IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 October 2025, accepted 26 October 2025, date of publication 6 November 2025, date of current version 14 November 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3629617

RESEARCH ARTICLE

BiAgriNet: Binarized Knowledge-Distilled Network for Real-Time Semantic Segmentation in Agriculture

HASSAN KHAN[1], SUNBAL IFTIKHAR[2], RORY WARD[1], (Member, IEEE), STEVEN DAVY[2], (Member, IEEE), AND JOHN G. BRESLIN[1], (Senior Member, IEEE)

[1]Data Science Institute, University of Galway, Galway, H91 TK33 Ireland
[2]Technological University (TU) Dublin, Dublin, D07 EWV4 Ireland

Corresponding author: Hassan Khan (h.khan5@universityofgalway.ie)

This work was supported by Taighde Éireann-Research Ireland under Grant 21/FFP-A/9174 (SustAIn), Grant 21/RC/10303_P2(VistaMilk), and Grant 12/RC/2289_P2 (Insight).

ABSTRACT Semantic segmentation is critical for agricultural applications such as crop-weed classification and precision farming. However, real-time segmentation in resource-constrained environments requires lightweight and efficient networks. This paper introduces **BiAgriNet**, a novel agricultural segmentation framework that combines full weight-and-activation binarization with a teacher–student knowledge-distillation paradigm. The proposed student network employs a 1bit ResNet18 encoder and a grouped dilated Atrous Spatial Pyramid Pooling bottleneck, enabling multiscale context capture while maintaining a lightweight design. Trained using advanced teacher-student knowledge distillation with ResNet18+DeepLabV3 as the teacher, BiAgriNet achieves **85.6% mIoU**, with a memory footprint of only **0.8 MB**, which is **27× smaller than DeepLabv3** while maintaining competitive accuracy. With an inference speed of **180 FPS**, nearly **5× faster than FCN8s (37 FPS)**, BiAgriNet demonstrates its practicality for real-time precision agriculture on embedded systems, offering a compelling balance between efficiency and performance.

INDEX TERMS Binary neural network (BNN), real-time semantic segmentation, computational efficiency.

I. INTRODUCTION

The growing demands for precision agriculture have emphasized the need for efficient and accurate computer vision models tailored to agricultural applications. While traditional convolutional networks (CNNs) have demonstrated exceptional performance in semantic segmentation tasks, their high computational cost and memory demands limit their deployment on resource-constrained systems in real-world agricultural environments.Monitoring crop growth, weeds, and disease at field scale increasingly relies on semantic segmentation of high-resolution UAV and proximal imagery [1]. Networks such as ResLMFFNet already reach 45 FPS while delineating tree-crowns and rust lesions [2], and multi-branch models have pushed real-time crop/weed discrimination below 20ms per frame on GPUs [3]. Transformer hybrids like TinySegformer further demonstrate that lightweight designs can maintain accuracy in cluttered pest scenes [4]. Field robots and UAVs operate under strict power budgets; on-board ARM or RISC-V SoCs rarely sustain 5W. Binary Neural Networks (BNNs) slash memory and multiply-accumulate counts by two orders of magnitude [5], enabling FPGA deployments for crop recognition with 17% of a Zynq-MPSoC's resources [6]. Dual-attention BNNs have already proven effective for plant-disease classification, closing much of the accuracy gap to 32-bit CNNs [7],while recent improvements in BNN training dynamics e.g. XNOR-Net++ and Riptide stabilise convergence on large-scale data sets [8], [9]. Parallel to quantisation, knowledge distillation (KD) transfers structural

The associate editor coordinating the review of this manuscript and approving it for publication was Ajit Khosla.

cues from a high-capacity teacher to a compact student. Label-assisted KD [10], query-level KD for Mask2Former transformers [11], raw-feature KD [12] and the recent I2CKD triplet-loss formulation [13] deliver notable mean Intersection over Union (mIOU) gains on urban and agricultural benchmarks. However, these methods assume at least 8-bit activations and are not optimised for 1-bit arithmetic. There is currently no framework that combines (i) segmentation-aware distillation, (ii) aggressive 1 bit binarization, and (iii) agriculture-specific data augmentation under a single training protocol. Lightweight CNN-Transformer hybrids continue to depend on INT8/FP16 computations [4], whereas current BNNs concentrate on classification tasks [5]. This creates a performance gap when implementing semantic segmentation on low power device, where memory bandwidth and battery longevity are critical. Addressing these challenges, we propose BiAgriNet, a binarized knowledge-distilled network designed to deliver real-time semantic segmentation with minimal computational overhead. We combine full binarization with knowledge distillation. This preserves segmentation accuracy while cutting memory and compute, as shown in Fig 1. The proposed framework suits crop–weed classification, where high precision and real-time inference are essential. With multi-scale features and a lightweight design, BiAgriNet provides an efficient, real-time solution for large-scale agricultural segmentation. We evaluate the proposed architecture on the PhenoBench dataset [14] and CWFID dataset [15], demonstrating its ability to achieve state-of-the-art efficiency while maintaining competitive accuracy. The main contributions of this work are as follows:

- **Fully binarised student network:** We present a novel architecture whose encoder, bottleneck, and decoder are all 1-bit, reducing the model footprint to **0.8 MB** and the computational cost to **0.62G FLOPs** per 1024 × 1024 image, **27**× smaller and ∼ **60**× lighter than the teacher.
- **Teacher-guided optimization** The student is trained with a combined **cross-entropy and Kullback–Leibler divergence loss**, enabling the transfer of fine-grained logit information from the teacher and thereby recovering much of the accuracy typically forfeited by binarised neural networks.
- **High accuracy and real-time throughput** The proposed BiAgriNet attains **85.6% mIoU on PhenoBench** and **86.7% mIoU on CWFID**, while maintaining **180 FPS** on an NVIDIA Jetson AGX Orin. This outperform ERFNet having 2× less computation and **5**× less parameters on PhenoBench dataset and surpasses other lightweight models such as BiseNetV2, delivering **5–20× higher energy efficiency**.
- **Comprehensive evaluation:** Ablation studies analyse the impact of distillation weight, bottleneck configuration, and binarisation scheme, whereas a detailed power–throughput assessment shows that BiAgriNet saves **10-35**× energy compared with full-precision counterparts.

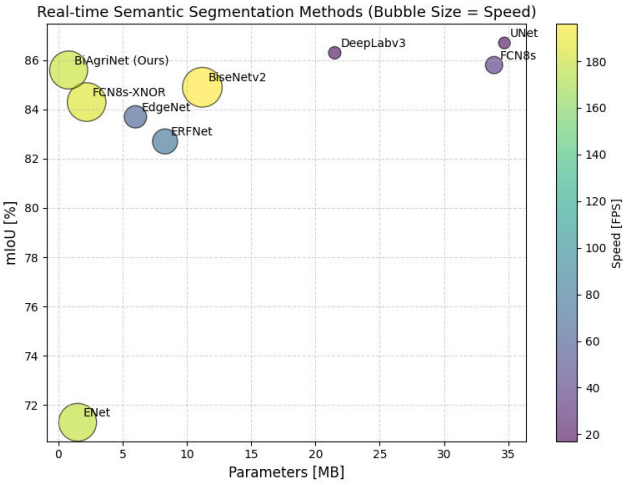


FIGURE 1. Accuracy–efficiency trade-off on *PhenoBench* at 1024×1024 input. Circle size encodes inference speed (FPS) on jetson AGX Orin (batch = 1). BiAgriNet attains competitive mIoU with real-time throughput and the smallest memory footprint among compared models.

The remainder of this paper is structured as follows: Section II surveys related work on efficient agricultural segmentation, binarisation, and distillation. Section III details the BiAgriNet architecture and training methodology. Section IV presents experimental setups, quantitative comparisons on PhenoBench and CWFID, and qualitative results. Section V Ablation Studies, and Section VI Conclusion.

II. RELATED WORK

In recent years, semantic segmentation has seen significant advancements, with networks like DeepLabv3 [16], U-Net [17], and PSPNet [18] pushing the boundaries of accuracy and feature extraction capabilities. Motivated by the demonstrated efficacy of these general-purpose architectures, recent research has increasingly sought to adapt and optimise semantic-segmentation networks for the strict real-time demands and distinctive environmental conditions encountered in agricultural imagery.

A. REAL-TIME SEMANTIC SEGMENTATION IN AGRICULTURE

Recent advancements in semantic segmentation for agriculture have addressed challenges like data scarcity, small object segmentation, and environmental variability. For instance, Heschl et al. introduced SynthSet, a generative model for creating annotated agricultural data [19]. Similarly, PD-SegNet enhances the segmentation of small agricultural targets in complex environments [20]. Moreover, approaches utilizing style transfer and Generative Adversarial Networks (GANs) have improved image-label datasets for dense vegetation [21]. This model achieved state-of-the-art results on the challenging MinneApple dataset by combining dynamic kernel updating with edge-aware refinement. Similarly,

Yu et al. designed DCSAnet for crop–weed segmentation in soybean fields, using MobileNetV3 blocks and asymmetric convolutions to reach 85.95% mIoU with just 0.57M parameters [22]. Beyond weeds, researchers have tackled disease symptom segmentation on leaves and canopy delineation: transformer-based models have been applied to segment leaf blight regions on crop imagery [23], while multispectral approaches like Wu et al. fused RGB and thermal data to accurately extract wheat canopies at tillering stage (mIoU ∼84–85%) [24]. These efforts underscore a broader trend for real-time segmentation networks in agriculture, emphasize both efficiency and task-specific robustness, enabling on the fly weed detection, disease localization, and canopy analysis under field conditions. Surveys also highlight progress in crop cover analysis, pest detection, and disease identification, underscoring the expanding role of segmentation in precision agriculture [1]. However, these models' heavy computational demands hinder their use in real-time or resource-limited settings like mobile and embedded systems. The demand for computational efficiency has driven the exploration of lightweight architectures like MobileNets [25], ShuffleNet [26], BiSeNet [27], BiseNet-v2 [28], EdgeNet [29], MSCFNet [30], and LETNet [31].

B. BINARIZED NETWORKS AND TINY TRANSFORMERS

In parallel, extreme model compression and compact transformer designs have been explored to meet strict speed and memory constraints. Binarized neural networks (BNNs) [8], which use 1-bit weights/activations, can drastically reduce model size and computations. While BNNs have shown promise in classification tasks (e.g., XNOR-Net [5]), their use in dense segmentation is growing, especially in agriculture, due to accuracy degradation challenges. On the other hand, tiny Transformers have gained attraction as lightweight backbones that retain strong representation power. Xie et al. introduced SegFormer-B0 (''TinySegFormer''), a Transformer model with only 3.8M parameters and ∼ 8.4G FLOPs that still achieved 37.4% mIoU on ADE20K [32]. Such designs combine efficient self-attention with convolutional stems [23], and have been adopted in agriculture: for example, TinySegFormer has been used as a baseline for segmenting fruit trees and diseased plants, often outperforming heavier CNNs on a FPS-per-accuracy basis. Purpose-built networks also merge transformers with CNNs for speed: Zhang et al. developed LACTNet, a hybrid encoder–decoder with a gated Transformer module, reaching 74.8% mIoU at 90 FPS on Cityscapes [33]. In the agricultural domain, Feng et al. proposed Lodging-U2NetP (L-U2NetP) an ultra-lightweight U-Net variant (1.1 M params) for wheat lodging segmentation. By simplifying U2Net with channel-wise attention, it accurately segmented lodged crop areas from UAV images while being deployable on edge devices (param count 1/40 of the original) [34]. These advances illustrate how model size and complexity have been aggressively reduced through binary convolutions, ensemble pruning, or miniaturized transformers to meet real-time requirements in precision agriculture. Nevertheless, no other efficient model has achieved full network binarization and knowledge distillation leveraging to the extent that we have.

