

# WeedRepFormer: Reparameterizable Vision Transformers for Real-Time Waterhemp Segmentation and Gender Classification

Toqi Tahamid Sarker, Taminul Islam, Khaled R. Ahmed,  
 Cristiana Bernardi Rankrake, Kaitlin E. Creager, Karla Gage  
 Southern Illinois University Carbondale, USA  
 {toqitahamid.sarker, taminul.islam, khaled.ahmed,  
 cris.rankrake, kaitlin.creager, kgage}@siu.edu

## Abstract

*We present WeedRepFormer, a lightweight multi-task Vision Transformer designed for simultaneous waterhemp segmentation and gender classification. Existing agricultural models often struggle to balance the fine-grained feature extraction required for biological attribute classification with the efficiency needed for real-time deployment. To address this, WeedRepFormer systematically integrates structural reparameterization across the entire architecture—comprising a Vision Transformer backbone, a Lite R-ASPP decoder, and a novel reparameterizable classification head—to decouple training-time capacity from inference-time latency. We also introduce a comprehensive waterhemp dataset containing 10,264 annotated frames from 23 plants. On this benchmark, WeedRepFormer achieves 92.18% mIoU for segmentation and 81.91% accuracy for gender classification using only 3.59M parameters and 3.80 GFLOPs. At 108.95 FPS, our model outperforms the state-of-the-art iFormer-T by 4.40% in classification accuracy while maintaining competitive segmentation performance and significantly reducing parameter count by 1.9×.*

## 1. Introduction

Waterhemp (*Amaranthus tuberculatus* (Moq.) Sauer) has emerged as one of the most troublesome weed species in North American agriculture, particularly threatening corn and soybean production across the Midwestern United States [2, 52, 61]. This dioecious species exhibits remarkable adaptability, with separate male and female plants requiring cross-pollination, generating extensive genetic diversity that facilitates rapid evolution of herbicide resistance [12, 33, 46]. Currently, waterhemp populations have developed confirmed resistance to seven herbicide sites of action (SOAs), with an eighth resistance to glufosinate (SOA Group 10) currently being confirmed [18, 40]. This

eighth resistance is particularly concerning as glufosinate represents the last remaining postemergence herbicide option in soybean production [40]. The severity of this problem is compounded by the species’ exceptional fecundity, with female plants producing between 300,000 to over 2 million seeds per plant [16], contributing to rapid population expansion and yield losses up to 74% in corn [53] and 43–73% in soybean [15, 58] under heavy infestation.

The dioecious nature of waterhemp presents both a challenge and an opportunity for precision weed management. Unlike monoecious weeds, waterhemp requires both male and female plants for reproduction, with only female plants producing seeds [35]. Male plants produce wind-borne pollen, leading to extensive gene flow that carries herbicide resistance genes up to 800 m from source plants [27, 55]. Recent studies have shown that herbicide exposure can alter population sex ratios [44, 45], with PPO-inhibitor treatments shifting male-to-female proportions in ways that could affect seedbank composition and resistance evolution. These dynamics suggest that targeted removal of female plants before seed set could significantly reduce population growth and slow herbicide resistance evolution. However, manual identification of waterhemp gender in field conditions is labor-intensive and impractical at scale, while visual differentiation between male and female plants requires expertise, as morphological differences are subtle and vary with growth stage. Automated gender classification through computer vision could enable selective herbicide application or mechanical removal strategies, potentially reducing chemical inputs while improving long-term management efficacy.

Recent advances in deep learning have demonstrated remarkable success in agricultural computer vision tasks [24, 47], with semantic segmentation methods achieving state-of-the-art performance in crop-weed discrimination and identifying weed plant species [23, 49, 51]. Multi-task learning frameworks that jointly optimize related tasks

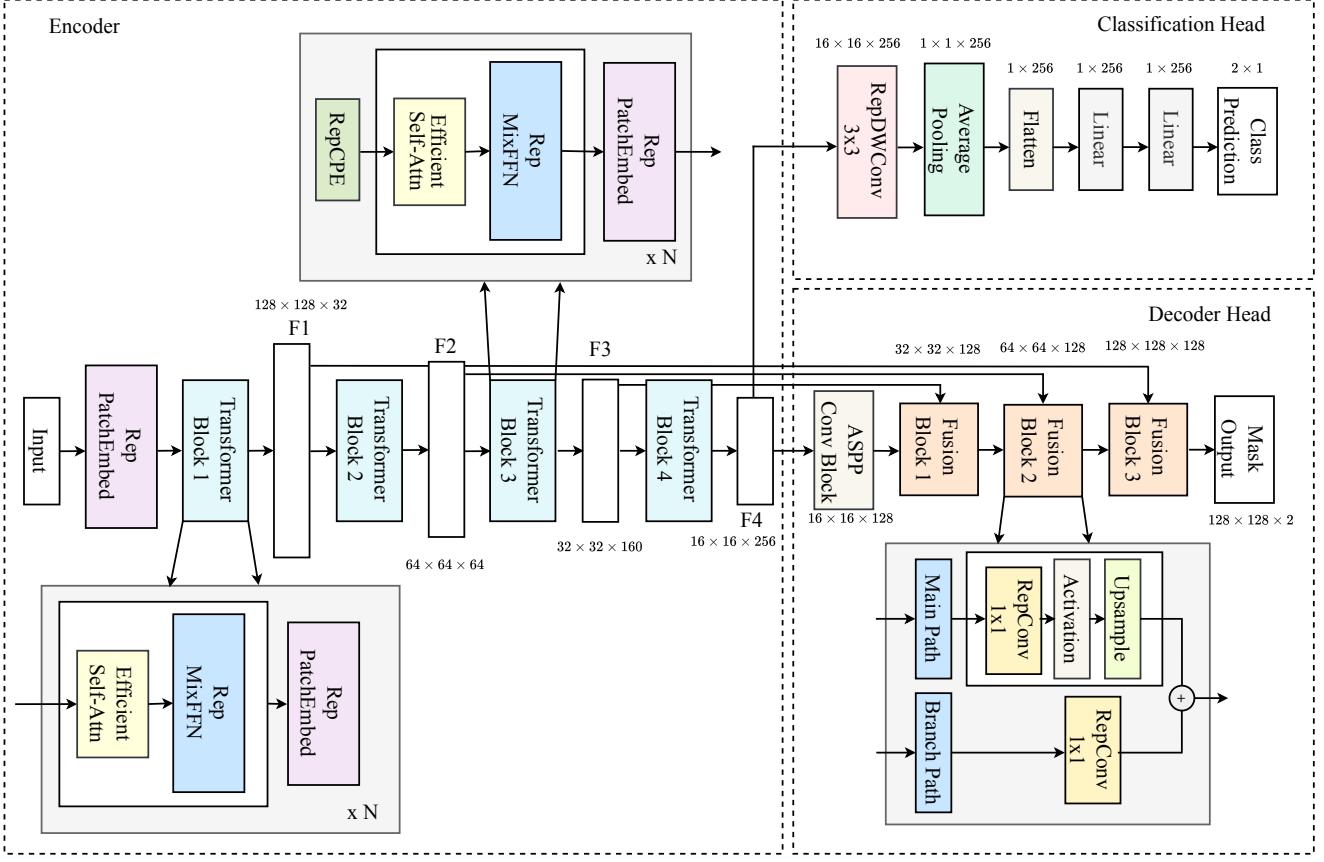


Figure 1. Overview of our multi-task architecture. (a) Network consists of four-stage hierarchical Vision Transformer backbone with reparameterizable components. (b) Reparameterizable patch embedding with multi-branch convolutions. (c) LRASPP decoder with reparameterizable convolutions. (d) Classification head with optional SE attention.

