Smart Agricultural Technology 13 (2026) 101875

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Unveiling the hidden: Early detection of invasive vegetation in crops with UAV multispectral imaging

Henry O. Velesaca [a,b,*], Andrea Mero [a], Hector Villegas [a], Angel D. Sappa [a,c,*]

[a] Center for Research, Development and Innovation of Computer Systems (CIDIS), ESPOL Polytechnic University, Campus Gustavo Galindo, Km. 30.5 Via Perimetral, 090902, Guayaquil, Ecuador
[b] Software Engineering Department, Research Center for Information and Communication Technologies (CITIC-UGR), University of Granada, 18071, Granada, Spain
[c] Computer Vision Center, Universitat Autònoma de Barcelona, 08193-Bellaterra, Barcelona, Spain

ARTICLE INFO

Keywords:
Camouflaged object detection
Unmanned aerial vehicle
Weed detection
Pest detection
Precision agriculture.

ABSTRACT

This work presents a practical framework for early detection of camouflaged weeds in agricultural production environments using UAV-based near-infrared (NIR) imagery. The proposed approach integrates RGB-NIR image fusion with advanced camouflaged object detection (COD) models to support precision agriculture applications at farm scale. Fourteen state-of-the-art (SOTA) fusion techniques are evaluated, and the three best-performing methods are selected for multispectral integration, while nine SOTA COD models are trained and tested on RGB and RGB-NIR fused data. Experimental validation on two UAV-based case studies–a commercial banana plantation (Weeds-Banana) and a maize field (WeedsGalore)–demonstrates that RGB-NIR fusion consistently improves weed segmentation accuracy compared with RGB-only inputs. Overall, the proposed framework provides a transferable and application-oriented solution for weed detection, enabling targeted interventions and contributing to reduced herbicide use and more sustainable crop management. The dataset is publicly available at GitHub: https://cod-espol.github.io/COD-Weeds/ .

1. Introduction

The increasing global demand for agricultural production, driven by population growth, presents significant challenges for the implementation of intelligent systems in agriculture (e.g., [1,2]). Moreover, the adoption of automated inspections in agriculture and the food industry has become an increasingly attractive solution for the final assessment of product quality [3]. Modern precision agriculture, without a doubt, relies on computer vision techniques to address critical tasks such as detecting plant diseases, predicting crop yields, recognizing species, identifying pests—-including weeds—-and managing irrigation [4].

One of the key tasks in computer vision is object detection, which is categorized into subfields: Generic Object Detection, Salient Object Detection, and Camouflaged Object Detection (COD) [5]. COD focuses on segmenting concealed objects, typically framed as a binary segmentation task [6]. In agriculture, pests often rely on camouflage as a survival mechanism [7], posing a real-world challenge due to variations in lighting, viewpoint, and environmental complexity. In general, pests blend into their environment, making them difficult to identify [8]. Similarly, weeds, which significantly hinder crop productivity, are often targeted by the widespread and unregulated application of chemical herbicides, leading to environmental degradation and contamination of agricultural land [4]. In such cases, the need for real-time COD becomes highly beneficial.

In recent years, several studies have explored the spectral signatures of camouflaged objects in greater depth by expanding spectral detection ranges and improving spectral resolution capabilities. According to [9], conventional approaches often rely on manually designed low-level features to identify essential image characteristics, while deep learning techniques autonomously learn hierarchical feature representations directly from data, consistently outperforming traditional methods across various computer vision applications. Furthermore, [10] classifies COD techniques into several categories, including Human Vision, Infrared, Multispectral, Hyperspectral, and LiDAR. It is also important to highlight that exploring the other spectral bands, such as the thermal or near-infrared (NIR) ranges, offers significant advantages in low-visibility conditions. NIR imaging, in particular, can provide valuable insights into the material composition of objects [11]. Also, camouflage detection systems have become increasingly compatible with multispectral technology, intelligent adaptation, and integrated design. However, despite advances in multispectral composite detection technology, single-band camouflaged object detection remains inadequate to meet complex

* Corresponding authors.
E-mail addresses: hvelesac@espol.edu.ec (H.O. Velesaca), anmero@espol.edu.ec (A. Mero), hvillega@espol.edu.ec (H. Villegas), asappa@espol.edu.ec (A.D. Sappa).

https://doi.org/10.1016/j.atech.2026.101875
Received 19 September 2025; Received in revised form 10 February 2026; Accepted 12 February 2026
Available online 24 February 2026
2772-3755/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

requirements [12]. Spectral imaging creates a multidimensional data cube that includes both spatial and spectral data, with each pixel represented as a spectral vector. This spectral vector, which contains radiance or reflectance values across multiple spectral bands, enhances the representation of targets compared to traditional RGB images, which only capture two-dimensional spatial information [13].

The main motivation of this study is to explore COD methodologies within the visible and NIR spectrum, addressing a gap in the current computer vision literature, as this area remains underexplored. The paper explores the fusion of RGB-NIR images for weed detection using camouflaged object detection techniques for precise segmentation.

The key contributions of this work include:

- We introduce and release Weeds-Banana, a UAV-based RGB–NIR benchmark dataset for early camouflaged weed segmentation in a commercial banana plantation.
- We propose a practical multispectral pipeline that combines RGB–NIR fusion with COD models to improve weed segmentation under camouflage conditions.
- We conduct a large-scale evaluation of 14 state-of-the-art RGB–NIR fusion methods and identify the top-performing fusion strategies per dataset.
- We benchmark 9 state-of-the-art COD architectures on RGB and fused RGB–NIR inputs across two UAV case studies (Weeds-Banana and WeedsGalore), showing consistent gains from incorporating NIR information.

The manuscript is organized as follows. Section 2 provides a review of related work on RGB-NIR fusion methods and COD techniques. Section 3 details the proposed methodology for enhancing weed segmentation in crops through the integration of RGB-NIR fusion and COD approaches. Section 4 describes two case studies where the methodology is applied to UAV imagery from banana and maize plantations, along with the presentation of experimental results, comparative analyses, and discussion of various COD methods. Finally, Section 5 discusses the results and Section 6 summarizes the main conclusions of the study.

2. Background

This section reviews the most relevant existing techniques for detecting camouflaged agricultural pests, with a particular focus on fusion detection methods and camouflaged object detection (COD) strategies. These approaches form the foundation for the proposed methodology and support its development.

2.1. Fusion approaches

Fusion detection methods combine information from multiple image modalities to generate a single, more informative representation, thereby enhancing application performance [14]. As described in [15], early fusion methods include the Cross-Bilateral Filter (CBF), which merges images by leveraging both intensity and geometric similarities. By using one image to guide the filtering of another, CBF preserves edges and reduces noise, making it especially effective for multi-focus and multisensor image fusion tasks. Building on the concept of edge-preserving fusion, [16] introduced the hybrid multiscale decomposition with a guided filter (HMSD-GF), specifically designed for fusing infrared and visible images to improve night vision. By integrating guided filtering within a multiscale decomposition framework, HMSD-GF effectively extracts complementary information from both spectral domains, excelling at enhancing visibility in low-light conditions while maintaining the structural integrity of features from the visible spectrum.

In addition to these methods, Latent Low-Rank Representation (LatLRR) [17] provides a decomposition-based approach for IR-visible fusion. This technique separates each image into a low-rank component (capturing global structures) and a salient component (preserving local details). The low-rank layers are fused using weighted averaging to minimize redundancy, while salient layers are merged by summation to retain fine-grained features. This dual-branch fusion architecture achieves a balance between structural consistency and detail enhancement.

Expanding on these paradigms, the Guided Filter-based Context enhancement and fusion model (GFCB) [16] advances night-vision image fusion by employing the guided filter for both adaptive visibility enhancement and multi-scale image decomposition. Initially, the GFCB model enhances the visible image through high dynamic range-inspired compression and contrast restoration, improving details in poorly illuminated regions while preserving edges. Subsequently, a hybrid multi-scale decomposition based solely on the guided filter extracts texture and edge features from both IR and visible images. These components are then fused using a perceptual saliency-driven weighting mechanism that dynamically adjusts the contribution of IR information, ensuring that important thermal features enhance the visible scene without compromising background details. This approach effectively balances noise suppression and detail preservation, thanks to the edge-aware smoothing properties of the guided filter, resulting in fused images with improved contextual clarity and suitability for human perception. Additionally, a perceptual-based parameter selection strategy, inspired by human visual system models, automatically adjusts fusion parameters, increasing robustness and reducing the need for manual tuning. Experimental results show that GFCB outperforms both traditional and other state-of-the-art methods in terms of visual quality and computational efficiency, making it a powerful tool for IR-visible image fusion in challenging low-light scenarios [16].

All these fusion schemes highlight the transformative potential of integrating multimodal data, particularly in IR-visible fusion, to address challenges such as low-light conditions. Each spectral band contributes unique and complementary information essential for effective problem-solving in real-world applications. According to the literature, each reviewed method offers distinct advantages. Therefore, in this work, several state-of-the-art fusion techniques are evaluated to identify the most effective ones.