C. KNOWLEDGE DISTILLATION FOR VISION MODELS

Knowledge distillation (KD) has become a powerful technique to transfer knowledge from large ''teacher'' models to compact ''student'' models, often boosting the student's performance. In computer vision, recent distillation methods have pushed the state-of-the-art. Qiu et al. proposed Label-Assisted Distillation (LAD) for semantic segmentation, which ingeniously injects perturbed ground-truth labels into the teacher's input to strengthen a lightweight teacher model [10]. This yields a more informative teacher without relying on an extra heavy model or modalities, and experiments on Cityscapes and ADE20K showed consistent student improvements with LAD. Another frontier is ensemble-based distillation: Landgraf et al. introduced DUDES (Deep Uncertainty Distillation using Ensembles for Segmentation), wherein a deep ensemble's predictive distribution is distilled into a single student [35]. DUDES preserved not only accuracy but also uncertainty estimates, accurately mimicking the ensemble's confidence while simplifying deployment. This is especially relevant for safety-critical agro-robotics, where knowing when the model is unsure (e.g. in novel field conditions) is valuable. A further innovation is cross-modal distillation. Nair and Hänsch demonstrated that a model trained on abundant RGB satellite imagery can teach a model on scarce SAR radar imagery via cross-modal KD [36]. Their framework, winner of the EarthVision 2024 challenge, used an RGB-segmentation teacher to generate pseudo-labels for unlabeled SAR images, boosting IoU on SAR building segmentation by 5–20%. This cross-domain knowledge transfer opens doors for agricultural applications where one sensing modality (e.g. RGB) has rich data to guide another (e.g. thermal or LiDAR) in a student model. In summary, modern KD techniques, from label-guided training to ensemble and multi-modal distillation, offer versatile tools to compress models without losing accuracy. Table 1 compares segmentation models applied to various agricultural vision tasks in the last few years. Although numerous studies have advanced either model binarization or knowledge-distillation techniques, none has yet integrated full network binarization with teacher–student distillation in the context of agricultural semantic segmentation. Existing BNN work is predominantly confined to image-classification tasks or general-purpose benchmarks, whereas distillation methods typically compress floating-point networks without modifying numerical precision. BiAgriNet constitutes the first agriculture-specific architecture that unifies these two paradigms: a fully binarized student network trained under the guidance of a high-capacity, full-precision teacher. This dual strategy proves essential, while binarization alone tends to degrade accuracy, the distillation process compensates by

TABLE 1. Comparative summary of agricultural semantic-segmentation studies, with an emphasis on real-time or otherwise efficiency-oriented designs.

Papers	Dataset(s)	Image Res.	mIoU (%)	Strengths	Limitations
Anand.et al [37] (AgriSegNet)	UAV anomaly	512×512	82–85	Multi-scale attention; detects field anomalies	Relies on large annotated UAV sets
Khan.et al [38] (CED-Net)	CWFID	1296×966	77.6	Cascaded encoder–decoder; binary crop–weed	Narrow two-class scope; moderate IoU
Kim. et al [39] (MTS-CNN)	CWFID	1296×966	91.6	Multi-task learning boosts weed accuracy	Heavier than CED-Net; slower inference
Janneh. et al [40] (HybridF.S.)	CWFID	1296×966	86.1	Multi-level feature fusion (SOTA CWFID)	Large (32MB); diminishing gains
Zhu.et al [20] (PD-SegNet)	MinneApple	1280×720	78–81	SegFormer backbone; excels at small fruit	High memory; domain-specific
Yu.et al [22] (DCSAnet)	Soybean weed	512×512	85.95	0.57M params; MobileNetV3 blocks	Limited cross-crop generalization
Kim.et al [41] (LCW-Net)	CWFID (low-light)	1260×966	$\approx$80	Attention dual decoder; night robustness	Two decoders add overhead
Wei.et al [42] (Attn-LightNet)	BoniRob, WeedMap	1296×966, 480×360	75–88	0.11M params; 55FPS Jetson NX	Lower accuracy in complex scenes
Islam.et al [43] (MobileNetV4-Seg)	Corn&soy UAV	256×256	69.9–76.8	44FPS Orin; edge-oriented pipeline	Significant mIoU loss for speed
Feng.et al [34] (L-U2NetP)	Wheat lodging UAV	512×512	~ 88	1.1M params; high lodging IoU	Tailored to fixed altitude UAV data
Galymzhankyzy.et al [44] (MultiSpecter)	WeedsGalore	600×600	78.9	RGB+NIR+RE fusion; +15mIoU vs RGB	Requires multispectral sensors
Wu.et al [24] (Tiff-SegFormer)	WheatRGB+TIR	640×480	84.3	Thermal+RGB canopy; high mPA (91%)	Dual-sensor hardware needed

transferring fine-grained representational knowledge from the teacher. Consequently, BiAgriNet attains unprecedented computational efficiency (0.62 GOP) and real-time throughput (180 FPS) without measurable loss in segmentation accuracy, thereby filling a critical gap in the literature and establishing a foundation for future systems that simultaneously optimise model size, speed, and precision in real-world precision-agriculture deployments. Full network binarization restricts the function class by quantizing weights/activations to $\{-1, +1\}$, which tightens Lipschitz bounds but risks underfitting high-frequency structure in dense prediction. In this constrained regime, distillation supplies informative targets p_T that shape the student's probability simplex beyond hard labels. Formally, minimizing $\mathrm{KL}(p_S \| p_T)$ at temperature τ aligns the student's logits with the teacher's relative class ordering, preserving inter-class margins even under 1-bit quantization. This coupling differs from prior efficient segmentation that relies on INT8/FP16 activations or FP students, and from BNN literature focused on classification; BiAgriNet unifies complete binarization with segmentation-aware KD, yielding stable training and accuracy at ultra-low compute/memory. Unlike prior efficient segmentation that retains $\geq$8-bit activations, our novelty is the first agricultural segmentation framework that fully binarizes encoder–bottleneck–decoder and couples it with KD, narrowing the accuracy gap while preserving the 1-bit efficiency envelope.

III. PROPOSED METHOD

The **BiAgriNet** framework Fig. 2 integrates a full-precision teacher model with a binarized student network for efficient semantic segmentation. The architecture includes a binary encoder, bottleneck, and decoder, which leverage skip connections and multi-scale context modules to reduce parameters while retaining accuracy. During training, a knowledge distillation mechanism transfers knowledge from the teacher to the student using supervised and distillation losses. Optimized for real-world deployments, the binarized student achieves real-time segmentation with minimal memory and computation requirements.

A. TEACHER MODEL (FULL PRECISION)

To provide robust "dark knowledge," we adopt a **full-precision** encoder–decoder network as the teacher. We trained five encoder–decoder combinations (VGG16–FCN8, VGG16–U-Net, ResNet18–FCN8, ResNet18–U-Net, and ResNet18–DeepLabV3) as potential teacher models. The VGG16-based networks serve as classic, parameter-heavy baselines, showcasing how earlier architectures (e.g., FCN8) compare in modern contexts. ResNet18, a more efficient backbone, is paired with simple (FCN8, U-Net) and a state-of-the-art decoder (DeepLabV3) featuring atrous convolutions and multi-scale context. This range of configurations allows us to evaluate trade-offs between model complexity, accuracy, and inference speed, ensuring that comprehensive empirical insights drive our final teacher selection. In particular, the ResNet18–DeepLabV3 model typically achieves strong segmentation performance while maintaining a relatively modest memory footprint, making it an appealing teacher for knowledge distillation. At the same time, examining mid-range or older baselines (VGG16–U-Net, ResNet18–U-Net) offers insight into how simpler backbones and decoders fare in current benchmarks. This diversity of teacher networks ultimately facilitates a more robust analysis of teacher-student knowledge transfer.

B. BINARIZED STUDENT MODEL

1) BINARIZED ENCODER FOR FEATURE REPRESENTATION

a: BINARY CONVOLUTION

To maintain general applicability, consider an input activation $A_{l-1} \in \mathbb{R}^{h \times w \times c}$ feeding into layer l. In a float-precision network, the trainable weights W_l are real-valued, but here they are projected into $\{-1, +1\}$ to form binary tensors B_l. In our approach, the encoder is based on ResNet18, yet every convolutional layer beyond the first adopts binary weights.

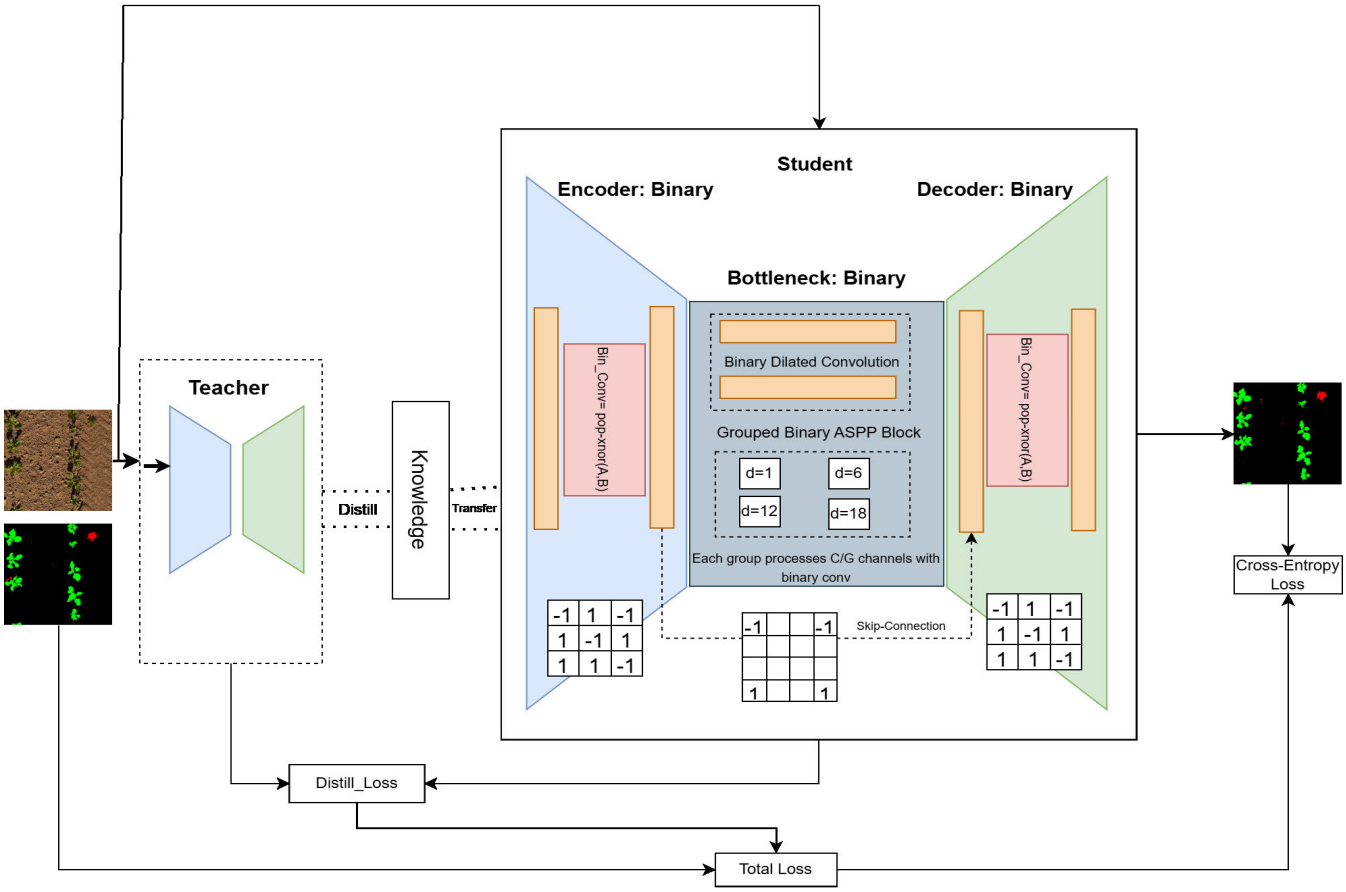


FIGURE 2. Overview of BiAgriNet: A full-precision teacher guides a binarized student with a binary encoder, bottleneck (dilated blocks and ASPP), and decoder. Knowledge distillation and combined loss ensure efficient and accurate segmentation.

