DNN Models for WiFi CSI-based Underground Image Reconstruction

Said Elhadi[1] and Yang Zhao[1]

Harbin Institute of Technology, Shenzhen, China
said.elhadi1080@gmail.com, yang.zhao@hit.edu.cn

Abstract. Non-invasive monitoring of underground biomass, such as root tubers, is crucial for advancing precision agriculture. This paper presents a wireless sensing system utilizing Wi-Fi Channel State Information (CSI) collected from a low-cost ESP32-C3-based mesh network for reconstructing high-resolution cross-sectional images of underground tubers. We developed an automated data acquisition testbed incorporating a synchronized Time Division Multiple Access (TDMA) protocol, multi-channel hopping, automated platform rotation, and camera-based ground truth generation, yielding a dataset of 256 labeled samples from 8 morphologically diverse potato tubers across 32 rotational positions, using which a WiFi CSI dataset is built for training and evaluating underground image reconstruction models. We evaluate five deep neural network (DNN) architectures—including U-Net variants, FCN, and DeepLabV3+—for mapping wireless signal tensors (CSI, RSSI, and fused CSI+RSSI) to tuber images. Statistical analysis across three repeated runs confirms that CSI significantly outperforms RSSI ($p < 0.005$ for all architectures). Fusing CSI and RSSI yields the best overall performance, with a DeepLabV3+ model achieving a mean Dice score of 0.8248 ± 0.0037 and an IoU of 0.7179 ± 0.0110. Ablation studies quantify the impact of channel diversity, node density, and subcarrier utilization, providing design guidelines for practical deployments.

Keywords: Wireless Sensor Network · Wi-Fi Sensing · Deep Neural Networks · Underground Imaging · Precision Agriculture.

1 Introduction

The ability to monitor below-ground biomass development is fundamental to optimizing crop yields and resource management in smart agriculture [7]. Root tubers, such as potatoes, represent a significant portion of global food production, making their non-invasive monitoring particularly valuable for yield prediction and growth assessment. However, sensing underground structures presents considerable challenges due to signal attenuation, scattering, and the heterogeneous nature of soil environments.

Traditional methods often involve destructive sampling or expensive equipment like Ground Penetrating Radar (GPR) [5], which suffers from high cost and limited scalability. Low-cost wireless sensing offers a promising alternative.

Early work demonstrated the feasibility of using Received Signal Strength Indicator (RSSI) from Zigbee networks for tuber imaging [12]. However, RSSI is a scalar power metric highly susceptible to multipath fading, lacking the resolution required for detailed underground imaging.

Wi-Fi Channel State Information (CSI), now accessible on low-cost devices such as the ESP32-C3, is a more suitable sensing primitive because it exposes the subcarrier-level frequency response of the wireless channel rather than a single received-power summary. This richer representation is sensitive to small dielectric disturbances caused by buried objects [9], and similar fine-grained CSI measurements have already supported demanding inference tasks such as activity recognition [10] and fruit sensing [11, **?**]. Yet underground image reconstruction is substantially more challenging: the object signatures are weak, the soil medium is lossy and heterogeneous, and training a reliable model requires stable multi-node acquisition together with trustworthy image labels. As a result, the key questions are still open: can low-cost Wi-Fi CSI support useful underground image reconstruction, how large is the gain over RSSI, does RSSI still offer complementary information when fused with CSI, and which deep architecture is most effective under this sensing geometry? This paper addresses these questions through a complete sensing-and-learning pipeline built around a 16-node ESP32-C3 platform, a labeled rotational dataset, and a systematic comparison of reconstruction models and sensing modalities.

To summarize, our contributions are:

- We develop a complete low-cost underground sensing platform built around a 16-node ESP32-C3 ring, synchronized Time Division Multiple Access (TDMA), 14-channel hopping, motorized rotation, and Segment Anything Model (SAM)-based label generation. This system produces a reproducible 256-sample labeled dataset spanning 8 morphologically diverse potato tubers and 32 angular positions, providing the experimental basis needed for supervised underground image reconstruction.
- We carry out a systematic benchmark of five DNN architectures using RSSI, CSI, and fused CSI+RSSI tensors to determine how sensing modality and model design affect reconstruction quality. The results clearly show that CSI is consistently more informative than RSSI alone, while fusing CSI and RSSI with DeepLabV3+ yields the strongest overall reconstruction accuracy.
- We perform targeted ablation studies on channel diversity, node density, subcarrier utilization, and fusion strategy to identify which sensing resources actually drive performance. These analyses turn the benchmark into practical design guidance by showing how much accuracy can be retained, or lost, when the hardware or spectral resources are reduced.