2.2. Camouflaged object detection approaches

While fusion techniques leverage complementary information from different modalities to improve detection, COD presents unique challenges that require specialized solutions. Unlike conventional object detection, COD must address objects that intentionally blend into their surroundings through similarities in color, texture, and shape. This has led to the development of tailored methods that focus on detecting subtle visual cues and fine boundaries. Recent surveys, such as the comprehensive review by Xiao et al. [18], classify COD methodologies into traditional and deep learning-based approaches, illustrating the field's evolution. Yang et al. [19] introduced PlantCamo, the first dataset dedicated to camouflaged plant detection, containing 1250 images across 58 plant categories and four camouflage types: background matching, disruptive coloration, masquerade, and decoration.

To address challenges such as irregular shapes and complex boundaries, PCNet introduces a Multi-scale Global Feature Enhancement module for bottom-up feature fusion and a Multi-scale Feature Refinement module with top-down iterative feedback for precise boundary delineation [19]. This Transformer-based architecture effectively captures both global context and fine details. Notably, PCNet accurately detects lithops–succulent plants that mimic stones–supporting biodiversity conservation and targeted weed control [19].

Another notable transformer-based approach is the High Resolution Iterative Feedback Network (HitNet) [20], which addresses COD by preserving and refining high-resolution details often lost during down-sampling. HitNet's architecture includes three main modules: a Transformer-based Feature Extraction module utilizing a Pyramid Vision Transformer backbone [21] for efficient multi-scale feature extraction; a Multi-Resolution Iterative Refinement module that recursively

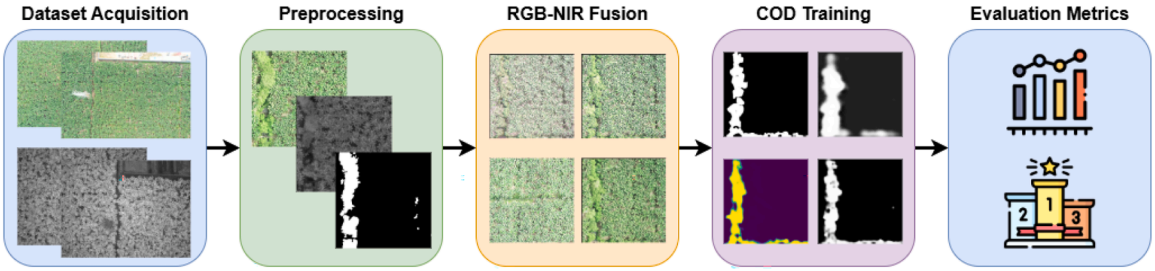


Fig. 1. Overall pipeline of the proposed methodology.

enhances low-resolution features using high-resolution information via a global feedback mechanism; and an Iteration Feature Feedback (IFF) module that ensures progressive improvements in segmentation accuracy across iterations. Additionally, HitNet [20] incorporates a graph fusion module to integrate multi-scale features for precise boundary localization. An innovative application extends HitNet's capabilities by converting salient objects into camouflaged objects through cross-domain learning, thereby enriching the diversity of training data.

In contrast to transformer-based models like PCNet and HitNet, CNN-based methods such as Edge-Aware Mirror Network (EAMNet) and Boundary-Guided Network (BGNet) focus on explicit edge and boundary refinement for COD. EAMNet employs a dual-branch architecture featuring (1) a segmentation-induced edge aggregation module and (2) an edge-induced integrity aggregation module, which mutually refine each other through cross-branch guidance. The framework also includes a guided residual channel attention module that combines residual connections with gated convolution to enhance the extraction of structural details from low-level features. Similarly, BGNet [22] uses a Res2Net-50 backbone [23] to extract multi-level features from input images. It introduces three key modules: the Edge-Aware Module (EAM), Edge-Guidance Feature Module (EFM), and Context Aggregation Module (CAM). The EAM integrates low-level local edge details with high-level global location information to extract object-related edge semantics under explicit boundary supervision. EFM injects these edge cues into multi-level backbone features using local channel attention, enhancing feature representation with object structure information. CAM progressively aggregates multi-level fused features through cross-scale interactions using expanded convolutions to strengthen contextual semantics for accurate camouflaged object detection. The model employs multi-task supervision, combining mask and edge losses to improve boundary localization and object structure preservation. BGNet achieves strong performance on multiple benchmarks, delivering precise detection with fine and complete object boundaries, while maintaining balanced model complexity and inference speed. Its design mitigates the effects of edge disruption and background ambiguity in COD.

Transformer-based and CNN-based models thus offer complementary strategies for addressing COD through advanced feature fusion and boundary refinement. These methods improve detection accuracy by preserving high-resolution details and refining object boundaries in complex environments, making them especially valuable for agricultural applications. Most of the approaches reviewed in this section are evaluated in the present work to identify those best suited for the specific challenge of camouflaged weed detection.

3. Proposed methodology

The proposed methodology introduces a comprehensive and systematic framework for the early detection of camouflaged weeds in agricultural environments, leveraging both visible (RGB) and near-infrared (NIR) imaging. The framework is composed of five main stages: (1) dataset acquisition, (2) preprocessing, (3) RGB-NIR image fusion, (4) training COD techniques, and (5) evaluation of multiple metrics. Fig. 1 shows the overall pipeline of the proposed methodology. Each stage is described in detail below.

Table 1
Comparison of available real-world agricultural datasets for weed segmentation.

Dataset	Year	Crop	Modality	# images	Resolution
Carrot-Weed [24]	2015	Carrot	RGB	60	1296 × 966
CoFly-WeedDB [25]	2022	Cotton	RGB	201	1280 × 720
WeedsGalore [26]	2025	Maize	MSI	156	600 × 600
Potato-Weed [27]	2025	Potato	RGB	150	640 × 640
Weeds-Banana (Ours)	2025	Banana	MSI	272	1024 × 1024

3.1. Dataset

A high-quality and representative dataset is a key component for developing and validating automated weed detection and segmentation methods. To support this objective, data are collected in real agricultural field conditions to capture the variability typically encountered in operational scenarios, including changes in plant growth stages, weed density, soil appearance, and illumination.

Aerial surveys are conducted using an unmanned aerial vehicle (UAV) equipped with a multispectral sensor capable of acquiring both visible (RGB) and near-infrared (NIR) imagery. Flight plans are designed to ensure consistent ground sampling and sufficient image overlap for reliable mosaicking and georeferenced orthomosaic generation. Data acquisition is scheduled to promote stable lighting conditions and reduce shadow artifacts, improving the quality and consistency of the captured imagery across the surveyed area.

Table 1 summarizes the main real-world agricultural datasets with weed-segmentation annotations and positions our Weeds-Banana dataset within the current landscape. Carrot-Weed [24], one of the earliest works in this area (2015), focuses on carrot crops with 60 RGB images at a resolution of 1296×966. Later, CoFly-WeedDB [25] (2022) extends the domain to cotton crops, also in RGB modality, with 201 images of 1280×720. More recently, WeedsGalore [26] (2024) introduces multispectral (MSI) imagery for maize, comprising 156 samples at 600×600, while Potato-Weed [27] (2025) covers potato crops with 150 RGB images at 640×640. In this context, our Weeds-Banana dataset (2025) stands out by targeting banana crops–previously unexplored in this line of work–using multispectral modality and providing the largest number of images (272) with higher resolution (1024×1024), thus offering a richer and more detailed basis for developing and evaluating weed segmentation methods under real field conditions.

3.2. Preprocessing

The raw aerial imagery is first processed using standard photogrammetric and remote-sensing workflows to obtain analysis-ready products. This stage typically includes radiometric normalization, geometric correction and alignment, and the generation of orthomosaics for each spectral band. The resulting orthomosaics are co-registered to

Table 2
Comparison between different characteristics of the camouflage techniques used.

Technique	Source	Source Type	Year	Image Size (px)	Backbone	#Par. (M)
BASNet [44]	CVPR	Conference	2019	256×256	ResNet-34 [50]	87.06
SINet-v2 [45]	TPAMI	Journal	2021	352×352	Res2Net-50 [23]	24.93
BGNet [22]	IJCAI	Conference	2022	416×416	Res2Net-50 [23]	77.80
C^2F-Net [46]	TCSVT	Conference	2022	352×352	Res2Net-50 [23]	26.36
OCENet [47]	WACV	Conference	2022	352×352	ResNet-50 [50]	58.17
EAMNet [48]	ICME	Conference	2023	384×384	Res2Net-50 [23]	30.51
DGNet [49]	MIR	Journal	2023	352×352	EfficientNet [51]	8.30
HitNet [20]	AAAI	Conference	2023	352×352	PVTv2 [52]	25.73
PCNet [19]	arXiv	-	2024	352×352	PVTv2 [52]	27.66

Table 3
Evaluation metrics used with SOTA fusion techniques. '+' means that a high value is a good performance, while '-' means that a small value denotes a good performance.