Specifically, each filter W is quantized into $\{-1, +1\}$ through a sign function, while a per-channel scaling factor α approximates magnitude information. Thus, the binarized convolution computes:

$$\alpha \cdot (x \otimes \text{sign}(W))$$

Where $\otimes$ represents XNOR multiplication in 1-bit space. These binarized residual blocks preserve skip connections vital for stable gradient flow while drastically shrinking parameter storage and Multiply-Accumulate Operations (MACs) operations. The first convolution (along with its associated batch normalization) remains in a full float precision to avoid early training instabilities, whereas subsequent stages are fully binarized. ResNet18's well-known efficiency and residual shortcuts help mitigate the representational gap introduced by ± 1 weights. By leveraging these 1-bit filters, the network preserves the essential capacity for discriminative feature extraction yet remains lightweight, making it ideal for scenarios like large-scale agricultural imagery.

2) BOTTLENECK WITH BINARY DILATED CONVOLUTIONS

BiAgriNet adapts dilated convolutions in the bottleneck to expand the receptive field without additional parameters. Specifically, each dilated convolution in the bottleneck is subject to the same binarization procedure—both weights and activations are quantized into ± 1. Batch normalization and a non-linear activation ReLU are applied after each binary convolution to maintain distributional stability. Inspired by DeepLab's multi-dilation concept, BiAgriNet employs a Grouped Binary ASPP, where channels are partitioned into groups; each group uses a different dilation rate (e.g., $\{1, 6, 12, 18\}$). These parallel branches capture diverse scales to detect partial regions in agricultural imagery. Concatenating these channels yields a robust mid-level representation. Unlike a float-based approach, binarized dilated layers keep overhead low, even when the network processes large input sizes (e.g., 1024×1024 fields).

3) BINARIZED DECODER FOR SEGMENTATION

Our method incorporates a binarized DeepLab decoder to complete the segmentation pipeline. Traditionally, DeepLab leverages atrous (dilated) convolutions and an Atrous Spatial Pyramid Pooling (ASPP) block for multi-scale context. In our binarized variant, each dilated 3×3 convolution within the decoder also undergoes sign(W)-based quantization, with a learned scaling factor appended post-XNOR. This arrangement permits robust multi-scale feature fusion

without increasing the FLOPs. The final classifier layer typically remains in full precision to stabilize the generation of class logits, ensuring that subtle boundary probabilities are not lost to binarization. Consequently, the binarized DeepLab decoder complements the lightweight encoder while retaining enough representational power to refine coarse feature maps into high-resolution, per-pixel predictions.

C. KNOWLEDGE DISTILLATION FOR BINARIZATION IN BIAGRINET

Binarizing the network (weights and activations to ± 1 via the sign function) significantly reduces model precision and disrupts gradient flow. Knowledge distillation is employed in BiAgriNet to overcome these challenges by using the teacher model's *soft logits* as additional guidance. The full-precision teacher's output probabilities (soft targets) retain rich class information (relative confidences for each class) that would be lost in the hard binary activations. By training the binarized student to match these soft logits, we mitigate the information loss introduced by the non-differentiable sign function. In essence, the teacher's continuous output distribution provides gradient signals and "dark knowledge" that help the binary student learn nuanced decision boundaries that a one-hot label alone cannot convey. This extra supervision directs the student to mimic the teacher's feature representations despite the 1-bit weight constraint, thereby addressing the accuracy drop typically caused by binarization. We optimize a joint loss consisting of two terms: the standard cross-entropy loss L_{CE} with ground-truth labels, and a distillation loss L_{KD} defined as the Kullback–Leibler divergence between the student's prediction distribution p_S and the teacher's distribution p_T. Let $z_T(x_i), z_S(x_i) \in \mathbb{R}^C$ denote teacher and student logits at pixel i for C classes. With temperature $\tau > 0$, define softened distributions

$$q_T(c \mid x_i) = \mathrm{softmax}\big(z_T(x_i)/\tau\big)_c, \tag{1}$$

$$q_S(c \mid x_i) = \mathrm{softmax}\big(z_S(x_i)/\tau\big)_c. \tag{2}$$

We adopt the student-to-teacher Kullback–Leibler divergence, averaged per pixel:

$$\mathcal{L}_{\mathrm{KD}} = \tau^2 \frac{1}{|\Omega|} \sum_{i \in \Omega} \sum_{c=1}^{C} q_S(c \mid x_i) \log \frac{q_S(c \mid x_i)}{q_T(c \mid x_i)}. \tag{3}$$

The full objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{CE}} + \lambda \mathcal{L}_{\mathrm{KD}}, \quad \mathcal{L}_{\mathrm{CE}} = -\frac{1}{|\Omega|} \sum_{i \in \Omega \setminus \{y_i = \xi\}} \log p_S(y_i \mid x_i), \tag{4}$$

where $p_S = \mathrm{softmax}(z_S)$, ξ is the ignore-index, and λ balances supervision and distillation. We use the standard τ^2 factor to maintain gradient scale at higher τ. All losses use natural logarithms and are averaged per image and per pixel. In our best configuration, we fix $\tau = 4$ and $\lambda = 1.0$. Without distillation ($\lambda = 0$), the binarized student's accuracy drops sharply, whereas an optimal $\lambda \approx 1.0$ can boost accuracy by $\sim$ 5% absolute. Too large a λ may overemphasize the teacher at the expense of true labels. The knowledge-distillation scheme for BiAgriNet is outlined in Algorithm 1. In each iteration:

1) The *frozen teacher* computes logits p_T for the input mini-batch.
2) The *binarized student* computes logits p_S (using the sign function in each layer).
3) Compute $L_{\mathrm{CE}} = \mathcal{H}(p_S, Y)$ and $L_{\mathrm{KD}} = \mathrm{KL}(p_S \parallel p_T)$.
4) Form $L = L_{\mathrm{CE}} + \lambda L_{\mathrm{KD}}$ and backpropagate via the straight-through estimator (STE), updating only the student's parameters.

After training, the teacher is discarded, the student runs independently with binary weights for efficient inference. Knowledge distillation markedly improves the alignment between the teacher's and student's internal representations, even under binarization. Fig 3 illustrates the layer-wise cosine similarity between feature maps of the teacher and those of two variant students (one with a ResNet-18 backbone, the other with a MobileNet-V3 backbone). The ResNet-18 based student, which closely matches the teacher's architecture, achieves an exceedingly high cosine similarity (around 0.9) in deeper layers. This indicates that by the final layers, the binarized ResNet-18 student's feature representations are almost indistinguishable from the teacher's, a direct result of the distillation pressure to mimic the teacher. In contrast, the MobileNet-V3 student shows lower similarity values at corresponding layers (i.e. weaker representation alignment). This reduced alignment for MobileNet-V3 correlates with

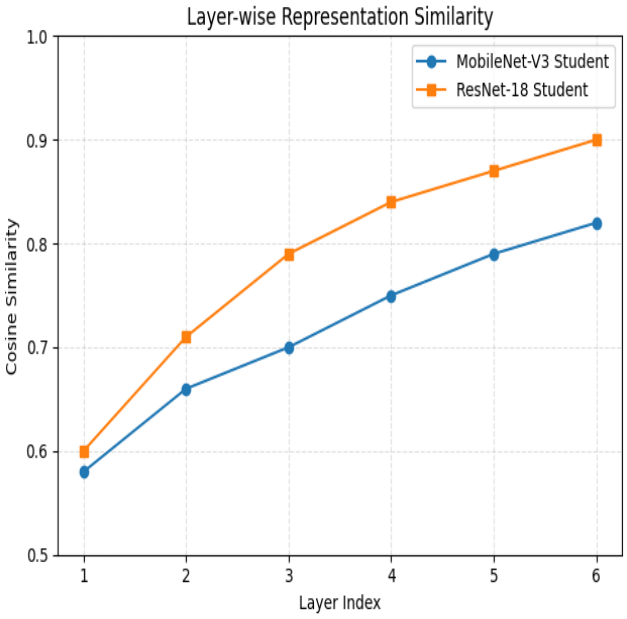


FIGURE 3. Layer-wise cosine similarity between the teacher (ResNet-18 + DeepLabV3) and two binarized students. The ResNet-18 student maintains high alignment with the teacher, achieving a cosine similarity of ∼0.9 in deeper layers due to architectural affinity and effective knowledge distillation. In contrast, the MobileNet-V3 student exhibits lower similarity scores, indicating weaker feature transfer and reduced representational fidelity.

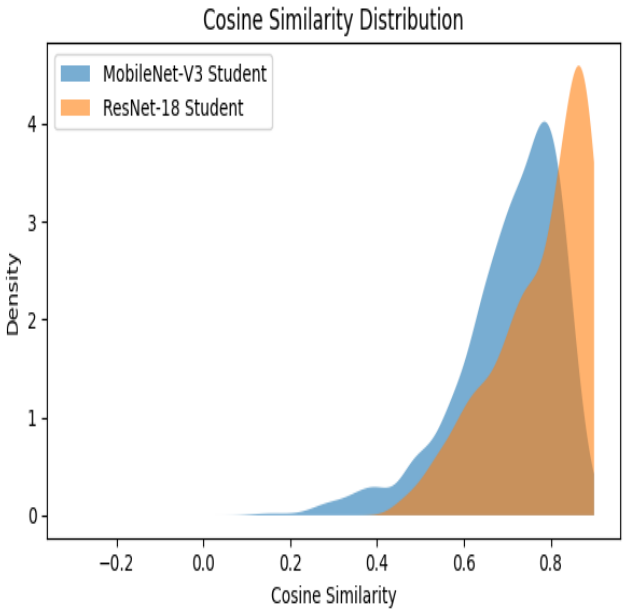


FIGURE 4. Cosine similarity distribution of feature representations across all layers between the teacher and binarized students. The ResNet-18 student shows a narrow, high-peaked distribution close to 1.0, demonstrating consistent feature alignment and minimal information loss despite binarization. Conversely, the MobileNet-V3 student exhibits a broader, lower-centered distribution, suggesting less effective knowledge transfer and diminished capacity to mimic the teacher's representations.

its larger drop in segmentation accuracy (a 3% lower mIoU than the ResNet student under the same teacher), underscoring that architectural disparity can make it harder for the student to absorb the teacher's knowledge. Fig 4 provides the distribution of cosine similarities across all layers (or feature channels) for the two students. The ResNet-18 student's cosine similarity scores are not only higher on average but also tightly clustered near 1.0, indicating consistently strong alignment with the teacher's features. In comparison, the MobileNet-V3 student's similarity scores are more spread out and centered at a lower value, reflecting more variable and generally poorer alignment. This tighter representation alignment in the ResNet-18 case demonstrates the effectiveness of knowledge distillation in guiding a binarized model. When the student network has sufficient capacity and a compatible architecture, the distillation process can drive it to mirror the teacher's learned representations. Fig 3 and 4 shows that distillation helps the binary student retain important feature information that the sign quantization alone would have destroyed. Consequently, the BiAgriNet student, aided by distillation, achieves significantly improved accuracy compared to a binarized model trained only with hard labels, closing much of the gap towards its full-precision teacher. Knowledge distillation serves as a critical bridge in BiAgriNet, transferring nuanced knowledge to the binarized network and thereby alleviating the detrimental effects of binarization on model performance.