The paper is organized as follows: Section 2 formulates the imaging problem. Section 3 details the hardware, TDMA protocol, and dataset. Section 4 describes the preprocessing pipeline and models. Section 5 presents experimental results. Section 6 concludes.

2 Problem Statement

The objective is to reconstruct a 2D cross-sectional occupancy mask $\mathbf{M} \in \{0,1\}^{H\times W}$ representing an underground tuber from wireless measurements. The mask dimensions are $H = W = 360$ pixels. The reconstruction maps the collected signal tensor $\mathbf{X}$ to $\hat{\mathbf{M}}$ via a DNN f:

$$\hat{\mathbf{M}} = f(\mathbf{X}; \Theta), \tag{1}$$

where Θ represents the trainable parameters. The input $\mathbf{X}$ is derived from RSSI, CSI, or a fusion of both.

2.1 Received Signal Strength Indicator (RSSI)

In a multipath environment, the complex baseband voltage at the receiver is the superposition of N_{paths} propagation components [6]:

$$V = \sum_{i=1}^{N_{\text{paths}}} |V_i|\, e^{-j\theta_i}, \tag{2}$$

where $|V_i|$ and θ_i are the amplitude and phase of the i-th path. RSSI collapses this multipath structure into a single power value:

$$\text{RSSI} = 10\log_{10}(|V|^2). \tag{3}$$

This irreversible aggregation discards all phase and frequency-selective information, making RSSI fundamentally limited for high-resolution imaging.

2.2 Channel State Information (CSI)

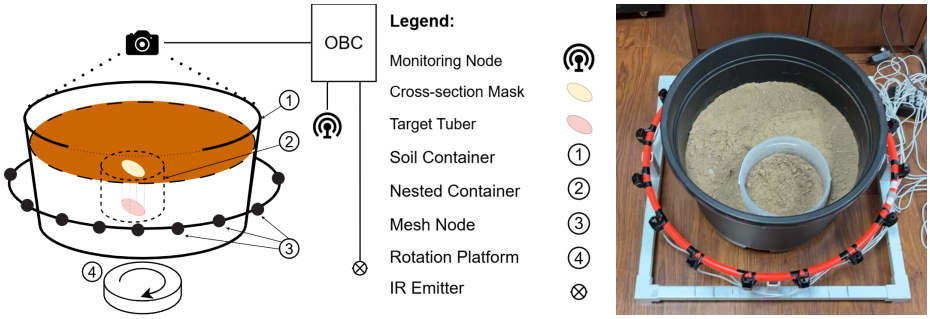
CSI resolves the frequency-selective nature of the wireless channel across K active OFDM subcarriers. For the k-th subcarrier at frequency f_k, the complex channel estimate is:

$$H(f_k) = |H(f_k)|\, e^{j\angle H(f_k)}, \tag{4}$$

where $|H(f_k)|$ captures frequency-dependent attenuation and $\angle H(f_k)$ captures phase shift due to multipath. Unlike RSSI, CSI provides a multi-dimensional fingerprint sensitive to the dielectric contrast between a buried tuber (approximately 80% water, $\epsilon_r' \approx 80$) and surrounding dry soil ($\epsilon_r' \approx 3$–15 at 2.4 GHz). The ESP32-C3 provides 64 raw subcarrier estimates per packet, of which **52 active data subcarriers** are used in our feature tensors.

3 System Architecture and Data Acquisition

The overall architecture is illustrated in Fig. 1. The system consists of three subsystems: (1) a hardware sensing platform, (2) a communication and control subsystem, and (3) an offline data processing and reconstruction pipeline.



(a) System overview schematic.

(b) Physical testbed with the 16-node sensing ring.

Fig. 1. The automated underground sensing testbed. Sixteen ESP32-C3 nodes are arranged in a circular ring (1.2 m diameter) around the soil container. A Raspberry Pi coordinator manages the network, an overhead camera generates ground truth, and a motorized platform provides 32 angular positions.

3.1 16-Node ESP32-C3 Sensing Ring

We constructed a sensing ring of 16 **ESP32-C3** microcontrollers distributed uniformly on a circle of **60 cm radius (1.2 m diameter)** surrounding a centrally positioned soil container. The ESP32-C3 was chosen for its low unit cost ($< \$3$), integrated 2.4 GHz Wi-Fi, and raw CSI access via the Espressif ESP-IDF API. Standard omnidirectional 2.4 GHz antennas were mounted horizontally on all nodes to ensure consistent vertical E-field polarization and uniform azimuthal coverage.

The circular topology provides (i) uniform 360ř angular coverage without directional bias, and (ii) the highest density of link crossings concentrated at the center of interest. The 16-node ring yields $16 \times 15 = 240$ unique unidirectional links per Wi-Fi channel.