Category	Metric	Description	+/-
Information theory-based	CE	Cross entropy [53]	-
	EN	Entropy [54]	+
	MI	Mutual information [55]	+
	$PSNR$	Peak signal-to-noise ration [56]	+
Structural similarity-based	$SSIM$	Structural similarity index measure [57]	+
	$RMSE$	Root mean squared error [56]	-
Image feature-based	AG	Average gradient [58]	+
	EI	Edge intensity [59]	+
	SD	Standard deviation [60]	+
	SF	Spatial frequency [61]	+
	$Q^{AB/F}$	Gradient-based fusion performance [61]	+
Human perception inspired	Q_{CB}	Chen-Blum metric [62]	+
	Q_{CV}	Chen-Varshney metric [63]	-

ensure pixel-level correspondence between modalities, enabling consistent multi-band analysis. Weed presence is then delineated on the orthomosaic imagery through manual annotation using a dedicated labeling tool, producing segmentation masks aligned with the imagery.

To make the data suitable for deep learning training and evaluation, the large orthomosaics and their corresponding masks are converted into fixed-size samples. An automated preprocessing routine extracts image patches using a sliding-window strategy with configurable patch size and stride, producing a set of (multi-modal image, mask) subimages with optional overlap. Patches are stored in an organized structure by modality, yielding a standardized and manageable dataset format. This patch-based preprocessing facilitates efficient batch loading during training, improves coverage of spatial variability, and increases the number of usable training samples without altering the underlying field content.

3.3. RGB-NIR fusion techniques

In agricultural scenes, RGB and NIR capture different physical drivers of vegetation appearance: visible bands are largely governed by pigment absorption, whereas NIR is dominated by scattering processes and canopy structure effects. This complementarity motivates multispectral fusion as a preprocessing step to form a more informative representation before training segmentation/detection models [28–30].

To exploit the complementary information provided by RGB and NIR bands, a comprehensive evaluation of RGB-NIR image fusion techniques is performed. The fusion process is implemented using the VIFB framework proposed by [31], which supports a wide range of SOTA fusion algorithms. Fourteen fusion methods are selected for evaluation, representing diverse algorithmic families:

- Edge-preserving and filtering-based methods: Cross-Bilateral Filter (CBF) [32], Guided Filter-based Context Enhancement (GFCE) [33], Hybrid Multiscale Decomposition with Guided Filter (HMSD-GF) [33], Hybrid_MSD [34], Anisotropic Diffusion Fusion (ADF) [35].
- Decomposition-based methods: Latent Low-Rank Representation (LatLRR) [17], Multi-scale Guided Filter Fusion (MGFF) [36], Multi-resolution Singular Value Decomposition (MSVD) [37], Gradient Transfer Fusion (GTF) [38].
- Saliency and frequency-based methods: Two-scale Image Fusion (TIF) [39], Fourth Order Partial Differential Equation Fusion (FPDE) [40], Visual Saliency Map Weighted Least Squares (VSMWLS) [41].
- Deep learning-based methods: Convolutional Neural Network (CNN) fusion [42], Infrared Feature Extraction and Visual Information Preservation (IFEVIP) [43].

Each fusion method is applied to the registered RGB and NIR image pairs to generate fused images that combine spatial detail from the visible spectrum with spectral information from the NIR band.

3.4. COD techniques

The fused images, along with the original RGB images, are used to train and evaluate nine state-of-the-art deep learning models for COD. These models are chosen to represent a variety of architectural paradigms, including convolutional neural networks (CNNs), transformer-based networks, and hybrid approaches. Table 2 shows the techniques to be used and the main characteristics. The evaluated models are:

- BASNet [44]: A boundary-aware salient object detection network using a ResNet-34 backbone.
- SINet-v2 [45]: A Res2Net-50-based model designed for concealed object detection.
- BGNet [22]: A boundary-guided network with a Res2Net-50 backbone, optimized for precise edge detection.
- C^2F-Net [46]: A context-aware cross-level fusion network using Res2Net-50.
- OCENet [47]: A ResNet-50-based model that incorporates aleatoric uncertainty modeling.
- EAMNet [48]: An edge-aware mirror network with a Res2Net-50 backbone.
- DGNet [49]: A deep gradient learning network using EfficientNet.
- HitNet [20]: A high-resolution iterative feedback network based on PVTv2.
- PCNet [19]: A camouflage plant detection network using PVTv2, specifically designed for agricultural applications.

Each model is trained and validated on both the original RGB dataset and the three best-performing fusion datasets (CBF, HMSD-GF, LatLRR). Training is performed using standard data augmentation techniques to improve generalization, and hyperparameters are optimized for each model to ensure fair comparison. The evaluation is conducted on a held-out test set with manually annotated ground truth masks.

3.5. Evaluation metrics

For RGB-NIR fusion quality evaluation of the fused images are quantitatively assessed using thirteen metrics, each capturing a different aspect of fusion performance. A point-based ranking system is used to identify the top-performing fusion methods, which are then selected for subsequent COD experiments. Table 3 shows each of the metrics and the category to which they belong.

On the other hand, for COD evaluation performance and to provide a rigorous assessment, five widely adopted metrics are used:

- S-measure (S_α) [64]: Measures the structural similarity between the predicted segmentation and the ground truth, capturing both region-aware and object-aware similarities.



Fig. 2. *(left)* DJI Mavic 2 Pro-UAV. *(right)* MicaSense RedEdge-M multispectral camera.

Table 4
Details of the training parameters used in evaluated SOTA COD techniques. Learning rate (LR); Batch size (BS).

Technique	Optimizer	LR	BS	Epochs	Scheduler	Loss function
BASNet [44]	Adam	1e-3	8	1000	ReduceLROnPlateau	BCE + SSIM + IOU (multi-stage fusion)
SINet-v2 [45]	Adam	1e-4	16	150	Custom (Adjust LR)	Structure loss (weighted BCE + weighted IOU)
BGNet [22]	Adam	1e-4	12	100	Custom (Poly LR)	Structure loss (weighted BCE + weighted IOU) + Dice loss (edge)
C^2F-Net [46]	AdaXW	1e-4	32	50	Custom (Poly LR)	Structure loss (weighted BCE + weighted IOU)
OCENet [47]	Adam	1e-5	4	50	StepLR	Uncertainty aware structure loss (weighted BCE + weighted IOU)
EAMNet [48]	AdamW	5e-5	16	150	Custom (Adjust LR)	Hybrid loss (weighted BCE + weighted IOU) + Edge loss (edge)
DGNet [49]	AdamW	5e-5	16	150	CosineAnnealingLR	Hybrid loss (weighted BCE + weighted IOU) + MSE loss (grad)
HitNet [20]	AdamW	1e-4	8	150	Custom (Adjust LR)	Structure loss (weighted BCE + weighted IOU)
PCNet [19]	AdamW	1e-4	8	150	Custom (Adjust LR)	Structure loss (weighted BCE + weighted IOU)

- Weighted F-measure (F_β^w) [65]: An extension of the traditional F-measure that incorporates spatial weights, emphasizing boundary accuracy and the spatial relevance of detected pixels.
- Mean Absolute Error (M) [66]: Computes the average pixel-wise error between the normalized prediction and the ground truth mask.
- E-measure (E_ϕ) [67]: Simultaneously evaluates global and local accuracy based on human visual perception.
- F-measure (F_β) [68]: Provides a balanced measure of precision and recall.

To capture different aspects of model performance, adaptive, mean, and maximum variants of both F-measure and E-measure (F_β^{adp}, F_β^{mean}, $F_\beta^{\max}$, E_ϕ^{adp}, E_ϕ^{mean}, $E_\phi^{\max}$) are also computed. This comprehensive evaluation framework ensures that both the accuracy and the robustness of each model are thoroughly assessed across multiple datasets and fusion strategies.

3.6. Training details

This section summarizes the training configuration used across all COD techniques. Table 4 reports the optimizer, learning rate, batch size, number of epochs, scheduler, and loss functions employed per model. All models are trained under the same consistent splits and metrics described in Section 3.5 are used to enable a fair comparison.

4. Case studies

This section presents the experimental results from the two case studies evaluated with the proposed methodology. The first case study presented uses our own dataset (Weeds-Banana) while the second case study uses a state-of-the-art dataset (WeedsGalore [26]). For the performance evaluation of this work, the metrics described in Section 3.5 are used. All experiments are conducted on a workstation configured with an Intel Core i9 processor and an NVIDIA Titan XP GPU. The framework proposed by [31] is used for the fusion techniques. This code has been implemented using MATLAB and is available at the following URL https://github.com/xingchenzhang/VIFB. For the nine COD techniques, all implementations are written in Python by the authors.