1) TRAINING SCHEME OF BIAGRINET

The proposed BiAgriNet model is trained on 3 A4000 GPU (16 GB), and an NVIDIA AGX ORIN is used for inference. Training BiAgriNet begins with a pre-trained teacher network (e.g., a full-precision ResNet+DeepLab model) whose weights remain frozen. A dataset of images I and pixel-level labels Y is fed in mini-batches; each batch is processed by both the teacher (frozen) and the student (BiAgriNet). The teacher produces "soft" logits p_{teacher}, while BiAgriNet, composed of binarized convolution layers, generates its predictions $\hat{Y}$. For the student's forward pass, each binarized convolution uses sign(W_l) for the weights, along with normalization and ReLU. During backpropagation, only the student's parameters are updated. A straight-through estimator (STE) handles the non-differentiable sign function, effectively allowing real-valued latent weights to receive gradients and be binarized anew each iteration. We employ a combined loss L_{CE} enforces alignment with ground-truth labels, and L_{KD} encourages the binarized student to mimic the teacher's richer predictions. The teacher itself remains unchanged throughout. BiAgriNet gradually refines its 1-bit convolution filters by iterating over these mini-batches and losses, balancing semantic accuracy (from cross-entropy) with teacher knowledge (from distillation). Once trained, the student is deployed alone, capitalizing on binarized weights for fast, memory-efficient inference, an ideal characteristic for real-time agricultural applications. We use STE for sign($\cdot$) with gradient clipping (norm=1), cosine LR with warm-up, BN momentum 0.9, and KD with τ=4. These mitigated training instability and boundary artifacts in the 1-bit regime.

IV. EXPERIMENTS AND RESULTS

A. DATASETS AND EVALUATION METRICS

1) PHENOBENCH

The PhenoBench dataset [14] comprises 1,407 training images and 772 validation images, each with corresponding ground truth labels. Ground truth labels for the test set are not publicly available. Each image, sized at 1024×1024 pixels, includes pixel-wise semantic masks with label IDs assigned as follows: background (0), crop (1), and weed (2).

2) CWFID (CROP/WEED FIELD IMAGE DATASET)

The CWFID dataset [15] consists of 60 RGB images captured in organic carrot fields using a BoniRob mobile robot. Early-growth carrot plants and inter-row weeds are present near, making pixel-level crop–weed discrimination necessary. Each image is annotated with a binary vegetation mask, further labeled as crop vs. weed (background soil is implicitly the third class). In total, CWFID includes 162 annotated crop plants and 332 weeds. The original image resolution is on the order of 1 megapixel (approximately 1296×966 pixels), which we rescaled to 1024×1024 for consistency with our model's input. We split the 60 images for training and validation (e.g., an 80/20 split, ensuring a mix

Algorithm 1 Training BiAgriNet With Knowledge Distillation (frozen Teacher; Binarized student)

Require: Teacher T (full precision; frozen), student S (binarized with straight-through estimator, STE), training set $\{(x, y)\}$ with ignore-index ξ, temperature $\tau = 4$, distillation weight $\lambda = 1.0$, optimizer $\mathcal{O}$

Ensure: Trained student parameters θ_S (teacher discarded at inference)

for epoch $= 1, \ldots, E$ **do**
  for mini-batch $\mathcal{B} = \{(x_b, y_b)\}$ **do**
    Teacher forward (no gradients): $z_T \leftarrow T(x_b)$; $q_T \leftarrow \mathrm{softmax}(z_T/\tau)$
    Student forward (binarized; STE in backprop): $z_S \leftarrow S(x_b)$; $q_S \leftarrow \mathrm{softmax}(z_S/\tau)$; $p_S \leftarrow \mathrm{softmax}(z_S)$
    Losses (pixel-wise means; ignore $y = \xi$):
    $$\mathcal{L}_{\mathrm{CE}} \leftarrow -\frac{1}{|\Omega|} \sum_{i \notin \{y_i = \xi\}} \log p_S(y_i \mid x_i)$$
    $$\mathcal{L}_{\mathrm{KD}} \leftarrow \tau^2 \frac{1}{|\Omega|} \sum_i \sum_c q_S(c \mid x_i) \log \frac{q_S(c \mid x_i)}{q_T(c \mid x_i)}$$
    $\mathcal{L} \leftarrow \mathcal{L}_{\mathrm{CE}} + \lambda \mathcal{L}_{\mathrm{KD}}$
    Update student only: $\theta_S \leftarrow \mathcal{O}(\theta_S, \nabla_{\theta_S} \mathcal{L})$ ▷ STE handles sign(·)
  end for
end for
return S

of field conditions in each) when evaluating our models on this dataset. Despite its smaller scale, CWFID is a standard benchmark in precision agriculture vision, allowing us to validate BiAgriNet's generalization to a different crop type and sensor scenario.

3) PERFORMANCE METRICS

The metrics reported in these experiments include Mean Intersection-over-Union (mIoU), Average Precision (AP), False Positive Rate (FPR), and False Negative Rate (FNR). For agriculture field applications, perception models must operate in real time.

B. IMPLEMENTATION DETAILS

1) TRAINING PROCEDURE

The binarized student network is trained with the Adam optimizer, using a base learning rate $\eta = 0.001$ and weight decay $\lambda = 0.0005$. We set momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is decayed by a factor of 0.8 every 10 epochs. The network is trained for 200 epochs, after which we reinitialize and fine-tune the batch normalization statistics for stability.

C. TEACHER MODEL SELECTION

The following analysis aims to identify suitable modules for BiAgriNet, including the encoder and decoder, a local binary approximation method (e.g., XNOR [5], CompactBNN [45], and ABC [46]), and structural approximation strategies for the bottleneck. Table 2 demonstrates on the PhenoBench dataset how various encoder–decoder pairs perform on a 1024 × 1024 input resolution when used as teacher models, measuring mIoU, GOPs, and memory. Notably, ResNet18 + DeepLabV3 achieves the highest mIoU (89.97%) while maintaining a relatively modest 44.7 GOPs and a memory footprint of 21.6 MB. In contrast, VGG16-based configurations (e.g., FCN8, U-Net) offer competitive accuracy but at a significantly higher computational and memory cost, this is partly due to VGG16's parameter-heavy structure. ResNet18-based FCN8 or U-Net remain more efficient in compute (31.4–35.7 GOPs) but delivers slightly lower mIoU (88.95–89.23%). Thus, ResNet18 + DeepLabV3 emerges as a strong "teacher" candidate, balancing advanced multi-scale features (via DeepLab) with a lighter backbone (ResNet18).

TABLE 2. Comparison of State-of-the-Art encoder/decoder configurations at 1024×1024 input resolution for teacher model on phenobench dataset.

Encoder	Decoder	mIoU [%]	GOPs	Mem. [MB]
VGG16	FCN8	88.57	334.2	402.8
VGG16	U-Net	87.12	451.8	408.3
ResNet18	FCN8	88.95	**31.4**	33.9
ResNet18	U-Net	89.23	35.7	34.9
ResNet18	**DeepLabV3**	**89.97**	44.7	**21.6**

D. BINARISATION SCHEMES AND STUDENT CAPACITY

Tables 3 and 4 indicate that both student capacity and teacher strength influence distillation performance. We chose ResNet-18+DeepLabV3 as the final student due to its optimal balance of accuracy and efficiency, while preserving a robust architectural affinity with the teacher. When derived from the more robust teacher (ResNet-18+DeepLabV3), the ResNet-based student achieves a mIoU of 85.46%, while the MobileNet-V3 student attains 82.72%, reflecting a decrease of 3.2% in mIoU. With the less effective teacher (ResNet-18+FCN8), the loss increases to 5.34% mIOU, suggesting that MobileNet-V3 is unable to completely assimilate the teacher's knowledge when the supervision is less informative. The margins, although quantitatively low, result in misclassified pixels on standard field images, which is essential for subsequent operations like precision spraying, where each inaccurate weed identification leads to herbicide wastage. The ResNet-18 backbone preserves skip connections and channel widths analogous to the teacher, facilitating a high layer-wise cosine similarity (approximately 0.9 in deeper blocks) without additional memory (0.86 MB) or computational cost (0.62 GFLOPs). The MobileNet-V3 variant is 0.36 MB smaller and 0.24 GFLOPs lighter; however, its weaker feature alignment results in the observed accuracy decline. The Jetson AGX Orin readily supports the extra 0.36 MB and 0.24 GFLOPs while functioning well within its thermal limits.

TABLE 3. Local binary approximation in teacher–student frameworks(ResNet18 + DeepLabV3 student).

Student (Binarized)	Scheme	mIoU [%]	Flops[G]	Mem. [MB]
(1) Teacher: ResNet18 + FCN8				
ResNet18 + DeepLabV3	XNOR	83.67	0.62	0.86
ResNet18 + DeepLabV3	Compact	81.03	1.34	0.86
ResNet18 + DeepLabV3	ABC	81.92	4.86	2.57
(2) Teacher: ResNet18 + DeepLabV3				
ResNet18 + DeepLabV3	XNOR	**85.46**	**0.62**	**0.86**
ResNet18 + DeepLabV3	Compact	82.39	1.34	0.86
ResNet18 + DeepLabV3	ABC	83.17	4.86	2.57

TABLE 4. Local binary approximation in teacher–student frameworks (MobileNet-V3 student).

Student (Binarized)	Scheme	mIoU [%]	Flops[G]	Mem. [MB]
(1) Teacher: ResNet18 + FCN8				
MobileNet-V3	XNOR	79.20	0.38	0.50
MobileNet-V3	Compact	76.54	0.82	0.50
MobileNet-V3	ABC	77.40	2.25	1.28
(2) Teacher: ResNet18 + DeepLabV3				
MobileNet-V3	XNOR	**82.72**	**0.38**	**0.50**
MobileNet-V3	Compact	79.98	0.82	0.50
MobileNet-V3	ABC	81.56	2.25	1.28

The BiAgriNet with Dilations and ASPP achieves superior performance due to its ability to capture multi-scale context using the ASPP module and expanded receptive fields through dilated convolutions. The other configurations prioritize computational efficiency by excluding ASPP or dilations, which limits their feature extraction capabilities. This underscores the importance of advanced bottleneck designs in balancing performance and efficiency.

E. BOTTLENECK CONFIGURATION STUDY

Table 5 focuses on bottleneck design within the binarized student (BiAgriNet). It compares a BiAgriNet without dilations and no ASPP, a BiAgriNet with dilations but no ASPP, and a BiAgriNet with dilations and an ASPP block. We see a progression in mIoU: 83.37% rises to 84.26%, then 85.46% for the full "BiAgriNet" approach. At the same time, the number of flops increases slightly (0.54–0.62) while memory usage remains relatively constant (0.58–0.86 MB). Hence, adding dilations and an ASPP module improves segmentation quality (by 2% absolute mIoU), though it slightly increases memory usage.

TABLE 5. Different bottleneck configuration for BiAgriNet.

Bottleneck Config.	mIoU	Flops	Mem.
BiAgriNet w/o Dilations w/o ASPP	83.37	0.54	0.58
BiAgriNet w Dilations w/o ASPP	84.26	**0.54**	**0.58**
BiAgriNet w Dilations w ASPP	**85.46**	0.62	0.86

TABLE 6. Performance comparison of BiAgriNet to baseline models (phenobench dataset).