3.2 TDMA Protocol, Channel Hopping, and System Reliability

A dedicated Monitor Node (also an ESP32-C3) acts as the TDMA master, synchronizing all 16 sensing nodes. A single data collection cycle consists of **16 time slots of 100 ms each** (Fig. 2). In slot i, Node i broadcasts probe packets while the remaining 15 nodes log CSI and RSSI. The Monitor Node aggregates all packets and forwards them to the Raspberry Pi coordinator over USB/Serial.

After each complete 16-slot cycle, the network synchronously hops to the next Wi-Fi channel across **14 channels** in the 2.4 GHz ISM band, yielding $16 \times 15 \times 14 = 3{,}360$ distinct spatial-frequency link measurements per angular position.

Measured Protocol Reliability. Across multiple 2-hour experimental runs, the maximum inter-slot timing jitter was **below ±5 ms** (well within the 100 ms

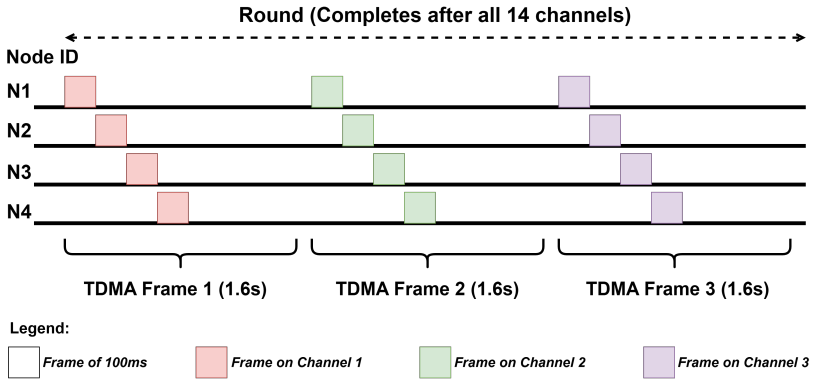
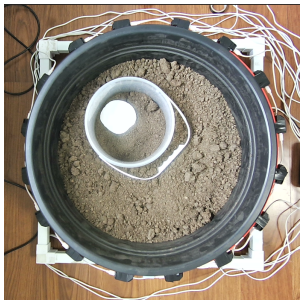
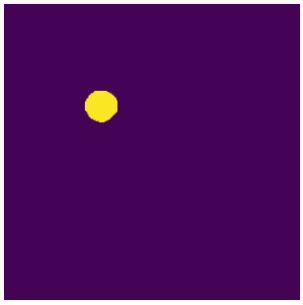


Fig. 2. TDMA timing diagram for one 16-slot collection cycle on a single Wi-Fi channel. Each slot is 100 ms; Node i transmits probe packets while the remaining 15 nodes receive and log CSI/RSSI data. After all 16 slots, the network hops synchronously to the next channel.



(a) Camera image of surface marker.



(b) SAM-generated binary ground truth (360×360 px).

Fig. 3. Automated ground truth generation. A cross-section cutout matching the tuber is placed on the soil surface and segmented by the Segment Anything Model (SAM) to produce the training label.

slot), and the end-to-end **Packet Delivery Rate (PDR) exceeded 98%**. The sub-2% packet loss, attributed to transient RF interference, is handled by the preprocessing pipeline (Section 4.1).

3.3 Testbed, Biological Samples, and Dataset

Soil and Burial Conditions. The testbed uses a cylindrical non-metallic container (40 cm diameter) filled with standardized potting mix (sphagnum peat, perlite, vermiculite). Volumetric water content was controlled at $\theta_v \approx 10\%$ and verified periodically with a calibrated soil probe. All tubers were buried at a standardized depth of 15 cm with their longest axis horizontal. A nested inner

container enabled rapid, reproducible sample swaps without disturbing the bulk soil.

Biological Samples. Eight potato tubers sourced from a local supplier were selected to cover a wide range of morphologies. Table 1 lists their characteristics. The diversity in mass (108–310 g), length (5.5–14.1 cm), and aspect ratio is essential for training generalizable models.

Table 1. Characteristics of Potato Tuber Samples Used in Experiments.

Tuber ID	Mass (g)	Max Length (cm)	Max Width (cm)
01	155	10.2	6.5
02	110	7.1	5.8
03	140	5.5	4.9
04	250	12.8	7.9
05	180	9.5	7.0
06	108	8.9	3.5
07	220	8.5	8.0
08	310	14.1	8.3

Data Collection. A motorized rotation platform advances the container through **32 discrete steps of 11.25ř each** (full 360ř). At each angular position: (1) the overhead camera captures a high-resolution image processed by SAM to generate the 360×360 binary ground truth mask (Fig. 3); (2) wireless data is logged for **2 minutes**, covering multiple complete 14-channel hopping cycles. One full run takes approximately 1 hour per tuber. The dataset contains **256 labeled samples** (8 tubers $\times$ 32 positions), split into training/validation (85%) and test (15%) sets stratified by tuber ID.