through shared representations have shown improved efficiency and generalization compared to single-task approaches [9, 43]. For waterhemp management, simultaneous segmentation and gender classification are naturally related tasks that benefit from shared feature learning, as both require understanding of plant morphology and spatial structure.

However, practical deployment of deep learning models for agricultural robotics faces significant constraints. Real-time operation requires efficient processing on resource-constrained edge devices such as those mounted on autonomous tractors or unmanned aerial vehicles [38]. While Vision Transformers have demonstrated superior performance in semantic segmentation due to their global receptive fields [62], their computational cost has limited adoption in resource-constrained agricultural scenarios. Standard models like DeepLabV3+ [5] and U-Net [42] achieve high accuracy but require substantial computational resources [34]. Recent lightweight architectures using depthwise separable convolutions [21], neural architecture search [54], or knowledge distillation [19] often sacrifice accuracy for speed or require complex training proce-

dures.

Structural reparameterization offers a promising alternative, enabling networks to train with multiple parallel branches for increased capacity while deploying as efficient single-path models through algebraic fusion [10, 57]. This approach has achieved state-of-the-art efficiency in image classification and detection [56], and has been successfully applied to agricultural CNN-based architectures for disease detection, weed identification, and multi-task learning [14, 37, 50, 67]. However, to the best of our knowledge, structural reparameterization has not yet been applied to Vision Transformers for multi-task dense prediction combining segmentation with fine-grained biological attribute classification in agriculture.

We propose WeedRepFormer, a Vision Transformer architecture that achieves an optimal balance of accuracy and efficiency for simultaneous waterhemp segmentation and gender classification via systematic structural reparameterization. In summary, our contributions are as follows:

- We propose a fully reparameterizable multi-task Vision Transformer that systematically applies structural reparameterization across backbone, segmentation head, and

classification head.

- We introduce a waterhemp dataset with 10,264 annotated frames from 23 plants, including pixel-level segmentation masks and plant-level gender labels, and establish baseline results for this task.

## 2. Related Work

**Weed Detection and Segmentation.** Deep learning has revolutionized weed detection in precision agriculture. Early work adapted general semantic segmentation architectures like FCN [30] and U-Net [42] for agricultural applications, with successful deployment for real-time crop-weed classification [1, 31]. Recent work has explored lightweight architectures for edge devices [26, 60], but these methods primarily address binary crop-weed segmentation without species-specific biological attribute classification.

**Vision Transformers for Segmentation.** Vision Transformers [11] capture long-range dependencies through self-attention but incur high computational costs. Hierarchical variants address this limitation: SegFormer [62] combines efficient Transformer encoders with lightweight MLP decoders, while Swin Transformer [28] uses shifted windows for complexity reduction. These architectures have shown success in general computer vision and are increasingly adopted for crop disease detection and weed species classification, though their application to fine-grained multi-task agricultural problems remains limited.

**Multi-Task Learning.** Multi-task learning (MTL) improves efficiency by sharing representations across related tasks [3, 43]. In agriculture, MTLSegFormer [13] performs multi-task semantic segmentation with attention-based feature sharing between tasks, while WeedSense [49] jointly performs semantic segmentation, height estimation, and growth stage classification. Other work explores joint classification and regression for hyperspectral images [6]. These methods demonstrate the benefits of shared feature learning for related agricultural tasks.

**Structural Reparameterization.** Structural reparameterization enables multi-branch training with single-path inference through algebraic fusion. RepVGG [10] pioneered this approach for CNNs, MobileOne [57] achieved millisecond-scale mobile latency, and FastViT [56] extended the technique to Vision Transformers achieving significant speedups over hybrid architectures. In agriculture, reparameterization has been successfully applied to CNN-based architectures for disease detection [37, 67], weed detection [14], instance segmentation [32], and multi-task detection [50]. General vision work has also explored reparameterization for multi-task learning [25]. However, existing agricultural applications primarily use CNN-based backbones such as YOLO and RepVGG variants. We extend this paradigm by introducing reparameterizable Vision Transformers for joint segmentation and gender classifica-

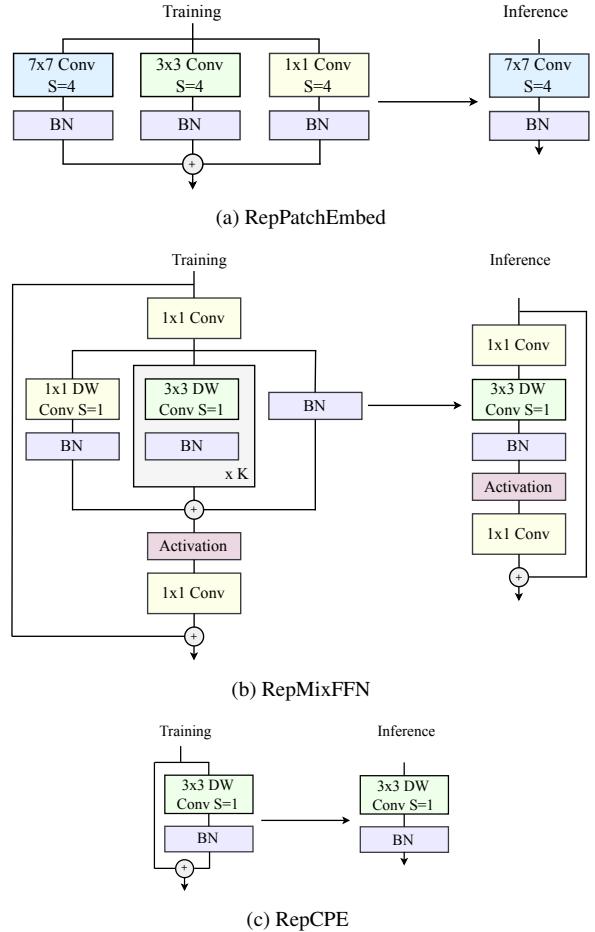


Figure 2. Reparameterizable components with train-time overparameterization. (a) RepPatchEmbed uses three parallel branches that fuse at inference. (b) RepMixFFN employs  $K$  parallel depthwise convolutions with identity connection. (c) RepCPE combines depthwise convolution with identity for positional encoding.

tion, combining the global receptive fields of transformers with the efficiency benefits of structural reparameterization.

