus on fine-grained edge and texture details critical for furrow line localization, while deeper layers (4th and 5th) capture more abstract and coarse scene-level features. The final attention map integrates both low-level and high-level information to produce accurate furrow line predictions. The directional arrows illustrate the hierarchical feature refinement from deep to shallow layers in the BiFPN decoder.

for each line pair is given by:

$$R_k = \frac{L_{pred_k}}{L_{gt_k}} \quad (6)$$

The Mean Length Ratio (MLR) is then calculated as the average of these ratios across all line pairs:

$$mLR = \frac{1}{N} \sum_{k=1}^N R_k \quad (7)$$

where N is the total number of line pairs, and R_k is the length ratio for the k th line pair. An *mLR* close to 1 indicates that the predicted lines closely match the lengths of the ground truth lines, while significant deviations suggest discrepancies.

These metrics collectively ensure a transparent and thorough evaluation of our model's performance, providing a solid benchmark for its effectiveness in detecting and delineating furrow centerlines in various agricultural settings. Figure 6 illustrates the methodology used for computing the Average Lateral Distance (ALD) and Length Ratio (LR).

G. OBJECTIVE FUNCTION

The furrow detection dataset presents a significant class imbalance between the furrow line and the background, leading to challenges in network training and the ability to effectively capture semantically rich features. To address this issue, we employ focal loss [52], which reduces the impact of well-classified pixels on the overall loss, allowing the

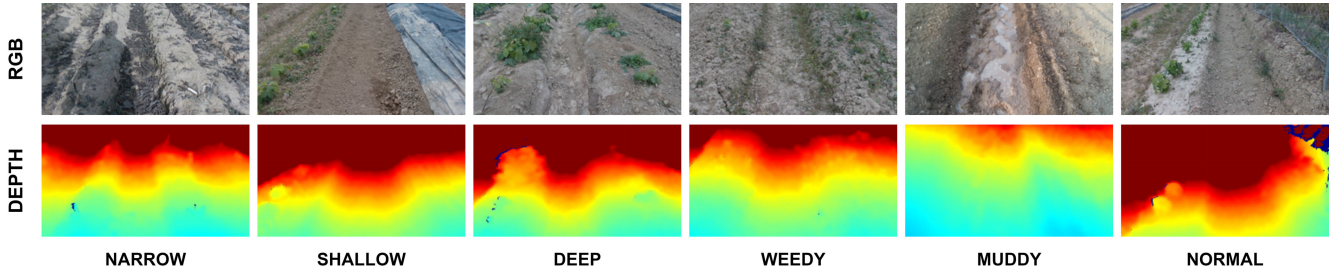


FIGURE 11. Sample images of the dataset: the top row displays RGB images, and the bottom row shows the corresponding depth maps for various types of furrows, including normal, weedy, shallow, deep, narrow, and muddy furrows.

network to focus more on difficult examples. The focal loss is defined as:

$$\mathcal{L}_{FL}(p_i) = -\alpha(1 - p_i)^\gamma \log(p_i) \quad (8)$$

where p_i is the probability of the predicted class, and α and γ are hyperparameters that control the influence of well-classified examples on the loss function. Empirically, we found this loss function to be particularly effective for furrow segmentation.

For furrow line detection, to achieve faster training and improved convergence, we use a weighted combination of three loss functions at the output of the semantic segmentation branch. The total loss for the furrow line detection branch is computed as:

$$\mathcal{L}_{line} = \mathcal{L}_{FT} + \mathcal{L}_{Combo} + \mathcal{L}_{Lovász} \quad (9)$$

where $\mathcal{L}_{FT}$ is the focal Tversky loss [53], $\mathcal{L}_{Combo}$ is the combo loss [54] and $\mathcal{L}_{Lovász}$ is the Lovász loss [55].

The focal Tversky loss is designed to address data imbalance in semantic masks, while the combination of the focal Tversky and combo losses aids in faster convergence by focusing on the most informative samples. The focal Tversky loss is defined as:

$$\mathcal{L}_{FT} = \sum_c \left(1 - \frac{\sum_{i=1}^N p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^N (p_{ic} g_{ic} + \alpha p_{ic} \bar{g}_{ic} + \beta p_{ic} \bar{g}_{ic})} \right)^{\frac{1}{\gamma}} \quad (10)$$

where p_{ic} and g_{ic} represent the predicted probability and ground truth label for the i th pixel in class c , N is the total number of pixels, ϵ is a smoothing constant, and α , β , and γ are the hyperparameters.

Given the pronounced foreground-background data imbalance in furrow line detection compared to furrow segmentation, we combine the focal Tversky loss ($\mathcal{L}_{FT}$) with the combo loss to further enhance learning. The combo loss, specifically designed to avoid suboptimal local minima and appropriately penalize false positives and false negatives, is defined as:

$$\mathcal{L}_{Combo} = \alpha \cdot \mathcal{L}_{BCE} + (1 - \alpha) \cdot \mathcal{L}_{Dice} \quad (11)$$

where α is a hyperparameter of the combo loss, $\mathcal{L}_{BCE}$ represents the standard binary cross-entropy loss, and $\mathcal{L}_{Dice}$ is the Dice loss component.

IV. EXPERIMENTAL SETUP

The experimental video, captured with an Intel Real Sense Depth Camera D435i, had an initial resolution of 1080×520 . Both the height and width of the images were subsequently resized to 384×768 to meet specific computational requirements, optimizing processing efficiency while maintaining sufficient image quality. This USB-powered cable camera features RGB and depth sensors, with a notable inclusion being the BMI055 Inertial Measurement Unit (IMU), enhancing depth analysis even when the depth camera is in motion. The video data, stored in bag files, underwent extraction to isolate only the RGB and depth images.

The RGB and depth images, illustrated in Figure 11, showcase a robust complementary connection, delivering a comprehensive understanding of furrow scenes. Each furrow type is distinguished by unique features that contribute to the overall dataset diversity. Depth information is integral, offering a three-dimensional representation of furrow scenes and enabling precise measurement of their actual depth through depth images.