4.1. Datasets

Weeds-Banana. For the first case study, a custom dataset created to validate the proposed methodology is used, named Weeds-Banana. Dataset acquisition is systematically conducted in the Milagro canton, located in Guayas, Ecuador, a region characterized by a tropical climate with average temperatures ranging between 25 °C and 27 °C and annual precipitation between 1,100 and 1,800 mm. The study focused on an experimental plot of approximately 4.3 hectares within a banana plantation, providing a representative agricultural environment for research on the detection of camouflaged weeds. Data collection utilized a DJI Mavic 2 Pro-UAV equipped with MicaSense RedEdge-M multispectral camera, enabling simultaneous capture of visible and near-infrared spectral information (See Fig. 2). Flight operations are programmed with 85% lateral and frontal overlap at an altitude of 120 m above the canopy, ensuring comprehensive coverage and sufficient image overlap for accurate mosaic generation. Image capture is conducted in April 2024 between 11:00 and 12:00 to minimize shadow effects and optimize lighting conditions for both RGB and NIR imaging.

WeedsGalore [26]. For the second case study, a state-of-the-art dataset, WeedsGalore, is used. This dataset is systematically acquired in an experimental maize field near Marquardt, Potsdam, Germany (approx. 1,840 m^2), representative of conventional row-crop systems in Central Europe. Maize (Zea mays L.) is sown on 9 May 2023 under standard management practices, while naturally emerging weeds are allowed to develop, resulting in a dense and diverse weed community typical of regional maize production. Data collection is carried out using a DJI Phantom 4 Multispectral UAV, which provides RGB imagery together with five monochromatic multispectral bands (blue, green, red, red-edge, and near-infrared). Four flight campaigns are conducted on 25 May, 30 May, 6 June, and 15 June 2023, covering phenological stages from early maize emergence with sparse vegetation to almost complete ground cover by crops and weeds. Flights are performed at 5 m above ground level in Hover&Capture mode with 70% side and 60% front over-

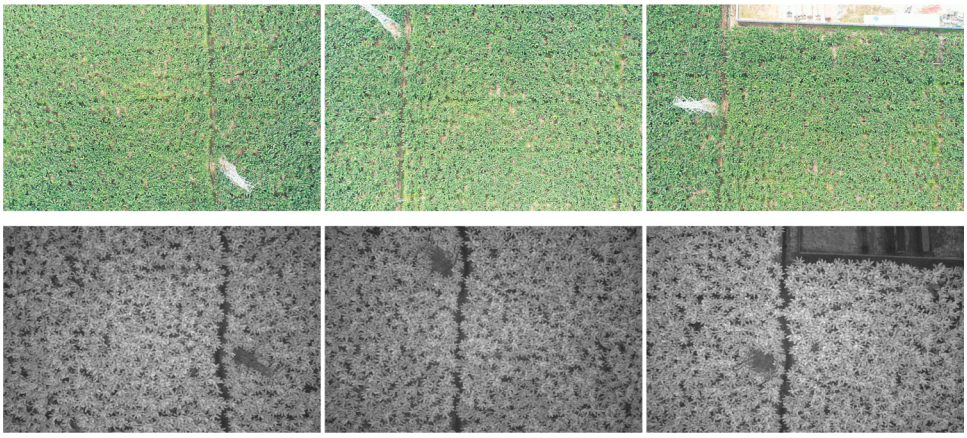


Fig. 3. Example images used to build the RGB and NIR orthomosaics.

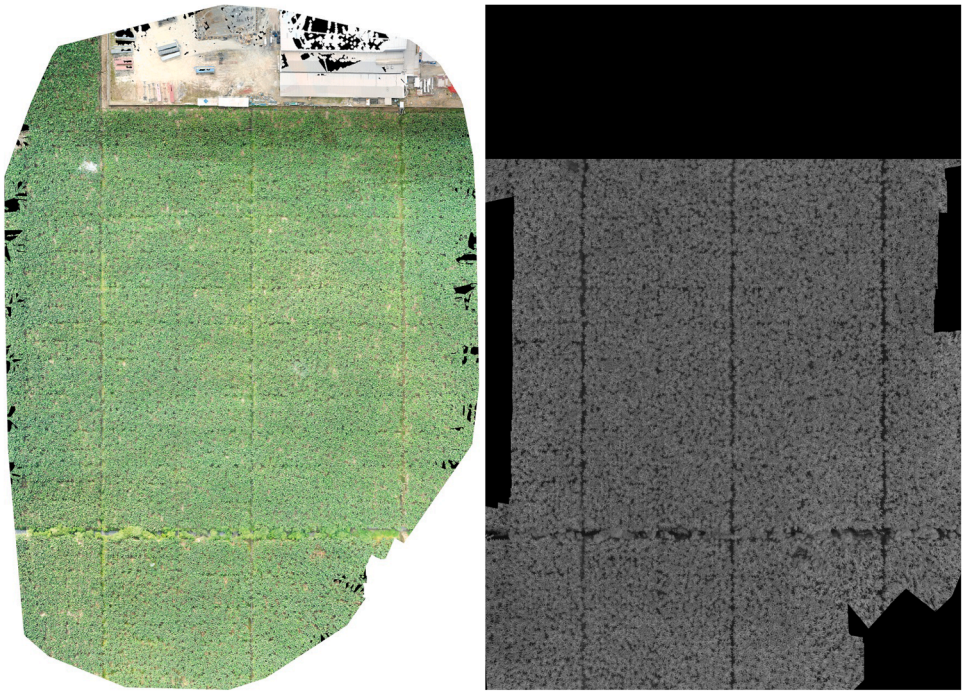


Fig. 4. RGB and NIR orthomosaics registered using QGIS.

lap, yielding a ground sampling distance of approximately 2.5 mm/pixel and enabling individual plant discrimination. From roughly 1,150 raw images per campaign, 48 geo-referenced locations are selected without spatial overlap, and for each location, a central 600×600-pixel window per band is extracted to form the annotated tiles used for semantic segmentation.

4.2. Preprocessing

Weeds-Banana. For our Weeds-Banana dataset, the acquired images underwent systematic processing using PIX4Dmapper[1] software, involving radiometric calibration to ensure spectral accuracy, photographic alignment for geometric consistency, image geometry construction for spatial precision, and orthophotography generation to produce high-quality RGB and multispectral mosaics suitable for subsequent fusion and analysis processes. Fig. 3 shows example images used to build the RGB and NIR orthomosaics. On the other hand, Fig. 4 shows the orthomosaic generated using RGB (*left*) and NIR (*right*) images.

In our Weeds-Banana dataset, the images have been registered using the QGIS[2] platform, a crucial step that will be useful in the subsequent stages. The resulting registered RGB and NIR image sizes are 5571×5855 pixels. After registering the RGB-NIR images, they are labeled using the RGB image as a reference to identify areas with weeds. To carry out the labeling task, the Roboflow[3] tool is used. Fig. 5 (left) shows the labeling of the RGB orthomosaic image; the result of the annotation process is a binary mask of 5571×5855 pixels (see Fig. 5 (right)). The registered RGB-NIR images and the binary mask resulting from the annotation process will be available in the GitHub repository. Finally, after completing the annotation stage of the RGB orthomosaic image, a Python script called *"split_dataset.py"* (available in GitHub repository) is applied to generate 272 patches. The distribution of the images used to train and test the different models is training = 218, validation = 27, and

[1] https://www.pix4d.com/

[2] https://qgis.org/

[3] https://app.roboflow.com/

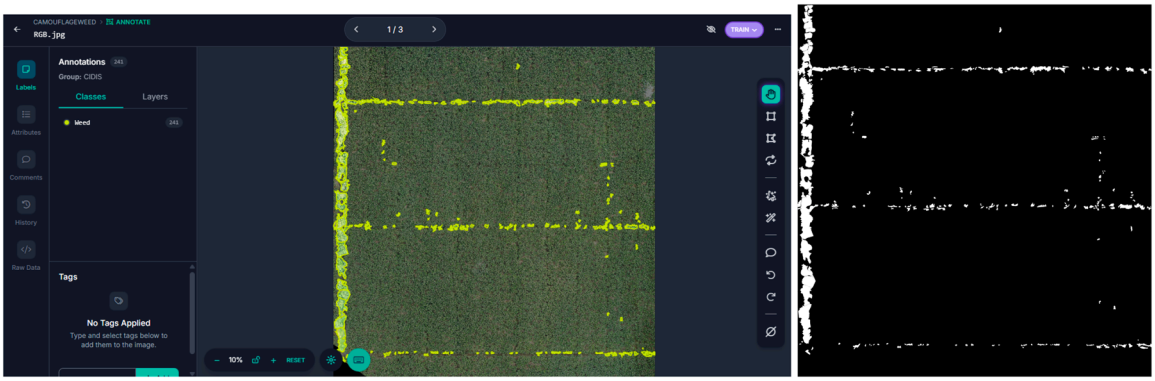


Fig. 5. (Left) View of the Roboflow annotation tool for the RGB orthomosaic image. (Right) Labeled orthomosaic binary mask using Roboflow.

testing = 27. Fig. 6 (left) shows examples of the patches created from the registered RGB-NIR orthomosaic.