## 3. Architecture

We propose a fully reparameterizable architecture for real-time waterhemp segmentation and gender classification as shown in Figure 1. Following structural reparameterization [10, 57], we decouple training and inference by using multi-branch overparameterization during training that fuses into efficient single-path inference with zero overhead. Our architecture comprises: (1) a Vision Transformer backbone with RepPatchEmbed, RepCPE, and RepMixFFN, (2) a reparameterizable Lite R-ASPP decoder, and (3) a reparameterizable classification head.

### 3.1. Vision Transformer Backbone

We extend the hierarchical Mix Transformer (MiT) backbone from SegFormer [62] with systematic repara-

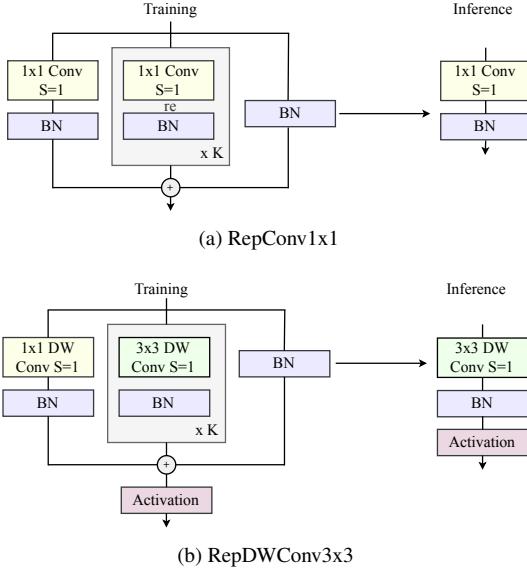


Figure 3. Reparameterizable convolution modules used in decoder and classifier. (a) RepConv1x1 with  $K$  parallel  $1 \times 1$  branches for efficient channel mixing. (b) RepDWConv3x3 with  $K$  parallel  $3 \times 3$  depthwise branches for spatial feature refinement.

terization across all components. The backbone consists of four stages with progressively downsampled feature maps at resolutions  $\{\frac{H}{4}, \frac{H}{8}, \frac{H}{16}, \frac{H}{32}\}$ , producing multi-scale features  $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4\}$  with channel dimensions  $[32, 64, 160, 256]$ . We use efficient multi-head self-attention with spatial reduction ratios  $r \in \{8, 4, 2, 1\}$  across stages to reduce complexity at higher resolutions.

**RepPatchEmbed.** Following MobileOne [57], we use train-time overparameterization for patch embedding layers as shown in Figure 2a. RepPatchEmbed employs three parallel branches: a large kernel conv ( $p_i \times p_i$  where  $p_i \in \{7, 3, 3, 3\}$  for stages 1-4), a  $3 \times 3$  conv, and a  $1 \times 1$  conv. The multi-scale branches improve low-level feature learning. At deployment, all branches fuse into a single  $p_i \times p_i$  convolution with zero computational overhead.

**RepCPE.** Unlike absolute positional encodings, conditional positional encodings (CPE) [7, 8] are dynamically generated based on local input context. We apply reparameterizable CPE [56] in stages 3-4 as shown in Figure 2c. RepCPE adds a  $3 \times 3$  depthwise convolution to the input:  $\text{RepCPE}(\mathbf{F}) = \mathbf{F} + \text{DWConv}_{3 \times 3}(\mathbf{F})$ . At inference, the identity and convolution fuse into a single operation.

**RepMixFFN.** We enhance the MixFFN with multi-branch depthwise convolutions for spatial mixing as shown in Figure 2b. RepMixFFN uses  $K$  parallel  $3 \times 3$  depthwise branches plus a  $1 \times 1$  branch and identity connection between two pointwise layers. Layer scale [29] initialized to  $10^{-5}$  stabilizes training. All branches fuse at deployment for efficient inference.

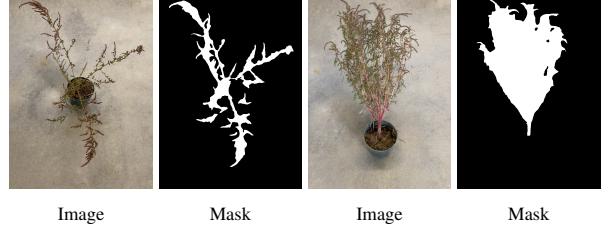


Figure 4. Example image-mask pairs from the waterhemp dataset. Left pair shows a female specimen, right pair shows a male specimen. Binary segmentation masks delineate complete plant architecture against background.

### 3.2. Segmentation Decoder and Classification Head

**RepLR-ASPP Decoder.** For efficient multi-scale feature aggregation, we adapt Lite R-ASPP [20] by replacing all convolutions with reparameterizable blocks. LR-ASPP provides a lightweight alternative to ASPP [4] by using global pooling instead of dilated convolutions. The decoder uses RepConv1x1 blocks as shown in Figure 3a, which employ  $K$  parallel  $1 \times 1$  branches with layer scale that fuse at deployment. We first apply global context attention to  $\mathbf{F}_4$  via global average pooling and sigmoid gating. Then, features are progressively upsampled and fused with lateral connections from  $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3\}$  through channel concatenation and RepConv1x1 fusion to produce the final segmentation output.

**RepClsHead.** We propose RepClsHead, a reparameterizable classification head inspired by FastViT [56]. Following FastViT’s design, RepClsHead refines features before global pooling using RepDWConv3x3 as shown in Figure 3b, which employs  $K$  parallel  $3 \times 3$  depthwise branches plus a  $1 \times 1$  depthwise branch and identity connection. After refinement, features undergo global average pooling followed by a two-layer MLP with hidden dimension 256 and dropout rate 0.5 to produce binary gender classification logits. All branches fuse into a single  $3 \times 3$  depthwise convolution at deployment.

**Multi-Task Loss.** We jointly optimize the network for segmentation and classification using a multi-task loss  $\mathcal{L} = \mathcal{L}_{\text{seg}} + \lambda \mathcal{L}_{\text{cls}}$ , where  $\lambda = 0.5$ .  $\mathcal{L}_{\text{seg}}$  denotes pixel-wise cross-entropy over background and plant classes, while  $\mathcal{L}_{\text{cls}}$  is binary cross-entropy for gender classification.

## 4. Dataset

We introduce a novel waterhemp plant dataset designed for joint semantic segmentation and gender classification tasks. The dataset comprises 23 individual mature waterhemp plants, specifically 13 females and 10 males identified with the assistance of an agricultural expert, which were collected from a field site and transplanted into pots for imaging. All plants were at the flowering stage, exhibiting fully developed morphological characteristics essential

for gender differentiation. The complete dataset statistics are presented in Table 1. Representative samples from the dataset are shown in Figure 4.