As shown in Figure 12, the bar graph visually encapsulates the distribution of distinct furrow types within our dataset, offering a quick and comprehensive overview for efficient analysis and interpretation. Each bar corresponds to a specific furrow type including Deep, Muddy, Narrow, Normal, Shallow, and Weedy. The height of the bars reflects the

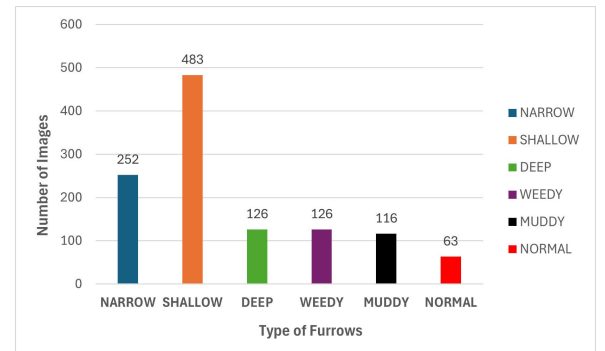


FIGURE 12. Bar graph illustrating the distribution of different types of furrows in the dataset. The x-axis represents the types of furrows (narrow, shallow, deep, weedy, muddy, and normal), and the y-axis shows the number of images for each type of furrow.

TABLE 2. Comparison of the proposed model with other pre-trained models. For the proposed model, *ST* indicates training on a single task, either furrow line detection or furrow segmentation, while *MT* represents multi-task (joint) training.

Model	Furrow Line			Furrow Segment		
	mIOU	mALD	mLR	mIOU	Precision	Recall
	(%)	pixels	(%)	(%)	(%)	(%)
U-Net [56]	41.2	14.22	47.45	61.57	60.68	63.59
FPN [57]	37.23	16.13	46.79	57.43	59.01	58.37
PSP-Net [58]	31.97	8.29	17.08	54.26	55.77	56.02
U-Net++ [59]	37.34	11.58	37.55	58.97	60.33	59.88
Link-Net [60]	39.63	10.31	37.89	60.14	61.22	62.05
MA-Net [61]	38.99	9.97	41.84	59.85	60.98	61.34
DeepLabV3 [62]	45.97	7.99	48.35	75.32	74.89	76.04
DeepLabV3+ [39]	41.36	13.25	47.39	79.46	80.65	78.47
Proposed-ST	49.36	8.05	57.13	82.63	82.71	79.82
Proposed-MT	59.44	7.83	71.13	89.04	91.46	91.14

respective number of images associated with each furrow category.

The integration of depth information with RGB images enhances our ability to analyze furrow structures, aiding in discerning soil surfaces, weed presence, muddy areas, and the characteristics of normal furrows. This collaborative approach proves particularly valuable in precision agriculture, where accurate depth measurements contribute to optimizing planting depth and assessing soil characteristics. The collected dataset's detailed examination of deep, narrow, muddy, shallow, normal, and weedy furrows underscores the crucial role of depth information in unraveling the intricate details of furrow landscapes.

Creating a diverse and representative dataset for furrow detection using a RealSense camera presented several challenges. Ensuring balanced coverage of different types of furrows (deep, normal, muddy, shallow, weedy, narrow) required extensive fieldwork amid unpredictable factors such as changing weather conditions and varying soil moisture levels. The three-dimensional complexities introduced by the RealSense camera demanded meticulous calibration and alignment efforts. Additionally, capturing dynamic, real-world furrow scenarios with the presence of weeds posed challenges. To address these issues, a systematic approach involving documentation, metadata tracking, and iterative refinement was crucial for building a high-quality dataset. Overcoming these challenges laid the groundwork for subsequent model development and training phases.

A. DATASET DESCRIPTION

The quality and diversity of our dataset play a key role in obtaining optimal results. The images were captured in Jeonju, South Korea, forming a dataset that encompasses various furrow types, including narrow, deep, shallow, muddy, weedy, and normal furrows. The annotation process was meticulously executed using the Computer Vision Annotation Tool (CVAT) to label furrow segments accurately.

Additionally, the center lines of the furrows were delineated using CVAT. To facilitate comprehensive model training and evaluation, the dataset underwent a meticulous division into distinct training and testing sets. The training set consisted of 1166 samples, while the test set comprised 420 samples. This systematic organization ensures a robust assessment of the model's generalization and predictive capabilities across diverse datasets, affirming the reliability of our results and the model's applicability in real-world scenarios. Throughout the experiments, PyTorch was employed, utilizing a Graphics Processing Unit (GPU) - the NVIDIA TITAN RTX 2080 with 24 GB of memory space. A batch size of 2 was utilized during the experimental runs, optimizing the processing efficiency of the neural network models. The learning rate for the training process was set at 0.001, a crucial parameter influencing the convergence and performance of the model during the optimization process. The experiments were conducted over 100 epochs, providing a comprehensive evaluation of the model's performance, and learning capabilities. Moreover, the proposed method is compared with previous techniques on the different pre-train models UNet [56], FPN [57], PSPNet [58], UNet++ [59], LinkNet [60], MANet [61], DeepLabV3 [62] and DeepLabV3+ [39] applied in datasets.

B. DATASET PREPROCESSING AND AUGMENTATION

To ensure the diversity and robustness of our dataset, images were captured under varying field conditions, including sunny, cloudy, and shaded environments, at different times of the day in Jeonju, South Korea.

The preprocessing pipeline includes the following steps:

- **Image Extraction:** RGB and depth images were extracted from bag files recorded with Intel RealSense D435i.
- **Image Resizing:** All images were resized from their original 1080×520 resolution to 384×768 pixels to balance computational cost and spatial resolution.

- **Annotation:** Furrow segments and centerlines were manually annotated using the CVAT tool.
- **Normalization:** RGB images were standardized and normalized to a [0, 1] range. Depth images were clipped at a maximum threshold of 10 meters to remove irrelevant far-field data and normalized similarly.

C. DATA AUGMENTATION PIPELINE

To further improve robustness to real-world variability, we applied geometric and noise-based augmentations:

- **Geometric Augmentations:** Horizontal flipping, random scaling, translation, rotation, shearing, and cropping were applied to both RGB and depth images as well as their corresponding furrow masks. These augmentations simulate camera perspective changes and field variability.
- **Noise Augmentations:** Gaussian blur, motion blur, additive Gaussian noise, channel shuffling, and brightness/saturation adjustments were applied to RGB images only, simulating sensor noise and varying lighting conditions.