WeedsGalore [26]. For the WeedsGalore [26] dataset, several preprocessing steps are applied to transform the raw UAV imagery into consistent tiles suitable for semantic and instance segmentation. For each of the four flight campaigns, the images captured by the DJI Phantom 4 Multispectral are processed at the level of individual frames rather than full-field orthomosaics. The five monochromatic multispectral bands (blue, green, red, red-edge, and near-infrared) are geometrically aligned and stacked, and, for every selected frame, a central 600×600-pixel window is cropped in all bands. From approximately 1,150 raw images per campaign, 48 geo-referenced locations are defined across the field, and the corresponding tiles are extracted such that there is no spatial overlap between annotated images, either within a date or across dates. This procedure results in 156 multispectral tiles at a ground sampling distance of about 2.5 mm/pixel, allowing individual crops and weeds to be clearly resolved. For each tile, paired semantic and instance annotations are generated using consistent class indices for maize and four weed categories, and the final dataset is organized into spatially disjoint training, validation, and test splits (104 / 26 / 26 tiles) using the official split files provided with the dataset. Fig. 6 (right) shows patch examples used for the WeedGalore dataset.

4.3. Fusion techniques results

To establish the best fusion techniques, a point system is defined based on ranking position (i.e., $1st = 3points$, $2nd = 2points$, $3rd = 1point$). The analysis is based on multiple complementary metrics that evaluate different aspects of fusion quality. Among them, some seek lower values, such as CE, $RMSE$, and Q_{CV}, while others, such as EN, MI, $PSNR$, AG, FI, $Q^{AB/F}$, SD, SF, $SSIM$, and Q_{CB}, are improved with higher values. This comparison is essential for selecting the most appropriate fusion method for the specific application, particularly in the context of the COD task. Also, Table 3 shows, for each metric, a description and the category to which it belongs.

Weeds-Banana. Table 5 presents a comprehensive evaluation of SOTA fusion techniques using our Weeds-Banana dataset. CBF [32] stands out as the most effective method with 15 total points. HMSD-GF [33] follows it, and LatLRR [17], both with 12 points, are the top-3 of the best techniques, while the remaining methods obtain progressively lower scores, demonstrating a clear hierarchy in the performance of these techniques. CBF demonstrates superior performance across most metrics, justifying its dominant position in the overall ranking, although other methods, such as HMSD-GF, LatLRR, and GFCE, also exhibit significant strengths in specific metrics, offering viable alternatives depending on the priority requirements in each use case.

WeedsGalore [26]. On the other hand, Table 6 presents a comprehensive evaluation of SOTA fusion techniques using the WeedsGalore [26] dataset. Contrary to the results presented in Table 5, the ranking position of each technique is different, indicating that the performance of each fusion technique depends on the characteristics of each dataset. LatLRR [17] had the best overall performance with a total of 18 points. GFCE [33] followed at a considerable distance with a total of 10 points, while FPDE [40] with 9 points completed the top three best fusion techniques.

4.4. SOTA COD techniques results

Weeds-Banana. Using our Weeds-Banana dataset, we report Table 7, Fig. 7 and Fig. 8 show the results obtained. Table 7 presents a comprehensive quantitative evaluation of nine SOTA COD techniques (i.e., BASNet [44], SINet-v2 [45], BGNet [22], C^2F-Net [46], OCENet [47], EAMNet [48], DGNet [49], HitNet [20], PCNet [19]), evaluated across 1 RGB dataset and 3 datasets generated with fusion techniques (i.e., CBF, HMSD-GF, and LatLRR). The results show that PCNet [19] consistently stands out, obtaining the highest scores on F_β^w, M, E_ϕ^{adp}, E_ϕ^{mean}, $E_\phi^{\max}$, F_β^{adp}, and F_β^{mean} metrics. Furthermore, BASNet [44] shows the best result for the S_α metric, just as BGNet shows the best result for the $F_\beta^{\max}$ metric.

On the other hand, for our Weeds-Banana dataset, Fig. 7 shows qualitative results using two sample images from the test set evaluated for the top-3 techniques. It can be observed that RGB-NIR-based techniques outperform RGB-only techniques in small regions. Furthermore, corroborating the quantitative results, the PCNet [19] technique presents consistent results with well-defined segmented areas of weeds relative to the GT. In a more specific analysis focusing on the PCNet [19] COD technique Fig. 8 shows a comparison of the mask prediction results using RGB and fusion techniques where it can be observed that using RGB-NIR fusion techniques, masks with better contour definition and fewer false negatives are obtained, in this way the qualitative results presented in Table 7 can be confirmed.

WeedsGalore [26]. Using the WeedsGalore [26] dataset, we report Table 8, Fig. 9, and Fig. 10 summarize the quantitative and qualitative performance of the nine SOTA COD techniques across the RGB baseline and the three best-ranked fusion configurations for this dataset (i.e., FPDE, GFCE, and LatLRR, as identified in Section 4.3). Overall, the results confirm that incorporating NIR information through RGB-NIR fusion benefits camouflaged weed segmentation also in this second case study, although the magnitude of the improvement is more method- and metric-dependent than in Weeds-Banana. In particular, the best-performing techniques on WeedsGalore are BASNet, PCNet, and OCENet, which are therefore selected for qualitative inspection in Fig. 9.

From the quantitative results using the WeedsGalore dataset in Table 8, BASNet provides the most consistent trade-off across the evaluated metrics, achieving top-ranked scores when coupled with fusion inputs–especially under GFCE/FPDE, where improvements are observed in both region accuracy and boundary consistency. BASNet is highly competitive and yields very strong F-measure values on this dataset and

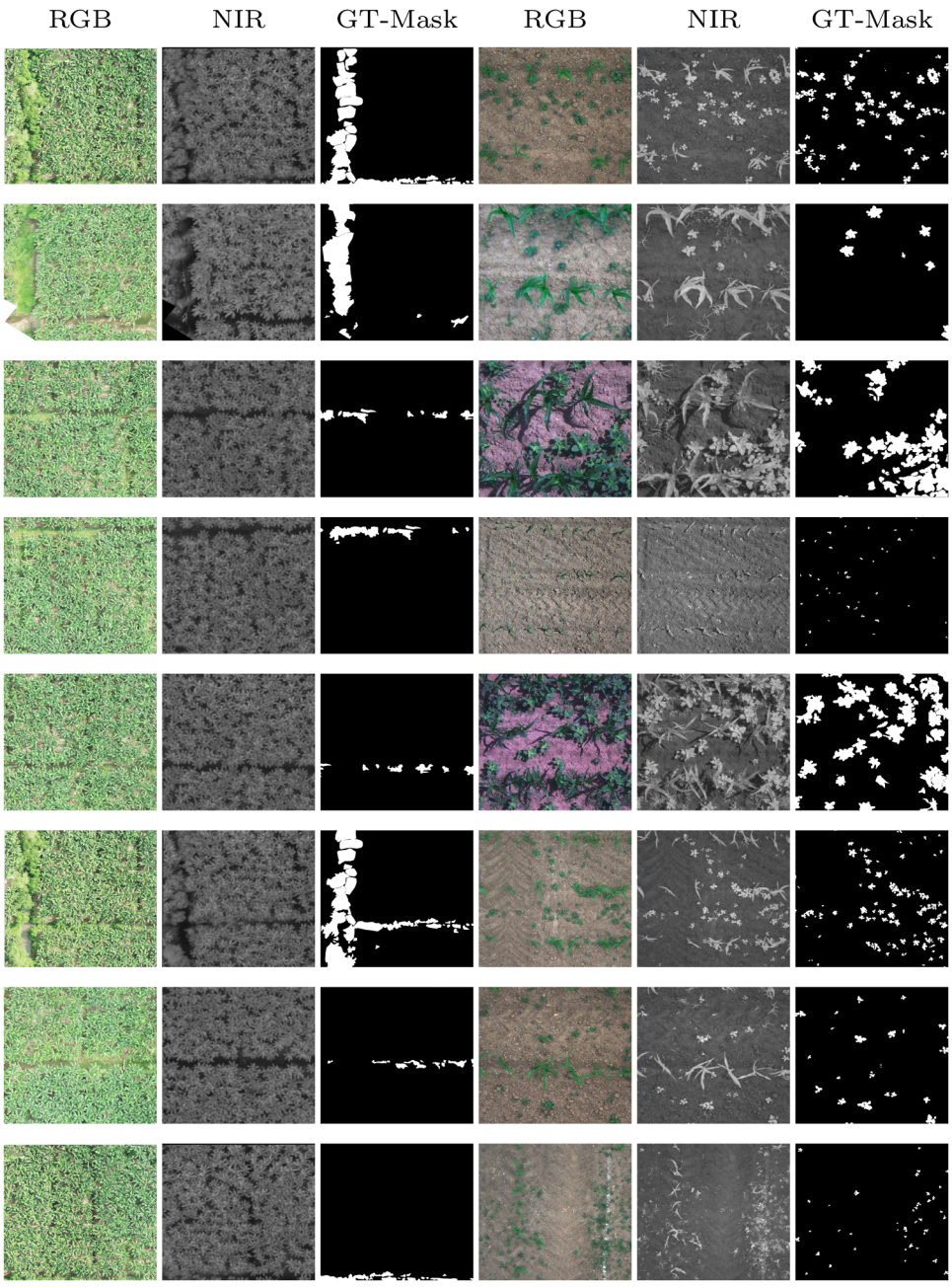


Fig. 6. Example images of the patches created from the registered orthomosaic using our Weeds-Banana and the WeedsGalore [26] datasets.