#### 4.1. Plant Collection

Waterhemp plants were collected from Field 27 at SIUC University Farms ( $37.706^{\circ}\text{N}$ ,  $89.253^{\circ}\text{W}$ ) on October 14, 2024 and transplanted into 6” diameter plastic pots for imaging. The field had been fertilized with 200 lbs per acre of potash in spring 2024. Soybeans (Pioneer P37A18E) were planted on June 13, 2024 at 180,000 seeds per acre with 15-inch row spacing. The field was tilled at planting to remove weed competition. A single early postemergence herbicide application was made on August 12, 2024, consisting of glufosinate (Liberty, 32 oz/acre), glyphosate (Roundup Powermax 3, 1 qt/acre), citric acid (2.4 oz/acre), and ammonium sulfate (2.4 lb/acre). Soybeans were harvested on October 21, 2024 with a yield of approximately 25 bu/acre.

#### 4.2. Data Collection and Preprocessing

We captured one video per plant using an iPhone 15 Pro Max, performing a complete 360-degree horizontal rotation to ensure comprehensive morphological coverage. To balance dataset size with temporal diversity, we sample every third frame, effectively reducing redundancy while preserving motion variation. Frames are resized to  $720 \times 960$  pixels, maintaining sufficient spatial resolution for fine-grained structural analysis while halving storage requirements. The final dataset comprises 10,264 frames across 23 plants.

#### 4.3. Annotation Protocol

We employ SAM2.1 Hiera Large [41] for efficient semi-automatic mask generation. For each video sequence, we manually annotate positive and negative prompt points on the initial frame using an interactive bounding box interface. SAM2.1 then propagates these sparse annotations across all subsequent frames via its video predictor, ensuring temporal consistency. The resulting binary masks delineate the complete plant structure (stems, leaves, flowers) as foreground, with all other regions classified as background. Gender labels are assigned at the plant level based on botanical verification during the flowering stage.

#### 4.4. Dataset Split Strategy

To prevent data leakage, we partition the dataset at the plant level rather than the image level. Specifically, all frames extracted from a single plant are assigned exclusively to one split—training, validation, or test. This ensures that no visual information from an individual plant appears across multiple splits, thereby providing a more realistic evaluation of model generalization to unseen specimens. Unlike image-level splitting, which can inadvertently leak visual

Gender Type	Frames	Percentage	Plants	Train (70.5%)	Val (15.5%)	Test (13.9%)
Male	4,158	40.5%	10	2,692	780	686
Female	6,106	59.5%	13	4,544	816	746
<b>Total</b>	<b>10,264</b>	<b>100.0 %</b>	<b>23</b>	<b>7,236</b>	<b>1,596</b>	<b>1,432</b>

Table 1. Waterhemp dataset statistics. Percentages show distribution within 10,264 annotated frames.

patterns across splits when consecutive frames share similar appearances, our plant-level strategy guarantees complete independence between training and evaluation data.

The plant-level split allocation is as follows: 15 plants for training (9 female, 6 male), 4 plants for validation (2 female, 2 male), and 4 plants for testing (2 female, 2 male). Table 1 summarizes the resulting frame-level distribution across splits. This yields a split ratio of approximately 70.5% for training, 15.5% for validation, and 13.9% for testing. The dataset exhibits a natural gender imbalance, with female plants contributing 59.5% of total frames. This distribution reflects inherent biological differences: female waterhemp plants typically develop denser foliage and more complex branching structures compared to their male counterparts, resulting in longer video sequences and more extracted frames per plant.

## 5. Experiments

**Implementation Details.** All models are trained for 80K iterations using AdamW optimizer with learning rate  $6 \times 10^{-5}$ , weight decay 0.01, and momentum (0.9, 0.999). The learning rate follows a polynomial decay (power 1.0) after a 1,500-iteration linear warmup with start factor  $10^{-6}$ . We use batch size 8 on an NVIDIA A100 80GB PCIe GPU at  $512 \times 512$  resolution with standard augmentations including random horizontal flip, resize, and photometric distortion. The multi-task loss uses  $\lambda = 0.5$  as described in Section 3.2. For fair comparison, all comparison methods utilize a standardized MLP classification head comprising global pooling followed by two linear layers with Xavier initialization, except Fast-SCNN which loads Cityscapes weights [39]. Our WeedRepFormer employs the proposed RepClsHead and initializes from SegFormer-B0 ImageNet weights [62], where standard components transfer directly, while reparameterizable modules inherit weights into their primary branch matching the source kernel size, with additional branches using Kaiming initialization [17]. The backbone learning rate is scaled by  $0.1 \times$  while the classification head uses  $1.0 \times$  the base rate for rapid adaptation. We report mean IoU (mIoU) and mean F-score (mFscore) for segmentation; mean accuracy (mAcc) and mean F1 (mF1) for classification; and inference FPS.

Method	Segmentation		Classification		Efficiency		
	mIoU (%)↑	mFscore (%)↑	mAcc (%)↑	mF1 (%)↑	FPS↑	GFLOPs↓	Params (M)↓
<i>Conv-based Models</i>							
BiSeNet [63]	92.27	95.87	67.25	66.87	<u>233.45</u>	14.821	13.455
BiSeNet V2 [64]	92.30	95.89	58.17	57.96	<u>159.61</u>	12.286	14.820
DDRNet [36]	92.81	96.18	45.74	44.94	151.58	4.560	5.766
Fast-SCNN [39]	88.27	93.51	59.57	59.41	225.70	<b>0.927</b>	<b>1.488</b>
ICNet [65]	91.58	95.48	56.08	56.02	134.73	15.426	47.859
MobileNetV2 [48]	91.50	95.43	70.88	69.34	179.47	39.261	9.793
MobileNetV3 [20]	92.28	95.88	68.99	67.92	121.55	8.692	3.529
MobileOne-S0 [57]	73.02	82.76	60.61	59.23	<b>296.97</b>	8.213	28.915
RepViT-M0.9 [59]	<b>94.31</b>	<b>97.01</b>	69.20	68.86	82.23	25.404	8.954
<i>Transformer-based Models</i>							
SegFormer-B0 [62]	94.10	96.90	74.72	75.34	115.27	7.885	3.782
iFormer-T [66]	94.05	96.87	<u>77.51</u>	<u>77.08</u>	99.05	24.267	6.804
FastViT-T8 [56]	93.33	96.47	<u>55.24</u>	<u>54.41</u>	189.71	23.806	7.382
<b>WeedRepFormer (Ours)</b>	92.18	95.82	<b>81.91</b>	<b>81.90</b>	108.95	<u>3.801</u>	<u>3.592</u>

Table 2. Comparison with state-of-the-art methods on waterhemp multi-task learning. All models trained with identical protocols. Best results in **bold**, second best underlined.