Approach	Parameters [MB]	Flops [G]	mIoU [%]	Accuracy [%]	FPR [%]	FNR [%]	Speed [FPS]
FCN8s	33.9	31.4	85.8	86.1	1.7	5.9	37
DeepLabv3	21.5	39.8	86.3	86.4	**1.4**	5.7	19
UNet	34.7	35.8	**86.7**	**86.9**	1.7	**5.6**	31
FCN8s-XNOR	2.2	31.4	84.3	84.7	4.1	8.9	186
ERFNet	4.3	1.2	82.7	83.1	5.3	9.4	166
BiseNetv2	11.2	31.8	84.9	85.2	2.6	6.7	**196**
ENet	1.5	5.7	71.3	71.9	13.4	16.8	178
EdgeNet	6.0	26.2	83.7	84.1	6.4	9.7	62
BiAgriNet(Ours)	**0.8**	**0.62**	85.5	85.8	2.1	6.5	180

F. DISCUSSION

Table 6 compares our BiAgriNet in terms of accuracy, computational efficiency, and memory usage against state-of-the-art segmentation architectures. Most prominently, BiAgriNet uses only 0.8 MB of parameters—an extreme reduction compared to classical models like FCN8s (33.9 MB) or DeepLabv3 (21.5 MB). It is 27× smaller than DeepLabv3, reflecting how binarized weights drastically cut memory requirements. Despite this compression, BiAgriNet maintains competitive mIoU (85.5%), comparable to FCN8s (85.8%) and outpacing model like ERFNet having 2× less computation and 5× less parameters. BiAgriNet achieves 180 FPS, placing it among the fastest approaches—nearly 5× faster than FCN8s (37 FPS) and 9× faster than DeepLabv3 (19 FPS). The accuracy (85.8%) remains robust, and FPR/FNR values (2.1% and 6.5%) are on par with other real-time networks such as BiseNetv2 confirming that such a dramatic reduction in parameters and compute does not incur a prohibitive drop in segmentation quality. Overall BiAgriNet shows ultra-low memory consumption, high speed, and competitive accuracy, underscoring the viability of binarized networks in demanding, real-time applications. We compare **BiAgriNet** against prior models on the Crop/Weed Field Image Dataset (CWFID), focusing on network size, computational cost, accuracy, and speed. **Table 7** summarizes the results. BiAgriNet matches the top mIoU on CWFID with far fewer parameters and FLOPs, yielding much higher FPS. BiAgriNet achieves state-of-the-art segmentation accuracy (mIoU) on CWFID while using significantly fewer parameters and FLOPs than earlier architectures, resulting in much higher inference speed. In Table 7, earlier deep CNNs like SegNet and U-Net delivered moderate accuracy on CWFID (around 74.06% mIoU for U-Net) but with large size(33 MB). Their inference speed on the Jetson Orin is relatively low (26–32 FPS) due to heavy computation. Even FCN-8s, with 31.4MB size (after converting VGG16 fully-connected layers to convolutions), only achieved 59.31% mIoU, indicating that mere model size did not translate to better performance on this task. DeepLabv3 (with 21.2MB size) reached 70.7% mIoU, slightly below U-Net's accuracy on CWFID, likely because

TABLE 7. Semantic segmentation on *CWFID*. BiAgriNet matches the best mIoU while using far fewer parameters and FLOPs, achieving 187 FPS on Jetson Orin.

Model	Parameters [MB]	FLOPs [G]	mIoU [%]	Speed [FPS]
SegNet (VGG16)	29.5	∼29	53.94	∼32
U-Net	33.7	∼33	74.06	∼26
FCN-8s(VGG16)	31.4	∼31	59.31	∼29
DeepLabV3	21.2	∼38	70.74	∼19
CED-Net	∼20.6	∼20	77.61	∼55
MTS-CNN	∼24.1	∼27	83.72	∼35
Hybrid F.S.	∼32.3	∼32	86.13	∼28
BiAgriNet (Ours)	**0.9**	**0.65**	**86.7**	**187**

CWFID's small training set (60 images) limits the benefit of very deep or complex backbones. Conventional models still struggle on CWFID despite their larger capacity. Specialized architectures tailored for crop-weed segmentation show clear improvements. CED-Net introduced a cascaded encoder-decoder design, boosting mIoU to 77.6% on CWFID about 3% higher than U-Net. This efficiency reflects CED-Net's focused design for two-class (crop vs weed) segmentation. Building on that idea, MTS-CNN added multi-task learning (jointly optimizing crop, weed, and combined classes) to better exploit class correlation. MTS-CNN achieved 83.7% mIoU on CWFID, a significant jump of 6% over CED-Net. This made MTS-CNN one of the top performers on CWFID, though with a slightly larger model. A recent approach by Janneh et al. [40] incorporated a hybrid feature selection (Hybrid F.S.) module, pushing mean IoU to 86.13% the highest reported on CWFID by effectively fusing multi-level features. This Hybrid F.S. model outperformed MTS-CNN by 2.4 percentage points, indicating diminishing returns as accuracy nears the dataset's ceiling. Its model size (32MB) is still moderate, and it was designed with some emphasis on efficiency. BiAgriNet exceeds this state-of-the-art accuracy (86% mIoU) on CWFID while being substantially lighter and faster with 0.9 MB model size, BiAgriNet (binarized) is approximately 20 × smaller than networks (FP32). Its FLOPs are correspondingly low (0.65 G), which, combined with an optimized architecture (e.g. efficient grouped ASPP and streamlined decoder), enables real-time inference. As seen in Table 7, BiAgriNet runs at 187 FPS on the AGX Orin – roughly 7-9× faster than U-Net or DeepLabv3. BiAgriNet achieves high precision (86.7% mIoU) on the CWFID dataset without the heavy computation and large memory footprint of earlier models. BiAgriNet's performance is on par with the crop-weed segmentation methods (exceeding older ones by a large margin) while offering superior efficiency. Recent works prioritized speed at the cost of accuracy. For example, Liu et al. [3] proposed a real-time multi-branch CNN for crop-weed recognition at resolution of 512 × 512 that achieved only 74.6% mIoU on CWFID having lower memory footprint (21.2MB) and faster (93.6 FPS) than U-Net(balancing speed/accuracy). In contrast, BiAgriNet manages to balance both, maintaining SOTA-level accuracy with real-time throughput. The comparison demonstrate how BiAgriNet's design (knowledge distillation, lightweight modules) yields an excellent trade-off: it outperforms older large models (U-Net, DeepLabv3) by 10–15% absolute mIoU, and is on par with the best modern specialized models, remain smaller and faster than all of them.

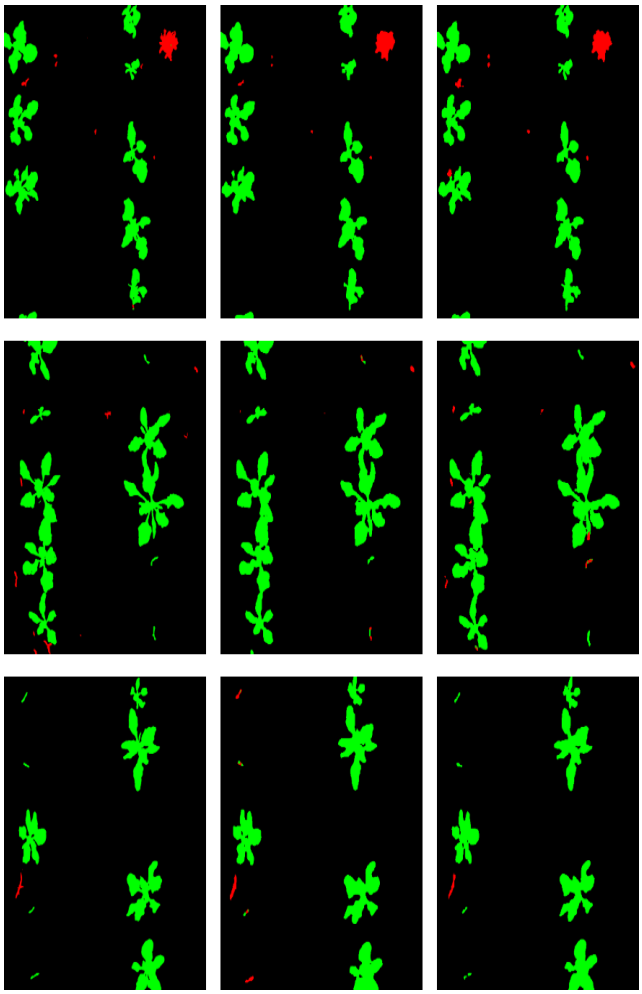


FIGURE 5. Qualitative Results (PhenoBench Dataset): DeepLabv3 [16] (first column), BiseNetv2 [28] (second column), and BiAgriNet (Ours, third column).

Fig 5 presents qualitative results on various images from the PhenoBench dataset. The predictions demonstrate that the slight accuracy degradation is negligible when weighed against the significant performance benefits.

G. RUNTIME AND ENERGY EFFICIENCY

The results (Table 8 and Fig 6, Fig 7) characterise how much computational work each network extracts from a fixed power budget. BiAgriNet consumes just 0.028 J frame^{-1}, whereas classical, full-precision U–Net, DeepLabV3 and FCN8s require 0.54–1.05 J frame^{-1} 20–35× more energy for the same 1024 × 1024 inference (Fig. 6). This gap stems from BiAgriNet's 1-bit weights and 0.62 G FLOPs workload, versus the multi-gigaflop workloads and heavier memory traffic of the VGG-16 and ResNet-based baselines.

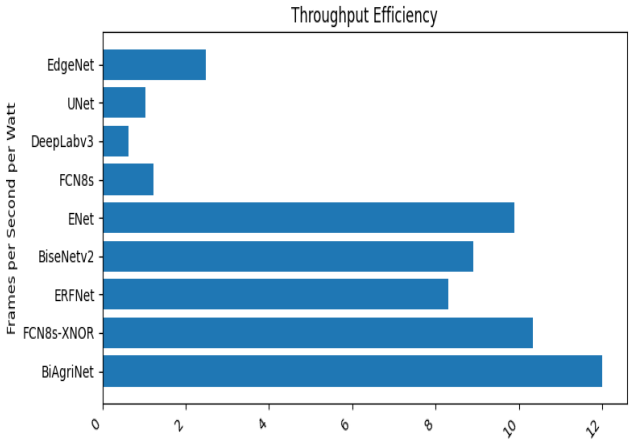


FIGURE 6. Throughput per watt on Jetson AGX Orin (idle→active power as in Table 8). BiAgriNet delivers 12 FPS W^{-1}, doubling ERFNet/BiseNetV2 and exceeding FP baselines by 10–20×.

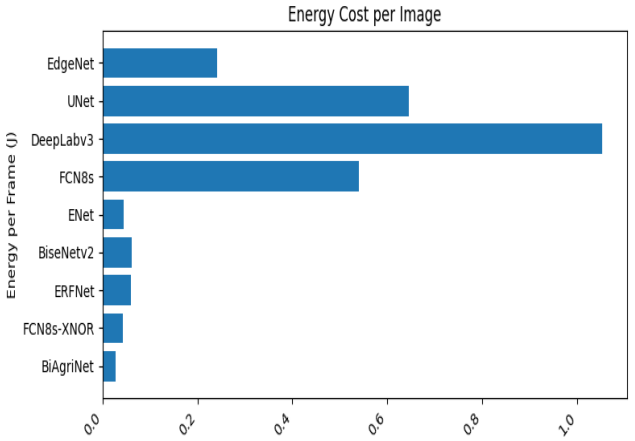


FIGURE 7. Energy per frame (J) at 1024×1024 on Jetson AGX Orin. BiAgriNet uses 0.028 J/frame, an order of magnitude lower than U-Net, DeepLabV3, and FCN-8s.