4 Algorithms and Deep Learning Framework

4.1 Data Preprocessing and Tensor Formulation

CSI Amplitude Extraction. The ESP32-C3 provides raw CSI as integer pairs $(\Re, \Im)$ for 64 subcarriers per packet. The amplitude for subcarrier k is computed as:

$$|H(f_k)| = \sqrt{\Re(H(f_k))^2 + \Im(H(f_k))^2}. \tag{5}$$

Raw CSI phase is discarded because low-cost chipsets suffer from Carrier Frequency Offset (CFO) and random packet detection delays, which introduce large uncorrelated phase offsets between packets [9]. Of the 64 raw subcarriers, **52 active data subcarriers** are retained. For each unique (Tx, Rx, channel) triplet over the 2-minute logging window, multiple measurements are aggregated by their median to reduce noise.

Tensor Formulation. Three input tensors are constructed per sample. Let N=16 (nodes), C=14 (channels), and S=52 (subcarriers). Because the 16 nodes are arranged in a circular ring, the pairwise links can be mapped to an adjacency-like matrix where element (i, j) corresponds to the signal path from transmitter node i to receiver node j. While this representation organizes spatial relationships topologically rather than in physical Euclidean coordinates, 2D convolutional layers can still extract localized patterns from adjacent node pairings. The tensors are defined as:

- **RSSI Tensor** $X_{\text{RSSI}} \in \mathbb{R}^{C\times N\times N}$ ($14\times16\times16$): element (c, i, j) is the median RSSI on the link $i\to j$ for channel c.
- **CSI Tensor** $X_{\text{CSI}} \in \mathbb{R}^{(C\cdot S)\times N\times N}$ ($728 \times 16 \times 16$): the channel dimension concatenates all 52 CSI amplitudes across all 14 channels.
- **Fused Tensor** $X_{\text{Fused}} \in \mathbb{R}^{C(1+S)\times N\times N}$ ($742 \times 16 \times 16$): for each channel, the RSSI scalar is prepended to the 52 CSI amplitudes to maintain modality alignment and frequency-selective ordering.

Diagonal entries representing self-links (i=j) are set to zero. All tensors are ℓ_2-normalized channel-wise prior to network ingestion to account for variation in transmit power and system gains.

4.2 Feature Fusion Strategy

To leverage the complementary nature of RSSI and CSI, we evaluated three fusion strategies on the DeepLabV3+ backbone: Simple Concatenation (the fused tensor processed by a single shared encoder), SE-Weighted Fusion [2] (separate branches combined via Squeeze-and-Excitation recalibration), and Cross-Attention Fusion (one branch attends to key-value representations from the other).

Simple Concatenation consistently outperformed the attention-based variants (Dice: 0.807 vs. 0.745 for SE, 0.722 for Cross-Attention). The most likely explanation is that the current low-data regime (256 samples, 16×16 spatial grid) penalizes the extra parameters of attention mechanisms, causing overfitting. Simple Concatenation is adopted as the final fusion strategy.

4.3 DNN Architectures

Five semantic segmentation architectures were adapted to the 16×16 spatial resolution of our tensors. Standard ImageNet backbones are unsuitable here—their aggressive pooling would collapse spatial information after only a few layers. All models therefore use a *custom shallow encoder*: an initial 1×1-preserving convolutional layer followed by two downsampling stages, reducing spatial dimensions from 16×16 to 4×4 before decoding.

FCN [3]: Skip connections from the 16×16 and 8×8 encoder stages feed transposed convolutions that upsample to 360×360.

U-Net variants [8]: Three variants—Simple (standard skips), Attention (attention-gated skips), and Transformer (Swin Transformer encoder).

DeepLabV3+ [1]: An ASPP module with dilation rates $\{1, 2, 4\}$ (adapted for the small 4×4 feature map) explicitly captures multi-scale spatial correlations, which is critical for tubers of varying size and shape.

All models terminate with a 1×1 convolution + Sigmoid activation, producing a per-pixel probability map.