The augmentation probabilities were dynamically adjusted based on training epochs, promoting geometric transformations in the early stages and noise robustness in later stages. This curriculum-based augmentation strategy ensures balanced learning across different types of data variability.

D. TRAIN-TEST SPLIT AND ENVIRONMENTAL DIFFERENCES

The training and testing sets were collected from separate field plots, ensuring no data from the same plot appears in both sets. The dataset was split into 75% for training and 25% for testing, with the test set containing unseen examples to ensure unbiased evaluation. This split introduces natural variations in soil texture, furrow shapes, and lighting conditions, making the evaluation more realistic and representative of deployment scenarios.

These preprocessing and augmentation strategies collectively address the challenges of dataset diversity, environmental variability, and overfitting. By systematically preparing and augmenting the dataset, we aim to validate the robustness and generalization capabilities of our proposed model across varying field conditions.

V. PERFORMANCE EVALUATION OF THE PROPOSED MODEL

The performance evaluation of the proposed model in comparison to several state-of-the-art (SOTA) models is summarized in Table 2. The results demonstrate that our proposed model, particularly under the multi-task (MT) learning setting, significantly outperforms all other pre-trained models across various metrics for both furrow line detection and furrow segmentation tasks.

For furrow line detection, the proposed model trained in a multi-task setting (Proposed-MT) achieved the highest Mean

Intersection Over Union (*mIOU*) of 59.44%, which is notably higher than the best-performing SOTA model, DeepLabV3, which recorded an *mIOU* of 45.97%. This indicates that the proposed model is more effective in accurately identifying and delineating furrow centerlines compared to other models. Additionally, the Proposed-MT model achieved the lowest Median Average Lateral Distance (*mALD*) of 7.83 pixels, which is comparable to DeepLabV3's 7.99 pixels but with the added benefit of superior line ratio performance. The Mean Length Ratio (*mLR*) for the Proposed-MT model was 71.13%, significantly outperforming the next best model, DeepLabV3, which achieved 48.35%. This superior *mLR* suggests that the proposed model is more reliable in maintaining the true length of the detected lines, a critical factor for practical applications in autonomous navigation.

In terms of furrow segmentation, the Proposed-MT model again demonstrated superior performance, achieving an *mIOU* of 89.04%, which is a substantial improvement over the best SOTA model, DeepLabV3+, which recorded an *mIOU* of 79.46%. The precision and recall metrics for the Proposed-MT model were 91.46% and 91.14%, respectively, both of which surpassed the SOTA models, indicating not only high accuracy but also consistency in furrow segmentation. This level of performance underscores the model's effectiveness in handling complex agricultural environments, where accurate segmentation is crucial for the autonomous navigation of agricultural robots.

The single-task (ST) version of the proposed model also showed commendable results, outperforming most SOTA models in both line detection and segmentation tasks. However, it is clear from the results that the multi-task learning approach provides a significant advantage, likely due to the shared learning across related tasks, which enhances the model's overall capability and robustness.

A. DISCUSSION ON MLR IMPROVEMENT

The notable improvement in the *mLR* for the Proposed-MT model compared to other models, can be described as below:

1. Data Diversity and Augmentation: The proposed model was trained on a diverse dataset which includes various types of furrows (e.g., deep, narrow, shallow, muddy, weedy, normal) and various field conditions (e.g., sunny, cloudy, shaded). This variety makes model the proposed model learn to detect furrow lines accurately, under diverse environmental conditions.

2. Multi-Task Learning (MT): In a multi-task learning approach, the proposed method can detect both furrow line and segmentation, which allows the model to share special features for the tasks. This sharing helps the model ability to accurately detect the furrow lines along with better detection in segmentation, which eventually results better as compared to other models.

3. Model Architecture: The structure of our proposed method includes superior procedures like Feature Pyramid Networks (FPN) and encoder-decoder structures, which improves the model's ability to maintain important features

TABLE 3. The ablation studies on varying Encoder Inputs for the proposed model under Multi-Task learning settings.

Encoder Input		Furrow Line			Furrow Segment		
RGB	Depth	mIOU	mALD	mLR	mIOU	Precision	Recall
✓	✗	57.32	8.87	59.6	85.61	85.17	81.26
✗	✓	58.54	9.41	60.93	86.74	83.61	82.47
✓	✓	59.44	7.83	71.13	89.04	91.46	91.14

TABLE 4. The ablation studies on varying Decoder outputs for the proposed model under Multi input (RGB+D) settings.

Decoder Output		Furrow Line			Furrow Segment		
Line	Segment	mIOU	mALD	mLR	mIOU	Precision	Recall
✓	✗	53.69	12.16	24.92	-	-	-
✗	✓	-	-	-	87.64	85.71	83.26
✓	✓	59.44	7.83	71.13	89.04	91.46	91.14

of the furrow lines. This method helps the model to preserve good quality results for the *mLR* metric as demonstrated in Table 2, indicating that it performs better in maintaining continuous furrow line detection.

4. Dataset Quality: The preprocessing of the dataset, which includes labelling the furrow lines, normalizing the depth data helped reduce noise and ensures the model performance well in furrow line detection. These procedures not only help the model to avoid overfitting to for various field conditions, but also allowing the model to generalize better to unseen data.

B. COMPARISON WITH SOTA MODELS

The better *mLR* performance of our proposed model clearly shows that it is more reliable in detecting furrow lines. This is very crucial in real world applications for autonomous navigation, where the robot's ability to follow the furrow accurately depends on how well the furrow lines are detected in shape and continuity. Our proposed model significantly outperforms other models such as DeepLabV3, which, despite being strong in general segmentation, struggles to maintain the continuity of the groin line, as illustrated in Table 2 for the values *mLR*.

1. Architectural Limitations of Baseline Models: Most of the baseline models we evaluated (e.g., vanilla UNet, DeepLabV3+) are designed for single-task semantic segmentation. These models do not explicitly optimize for centerline continuity or structural preservation. Without an inductive bias or dedicated decoder for centerline detection, they tend to produce coarse or broken predictions. Additionally, their pixel-wise prediction approach lacks the capacity to capture elongated, structurally coherent features such as furrows.

2. Advantages of the Proposed Model: In contrast, our multi-task dual-decoder architecture jointly learns to segment furrow regions and extract centerlines using shared encoded features. This synergy reinforces spatial and structural coherence in the output, leading to longer and more accurate centerline predictions, as reflected in the superior *mLR* metric (71.13% vs. 48.35%) (see Table 2).