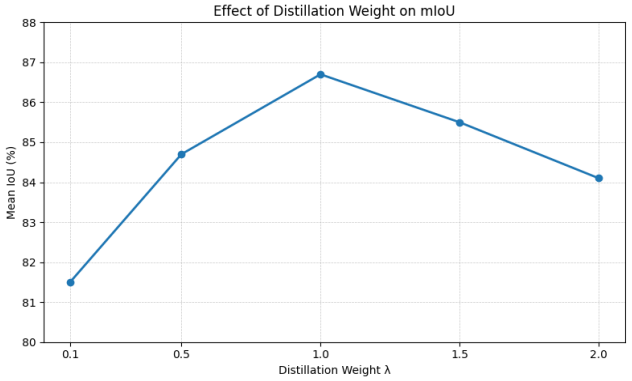


FIGURE 8. Effect of distillation weight λ on CWFID mIoU (teacher: ResNet-18 + DeepLabV3, $\tau = 4$). Moderate λ values are best; $\lambda \approx 1.0$ achieves peak accuracy.

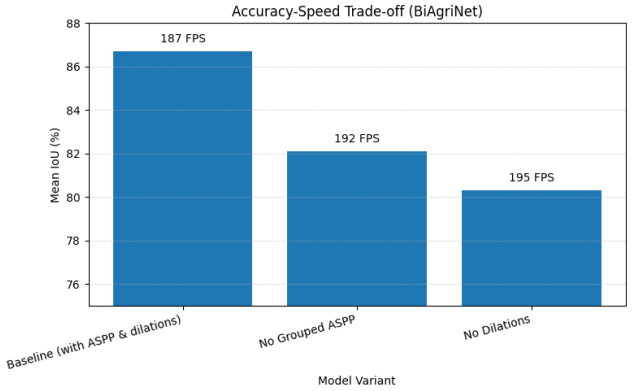


FIGURE 9. Accuracy–speed trade-off on *CWFID* when ablating BiAgriNet. Removing grouped-ASPP or dilations yields 2–5 FPS gains but lowers mIoU by 4–6 pp; the baseline model provides the best accuracy–speed operating point.

TABLE 8. Power–throughput metrics on Jetson AGX Orin (idle→active). BiAgriNet combines the smallest ΔPower (5 W) with 180 FPS, yielding the lowest energy per frame and the highest FPS/W among contenders.

Model	Δ Power (W)	FPS	Energy / Frame (J)	FPS / W
FCN8s-XNOR	8	186	0.043	10.3
ERFNet	10	166	0.060	8.3
BiseNetv2	12	196	0.061	8.9
ENet	8	178	0.045	9.9
FCN8s	20	37	0.54	1.2
DeepLabV3	20	19	1.05	0.6
U-Net	20	31	0.65	1.0
EdgeNet	15	62	0.24	2.5
BiAgriNet(Ours)	**5**	**180**	**0.028**	**12.0**

Throughput per watt favours binarised or heavily pruned models. All lightweight models outperform the legacy networks in FPS W^{-1} (Fig. 7), but BiAgriNet attains the highest score (12 FPS W^{-1}). ENet and FCN8s-XNOR also benefit from reduced precision, yet remain 15–20% less efficient because they either (i) lack multi-scale context modules, forcing deeper layers to process larger feature maps (ENet), or (ii) retain VGG-style 5 × 5 pooling paths that hurt cache locality (FCN8s-XNOR). BiseNetv2 adds guided-aggregation branches and dilated context blocks to boost accuracy, but the extra memory traffic raises active power to 22 W, making its FPS W^{-1} lower than BiAgriNet even though raw FPS is comparable (196 vs 180). Architectural overhead can erase the energy gains of lightweight backbones. BiAgriNet raises board power by only 5 W (10→ 15 W). DeepLabV3 and U–Net triple that increase (Δ20 W), saturating the Orin's thermal headroom and risking throttling. In field robotics where additional sensors and locomotion already tax the battery such deltas often determine whether autonomous operation lasts minutes or hours. The grouped-ASPP maintains multi-scale context without increasing parameter groups, while the binarized convolutions minimize computational load, which decreases energy consumption per frame and enhances frames-per-second per watt (FPS W^{-1}) on Jetson Orin. Empirically,

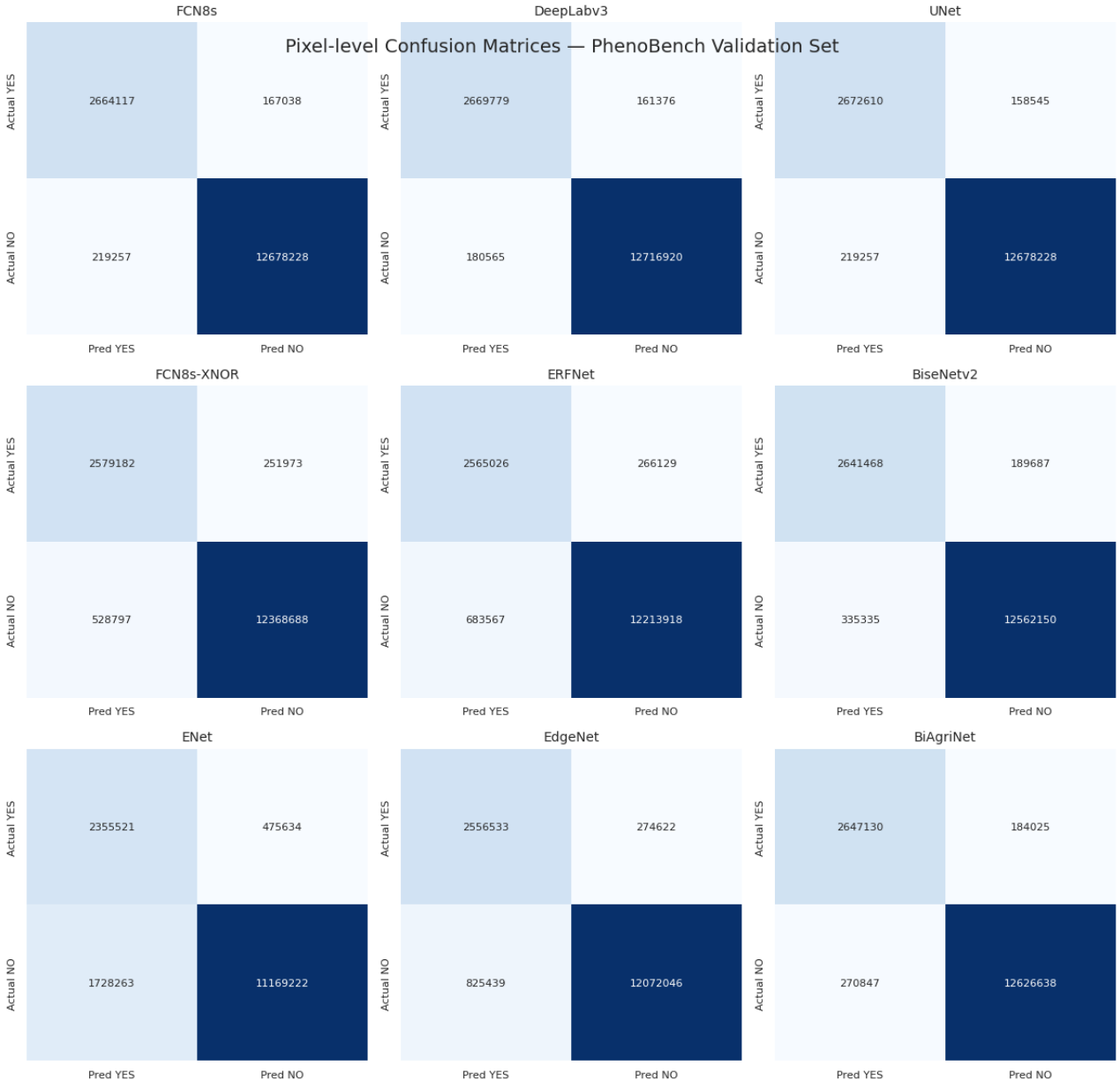


FIGURE 10. Pixel-level confusion matrices for segmentation models on the PhenoBench Validation set (15 images, 15,728,640 pixels).

TABLE 9. Comparison of dataset image sizes and total input pixels per model inference.

Dataset	Native image size	Inference rule (per image)	Test images	Tiles / image	**n pixels per model**
PhenoBench	1024×1024	none—process the whole image	15	1	$15 \times 1024^2 = \mathbf{15\,728\,640}$
CWFID	1296×966	**resize/pad to a single** 1024×1024 **tile**	10	1	$10 \times 1024^2 = \mathbf{10\,485\,760}$

BiAgriNet increases board power by only $\sim$ 5 W yet attains the highest FPS W^{-1} and the lowest energy per frame among contenders, whereas heavier baselines saturate thermal headroom (+20 W), risking throttling. In field robotics, added sensor and locomotion loads tax the battery. These power deltas can determine whether missions last minutes or hours.

H. PRACTICAL DEPLOYMENT CONSIDERATIONS

Agricultural environments present domain shifts from weather, illumination, occlusion, and sensor noise. BiAgriNet's binarized convolutions stabilize power draw, while grouped-dilated ASPP preserves multi-scale context without expanding parameter groups. Distillation with KL($p_S \,\|\, p_T$) maintains teacher-relative class ordering, which

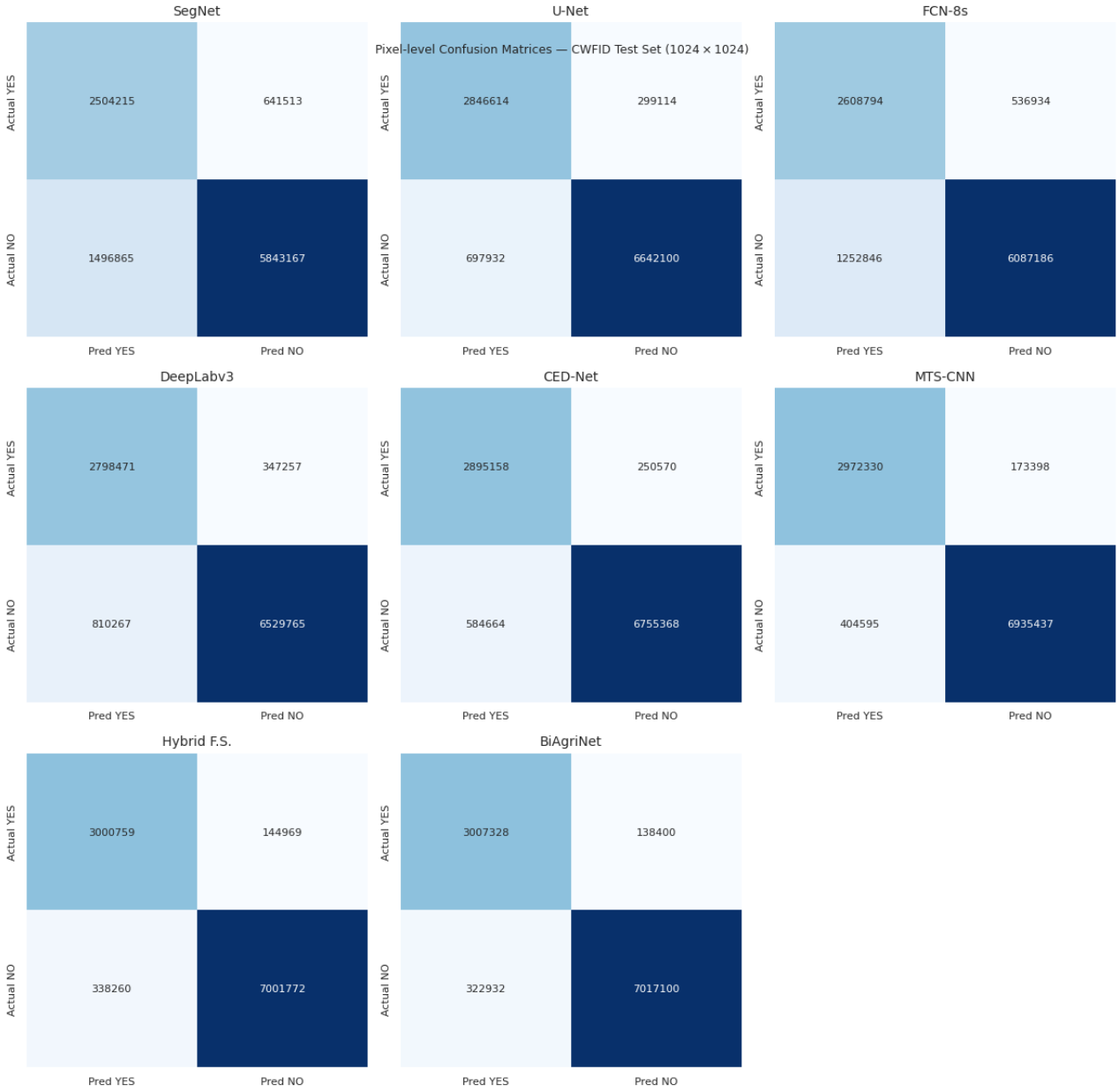


FIGURE 11. Pixel-level confusion matrices for segmentation models on the CWFID Test set (10 images, 10,485,760 pixels).

we observe to reduce overconfident errors under illumination changes. Combined with the measured energy-per-frame and FPS-per-watt advantages on Jetson Orin, these factors support operation in battery- and thermally-constrained field robots, where power headroom and runtime directly affect mission duration.