4.4 Training Protocol

All models were implemented using PyTorch Lightning and trained using the **AdamW** [4] optimizer with a cosine annealing learning rate scheduler. To mitigate the severe spatial class imbalance between the tuber target (occupying roughly 10%–20% of the pixel area) and the background soil, we evaluated both weighted loss formulations and standard Binary Cross-Entropy (BCE). Standard BCE was selected as it demonstrated superior empirical stability during initial validation. The loss is computed as:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N_{\text{px}}} \sum_{i=1}^{N_{\text{px}}} \big[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \big], \tag{6}$$

where $y_i \in \{0, 1\}$ is the ground truth pixel value and $\hat{y}_i \in [0, 1]$ is the predicted probability for pixel i. Early stopping was applied with a patience of 20 epochs based on validation loss. To obtain the final binary occupancy masks, the continuous probability maps were binarized at a threshold of τ=0.3, which was optimized via grid search on the validation subset to maximize the validation Dice score. All quantitative metrics are reported as the mean $\pm$ standard deviation across n=3 independent runs with randomized initial weights on the same stratified split.

5 Evaluation and Results

5.1 System Reliability Validation

Before reporting reconstruction accuracy, we validate the data acquisition platform. Across all experimental runs, the measured **PDR exceeded 98%** and the **maximum inter-slot timing jitter was below $\pm$5 ms** against a 100 ms slot duration. These numbers confirm collision-free TDMA operation and a high-fidelity dataset, which is the prerequisite for the statistical results below.

5.2 Quantitative Performance Comparison

Table 2 reports Dice@0.3, IoU@0.3, and BCE for all five architectures across three modalities (n=3). Fig. 4 visualizes peak Dice scores.

Table 2. Performance analysis of different data modalities on various models (mean ± std, $n = 3$).

Modality	Architecture	Dice@0.3 ↑	IoU@0.3 ↑	BCE ↓
Fused	DeepLabV3+	0.82 ± 4 e−3	0.72 ± 11 e−3	0.01 ± 1 e−3
	FCN	0.78 ± 12 e−3	0.68 ± 15 e−3	0.02 ± 1 e−3
	U-Net Simple	0.71 ± 18 e−3	0.61 ± 10 e−3	0.02 ± 3 e−3
	U-Net Attention	0.72 ± 12 e−3	0.61 ± 7 e−3	0.03 ± 5 e−3
	U-Net Transformer	0.70 ± 2 e−3	0.58 ± 1 e−3	0.02 ± 5 e−3
CSI	DeepLabV3+	0.78 ± 10 e−3	0.67 ± 18 e−3	0.02 ± 4 e−3
	FCN	0.80 ± 4 e−3	0.69 ± 3 e−3	0.01 ± 2 e−3
	U-Net Simple	0.74 ± 18 e−3	0.64 ± 17 e−3	0.03 ± 10 e−3
	U-Net Attention	0.75 ± 7 e−3	0.64 ± 7 e−3	0.04 ± 11 e−3
	U-Net Transformer	0.73 ± 15 e−3	0.61 ± 12 e−3	0.03 ± 7 e−3
RSSI	DeepLabV3+	0.62 ± 7 e−3	0.50 ± 4 e−3	0.02 ± 4 e−3
	FCN	0.65 ± 10 e−3	0.53 ± 7 e−3	0.02 ± 2 e−3
	U-Net Simple	0.46 ± 9 e−3	0.36 ± 6 e−3	0.05 ± 14 e−3
	U-Net Attention	0.51 ± 24 e−3	0.41 ± 20 e−3	0.04 ± 5 e−3
	U-Net Transformer	0.62 ± 10 e−3	0.51 ± 11 e−3	0.06 ± 46 e−3

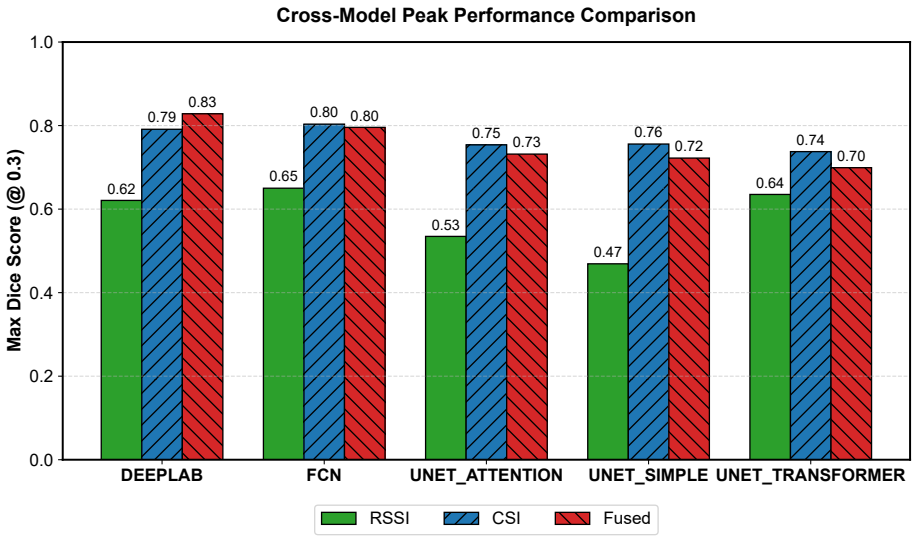


Fig. 4. Cross-model peak Dice comparison across RSSI, CSI, and Fused modalities. CSI and Fused consistently outperform RSSI across all architectures.