## 5.1. Comparison with State-of-the-Art Methods

We compare our WeedRepFormer against state-of-the-art efficient segmentation and multi-task models on the waterhemp dataset. Table 2 presents comprehensive results across segmentation, classification, and efficiency metrics.

**Classification Performance.** WeedRepFormer achieves 81.91% mAcc and 81.90% mF1, outperforming all other methods on gender classification. Compared to the second-best method iFormer-T with 77.51% mAcc, our approach improves accuracy by 4.40% and F1 score by 4.82%. Among transformer baselines, WeedRepFormer surpasses SegFormer-B0 by 7.19% mAcc and RepViT-M0.9 by 12.71% mAcc. In contrast, convolution-based models struggle with this task; for instance, BiSeNet and DDRNet achieve only 67.25% and 45.74% mAcc respectively, despite their competitive segmentation results. These findings suggest that our structural reparameterization approach effectively captures the discriminative features required for fine-grained classification.

**Segmentation Performance.** On segmentation, WeedRepFormer yields 92.18% mIoU and 95.82% mFscore. RepViT-M0.9 achieves the highest segmentation at 94.31% mIoU and 97.01% mFscore, followed by SegFormer-B0 at 94.10% mIoU and 96.90% mFscore. However, these methods require substantially higher computation while achieving weaker classification performance. Among efficient models under 5 GFLOPs, WeedRepFormer achieves competitive segmentation accuracy, comparable to DDRNet (92.81% mIoU) which drastically fails at classification (45.74% mAcc). The modest 2.13% mIoU gap compared to RepViT-M0.9 represents a favorable trade-off for superior classification and efficiency.

**Efficiency Analysis.** Table 2 highlights the efficiency advantages of WeedRepFormer. With only 3.59M param-

ters and 3.80 GFLOPs, our model achieves 108.95 FPS while maintaining high multi-task accuracy (92.18% mIoU, 81.91% mAcc). Compared to the widely used SegFormer-B0 (7.89 GFLOPs, 115.27 FPS), WeedRepFormer reduces computational cost by 51.7% and parameter count by 5.0% with negligible impact on throughput. While Fast-SCNN is more lightweight (0.93 GFLOPs), it suffers significant accuracy degradation. Conversely, higher-capacity models like iFormer-T (24.27 GFLOPs) and RepViT-M0.9 (25.40 GFLOPs) incur 6.4× and 6.7× higher computational costs, respectively. Furthermore, our structural reparameterization strategy yields a 1.54× inference speedup, collapsing from 4.63 GFLOPs (70.89 FPS) during training to 3.80 GFLOPs (108.95 FPS) at deployment. These results position WeedRepFormer as an optimal solution for resource-

Method	Segmentation				Classification			
	IoU (%)↑		F-score (%)↑		Acc (%)↑		F1 (%)↑	
	BG	Weed	BG	Weed	Male	Female	Male	Female
<i>Conv-based Models</i>								
BiSeNet	98.48	86.06	99.23	92.51	81.34	54.29	70.41	63.33
BiSeNet V2	98.51	86.10	99.25	92.53	68.08	49.06	60.93	55.00
DDRNet	98.63	86.99	99.31	93.05	60.35	32.31	51.59	38.28
Fast-SCNN	97.59	78.95	98.78	88.24	55.69	63.14	56.89	61.93
ICNet	98.34	84.82	99.16	91.79	62.10	50.54	57.53	54.52
MobileNetV2	98.34	84.66	99.16	91.69	<b>97.38</b>	46.51	76.21	62.47
MobileNetV3	98.51	86.05	99.25	92.50	91.11	48.66	73.79	62.05
MobileOne-S0	93.56	52.49	96.67	68.84	82.51	40.48	66.75	51.71
RepViT-M0.9	<b>98.93</b>	<b>89.69</b>	<b>99.46</b>	<b>94.56</b>	83.24	56.30	72.14	65.57
<i>Transformer-based Models</i>								
SegFormer-B0	98.88	89.31	99.44	94.36	90.09	60.59	77.35	71.41
iFormer-T	98.87	89.24	99.43	94.31	95.19	61.26	80.22	73.95
FastViT-T8	98.73	87.93	99.36	93.58	71.72	40.08	60.55	48.26
<b>WeedRepFormer</b>	98.45	85.91	99.22	92.42	88.05	<b>76.27</b>	82.34	<b>81.46</b>

Table 3. Class-wise performance breakdown. WeedRepFormer achieves the best female classification accuracy (76.27%) and F1-score (81.46%), with the smallest male-female accuracy gap (11.78%) compared to other methods.

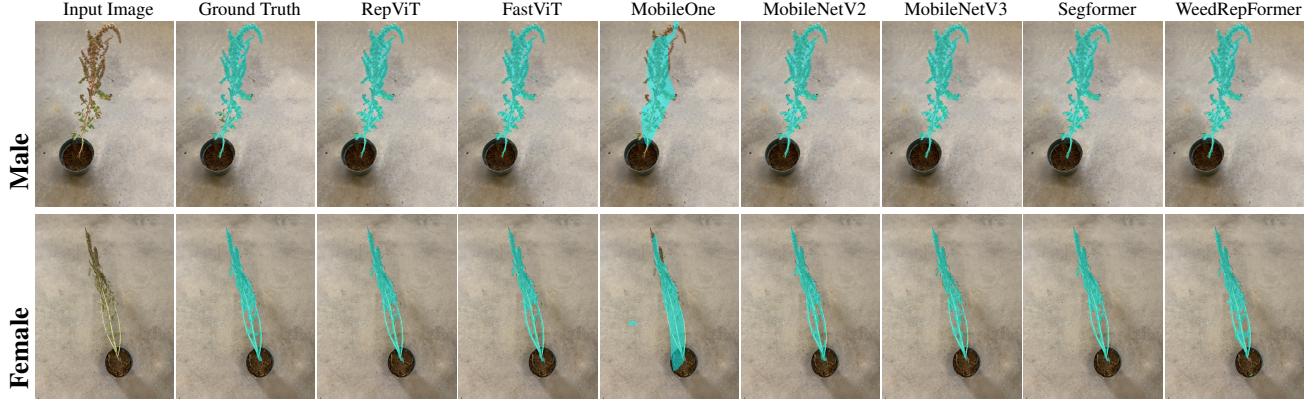


Figure 5. Qualitative comparison on male and female waterhemp. From left to right: input, ground truth, RepViT, FastViT, MobileOne, MobileNetV2, MobileNetV3, SegFormer-B0, and WeedRepFormer (ours). Best viewed on screen.

constrained agricultural robotics.