3. Modality Fusion and Feature Learning: Our proposed method leverages both RGB and depth data. The RGB input captures rich texture and edge information, while the depth modality provides geometric cues. The fusion of these two modalities helps the model more accurately interpret the spatial continuity of furrow lines under various environmental conditions.

4. Conclusion: Therefore, the performance gap in *mLR* arises from the inability of baseline models to model structured continuity and the strength of our model's multi-task, multi-modal learning design. The proposed architecture is better equipped to handle real world field variability and ensures robust furrow detection essential for autonomous navigation.

C. ABLATION STUDY

The ablation studies, as detailed in Tables 3, 4, and 5 provide deeper insights into the impact of various configurations and components on the performance of the proposed model.

1) ENCODER INPUT CONFIGURATION

Table 3 explores the effect of different encoder inputs on the model's performance under multi-task learning settings. The results show that using both RGB and depth images as inputs yields the best performance across all metrics for both furrow line detection and segmentation. Specifically, the RGB+D configuration achieved an *mIOU* of 59.44% for line detection and 89.04% for segmentation, along with an *mALD* of 7.83 pixels and an *mLR* of 71.13%. This highlights the importance of leveraging both color and depth information, as they provide complementary data that enhances the model's ability to accurately detect and delineate furrows in various environmental conditions.

2) DECODER OUTPUT CONFIGURATION

Table 4 evaluates the impact of different decoder outputs under the RGB+D input configuration. The results indicate that the combination of both line and segment outputs (Line+Segment) produces the best overall performance,

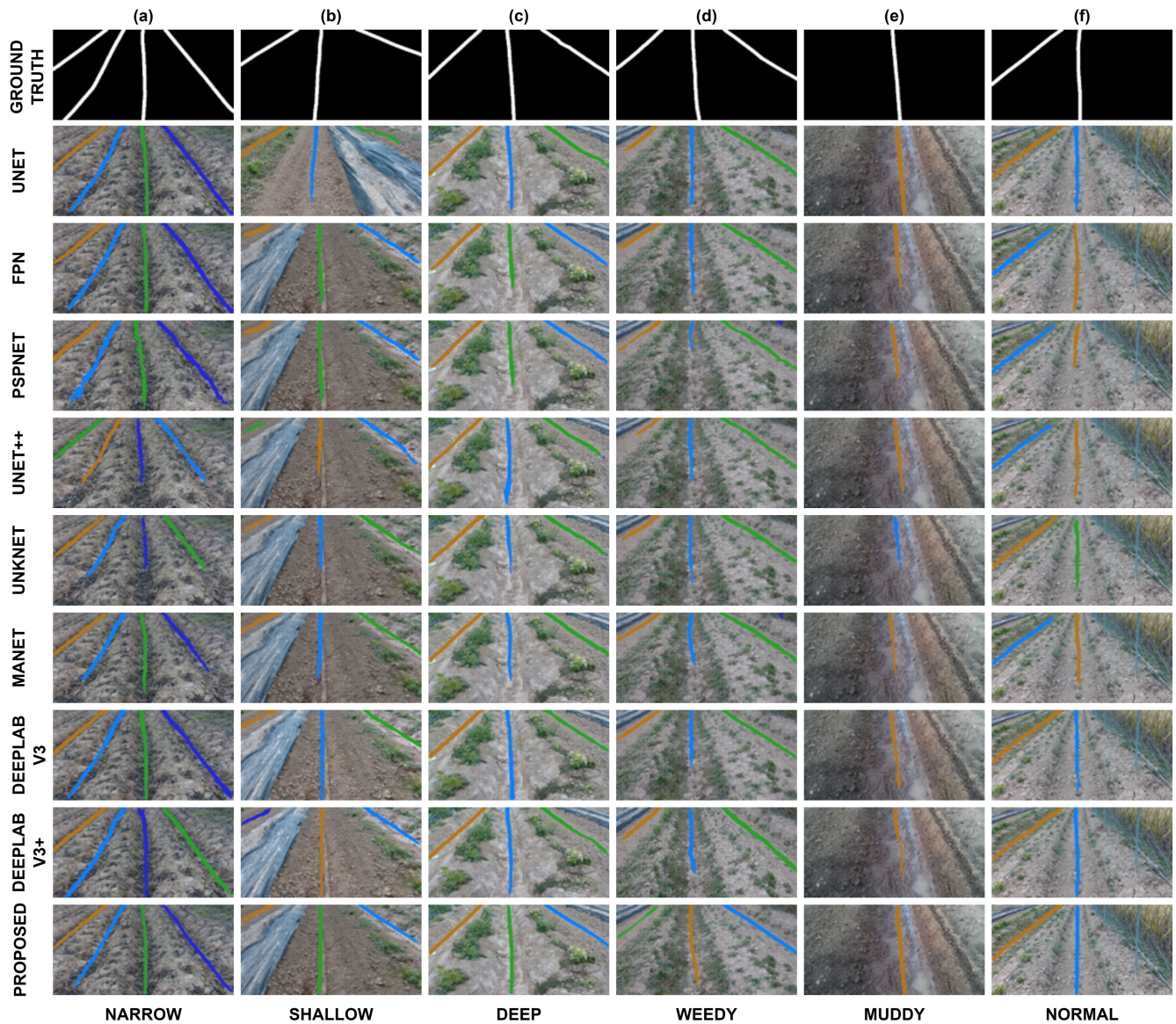


FIGURE 13. The comparison of the proposed method with other pre-trained models on different types of furrows. The figure shows the performance of various models across different furrow conditions (narrow, shallow, deep, weedy, muddy, and normal). Our proposed model accurately detects the center line of the furrow compared to other models.

with an $mIOU$ of 59.44% for line detection, an $mALD$ of 7.83 pixels, and an mLR of 71.13%. For segmentation, the Line+Segment configuration achieved an $mIOU$ of 89.04%, with precision and recall scores of 91.46% and 91.14%, respectively. This configuration not only improves segmentation accuracy but also enhances the detection of furrow lines by utilizing shared features across tasks, resulting in a more robust and consistent model performance.