CSI vs. RSSI. CSI-based models outperform RSSI for every architecture ($p < 0.005$, paired t-test). Comparing the best RSSI model (FCN: 0.65 ± 0.010) to the best CSI model (FCN: 0.80 ± 0.004) yields a **22.9% relative Dice improvement** and a 29.8% relative IoU improvement. This definitively validates

(a) Ground Truth (b) RSSI (c) CSI (d) Fused

Fig. 5. DeepLabV3+ reconstruction for the same tuber sample across modalities. RSSI produces a diffuse, misshapen mask; CSI recovers the main region; Fused achieves the sharpest boundary delineation closest to the ground truth.

the core hypothesis: multi-subcarrier CSI is fundamentally more informative for subsurface shape reconstruction than scalar RSSI.

Fused CSI+RSSI. The best overall configuration is **DeepLabV3+ on Fused data**: Dice $= 0.8248 \pm 0.0037$, IoU $= 0.7179 \pm 0.0110$. The improvement over CSI-only DeepLabV3+ (0.7836 ± 0.0102) is statistically significant ($p = 0.012$), confirming that RSSI provides complementary global context when exploited by a powerful multi-scale decoder. However, for FCN, U-Net Simple, and U-Net Transformer, the Fused improvement over CSI does not reach $p < 0.05$, indicating the ASPP module is key to extracting the added value of RSSI.

Architecture Comparison. DeepLabV3+ achieves the highest performance on the Fused dataset, attributed to the ASPP module's explicit multi-scale context aggregation. The Swin Transformer U-Net does not outperform simpler CNN variants, likely because the 16×16 spatial resolution is too coarse for global self-attention to be effective.

5.3 Qualitative Reconstruction

Fig. 5 shows reconstructions for the same tuber under all three modalities. The RSSI model fails to delineate the boundary; the CSI model recovers the main shape; the Fused model produces the sharpest, most accurate contour. This qualitative difference directly corroborates the quantitative metrics in Table 2.

5.4 Ablation Studies

Table 3 summarizes ablation results for the Fused DeepLabV3+ baseline (n=3), and Fig. 6 visualizes the trends.

Channel Diversity. Using all 14 channels yields the best mean Dice (0.83 ± 0.011). Restricting to a single channel is unreliable: most channels degrade performance significantly (e.g., Ch. 6: Dice ≈ 0.65), although two channels (Ch. 2 and Ch. 4)

Table 3. Ablation Study Highlights for the fused DeepLabV3+ model (mean ± std, $n = 3$).

Configuration	**Setting**	**Dice@0.3 ↑**	**IoU@0.3 ↑**
Baseline (All)	Ch:All, N:16, SC:All	$0.83 \pm 11\,\text{e}{-3}$	$0.73 \pm 11\,\text{e}{-3}$
Best Single Channel	Ch: 2	$0.83 \pm 12\,\text{e}{-3}$	$0.73 \pm 11\,\text{e}{-3}$
2nd Best Single Channel	Ch: 4	$0.82 \pm 13\,\text{e}{-3}$	$0.71 \pm 10\,\text{e}{-3}$
3rd Best Single Channel	Ch: 5	$0.81 \pm 12\,\text{e}{-3}$	$0.67 \pm 9\,\text{e}{-3}$
Channel Subset	Ch: {1,3,5,7,9,11,13}	$0.79 \pm 12\,\text{e}{-3}$	$0.67 \pm 11\,\text{e}{-3}$
Channel Subset	Ch: {1,5,9}	$0.77 \pm 10\,\text{e}{-3}$	$0.64 \pm 10\,\text{e}{-3}$
Channel Subset	Ch: {1,6}	$0.65 \pm 9\,\text{e}{-3}$	$0.52 \pm 8\,\text{e}{-3}$
Channel Subset	Ch: {3,6,9,12}	$0.71 \pm 10\,\text{e}{-3}$	$0.59 \pm 9\,\text{e}{-3}$
Node Subset (12)	N: {0,1,2,4,5,6,8,9,10,12,13,14}	$0.75 \pm 10\,\text{e}{-3}$	$0.63 \pm 10\,\text{e}{-3}$
Node Subset (8)	N: {0,2,4,8,10,12,14}	$0.71 \pm 10\,\text{e}{-3}$	$0.58 \pm 9\,\text{e}{-3}$
Node Subset (4)	N: {0,4,8,12}	$0.22 \pm 6\,\text{e}{-3}$	$0.17 \pm 6\,\text{e}{-3}$
Best Subcarrier Subset	SC: [30-40]	$0.72 \pm 12\,\text{e}{-3}$	$0.60 \pm 11\,\text{e}{-3}$
Subcarrier Subset	SC: [10-20]	$0.70 \pm 10\,\text{e}{-3}$	$0.57 \pm 10\,\text{e}{-3}$
Subcarrier Subset	SC: [40-50]	$0.69 \pm 11\,\text{e}{-3}$	$0.56 \pm 8\,\text{e}{-3}$