**Class-wise Performance Analysis.** Table 3 provides a detailed breakdown of per-class performance. For segmentation, all methods achieve near-saturated background IoU ( $>97\%$ ), with primary differentiation occurring on the weed class where RepViT-M0.9 leads at 89.69% IoU. For classification, most methods exhibit severe male-female accuracy disparities. MobileNetV2 achieves the highest male accuracy yet drops to only 46.51% on females, resulting in a gap of 50.87%. Similarly, iFormer-T and DDRNet show gaps of 33.93% and 28.04%, respectively. In contrast, WeedRepFormer achieves the best female accuracy of 76.27% and F1-score of 81.46% with the smallest gap of only 11.78%. This balanced performance is agriculturally significant: female waterhemp plants produce seeds and drive herbicide resistance spread, making their reliable detection essential for effective weed management.

**Qualitative Results.** Figure 5 illustrates segmentation results on representative male and female Waterhemp samples. MobileOne produces degraded segmentation with coarse boundaries and missed regions. Other methods achieve visually similar segmentation quality. WeedRepFormer produces comparable visual quality to higher-capacity models like RepViT-M0.9 and SegFormer-B0, validating that structural reparameterization enables effective multi-task learning without compromising segmentation quality.

## 5.2. Ablation Study

We conduct a systematic ablation to validate architectural decisions. All experiments utilize the MixVisionTransformer backbone and RepLR-ASPP decoder unless otherwise specified. Results are summarized in Table 4.

**Reparameterizable Backbone Components.** We first evaluate the impact of including RepMixFFN, RepPatchEmbed, and RepCPE in the backbone. As shown in Table 4, using RepCPE alone achieves the highest segmentation performance (93.74% mIoU) but yields suboptimal

classification accuracy (68.09% mAcc). While individual and pairwise combinations show varying trade-offs, combining all three components achieves the best multi-task balance, yielding 92.15% mIoU and 77.09% mAcc. This configuration significantly improves classification (+9.0% mAcc over RepCPE alone) with only a minor trade-off in segmentation, validating that multi-component reparameterization is essential for capturing both the spatial context and semantic features required for simultaneous segmentation and classification.

**Classification Head Design.** We challenge the standard MLP head design. Replacing the Standard MLP with our proposed RepClsHead yields immediate gains. Even with a single branch ( $K = 1$ ), accuracy improves by +2.1%. With  $K = 2$  branches, RepClsHead achieves 81.91% mAcc versus 77.72% for the simple baseline. This validates that the classification head benefits significantly from the increased capacity of multi-branch training, which is subsequently compressed for inference.

**Branch Count ( $K$ ) & Efficiency.** We investigate the training-time width of the reparameterizable blocks ( $K \in \{1, 2, 3, 4\}$ ). We observe a performance peak at  $K = 2$ . Increasing branches further to  $K = 3$  or  $K = 4$  leads to diminishing returns and overfitting, degrading accuracy to 75-78%. The  $K = 2$  configuration offers the optimal sweet spot: it incurs computational cost only during training, collapsing to the exact same inference cost (3.80 GFLOPs) as the  $K = 1$  baseline, effectively providing a +2.7% accuracy boost via structural reparameterization.

**Patch Embedding Factorization.** We evaluate depthwise-pointwise factorization (V3) against regular convolution (V2) within RepPatchEmbed. While V3 offers a theoretical reduction in GFLOPs ( $3.80 \rightarrow 3.50$ ), it causes a catastrophic drop in classification accuracy ( $> 20\%$  decline). This finding suggests that dense feature interaction, provided by regular convolutions, is critical for capturing the subtle visual cues required for weed gender classification.

**Conditional Positional Encoding Placement.** We analyze

Configuration	mIoU (%) $\uparrow$	mAcc (%) $\uparrow$	FPS $\uparrow$	GFLOPs $\downarrow$	Params (M) $\downarrow$
<i>Reparameterizable Backbone Components</i>					
RepMixFFN	92.76	62.85	112.02	3.797	3.582
RepPatchEmbed	91.93	61.24	114.58	3.797	3.582
RepCPE	93.74	68.09	105.02	3.801	3.590
RepCPE + RepMixFFN	92.72	52.72	111.60	3.801	3.590
RepCPE + RepPatchEmbed	91.66	66.27	109.67	3.801	3.590
RepMixFFN + RepPatchEmbed	91.89	73.95	117.35	3.797	3.582
All Components	92.15	77.09	109.79	3.801	3.590
<i>Classification Head Design</i>					
$K=1$ , Standard MLP	92.15	77.09	109.79	3.801	3.590
$K=1$ , RepClsHead	92.32	79.19	109.99	3.801	3.592
$K=2$ , Standard MLP	92.25	77.72	109.26	3.801	3.590
$K=2$ , RepClsHead	92.18	<b>81.91</b>	<b>108.95</b>	3.801	3.592
<i>Parallel Branch Count (<math>K</math>)</i>					
$K=1$ , RepClsHead	92.32	79.19	109.99	3.801	3.592
$K=2$ , RepClsHead	92.18	81.91	108.95	3.801	3.592
$K=3$ , RepClsHead	92.25	75.56	108.31	3.801	3.592
$K=4$ , RepClsHead	92.17	78.84	108.26	3.801	3.592
<i>RepPatchEmbed Factorization (V2: Regular Conv vs V3: Depthwise-Pointwise)</i>					
V2, $K=1$ , Standard MLP	92.15	77.09	109.79	3.801	3.590
V2, $K=1$ , RepClsHead	92.32	79.19	109.99	3.801	3.592
V2, $K=2$ , RepClsHead	92.18	81.91	108.95	3.801	3.592
V3, $K=1$ , Standard MLP	90.14	59.01	110.15	<b>3.495</b>	<b>3.162</b>
V3, $K=1$ , RepClsHead	90.30	59.01	109.45	3.496	3.165
V3, $K=2$ , RepClsHead	88.58	56.22	108.42	3.496	3.165
<i>CPE Placement Across Stages (Stage1, Stage2, Stage3, Stage4)</i>					
(T, T, T, T)	91.98	69.90	106.27	3.816	3.594
(T, T, F, F)	92.03	65.78	110.08	3.811	3.586
(F, F, T, T)	92.18	81.91	108.95	3.801	3.592
(T, F, T, F)	92.06	68.09	105.58	3.810	3.588
(F, T, F, T)	92.35	73.46	109.88	3.803	3.591
<i>RepPatchEmbed Kernel Size Configuration</i>					
(7, 3, 3, 3)	92.18	81.91	108.95	3.801	3.592
(7, 7, 7, 7)	92.08	81.08	109.01	4.976	5.722
<i>Squeeze-and-Excitation in Classification Head</i>					
Without SE	92.18	81.91	108.95	3.801	3.592
With SE	92.12	76.19	109.03	3.801	3.601

Table 4. Comprehensive ablation study of architectural components. Gray rows indicate selected configurations. Results demonstrate systematic optimization from individual reparameterizable components through branch count selection, factorization strategy, CPE placement, patch size configuration, and attention mechanisms.

the optimal insertion points for Conditional Positional Encodings (CPE) across the four backbone stages. We find that applying RepCPE in early stages is detrimental to classification performance. The best balance is achieved by restricting RepCPE to the deeper layers (Stages 3 & 4). This aligns with the intuition that early stages focus on local, translation-invariant texture features, whereas deeper stages capture high-level semantic information where conditional spatial context is most beneficial.