3) LOSS FUNCTION EVALUATION

The impact of different loss functions on the model's performance is presented in Table 5. The combination of Focal Tversky loss, Combo loss, and Lovász loss yielded the best results across all metrics, with an $mIOU$ of

59.44% for furrow line detection and 89.04% for furrow segmentation. This combination also achieved the lowest $mALD$ of 7.83 pixels and the highest mLR of 71.13%. These results suggest that integrating multiple loss functions helps the model to better handle the class imbalance and improve the learning of critical features, ultimately leading to more accurate and reliable predictions.

D. DISCUSSION

The comprehensive analysis of the results demonstrates that the proposed model, particularly in its multi-task configuration with RGB+D input, significantly outperforms SOTA models across all key metrics. The ablation studies confirm that the strategic combination of RGB and depth

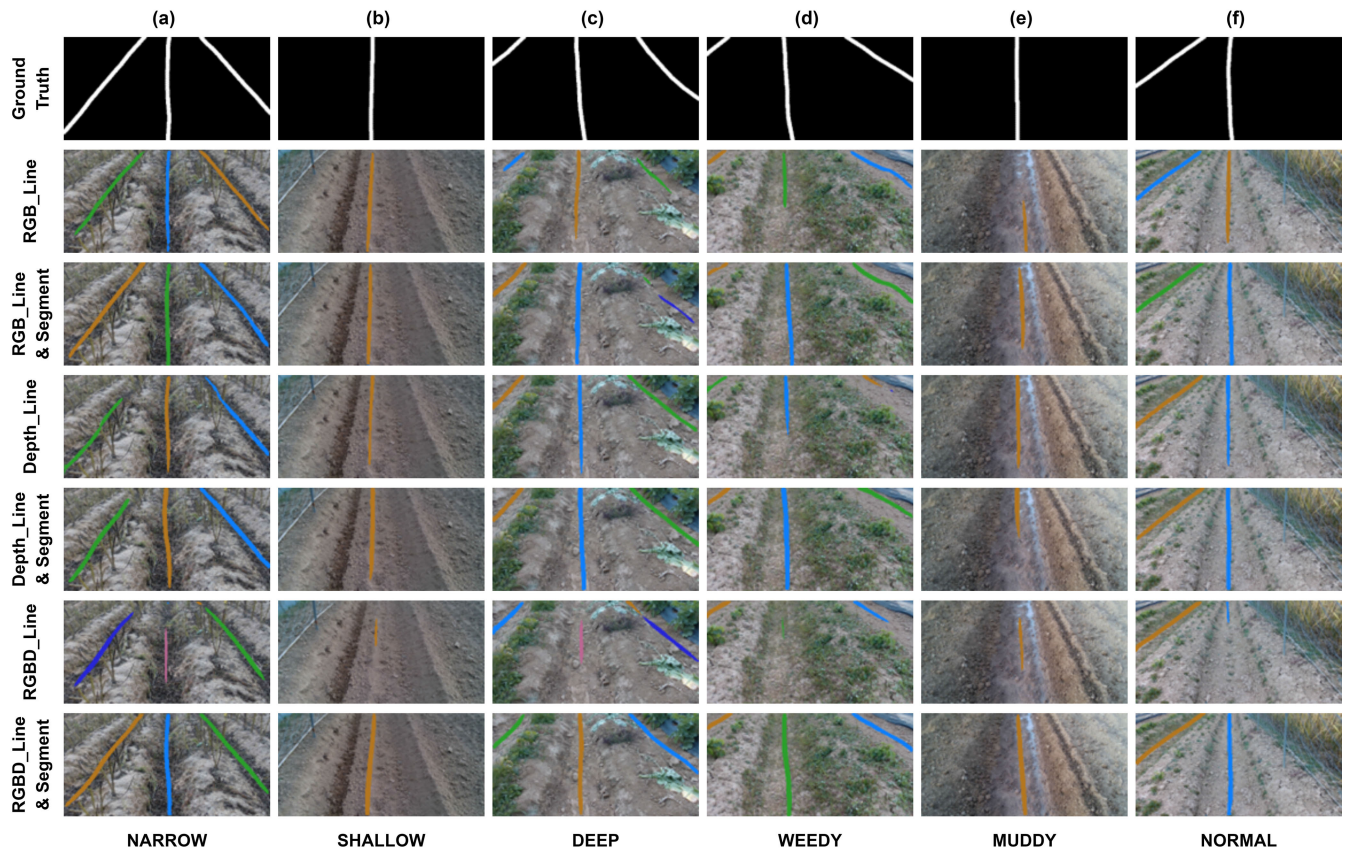


FIGURE 14. The figure illustrates how various configurations fare under different furrow conditions, including narrow, shallow, deep, weedy, muddy, and normal.

TABLE 5. Ablation study on performance improvements using different loss functions.

Loss		Furrow Line			Furrow Segment		
Seg.	Line	mIOU	mALD	mLR	mIOU	Precision	Recall
$\mathcal{L}_{BCE}$	$\mathcal{L}_{FT}$	49.67	9.75	58.33	80.6	82.56	81.96
$\mathcal{L}_{BCE}$	$\mathcal{L}_{combo}$	52.34	8.69	70.64	75.78	79.63	81.47
$\mathcal{L}_{BCE}$	$\mathcal{L}_{Lovász}$	55.68	8.03	65.11	79.54	80.43	79.88
$\mathcal{L}_{BCE}$	$\mathcal{L}_{FT} + \mathcal{L}_{combo} + \mathcal{L}_{Lovász}$	59.44	7.83	71.13	89.04	91.46	91.14

inputs, along with the use of both line and segment outputs in the decoder, contributes substantially to the model's superior performance. Additionally, the use of advanced loss functions plays a crucial role in optimizing the model's ability to accurately detect and segment furrows, even in challenging agricultural environments.

The findings of this study have important implications for the development of autonomous agricultural systems. The proposed model's ability to maintain high accuracy and reliability across different furrow types and environmental conditions makes it a valuable tool for precision agriculture, where the accurate detection of furrows is critical for tasks such as planting, weeding, and harvesting. The results also suggest that future research could further explore the

potential of multi-task learning and the integration of multiple data modalities to enhance the performance of autonomous systems in agriculture, as shown in Figure 14.

Overall, the proposed model not only demonstrates state-of-the-art performance in furrow detection and segmentation but also provides a robust and adaptable solution that is well-suited to the diverse and dynamic challenges of modern agriculture.