happened to match the baseline due to favorable propagation conditions. Systematic multi-channel hopping provides robustness by averaging out channel-specific fading artifacts.

Node Density. To assess performance under constrained hardware configurations, we simulated lower node densities by sub-sampling the original 16×16 link matrix to 12×12 (12 nodes) and 8×8 (8 nodes) topologies. The models were fully retrained on these reduced spatial configurations. Reducing the system to 12 nodes resulted in a manageable performance drop of approximately 8 Dice points (0.75 ± 0.010). At 8 nodes, the mean Dice score decreased to 0.71 ± 0.010, indicating that a moderate density is still capable of capturing the primary dielectric boundary. However, at 4 nodes, the spatial sampling density fell below the spatial Nyquist limit for the target dimensions, causing the reconstruction to fail catastrophically (0.22 ± 0.006). These results suggest that a minimum of 8 nodes is required to preserve basic morphological shape reconstruction.

Subcarrier Utilization. Restricting to any single 10-subcarrier sub-band reduces Dice by 11–14 points versus the full-subcarrier baseline. The most informative sub-band is [30–40] (Dice $= 0.72 \pm 0.012$), corresponding to the upper half of the 2.4 GHz channel. Using all 52 subcarriers consistently achieves the best performance, confirming that the full OFDM bandwidth is necessary.

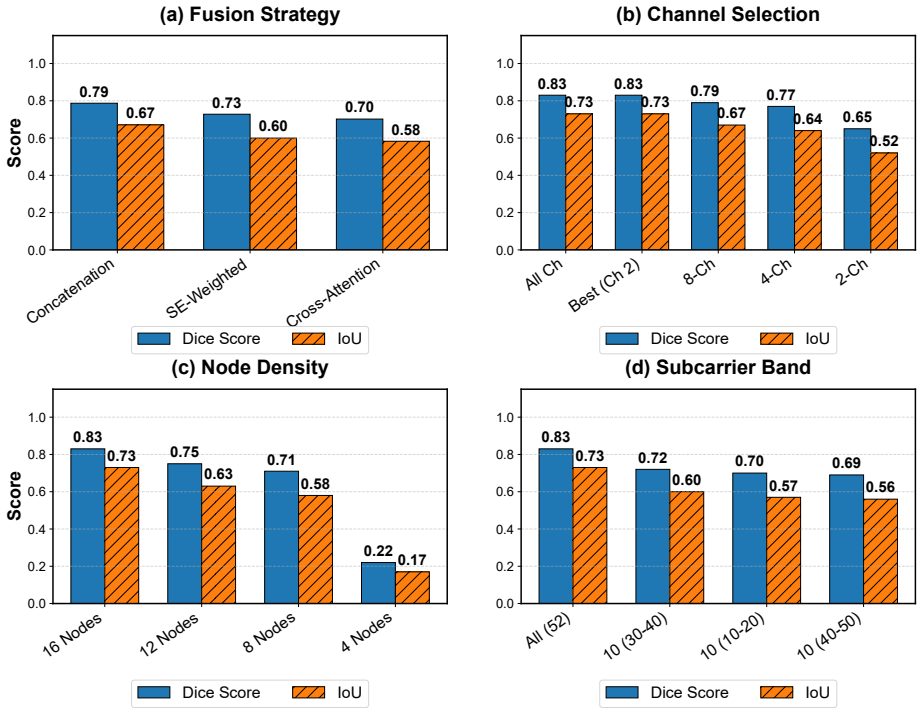


Fig. 6. Ablation study results for the Fused DeepLabV3+ model: (a) fusion strategy comparison, (b) single-channel performance per Wi-Fi channel, (c) node density, (d) subcarrier sub-band selection.