**RepPatchEmbed Kernel Size Configuration.** We investigate the impact of kernel size scaling within the embedding modules. We compare a uniform large-kernel strategy [7, 7, 7, 7] against a progressive reduction strategy [7, 3, 3, 3]. The uniform configuration incurs a significant pa-

rameter penalty (5.72M) without improving performance. In contrast, the progressive strategy achieves higher accuracy with only 3.59M parameters. This confirms that while a large initial receptive field is crucial for early tokenization, subsequent stages benefit from efficient, smaller kernels to refine features without unnecessary computational overhead.

**Squeeze-and-Excitation Attention.** Finally, we investigated the integration of Squeeze-and-Excitation (SE) attention [22] within the RepClsHead to potentially enhance feature selection. Contrary to expectations, adding SE resulted in a 5.72% degradation in classification accuracy. Given this significant performance penalty, we exclude attention mechanisms from the final architecture.

## 6. Conclusion

This paper introduces WeedRepFormer, which systematically integrates structural reparameterization throughout a Vision Transformer architecture for simultaneous waterhemp segmentation and gender classification. Unlike prior agricultural work that primarily uses CNN-based reparameterization, we apply it to hierarchical Vision Transformers across backbone, decoder, and task-specific heads for joint dense prediction and gender classification. WeedRepFormer outperforms all compared methods in classification accuracy while maintaining competitive segmentation with significantly reduced computational cost. Notably, our model achieves the best female classification performance with the smallest male-female accuracy gap, addressing a critical performance disparity. From an agricultural perspective, reliable female plant detection enables targeted weed management since female waterhemp are the primary seed producers responsible for population spread and herbicide resistance.

## References

- [1] Muhammad Hamza Asad and Abdul Bais. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Information Processing in Agriculture*, 7(4):535–545, 2020. 3
- [2] Michael S Bell, Aaron G Hager, and Patrick J Tranel. Multiple resistance to herbicides from four site-of-action groups in waterhemp (amaranthus tuberculatus). *Weed Science*, 61(3):460–468, 2013. 1
- [3] Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997. 3
- [4] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. In *arXiv preprint arXiv:1706.05587*, 2017. 4
- [5] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In

- Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 2
- [6] Koushikey Chhapariya, Alexandre Benoit, Krishna Mohan Buddhiraju, and Anil Kumar. A multitask deep learning model for classification and regression of hyperspectral images: Application to a large scale dataset. *IEEE Transactions on Geoscience and Remote Sensing*, 2025. 3
- [7] Xiangxiang Chu, Zhi Tian, Yuqing Wang, Bo Zhang, Haibing Ren, Xiaolin Wei, Huaxia Xia, and Chunhua Shen. Twins: Revisiting the design of spatial attention in vision transformers. *Advances in neural information processing systems*, 34:9355–9366, 2021. 4
- [8] Xiangxiang Chu, Zhi Tian, Bo Zhang, Xinlong Wang, and Chunhua Shen. Conditional positional encodings for vision transformers. *arXiv preprint arXiv:2102.10882*, 2021. 4
- [9] Michael Crawshaw. Multi-task learning with deep neural networks: A survey. *arXiv preprint arXiv:2009.09796*, 2020. 2
- [10] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13733–13742, 2021. 2, 3
- [11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021. 3
- [12] Karla L Gage, Ronald F Krausz, and S Alan Walters. Emerging challenges for weed management in herbicide-resistant crops. *Agriculture*, 9(8):180, 2019. 1
- [13] Diogo Nunes Goncalves, Jose Marcato Junior, Pedro Zamboni, Hemerson Pistori, Jonathan Li, Keiller Nogueira, and Wesley Nunes Goncalves. Mtlsegformer: Multi-task learning with transformers for semantic segmentation in precision agriculture. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6290–6298, 2023. 3
- [14] Ao Guo, Zhenhong Jia, Baoquan Ge, Wei Chen, Sensen Song, Congbing He, Gang Zhou, Jiajia Wang, and Xiaoyi Lv. Rlcfe-net: A reparameterization large convolutional kernel feature extraction network for weed detection in multiple scenarios. *Expert Systems with Applications*, 274:126941, 2025. 2, 3
- [15] Aaron G Hager, Loyd M Wax, Edward W Stoller, and Germán A Bollero. Common waterhemp (*amaranthus rudis*) interference in soybean. *Weed science*, 50(5):607–610, 2002. 1
- [16] Robert G Hartzler, Bruce A Battles, and Dawn Nordby. Effect of common waterhemp (*amaranthus rudis*) emergence date on growth and fecundity in soybean. *Weed Science*, 52(2):242–245, 2004. 1
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 5
- [18] Ian Heap. The international herbicide-resistant weed database. Online, 2025. Accessed: Sept. 30, 2025. [Online]. Available: <http://www.weedscience.org>. 1
- [19] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 2
- [20] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019. 4, 6
- [21] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 2
- [22] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018. 8
- [23] Taminul Islam, Toqi Tahmid Sarker, Khaled R Ahmed, Cristiana Bernardi Rankrake, and Karla Gage. Weedswin hierarchical vision transformer with sam-2 for multi-stage weed detection and classification. *Scientific Reports*, 15(1):23274, 2025. 1
- [24] Andreas Kamilaris and Francesc X Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90, 2018. 1
- [25] Menelaos Kanakis, David Bruggemann, Suman Saha, Stamatios Georgoulis, Anton Obukhov, and Luc Van Gool. Reparameterizing convolutions for incremental multi-task learning without task interference. In *European conference on computer vision*, pages 689–707. Springer, 2020. 3
- [26] Xiaotong Kong, Aimin Li, Teng Liu, Kang Han, Xiaojun Jin, Xin Chen, and Jialin Yu. Lightweight cabbage segmentation network and improved weed detection method. *Computers and Electronics in Agriculture*, 226:109403, 2024. 3
- [27] Jianyang Liu, Adam S Davis, and Patrick J Tranell. Pollen biology and dispersal dynamics in waterhemp (*amaranthus tuberculatus*). *Weed Science*, 60(3):416–422, 2012. 1
- [28] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pages 10012–10022, 2021. 3
- [29] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *CVPR*, pages 11976–11986, 2022. 4
- [30] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015. 3
- [31] Philipp Lottes, Jens Behley, Andres Milioto, and Cyrill Stachniss. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robotics and Automation Letters*, 3(4):2870–2877, 2018. 3
- [32] Rongxiang Luo, Rongrui Zhao, and Bangjin Yi. Enhanced yolo11n-seg with attention mechanism and geometric metric optimization for instance segmentation of ripe blueber-