VI. CONCLUSION

In this paper, we proposed a deep learning-based Multi-Input Multi-Output (MIMO) network designed to accurately detect the centerlines of various furrow types, effectively eliminating the need for post-processing.



FIGURE 15. (a) The robot setup that shows the view of the real sense camera (b) The real sense camera which can take both RGB and depth images that are combined and given to the model to get the prediction line on the furrow.

Our framework is composed of three key components. The first is a dual encoder that processes RGB and depth images, integrating their features by summing them at the network's end. These combined features are then passed through a channel-limiting network before being fed into the second component, the Deep Multi-scale Feature Fusion (DFF) module. This component scales the feature maps to different sizes while maintaining correlations across scales through cross-connections. The third component consists of two interlinked decoders that re-use the multiscale features learned in the DFF module to compute segmentations and lines. The interconnection between the decoders establishes a relationship between the segments and their corresponding centerline, improving the accuracy of the centerline predictions by leveraging the segmented furrow features.

To validate our approach, we constructed a diverse dataset featuring various types of furrows, including deep, normal, muddy, shallow, weedy, and narrow. We compared the performance of our proposed model against leading state-of-the-art methods, and our experiments demonstrated that the MIMO network, combined with our training strategy, outperforms all other models. We evaluated the algorithm using several metrics, including mean Average Lateral Deviation (*mALD*) and mean Line Ratio (*mLR*), achieving improvements of 7.8 pixels in (*mALD*) and 71.13% in (*mLR*), respectively, for furrow centerline detection.

Our results show that the proposed algorithm not only accurately predicts the centerlines of various furrow types across different environments but also outperforms previous methods. We believe that this methodology offers a significant advancement in precision agriculture, serving as a valuable reference for future research on autonomous agricultural robots.

A. FUTURE WORK ON AUTONOMOUS AGRICULTURE ROBOT

In a previous study [30], a robot was employed to effectively eliminate weeds. As shown in Figure 15, It is now possible through the implementation of a deep learning model on an autonomous agriculture vehicle equipped with a RealSense camera. This innovative solution makes use of both RGB and depth image capture, which is a breakthrough in the field of precision agriculture. As the vehicle moves through the field, the RealSense camera captures rich visual data,

which is then processed in real time by the deep learning model trained to detect furrow centerlines. By utilizing semantic segmentation, the model can accurately identify and map out the centerlines, allowing the vehicle to navigate autonomously along the furrows. This navigation capability, guided by the detected centerlines of furrows, makes it possible to perform various agricultural tasks, such as seeding, weeding, and harvesting, with unparalleled precision and efficiency.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Muhammad Ibrahim Zain Ul Abideen: Conceptualization, Investigation, Methodology, Validation, Formal analysis, Writing-original draft, Collected and labelled data, Investigation, Visualization, Discussed the results and contributed to the final manuscript. Dewa Made Sri Arsa: Methodology, Validation, Formal analysis, Investigation, Supervision, Visualization, Discussed the results and contributed to the final manuscript. Talha Ilyas: Methodology, Validation, Formal analysis, Supervision, review & editing, Investigation, Visualization, Discussed the results and contributed to the final manuscript. Hyunggi Jo: Discussed the results and contributed to the final manuscript. Sang Cheol Kim: Discussed the results and contributed to the final manuscript. Hyongsuk Kim: Conceptualization, Investigation, Methodology, Validation, Formal analysis, Supervision, Writing-review & editing, Funding acquisition, Discussed the results and contributed to the final manuscript.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

Data will be made available on request.

REFERENCES

- [1] Y. Morio, K. Teramoto, and K. Murakami, "Vision-based furrow line detection for navigating intelligent worker assistance robot," *Eng. Agricult., Environment Food*, vol. 10, no. 2, pp. 87–103, 2017.
- [2] N. S. Naik, V. V. Shete, and S. R. Danve, "Precision agriculture robot for seeding function," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, vol. 2, Aug. 2016, pp. 1–3.
- [3] G. A. Mesías-Ruiz, M. Pérez-Ortiz, J. Dorado, A. I. de Castro, and J. M. Peña, "Boosting precision crop protection towards agriculture 5.0 via machine learning and emerging technologies: A contextual review," *Frontiers Plant Sci.*, vol. 14, Mar. 2023, Art. no. 1143326.
- [4] A. T. Meshram, A. V. Vanalkar, K. B. Kalambe, and A. M. Badar, "Pesticide spraying robot for precision agriculture: A categorical literature review and future trends," *J. Field Robot.*, vol. 39, no. 2, pp. 153–171, Mar. 2022.
- [5] H. Zhou, X. Wang, W. Au, H. Kang, and C. Chen, "Intelligent robots for fruit harvesting: Recent developments and future challenges," *Precis. Agricult.*, vol. 23, no. 5, pp. 1856–1907, Oct. 2022.
- [6] G. Kootstra, X. Wang, P. M. Blok, J. Hemming, and E. van Henten, "Selective harvesting robotics: Current research, trends, and future directions," *Current Robot. Rep.*, vol. 2, no. 1, pp. 95–104, Mar. 2021.
- [7] A. Caruso, S. Chessa, S. Escobar, J. Barba, and J. C. López, "Collection of data with drones in precision agriculture: Analytical model and LoRa case study," *IEEE Internet Things J.*, vol. 8, no. 22, pp. 16692–16704, Nov. 2021.