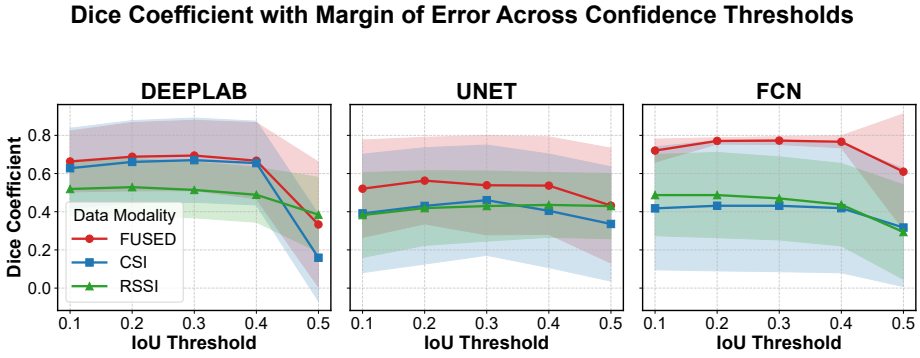


Fig. 7. Dice score vs. binarization threshold (0.1–0.5) with standard deviation envelopes (n=3). Fused and CSI models remain stable up to threshold 0.4; RSSI models collapse rapidly above 0.2.

5.5 Confidence Threshold Stability

Fig. 7 shows Dice across thresholds 0.1–0.5. Fused DeepLabV3+ and CSI models remain stable up to $\tau = 0.4$, while RSSI models collapse above $\tau = 0.2$. This

confirms that CSI-based predictions are well-calibrated (high-confidence predictions are correct), whereas RSSI predictions hover marginally above the decision boundary—a direct consequence of the information loss from scalar power aggregation.

5.6 Discussion and Practical Challenges

While these results demonstrate the feasibility of non-invasive reconstruction using low-cost hardware, translating this system from a controlled lab environment to open-field agricultural settings presents several challenges:

- **Soil Moisture Dynamics:** The dielectric properties of soil are heavily influenced by volumetric water content (θ_v). Higher moisture levels increase attenuation of 2.4 GHz signals, which limits the effective sensing depth. Future deployments will require calibration models that adapt to real-time moisture readings.
- **Heterogeneous Media:** Agricultural soil contains stones, root systems from weeds, and organic matter, which act as scattering objects. While deep networks are capable of learning representative features amidst clutter, training models on more heterogeneous backgrounds will be necessary to prevent false positives.
- **3D Volumetric Imaging:** The current 2D ring setup assumes a thin horizontal cross-section. Reconstructing full 3D tuber volumes will require stacking multiple rings vertically or using multi-static array configurations.

6 Conclusion

This paper presented a low-cost, non-invasive underground tuber imaging system based on a 16-node ESP32-C3 Wi-Fi mesh network. We collected a 256-sample dataset from 8 morphologically diverse potato tubers and demonstrated that CSI-based sensing yields a 22.9% Dice improvement over RSSI ($p < 0.005$). Fusing CSI and RSSI with DeepLabV3+ further improves performance to mean Dice 0.8248 ± 0.0037. Ablation studies establish practical design guidelines: a minimum of 8 nodes, all 14 Wi-Fi channels, and full 52-subcarrier utilization are needed for robust reconstruction. Future work will investigate leveraging sanitized CSI phase data and extending the system to multi-ring field deployments for large-scale agricultural monitoring.

References

1. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation (2018)
2. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks (2019)

3. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3431–3440 (2015)
4. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization (2019)
5. Lu, Y., Lu, G.: 3d modeling beneath ground: Plant root detection and reconstruction based on ground-penetrating radar. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 68–77 (2022)
6. Patwari, N., Wilson, J.: Rf sensor networks for device-free localization: Measurements, models, and algorithms. Proceedings of the IEEE **98**(11), 1961–1973 (2010). https://doi.org/10.1109/JPROC.2010.2052010
7. Roitsch, T., Cabrera-Bosquet, L., Fournier, A., Ghamkhar, K., Jiménez-Berni, J., Pinto, F., Ober, E.S.: New sensors and data-driven approachesa path to next generation phenomics. Plant Science **282**, 2–10 (2019)
8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 234–241. Springer (2015)
9. Shang, F., Yang, P., Yan, e.a.: Liquimager: Fine-grained liquid identification and container imaging system with cots wifi devices. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **8**(1), 1–29 (2024)
10. Strohmayer, J., Kampel, M.: Data Augmentation Techniques for Cross-Domain WiFi CSI-Based Human Activity Recognition, pp. 42–56. Springer Nature Switzerland (2024). https://doi.org/10.1007/978-3-031-63211-2_4, `http://dx.doi.org/10.1007/978-3-031-63211-2_4`
11. Tan, S., Yang, J.: Object sensing for fruit ripeness detection using wifi signals (2021), `https://arxiv.org/abs/2106.00860`
12. Wang, T., Zhao, Y., Liu, J., Zhuang, Y.: Underground potato root tuber sensing via a wireless network. In: 2024 23rd ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). pp. 251–252. IEEE (2024)