V. ABLATION STUDY

We conducted an **ablation study** to further justify BiAgriNet's choices. We examined two key aspects: (a) the effect of the **distillation weight** (λ) used in training BiAgriNet with a teacher model, and (b) the impact of removing certain architectural components (grouped ASPP module and dilated convolutions) on accuracy and speed. As shown in Fig 8, the distillation weight λ has a noticeable effect on final mIoU. Without distillation ($\lambda \approx 0$), BiAgriNet attained around 81.5% mIoU. As we increase λ to give more weight to teacher guidance, the student's precision improves substantially, peaking at $\lambda = 1.0$ with $\sim$ 86.7% mIoU. At this optimal setting, the distillation loss and the standard segmentation loss are balanced, BiAgriNet effectively learns from both the ground-truth labels and the teacher's softer predictions. This helps the student network generalize better despite its limited capacity. However, using an overly large distillation weight can hurt performance. At $\lambda = 2.0$ (twice the weight on distillation vs. classification loss), mIoU drops to $\sim$ 84.1%. In this regime, the student may overemphasize mimicking

TABLE 10. Architecture details for BiAgriNet. All convolutions in the encoder, bottleneck, and decoder are binarized (±1) with STE during backprop; the final classifier is full precision.

Block	Layer	Op (kernel / stride / dil.)	In → Out (C×H×W)	Activation / Norm
Stem (FP32)	Conv1 + MaxPool	$7 \times 7/2/1$ + pool $3 \times 3/2$	$3 \times 1024 \times 1024 \rightarrow 64 \times 256 \times 256$	ReLU, BN
Encoder (BNN)	Stage 1 (ResBlock×2)	$3 \times 3/1/1$	$64 \times 256 \times 256 \rightarrow 64 \times 256 \times 256$	Sign/STE, BN
	Stage 2 (ResBlock×2)	$3 \times 3/2/1$ (first), then 1	$64 \times 256 \times 256 \rightarrow 128 \times 128 \times 128$	Sign/STE, BN
	Stage 3 (ResBlock×2)	$3 \times 3/2/1$ (first), then 1	$128 \times 128 \times 128 \rightarrow 256 \times 64 \times 64$	Sign/STE, BN
	Stage 4 (ResBlock×2)	$3 \times 3/2/1$ (first), then 1	$256 \times 64 \times 64 \rightarrow 512 \times 32 \times 32$	Sign/STE, BN
Bottleneck (BNN)	Grouped Binary ASPP	parallel 3×3 dilations $\{1, 6, 12, 18\}$	$512 \times 32 \times 32 \rightarrow 512 \times 32 \times 32$	Sign/STE, BN
	Fusion 1×1	$1 \times 1/1/1$	$512 \times 32 \times 32 \rightarrow 256 \times 32 \times 32$	Sign/STE, BN
Decoder (BNN)	Up×2 + Conv	upsample×2, $3 \times 3/1/1$	$256 \times 32 \times 32 \rightarrow 128 \times 64 \times 64$	Sign/STE, BN
	Up×2 + Conv	upsample×2, $3 \times 3/1/1$	$128 \times 64 \times 64 \rightarrow 64 \times 128 \times 128$	Sign/STE, BN
	Up×2 + Conv	upsample×2, $3 \times 3/1/1$	$64 \times 128 \times 128 \rightarrow 32 \times 256 \times 256$	Sign/STE, BN
	Up×2 + Conv	upsample×2, $3 \times 3/1/1$	$32 \times 256 \times 256 \rightarrow 32 \times 512 \times 512$	Sign/STE, BN
Classifier (FP32)	1×1 Conv	$1 \times 1/1/1$	$32 \times 512 \times 512 \rightarrow \mathbf{C} \times 512 \times 512$	(linear)
	Final Upsample	bilinear ×2	$\mathbf{C} \times 512 \times 512 \rightarrow \mathbf{C} \times 1024 \times 1024$	—

Notes. Encoder based on ResNet-18 with binarized convs (except the very first conv). Bottleneck uses grouped, dilated binary branches (rates $\{1, 6, 12, 18\}$). Decoder mirrors DeepLab-style refinement with binary 3×3 layers; classifier kept full precision for stable logits. (See Sec. III-B and Fig. 2.)

the teacher at the expense of learning the true class signals, especially if the teacher model has its own biases or errors. The trend in the plot (λ vs. mIoU) suggests a concave shape, an optimal middle ground exists. In practice, $\lambda \approx 1.0$ gave the best results, indicating that knowledge distillation significantly boosted BiAgriNet's accuracy (by $\sim$ 5.2% absolute) when properly weighted, but too much reliance on the teacher can be counterproductive. Fig 9 shows the accuracy–speed trade-off. The baseline BiAgriNet operates at 86.7% mIoU and ∼187 FPS. Removing grouped-ASPP yields ∼192 FPS with ∼82.1% mIoU; removing dilations yields ∼195 FPS with ∼80.3% mIoU. We therefore keep both to maximize accuracy at real-time speed. Similarly, eliminating network dilations by, for instance, replacing dilated layers with standard convolution results in an even greater speedup (to about 195 FPS) at the expense of accuracy dropping to about 80.3% mIoU. The grouped ASPP contributed approximately +4.6% mIoU (from 82.1% → 86.7%) at a speed cost of about 2.6%, while the dilated convolutions contributed approximately +6.4% mIoU (from 80.3% → 86.7%) at a runtime-per-frame cost of about half. This represents a traditional balanced trade-off between accuracy and efficiency. The ablation study shows that BiAgriNet's complete architecture is strategically selected: the grouped ASPP and dilated convolutions each enhance IoU, and collectively they facilitate the model in achieving state-of-the-art accuracy. Removing grouped-ASPP or dilations yields small frame-rate gains ($\sim$ 2 − 5 FPS) but degrades mIoU by 4 − 6 pp; in weed-control robotics, this increases false actuations and downstream costs, so the baseline BiAgriNet configuration is the preferred accuracy–speed operating point. The trade-off analysis visually highlights how each architectural component alters the equilibrium. By maintaining both, BiAgriNet functions at the optimal intersection of this trade-off, providing high precision while simultaneously satisfying real-time speed demands.

VI. CONCLUSION

This paper introduces a novel BiAgriNet, a binarized knowledge distilled network designed for efficient, high-accuracy semantic segmentation in resource-constrained environments. By integrating advanced bottleneck designs, including dilated convolutions and the group ASPP module, and leveraging efficient teacher-student knowledge distillation with XNOR binarization, BiAgriNet achieves an excellent balance between performance and computational efficiency. Comprehensive evaluations demonstrate its superior speed, minimal memory usage, and competitive accuracy compared to state-of-the-art segmentation models, making it particularly suitable for real-time applications like precision agriculture. Future work will explore further enhancements in binarization techniques and deploy BiAgriNet on edge devices for practical validation.

TABLE 11. Training hyperparameters used for BiAgriNet.

Category	Setting
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Base learning rate	$\eta = 0.001$
LR schedule	Multiply by 0.8 every 10 epochs
Weight decay	5×10^{-4}
Epochs	200
Batch size	(set per GPU memory; e.g., $8 \sim 16$ on 16 GB)
Distillation weight	$\lambda = 1.0$
KD temperature	$\tau = 4$
Teacher	ResNet-18 + DeepLabV3 (frozen)
Student	Binarized ResNet-18 encoder + grouped-dilated ASPP + binarized decoder
Ignore index	ξ (as used in evaluation; keep consistent across loss/metrics)
Input resolution	1024×1024 (PhenoBench/CWFID experiments)
Losses	$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{KD}}$, $\mathcal{L}_{\text{KD}} = \text{KL}(p_S \| p_T)$
Precision	Binary weights/activations (STE); FP32 classifier head
Hardware (train / infer)	Train: A4000 (16 GB) $\times 3$; Infer: Jetson AGX Orin

APPENDIX A
CONSTRUCTION OF PIXEL LEVEL CONFUSION MATRICES

See Table 9, Fig 10 and 11. In the binary panels (rows: ground truth; columns: prediction), the two off-diagonals correspond to (i) *plant*→*background* errors (FN, top-right) and (ii) *background*→*plant* errors (FP, bottom-left). On *PhenoBench* (Fig. 10), BiAgriNet concentrates mass on the diagonal and the dominant residual is FN at thin boundaries and under difficult lighting; FP is comparatively smaller, indicating that the model is conservative rather than over-segmenting. On *CWFID* (Fig. 11), errors are more balanced, with a modest increase in FP due to soil/stubble textures that resemble vegetation, while boundary FN persists as the second main mode. Across baselines, larger FP blocks indicate over-segmentation; BiAgriNet reduces these FP errors while maintaining low FN, consistent with its higher mIoU and energy efficiency.

APPENDIX B
BIAGRINET ARCHITECTURE AND HYPERPARAMETER SETTINGS

See Table 10 and Table 11.

REFERENCES

[1] L. Lei, Q. Yang, L. Yang, T. Shen, R. Wang, and C. Fu, "Deep learning implementation of image segmentation in agricultural applications: A comprehensive review," *Artif. Intell. Rev.*, vol. 57, no. 6, p. 149, May 2024.

[2] I. Ulku, "ResLMFFNet: A real-time semantic segmentation network for precision agriculture," *J. Real-Time Image Process.*, vol. 21, no. 4, p. 101, Aug. 2024.

[3] Y. Liu, M. Liu, X. Zhao, J. Zhu, L. Wang, H. Ma, and M. Zhang, "Real-time semantic segmentation network for crops and weeds based on multi-branch structure," *IET Comput. Vis.*, vol. 18, no. 8, pp. 1313–1324, Dec. 2024.

[4] Y. Zhang and C. Lv, "TinySegformer: A lightweight visual segmentation model for real-time agricultural pest detection," *Comput. Electron. Agricult.*, vol. 218, Mar. 2024, Art. no. 108740.

[5] M. Rastegari, V. Ordóñez, J. Redmon, and A. Farhadi, "XNOR-Net: ImageNet classification using binary convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 525–542.