Table 5
Metric evaluation results of SOTA fusion techniques using our Weeds-Banana dataset. Results are presented using the metric notation defined in Section 3.5, "↑ / ↓" indicates that larger or smaller is better. The best three performing results are highlighted in red (first), blue (second), and green (third) respectively.

Technique	Ptos.	$CE \downarrow$	$EN \uparrow$	$MI \uparrow$	$PSNR \uparrow$	$AG \uparrow$	$EI \uparrow$	$Q^{AB/F} \uparrow$	$SD \uparrow$	$SF \uparrow$	$RMSE \downarrow$	$SSIM \uparrow$	$Q_{CB} \uparrow$	$Q_{CV} \downarrow$
CBF [32]	15	1.51	7.05	3.78	62.3390	14.07	141.00	0.8100	42.98	32.05	0.0391	1.0875	0.7248	152.00
HMSD-GF [33]	12	1.34	7.43	2.12	61.8620	15.41	154.45	0.7800	46.65	34.90	0.0432	1.0820	0.6320	148.50
LatLRR [17]	12	1.73	6.72	1.18	59.5820	24.58	246.15	0.4310	66.65	58.89	0.0717	0.8540	0.5706	223.47
GFCE [33]	11	1.55	7.62	1.06	60.2425	17.07	171.42	0.6191	54.80	38.83	0.0615	1.0051	0.5076	404.01
FPDE [40]	8	1.20	6.76	1.07	62.3835	9.74	94.39	0.5778	26.99	21.74	0.0387	1.0965	0.4654	339.28
Hybrid_MSD [34]	6	1.29	7.37	2.11	62.2400	14.35	143.94	0.7753	43.11	32.55	0.0398	1.0972	0.6419	152.11
MGFF [36]	3	1.13	7.46	1.13	61.9365	13.39	134.44	0.6724	44.00	30.56	0.0427	1.0940	0.5383	340.82
TIF [39]	3	1.05	7.33	1.35	62.0415	12.49	126.78	0.6694	40.32	28.44	0.0417	1.0730	0.5776	211.54
ADF [35]	2	1.24	6.89	1.07	62.3210	11.37	109.85	0.6507	29.45	26.00	0.0392	1.0862	0.4730	344.84
CNN [42]	2	1.32	7.40	1.68	61.9625	14.13	141.58	0.8043	43.76	32.21	0.0423	1.0939	0.5564	189.93
IFEVIP [43]	2	2.10	6.90	2.82	60.8045	10.26	103.25	0.5318	35.05	24.12	0.0542	1.0922	0.4802	378.28
VSMWLS [41]	2	1.09	7.20	1.21	62.1130	12.44	121.03	0.6252	36.55	29.20	0.0410	1.0795	0.4845	443.47
MSVD [37]	1	1.60	6.81	0.92	62.3170	9.59	89.80	0.4404	27.66	25.49	0.0392	1.0653	0.3997	426.00
GTF [38]	0	1.44	7.02	0.64	61.5470	11.90	115.70	0.6189	31.61	28.00	0.0468	0.9984	0.4687	836.72

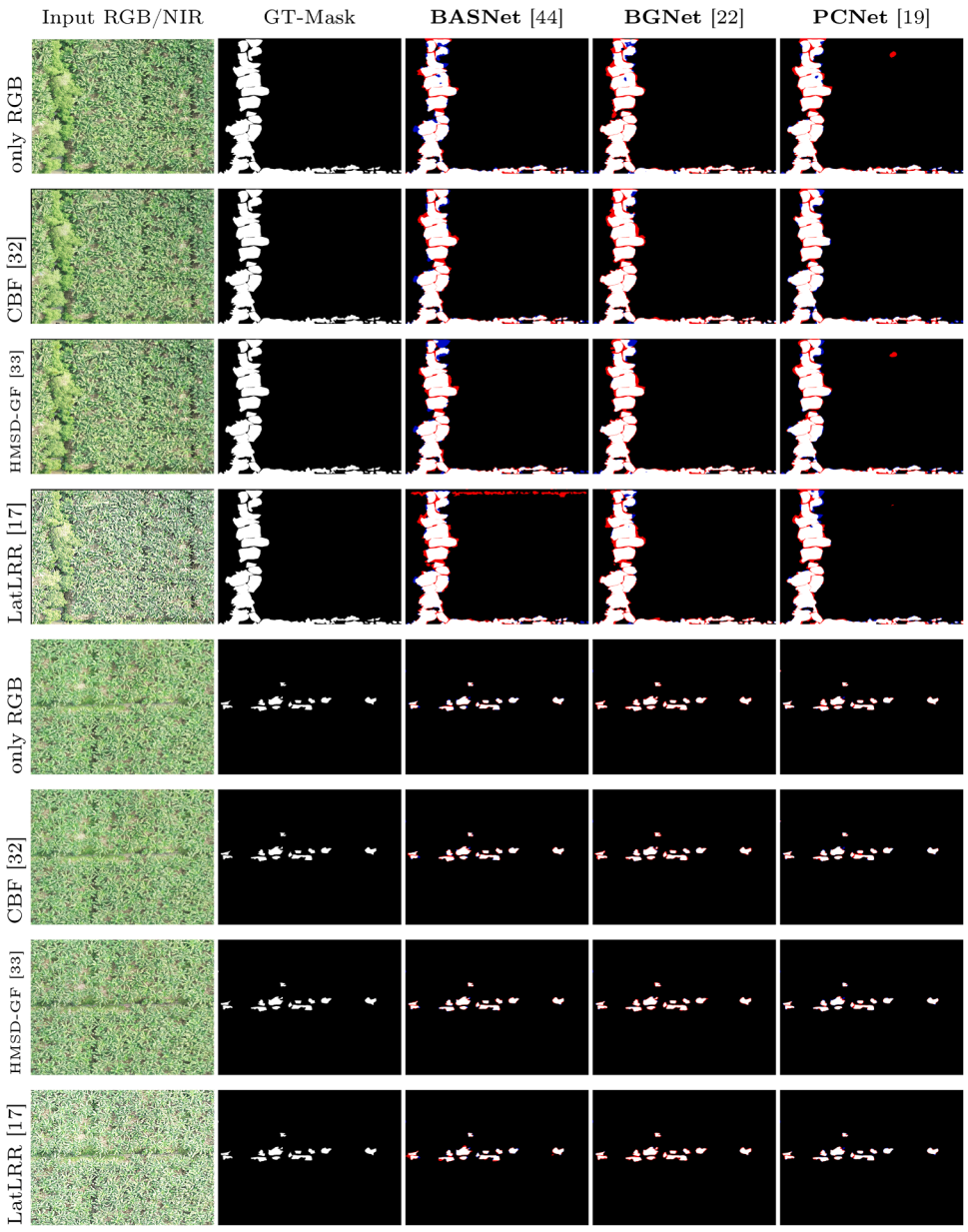


Fig. 7. Qualitative prediction comparison for the top-3 COD techniques on our Weeds-Banana test set using two representative samples. Rows correspond to RGB-only and the best-ranked fusion inputs for this dataset (CBF, HMSD-GF, LatLRR). Columns show the GT mask and predictions of BASNet, BGNet, and PCNet. White indicates correct overlap with GT, red false positives, and blue false negatives.

Table 6
Metric evaluation results of SOTA fusion techniques using the WeedsGalore [26] dataset. Results are presented using the metric notation defined in Section 3.5, "↑ / ↓" indicates that larger or smaller is better. The best three performing results are highlighted in red (first), blue (second green (third) respectively.