- [8] C. Yun, H.-J. Kim, C.-W. Jeon, M. Gang, W. S. Lee, and J. G. Han, "Stereo vision-based ridge-furrow detection and tracking for auto-guided cultivator," *Comput. Electron. Agricult.*, vol. 191, Dec. 2021, Art. no. 106490.
- [9] S.-J. Lin, Q. Wun Chen, and J.-J. Chen, "Ridge-furrow detection in glycine max farm using deep learning," in *Proc. Int. Conf. Pervasive Artif. Intell. (ICPAI)*, Dec. 2020, pp. 183–187.
- [10] S. P. Adhikari, H. Yang, and H. Kim, "Learning semantic graphics using convolutional encoder-decoder network for autonomous weeding in paddy," *Frontiers Plant Sci.*, vol. 10, p. 1404, Oct. 2019.
- [11] M. Schwarz, A. Milan, A. S. Periyasamy, and S. Behnke, "RGB-D object detection and semantic segmentation for autonomous manipulation in clutter," *Int. J. Robot. Res.*, vol. 37, nos. 4–5, pp. 437–451, Apr. 2018.
- [12] D. Li, C. Dong, B. Li, Y. Wang, and J. Wang, "Hybrid Rice row detection at the pollination stage based on vanishing point and line-scanning method," *Precis. Agricult.*, vol. 24, no. 3, pp. 921–947, Jun. 2023.
- [13] S. M. Abid Hasan and K. Ko, "Depth edge detection by image-based smoothing and morphological operations," *J. Comput. Des. Eng.*, vol. 3, no. 3, pp. 191–197, Jul. 2016.
- [14] Y. Nagasaka, H. Saito, K. Tamaki, M. Seki, K. Kobayashi, and K. Taniwaki, "An autonomous Rice transplanter guided by global positioning system and inertial measurement unit," *J. Field Robot.*, vol. 26, nos. 6–7, pp. 537–548, Jun. 2009.
- [15] K. H. Choi, S. K. Han, S. H. Han, K.-H. Park, K.-S. Kim, and S. Kim, "Morphology-based guidance line extraction for an autonomous weeding robot in paddy fields," *Comput. Electron. Agricult.*, vol. 113, pp. 266–274, Apr. 2015.
- [16] Q. Zhang, M. E. S. Chen, and B. Li, "A visual navigation algorithm for paddy field weeding robot based on image understanding," *Comput. Electron. Agricult.*, vol. 143, pp. 66–78, Dec. 2017.
- [17] D. Li, B. Li, S. Long, H. Feng, Y. Wang, and J. Wang, "Robust detection of headland boundary in paddy fields from continuous RGB-D images using hybrid deep neural networks," *Comput. Electron. Agricult.*, vol. 207, Apr. 2023, Art. no. 107713.
- [18] H. Mousazadeh, "A technical review on navigation systems of agricultural autonomous off-road vehicles," *J. Terramechanics*, vol. 50, no. 3, pp. 211–232, Jun. 2013.
- [19] T. Ilyas, D. M. S. Arsa, K. Ahmad, Y. C. Jeong, O. Won, J. H. Lee, and H. Kim, "CWD30: A comprehensive and holistic dataset for crop weed recognition in precision agriculture," 2023, *arXiv:2305.10084*.
- [20] J. Wilson, "Guidance of agricultural vehicles? A historical perspective," *Comput. Electron. Agricult.*, vol. 25, nos. 1–2, pp. 3–9, 2000.
- [21] L. Zhang, W. S. Mao, and C. B. Qi, "Crop-edge detection based on machine vision," *New Zealand J. Agricult. Res.*, vol. 50, no. 5, pp. 1367–1374, 2007.
- [22] B. Chen, S. Tojo, and K. Watanabe, "Machine vision based guidance system for automatic rice transplanters," *Appl. Eng. Agricult.*, vol. 19, no. 1, p. 91, 2003.
- [23] Y. Li, Z. Hong, D. Cai, Y. Huang, L. Gong, and C. Liu, "A SVM and SLIC based detection method for paddy field boundary line," *Sensors*, vol. 20, no. 9, p. 2610, May 2020.
- [24] T. Ilyas, J. Lee, O. Won, Y. Jeong, and H. Kim, "Overcoming field variability: Unsupervised domain adaptation for enhanced crop-weed recognition in diverse farmlands," *Frontiers Plant Sci.*, vol. 14, Aug. 2023, Art. no. 1234616.
- [25] D. Li, B. Li, S. Long, H. Feng, T. Xi, S. Kang, and J. Wang, "Rice seedling row detection based on morphological anchor points of Rice stems," *Biosyst. Eng.*, vol. 226, pp. 71–85, Feb. 2023.
- [26] Y. He, X. Zhang, Z. Zhang, and H. Fang, "Automated detection of boundary line in paddy field using MobileV2-UNet and RANSAC," *Comput. Electron. Agricult.*, vol. 194, Mar. 2022, Art. no. 106697.
- [27] M. Marshall, S. Crommelinck, D. Kohli, C. Perger, M. Y. Yang, A. Ghosh, S. Fritz, K. D. Bie, and A. Nelson, "Crowd-driven and automated mapping of field boundaries in highly fragmented agricultural landscapes of Ethiopia with very high spatial resolution imagery," *Remote Sens.*, vol. 11, no. 18, p. 2082, Sep. 2019.
- [28] L. Yan, J. Huang, H. Xie, P. Wei, and Z. Gao, "Efficient depth fusion transformer for aerial image semantic segmentation," *Remote Sens.*, vol. 14, no. 5, p. 1294, Mar. 2022.
- [29] L. Quan, H. Li, H. Li, W. Jiang, Z. Lou, and L. Chen, "Two-stream dense feature fusion network based on RGB-D data for the real-time prediction of weed aboveground fresh weight in a field environment," *Remote Sens.*, vol. 13, no. 12, p. 2288, Jun. 2021.
- [30] D. M. S. Arsa, T. Ilyas, S.-H. Park, O. Won, and H. Kim, "Eco-friendly weeding through precise detection of growing points via efficient multi-branch convolutional neural networks," *Comput. Electron. Agricult.*, vol. 209, Jun. 2023, Art. no. 107830.
- [31] H. Zhou, L. Qi, H. Huang, X. Yang, Z. Wan, and X. Wen, "CANet: Co-attention network for RGB-D semantic segmentation," *Pattern Recognit.*, vol. 124, Apr. 2022, Art. no. 108468.
- [32] S. E. A. Raza, L. Cheung, D. Epstein, S. Pelengaris, M. Khan, and N. M. Rajpoot, "MIMO-Net: A multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 337–340.
- [33] D. Igonin and Y. V. Tiumentsev, "Multi-input convolutional neural networks in real-time semantic segmentation tasks," in *Proc. Int. Conf. Neuroinform.*, 2022, pp. 381–394.
- [34] T. Ilyas, K. Ahmad, D. M. S. Arsa, Y. C. Jeong, and H. Kim, "Enhancing medical image analysis with unsupervised domain adaptation approach across microscopes and magnifications," *Comput. Biol. Med.*, vol. 170, Mar. 2024, Art. no. 108055.
- [35] A. Takagaki, R. Masuda, M. Iida, and M. Suguri, "Image processing for ridge/furrow discrimination for autonomous agricultural vehicles navigation," *IFAC Proc. Volumes*, vol. 46, no. 18, pp. 47–51, Aug. 2013.
- [36] R. Ji and L. Qi, "Crop-row detection algorithm based on random Hough transformation," *Math. Comput. Model.*, vol. 54, nos. 3–4, pp. 1016–1020, Aug. 2011.
- [37] T. Ilyas and H. Kim, "A deep learning based approach for strawberry yield prediction via semantic graphics," in *Proc. 21st Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2021, pp. 1835–1841.
- [38] D. Yashaswini, Sneha, and N. Kumar, "HTML code generation from website images and sketches using deep learning-based encoder-decoder model," in *Proc. IEEE 4th Int. Conf. Cybern., Cognition Mach. Learn. Appl. (ICCCMLA)*, Oct. 2022, pp. 133–138.
- [39] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with Atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 833–851.
- [40] Y. Liu, K. Chen, C. Liu, Z. Qin, Z. Luo, and J. Wang, "Structured knowledge distillation for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2599–2608.
- [41] N. U. Islam, S. Lee, and J. Park, "Accurate and consistent image-to-image conditional adversarial network," *Electronics*, vol. 9, no. 3, p. 395, Feb. 2020.
- [42] N. U. Islam and J. Park, "Depth estimation from a single RGB image using fine-tuned generative adversarial network," *IEEE Access*, vol. 9, pp. 32781–32794, 2021.
- [43] I. Armeni, Z.-Y. He, A. Zamir, J. Gwak, J. Malik, M. Fischer, and S. Savarese, "3D scene graph: A structure for unified semantics, 3D space, and camera," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5664–5673.
- [44] V. Balaska, L. Bampis, and A. Gasteratos, "Graph-based semantic segmentation," in *Proc. 27th Int. Conf. Robot. Alpe-Adria Danube Region (RAAD)*, 2018, pp. 572–579.
- [45] N. Pourian, S. Karthikeyan, and B. S. Manjunath, "Weakly supervised graph based semantic segmentation by learning communities of image-parts," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1359–1367.
- [46] P. Arbeláez, B. Hariharan, C. Gu, S. Gupta, L. Bourdev, and J. Malik, "Semantic segmentation using regions and parts," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3378–3385.
- [47] S.-J. Luo, C.-L. Liu, B.-Y. Chen, and K.-L. Ma, "Ambiguity-free edge-bundling for interactive graph visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 5, pp. 810–821, May 2012.
- [48] J. Lee, T. Ilyas, H. Jin, J. Lee, O. Won, H. Kim, and S. J. Lee, "A pixel-level coarse-to-fine image segmentation labelling algorithm," *Sci. Rep.*, vol. 12, no. 1, p. 8672, May 2022.
- [49] R. Yao, G. Lin, C. Shen, Y. Zhang, and Q. Shi, "Semantics-aware visual object tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 6, pp. 1687–1700, Jun. 2019.
- [50] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. C. Bovik, and Y. Li, "MaxViT: Multi-axis vision transformer," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 459–479.
- [51] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10781–10790.