[6] C.-H. Huang, "An FPGA-based hardware/software design using binarized neural networks for agricultural applications: A case study," *IEEE Access*, vol. 9, pp. 26523–26531, 2021.

[7] P. Ma, J. Zhu, and G. Zhang, "A binary neural network with dual attention for plant disease classification," *Electronics*, vol. 12, no. 21, p. 4431, Oct. 2023.

[8] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Binarized neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2022, pp. 4107–4115.

[9] J. Fromm, M. Cowan, M. Philipose, L. Ceze, and S. Patel, "Riptide: Fast end-to-end binarized neural networks," in *Proc. Mach. Learn. Syst. (MLSys)*, 2020, pp. 379–389.

[10] S. Qiu, J. Chen, X. Li, R. Wan, X. Xue, and J. Pu, "Make a strong teacher with label assistance: A novel knowledge distillation approach for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2024, pp. 371–388.

[11] M. Li, M. Halstead, and C. McCool, "Knowledge distillation for efficient instance semantic segmentation with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 5432–5439.

[12] T. Liu, C. Chen, X. Yang, and W. Tan, "Rethinking knowledge distillation with raw features for semantic segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2024, pp. 1155–1164.

[13] A. Karine, T. Napoléon, and M. Jridi, "I2CKD: Intra- and inter-class knowledge distillation for semantic segmentation," *Neurocomputing*, vol. 2025, May 2025, Art. no. 130791.

[14] J. Weyler, F. Magistri, E. Marks, Y. L. Chong, M. Sodano, G. Roggiolani, N. Chebrolu, C. Stachniss, and J. Behley, "PhenoBench—A large dataset and benchmarks for semantic image interpretation in the agricultural domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 9583–9594, Dec. 2024.

[15] S. Haug and J. Östermann, "A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks," in *Proc. Comput. Vis.-ECCV Workshops*, 2015, pp. 105–116.

[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[18] 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. 6230–6239.

[19] A. Heschl, M. Murillo, K. Najafian, and F. Maleki, "SynthSet: Generative diffusion model for semantic segmentation in precision agriculture," 2024, *arXiv:2411.03505*.

[20] Z. Zhu, M. Jiang, J. Dong, S. Wu, and F. Ma, "PD-SegNet: Semantic segmentation of small agricultural targets in complex environments," *IEEE Access*, vol. 11, pp. 90214–90226, 2023.

[21] D. Popescu, L. Ichim, and O. A. Sava, "Semantic segmentation of small region of interest for agricultural research applications," in *Proc. 13th Int. Conf. Comput. Collective Intell.*, 2021, pp. 585–598.

[22] H. Yu, M. Che, H. Yu, and Y. Ma, "Research on weed identification in soybean fields based on the lightweight segmentation model DCSAnet," *Frontiers Plant Sci.*, vol. 14, Dec. 2023, Art. no. 1268218.

[23] Z. Wang, W. Cui, C. Huang, Y. Zhou, Z. Zhao, Y. Yue, X. Dong, and C. Lv, "Framework for apple phenotype feature extraction using instance segmentation and edge attention mechanism," *Agriculture*, vol. 15, no. 3, p. 305, Jan. 2025.

[24] Y. Wu, L. Tang, and S. Yuan, "Semantic segmentation model of multi-source remote sensing images was used to extract winter wheat at tillering stage," *Sci. Rep.*, vol. 15, no. 1, pp. 1–14, Apr. 2025.

[25] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[26] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[27] C. Yu, J. Wang, C. Peng, C. Gao, and G. Yu, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 325–341.

[28] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "BiSeNet v2: Bilateral network with guided aggregation for real-time semantic segmentation," *Int. J. Comput. Vis.*, vol. 129, no. 11, pp. 3051–3068, Nov. 2021.

[29] H.-Y. Han, Y.-C. Chen, P.-Y. Hsiao, and L.-C. Fu, "Using channel-wise attention for deep CNN based real-time semantic segmentation with class-aware edge information," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1041–1051, Feb. 2021.

[30] G. Gao, G. Xu, Y. Yu, J. Xie, J. Yang, and D. Yue, "MSCFNet: A lightweight network with multi-scale context fusion for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 25489–25499, Dec. 2022.

[31] G. Xu, J. Li, G. Gao, H. Lu, J. Yang, and D. Yue, "Lightweight real-time semantic segmentation network with efficient transformer and CNN," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 15897–15906, Dec. 2023.

[32] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 12077–12090.

[33] X. Zhang, H. Li, J. Ru, P. Ji, and C. Wu, "LACTNet: A lightweight real-time semantic segmentation network based on an aggregated convolutional neural network and transformer," *Electronics*, vol. 13, no. 12, p. 2406, Jun. 2024.

[34] G. Feng, C. Wang, A. Wang, Y. Gao, Y. Zhou, S. Huang, and B. Luo, "Segmentation of wheat lodging areas from UAV imagery using an ultra-lightweight network," *Agriculture*, vol. 14, no. 2, p. 244, Feb. 2024.

[35] S. Landgraf, K. Wursthorn, M. Hillemann, and M. Ulrich, "DUDES: Deep uncertainty distillation using ensembles for semantic segmentation," *PFG-J. Photogramm., Remote Sens. Geoinformation Sci.*, vol. 92, no. 2, pp. 101–114, Apr. 2024.

[36] R. N. Nair and R. Hänsch, "Let me show you how it's done–cross-modal knowledge distillation as pretext task for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 595–603.

[37] T. Anand, S. Sinha, M. Mandal, V. Chamola, and F. R. Yu, "AgriSegNet: Deep aerial semantic segmentation framework for IoT-assisted precision agriculture," *IEEE Sensors J.*, vol. 21, no. 16, pp. 17581–17590, Aug. 2021.

[38] A. Khan, T. Ilyas, M. Umraiz, Z. I. Mannan, and H. Kim, "CED-Net: Crops and weeds segmentation for smart farming using a small cascaded encoder–decoder architecture," *Electronics*, vol. 9, no. 10, p. 1602, Oct. 2020.

[39] Y. H. Kim and K. R. Park, "MTS-CNN: Multi-task semantic segmentation-convolutional neural network for detecting crops and weeds," *Comput. Electron. Agricult.*, vol. 199, Aug. 2022, Art. no. 107146.

[40] L. L. Janneh, Y. Zhang, M. Hydara, and Z. Cui, "Deep learning-based hybrid feature selection for the semantic segmentation of crops and weeds," *ICT Exp.*, vol. 10, no. 1, pp. 118–124, Feb. 2024.

[41] Y. H. Kim, S. J. Lee, C. Yun, S. J. Im, and K. R. Park, "LCW-Net: Low-light-image-based crop and weed segmentation network using attention module in two decoders," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 106890.

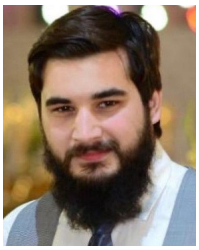
[42] Y. Wei, Y. Feng, X. Zhou, and G. Wang, "Attention-aided lightweight networks friendly to smart weeding robot hardware resources for crops and weeds semantic segmentation," *Frontiers Plant Sci.*, vol. 14, Dec. 2023, Art. no. 1320448.

[43] M. D. Islam, W. Liu, P. Izere, P. Singh, C. Yu, B. Riggan, K. Zhang, A. J. Jhala, S. Knezevic, Y. Ge, S. Pitla, J. Luck, and Y. Shi, "Towards real-time weed detection and segmentation with lightweight CNN models on edge devices," *Comput. Electron. Agricult.*, vol. 237, Oct. 2025, Art. no. 110600.

[44] Z. Galymzhankyzy and E. Martinson, "Lightweight multispectral crop-weed segmentation for precision agriculture," 2025, *arXiv:2505.07444*.

[45] W. Tang, G. Hua, and L. Wang, "How to train a compact binary neural network with high accuracy?" in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, 2017, pp. 1–8.

[46] X. Lin, C. Zhao, and W. Pan, "Towards accurate binary convolutional neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.



HASSAN KHAN received the Bachelor of Science degree in electronics engineering and the Master of Science degree in electrical engineering from the Capital University of Science and Technology (CUST), Islamabad, in 2017 and 2020, respectively, where he earned the Dean's Honor Awards for academic excellence. He was a Research Assistant with CUST, focusing on machine learning applications for wireless communication networks and energy-efficient systems. He is currently pursuing the Ph.D. degree in artificial intelligence from the University of Galway. His research interests include AI, deep learning, optimization models, and sustainability.

SUNBAL IFTIKHAR received the Bachelor of Science degree (Hons.) in electronics engineering from Mohammad Ali Jinnah University (MAJU), Islamabad, in 2017, and the Master of Science degree (Hons.) by research, majoring in communication networks and machine learning from the Capital University of Science and Technology (CUST), Islamabad, in 2020. She is currently pursuing the Ph.D. degree in sustainability of AI models. Since 2020, she has been a Research Assistant with the Department of Electrical Engineering, CUST, where she contributed to various projects. She has hands-on experience with Sun Microsystems' SunSPOT testbed, contributing to her solid foundation in research and development. She has conducted significant research in optimizing 6G wireless communication channels using deep learning for path loss prediction and machine learning-driven optimization of indoor 5G infrastructure deployment.

RORY WARD (Member, IEEE) received the bachelor's degree in electronic and computer engineering from the National University of Ireland Galway, in 2021. He is currently pursuing the Ph.D. degree with the University of Galway, where he is sponsored by the Centre for Research Training in Artificial Intelligence. His research interests include generative AI, neuro-symbolic AI, and computer vision. He received the Best Oral Presentation Award at the Irish Machine Vision and Image Processing (IMVIP) Conference 2022 for his paper "Towards Temporal Stability in Automatic Video Colorization". He has also won the Best Paper Award for the AI and Agents track at the ACM/SIGAPP Symposium on Applied Computing (SAC 2025) for his paper "RAGCol: RAG-Based Automatic Video Colorization Through Text Caption Generation and Knowledge Enrichment".

STEVEN DAVY (Member, IEEE) received the B.A. Mod. degree (Hons.) in computer science from Trinity College Dublin and the Ph.D. degree from Waterford Institute of Technology. He is an experienced researcher and leader in the field of sustainable digital technologies. He is currently the Director of the Centre for Sustainable Digital Technologies, TU Dublin. He has over two decades of experience in academia and industry. He has made significant contributions to the field, including developing the first formal model of the Policy Continuum and pioneering work on Edge-as-a-Service. He has led multiple high-profile European research projects and has a strong track record in securing research funding. Throughout his career, he has demonstrated a commitment to bridging the gap between academic research and industry applications. He has been instrumental in developing innovative services for businesses and has successfully commercialized research outcomes, including the creation of a spin-out company. As a leader, he has played a crucial role in developing research teams and mentoring early-career researchers. His research interests include policy-based network management, edge computing, and autonomous systems. He actively contributes to the Broader Research Community through his roles as a Reviewer for funding agencies and academic journals.

JOHN G. BRESLIN (Senior Member, IEEE) received the B.E. and Ph.D. degrees. He is currently a Personal Professor in electronic engineering with the College of Science and Engineering, University of Galway, where he is also the Director of the TechInnovate/AgInnovate programmes. He has taught electronic engineering, computer science, innovation and entrepreneurship topics over the past 25 years. He co-created the SIOC framework, implemented in hundreds of applications (by Yahoo, Boeing, Vodafone, etc.) on at least 65 000 websites with 35 million data instances. He has jointly written over 300 peer-reviewed academic publications, including books on *The Social Semantic Web* and *Social Semantic Web Mining*. Associated with two Taighde Éireann-Research Ireland Centres, he is a Principal Investigator at Insight (Data Analytics) and a Funded Investigator at VistaMilk (AgTech), and a Principal Investigator on the EDIH Data2Sustain.