Technique	Ptos.	$CE \downarrow$	$EN \uparrow$	$MI \uparrow$	$PSNR \uparrow$	$AG \uparrow$	$EI \uparrow$	$Q^{AB/F} \uparrow$	$SD \uparrow$	$SF \uparrow$	$RMSE \downarrow$	$SSIM \uparrow$	$Q_{CB} \uparrow$	$Q_{CV} \downarrow$
LatLRR [17]	18	1.52	7.44	1.86	58.2501	21.86	215.59	0.2501	52.30	48.32	0.0976	1.1310	0.6608	341.23
GFCE [33]	10	0.57	7.17	2.36	61.7139	13.68	133.48	0.5953	37.43	31.27	0.0459	1.5265	0.7006	144.28
FPDE [40]	9	0.32	6.43	2.52	64.7010	7.83	76.87	0.6511	22.31	17.52	0.0239	1.7425	0.7056	184.99
IFEVIP [43]	8	1.15	7.01	2.83	59.9904	11.64	115.48	0.6109	35.65	26.03	0.0662	1.5906	0.7194	209.53
ADF [35]	7	0.30	6.45	2.57	64.6922	8.04	78.61	0.6628	22.53	18.15	0.0239	1.7361	0.7061	189.01
CNN [42]	6	0.30	6.77	3.26	63.6096	10.16	97.71	0.6990	28.85	24.28	0.0307	1.6275	0.7072	118.90
TIF [39]	5	0.22	6.65	2.47	64.1743	9.03	88.90	0.6458	26.37	21.30	0.0266	1.6825	0.7185	135.33
MGFF [36]	4	0.28	7.01	2.20	63.3655	11.98	117.71	0.5874	33.06	27.32	0.0317	1.6027	0.6746	174.58
CBF [32]	3	0.32	6.60	2.63	64.2969	9.08	88.45	0.6788	25.09	21.18	0.0264	1.6808	0.6968	165.13
MSVD [37]	3	0.30	6.51	2.07	64.2872	8.89	83.54	0.5806	23.55	22.20	0.0258	1.6478	0.6772	199.47
VSMWLS [41]	3	0.21	6.72	2.52	64.0546	10.52	101.06	0.6456	27.40	24.91	0.0271	1.6661	0.6695	145.00
GTF [38]	2	0.29	6.63	2.46	63.6851	9.70	92.68	0.6859	26.65	23.42	0.0300	1.6232	0.6915	160.99
Hybrid_MSD [34]	2	0.26	6.81	2.64	63.7672	10.96	106.58	0.6558	29.14	25.86	0.0292	1.6580	0.7059	143.70
HMSD-GF [33]	1	0.38	6.90	2.56	62.8050	11.30	110.06	0.6517	31.00	26.12	0.0356	1.6396	0.7125	124.55

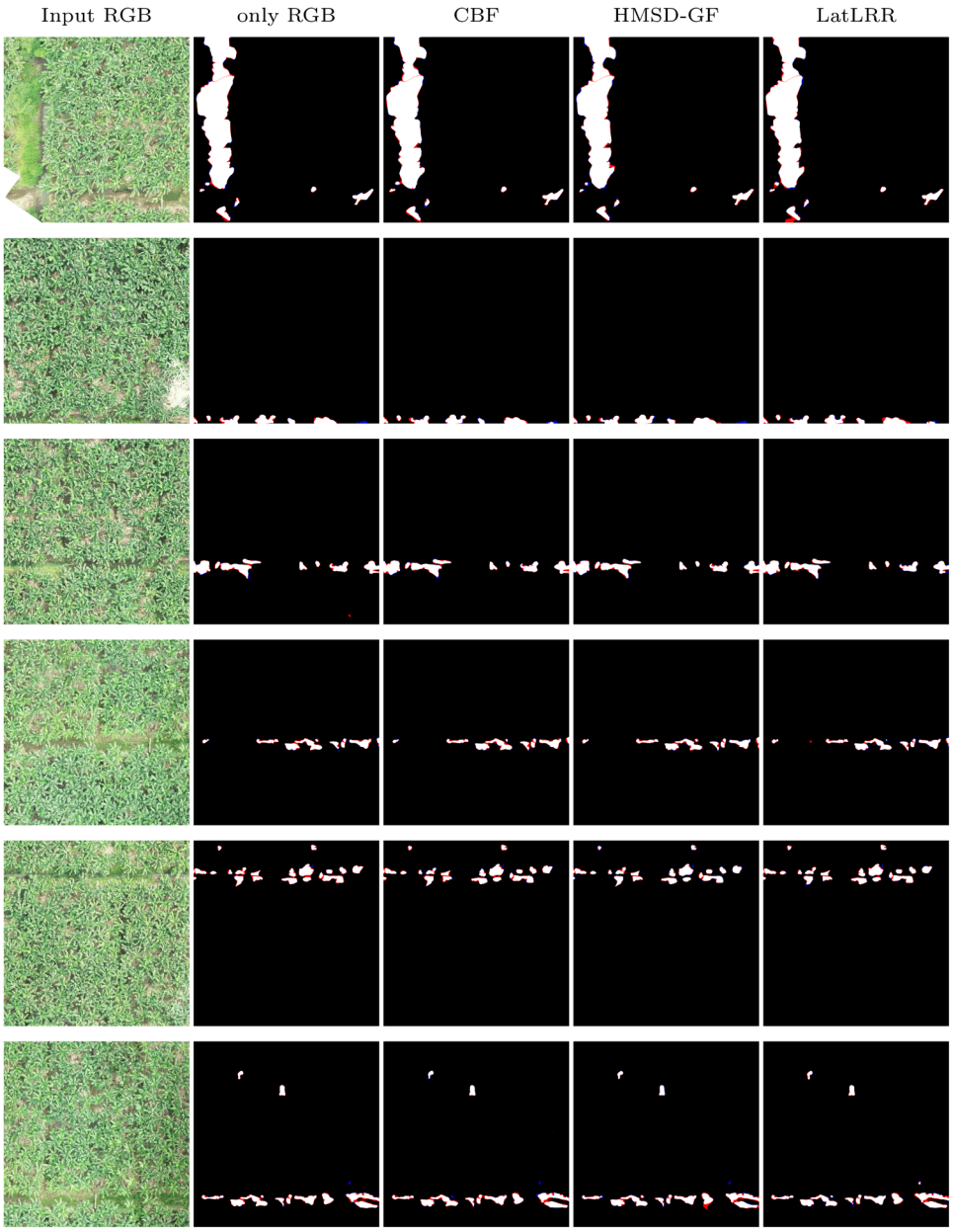


Fig. 8. Qualitative results for PCNet [19] on our Weeds-Banana [26] test set under RGB-only and fusion-based inputs (CBF, HMSD-GF, LatLRR). The first column provides the RGB input for reference. The error maps use white/red/blue to indicate correct overlap, false positives, and false negatives, respectively. Compared with RGB-only, fusion inputs typically produce less fragmented masks and improved alignment with GT.

remains robust and produces stable predictions, particularly in scenes where weeds are sparse and appear as small, low-contrast structures. PCNet and OCENet show good results but are far from the performance of BASNet. The qualitative examples in Fig. 9 show that fusion-based inputs (FPDE, GFCE, LatLRR) reduce miss-segmentation (false negatives) in difficult low-contrast weed regions compared to RGB-only inputs, while also helping mitigate over-segmentation in visually ambiguous areas. Finally, Fig. 10 focuses on BASNet, illustrating that fusion configurations generally lead to cleaner masks with fewer fragmented detections and improved alignment with the ground truth, corroborating the quantitative trends observed in Table 8.

In conclusion and in a general way, following the proposed methodology, it is shown that the approach using the fusion of the RGB and NIR images shows better results than only RGB images. Upon reviewing each camouflage technique, it can be observed that the datasets generated using the RGB-NIR fusion techniques exhibit better performance in all metrics (the first place for each COD technique is marked with an underline). In the case of our Weeds-Banana dataset, the combination of PCNet-LatLRR COD-Fusion technique presents the best overall result. While for WeedsGalore, the most competitive trade-offs are obtained with BASNet-FPDE, where this combination provides the strongest overall performance.

5. Discussion

The experimental results across the two case studies demonstrate that integrating NIR information with RGB imagery improves the segmentation of camouflaged weeds in real agricultural settings, while also showing that the optimal fusion strategy depends on the dataset characteristics.

In the banana plantation scenario, the fusion approaches CBF, HMSD-GF, and LatLRR consistently rank among the best, and the down-