- [52] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
- [53] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687.
- [54] S. A. Taghanaki, Y. Zheng, S. Kevin Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Med. Imag. Graph.*, vol. 75, pp. 24–33, Jul. 2019.
- [55] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4413–4421.
- [56] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 234–241.
- [57] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
- [58] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.
- [59] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.
- [60] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.
- [61] R. Li, S. Zheng, C. Zhang, C. Duan, J. Su, L. Wang, and P. M. Atkinson, "Multiattention network for semantic segmentation of fine-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.
- [62] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous convolution for semantic image segmentation," in *Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 1–10.



MUHAMMAD IBRAHIM ZAIN UL ABIDEEN

received the B.Sc. degree in electrical engineering from the FAST-National University of Computer and Emerging Sciences (NUCES), Peshawar, Pakistan, in 2018, and the M.S. degree in electronics and information engineering from Jeonbuk National University, Jeonju, South Korea, in 2024, where he is currently pursuing the Ph.D. degree in electronics and information engineering. His research interests include computer vision, deep

learning, precision agriculture, and the integration of artificial intelligence with robotics.



DEWA MADE SRI ARSA received the bachelor's degree from Universitas Udayana, Indonesia, and the master's degree from the Faculty of Computer Science, Universitas Indonesia, Indonesia. He is currently pursuing the Ph.D. degree with Jeonbuk National University, Republic of Korea. His research interests include computer vision, artificial intelligence, and image processing, including its application in agriculture and medical imaging.



ture, optical flow, CAM, GAN, and depth estimation.

TALHA ILYAS received the bachelor's degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2017, and the M.S. degree in electronics and information engineering from Jeonbuk National University, Jeonju, South Korea, in 2021, where he is currently pursuing the Ph.D. degree in electronics and information engineering. His research interests include object instance detection, medical image analysis, object tracking, precision agriculture, optical flow, CAM, GAN, and depth estimation.



HYUNGGI JO (Member, IEEE) received the B.S. and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2012 and 2020, respectively. He is currently an Assistant Professor with the Division of Electronic Engineering, Jeonbuk National University, Jeonju, South Korea. His research interests include smart mobility, sensor fusion, visual localization, and simultaneous localization and mapping.



Professor with the Intelligent Robot Research Institute, Jeonbuk National University. His research interests include developing smart farm automation systems and robots.

SANG CHEOL KIM received the Bachelor of Science degree from the Department of Electronic Calculation, Korea National Open University, in 1997, and the master's degree from the Department of Agricultural Machinery, Seoul National University, in 2002, and the Ph.D. degree from the Department of Agricultural Machinery, Gongju University, in 2011. From 1993 to 2019, he was a Researcher with the Rural Development Administration. Since 2020, he has been a Research



Professor with the Division of Electronics Engineering, Jeonbuk National University, Jeonju, South Korea. His current research interests include memristors and its application to the implementation of neural networks.

HYONGSUK KIM (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Missouri, Columbia, MO, USA, in 1992. From 2000 to 2002 and from 2009 to 2010, he was a Visiting Scholar with the Nonlinear Electronics Laboratory, Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA, USA. Since 1993, he has been a Professor with the Division of Electronics Engineering,

...