- ries in complex greenhouse environments. *Agriculture*, 15(15):1697, 2025. 3
- [33] Rong Ma, Joshua J Skelton, and Dean E Riechers. Measuring rates of herbicide metabolism in dicot weeds with an excised leaf assay. *Journal of Visualized Experiments: Jove*, (103):53236, 2015. 1
- [34] Andres Milioto, Philipp Lottes, and Cyrill Stachniss. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 2229–2235. IEEE, 2018. 2
- [35] Jacob S Montgomery, Ahmed Sadeque, Darci A Giacomini, Patrick J Brown, and Patrick J Tranel. Sex-specific markers for waterhemp (*amaranthus tuberculatus*) and palmer amaranth (*amaranthus palmeri*). *Weed Science*, 67(4):412–418, 2019. 1
- [36] Huihui Pan, Yuanduo Hong, Weichao Sun, and Yisong Jia. Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes. *IEEE Transactions on Intelligent Transportation Systems*, 24(3):3448–3460, 2022. 6
- [37] Guoquan Pei, Xueying Qian, Bing Zhou, Zigao Liu, and Wendou Wu. Research on agricultural disease recognition methods based on very large kernel convolutional network-replknet. *Scientific Reports*, 15(1):16843, 2025. 2, 3
- [38] Maurizio Pintus, Felice Colucci, and Fabio Maggio. Emerging developments in real-time edge aiot for agricultural image classification. *IoT*, 6(1):13, 2025. 2
- [39] Rudra PK Poudel, Stephan Liwicki, and Roberto Cipolla. Fast-scnn: Fast semantic segmentation network. *arXiv preprint arXiv:1902.04502*, 2019. 5, 6
- [40] Cristiana Bernardi Rankrape, E Lago, Eric J Miller, IS Werle, Patrick J Tranel, and Karla L Gage. Evaluating glufosinate resistance in a waterhemp (*amaranthus tuberculatus*) population from southern illinois. In *North Central Weed Science Society*, Kansas City, MO, 2024. 1
- [41] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädele, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. 5
- [42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2, 3
- [43] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017. 2, 3
- [44] Mafia M Rumpa, Ronald F Krausz, David J Gibson, and Karla L Gage. Effect of ppo-inhibiting herbicides on the growth and sex ratio of a dioecious weed species *amaranthus palmeri* (palmer amaranth). *Agronomy*, 9(6):275, 2019. 1
- [45] Mafia M Rumpa, Sirwan Babaei, Ronald F Krausz, David J Gibson, Eric J Miller, and Karla L Gage. Does exposure to ppo-inhibiting herbicides alter the male-to-female sex ratio of palmer amaranth? *Weed Technology*, 39:e58, 2025. 1
- [46] Reiofeli A Salas, Nilda R Burgos, Patrick J Tranel, Shilpa Singh, Les Glasgow, Robert C Scott, and Robert L Nichols. Resistance to ppo-inhibiting herbicide in palmer amaranth from arkansas. *Pest management science*, 72(5):864–869, 2016. 1
- [47] Muhammad Hammad Saleem, Johan Potgieter, and Khalid Mahmood Arif. Plant disease detection and classification by deep learning. *Plants*, 8(11):468, 2019. 1
- [48] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. 6
- [49] Toqi Tahmid Sarker, Khaled R Ahmed, Taminul Islam, Cristiana Bernardi Rankrape, and Karla Gage. Weedsense: Multi-task learning for weed segmentation, height estimation, and growth stage classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 7180–7190, 2025. 1, 3
- [50] Xiaojun Shen, Chaofan Shao, Danyi Cheng, Lili Yao, and Cheng Zhou. Yolov5-pos: research on cabbage pose prediction method based on multi-task perception technology. *Frontiers in Plant Science*, 15:1455687, 2024. 2, 3
- [51] Sovi Guillaume Sodjinou, Vahid Mohammadi, Amadou Tidjani Sanda Mahama, and Pierre Gouton. A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images. *information processing in agriculture*, 9(3):355–364, 2022. 1
- [52] Lawrence E Steckel. The dioecious *amaranthus* spp.: here to stay. *Weed Technology*, 21(2):567–570, 2007. 1
- [53] Lawrence E Steckel and Christy L Sprague. Common waterhemp (*amaranthus rudis*) interference in corn. *Weed Science*, 52(3):359–364, 2004. 1
- [54] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Int. Conf. Machine Learning*, pages 6105–6114, 2019. 2
- [55] Federico Trucco and Patrick J Tranel. Amaranthus. In *Wild Crop Relatives: Genomic and Breeding Resources: Vegetables*, pages 11–21. Springer, 2011. 1
- [56] Pavan Kumar Anasosalu Vasu, James Gabriel, Jeff Zhu, Oncel Tuzel, and Anurag Ranjan. Fastvit: A fast hybrid vision transformer using structural reparameterization. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5785–5795, 2023. 2, 3, 4, 6
- [57] Pavan Kumar Anasosalu Vasu, James Gabriel, Jeff Zhu, Oncel Tuzel, and Anurag Ranjan. Mobileone: An improved one millisecond mobile backbone. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7907–7917, 2023. 2, 3, 4, 6
- [58] JD Vyn, CJ Swanton, SE Weaver, and PH Sikkema. Control of herbicide-resistant common waterhemp (*amaranthus tuberculatus* var. *rudis*) with pre-and post-emergence herbicides in soybean. *Canadian journal of plant science*, 87(1):175–182, 2007. 1
- [59] Ao Wang, Hui Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Repvit: Revisiting mobile cnn from vit perspective.

- In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15909–15920, 2024. 6
- [60] Jun Wang, Zhengyuan Qi, Yanlong Wang, and Yanyang Liu. A lightweight weed detection model for cotton fields based on an improved yolov8n. *Scientific Reports*, 15(1):457, 2025. 3
- [61] Katherine E Waselkov, Nathaniel D Regenold, Romy C Lum, and Kenneth M Olsen. Agricultural adaptation in the native north american weed waterhemp, amaranthus tuberculatus (amaranthaceae). *PLoS one*, 15(9):e0238861, 2020. 1
- [62] Enze Xie, Wenhui Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34: 12077–12090, 2021. 2, 3, 5, 6
- [63] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 325–341, 2018. 6
- [64] Changqian Yu, Changxin Gao, Jingbo Wang, Gang Yu, Chunhua Shen, and Nong Sang. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *International journal of computer vision*, 129(11):3051–3068, 2021. 6
- [65] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia. Icnet for real-time semantic segmentation on high-resolution images. In *Proceedings of the European conference on computer vision (ECCV)*, pages 405–420, 2018. 6
- [66] Chuanyang Zheng. iformer: Integrating convnet and transformer for mobile application. *arXiv preprint arXiv:2501.15369*, 2025. 6
- [67] Jiye Zheng, Kaiyu Li, Wenbin Wu, and Huaijun Ruan. Repdi: A light-weight cpu network for apple leaf disease identification. *Computers and Electronics in Agriculture*, 212:108122, 2023. 2, 3