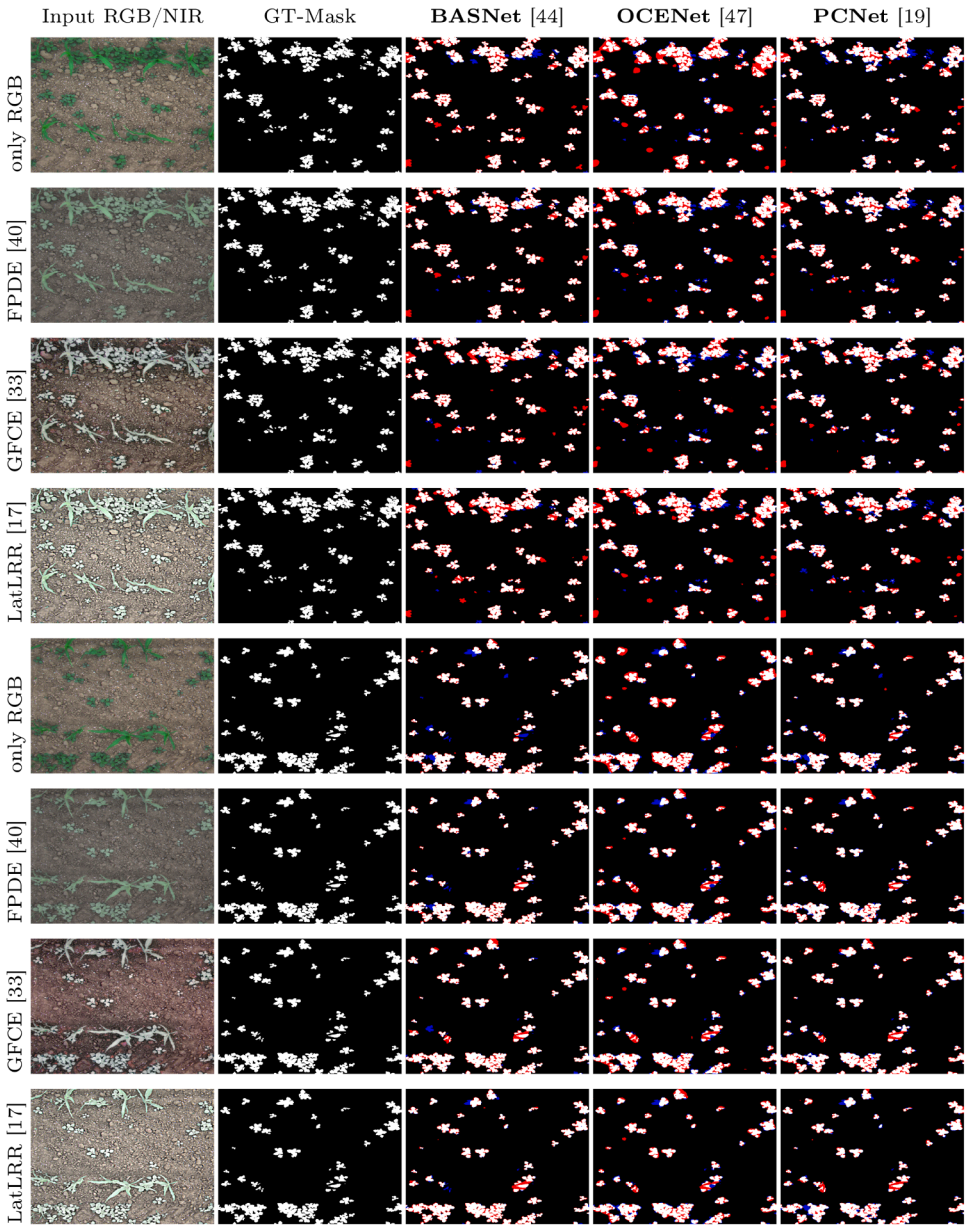


Fig. 9. Qualitative prediction comparison for the top-3 COD techniques on the WeedsGalore [26] test set using two representative samples. Rows correspond to RGB-only and the best-ranked fusion inputs for this dataset (FPDE, GFCE, LatLRR). Columns show the GT mask and predictions of BASNet, OCENet, and PCNet. White indicates correct overlap with GT, red false positives, and blue false negatives.

stream COD results show clear gains when models are trained with fused inputs rather than RGB alone. This behavior is particularly evident in small or thin weed regions, where RGB-only segmentation tends to miss low-contrast areas. Among the evaluated COD methods, PCNet stands out as the most reliable approach, producing masks with sharper contours and fewer false negatives, which is consistent with its design for plant camouflage detection.

In the maize field scenario, the fusion ranking changes, and the best-performing fusion configurations become LatLRR, GFCE, and FPDE, confirming that fusion performance is sensitive to acquisition conditions, scene structure, and target appearance. While fusion still provides benefits over RGB-only inputs, the improvement is more heterogeneous across methods and metrics than in Weeds-Banana. This can be attributed to the different imaging conditions (very high spatial resolution tiles, different UAV/sensor, multi-temporal variability) and to the fact that weeds often appear as small, scattered objects, making the task more sensitive to small false positives/negatives. In this dataset, BASNet provides the strongest overall trade-off and stable qualitative behavior, while PCNet and OCENet remain highly competitive depending on the metric emphasis (e.g., region overlap vs. boundary consistency).

Across both datasets, a consistent conclusion is that RGB-NIR fusion improves the representation available to COD models, enabling earlier and more accurate weed localization, which is crucial for precision agriculture workflows that aim to reduce herbicide usage through targeted interventions. At the same time, the cross-dataset behavior highlights two practical implications: (1) there is no universally optimal fusion method, and fusion should be selected (or learned) with awareness of the target domain; and (2) COD architectures exhibit dataset-dependent rankings, suggesting that model selection should consider the crop type, weed morphology, resolution, and density. Future work should therefore explore adaptive or learnable fusion strategies and broader validations across additional crops, sensors, and environmental conditions.

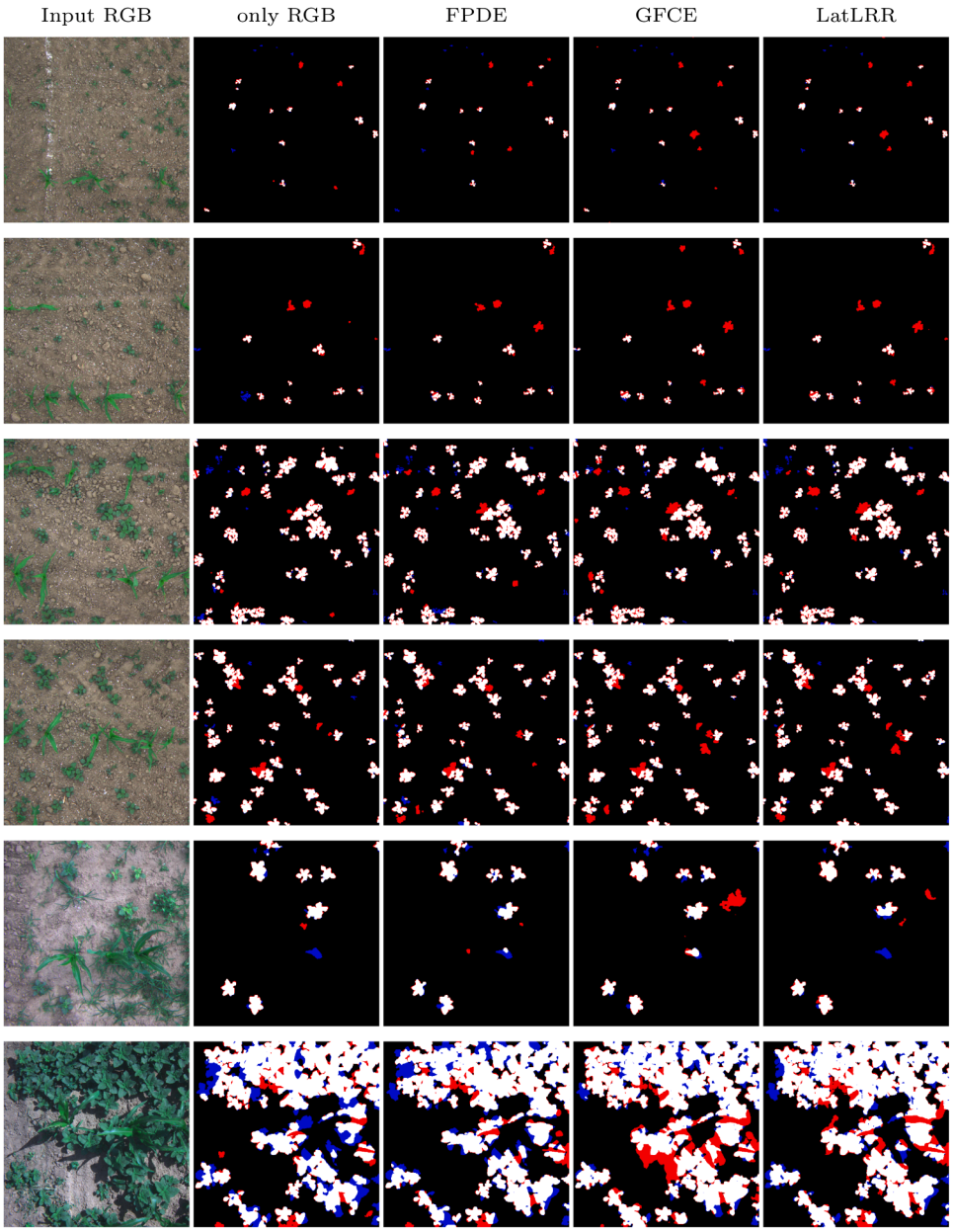


Fig. 10. Qualitative results for BASNet [44] on the WeedsGalore [26] test set under RGB-only and fusion-based inputs (FPDE, GFCE, LatLRR). The first column provides the RGB input for reference. The error maps use white/red/blue to indicate correct overlap, false positives, and false negatives, respectively. Compared with RGB-only, fusion inputs typically produce less fragmented masks and improved alignment with GT.

Despite the overall gains from RGB-NIR fusion, some failure modes are observed. First, very thin weed structures or extremely small weeds may be under-segmented (false negatives), particularly when their spatial footprint approaches the sensor resolution or when their NIR/RGB contrast is weak. Second, visually ambiguous transitions between soil residues and vegetation can produce localized over-segmentation (false positives), especially in high-texture backgrounds. Third, imperfect RGB-NIR registration and radiometric inconsistencies (e.g., brightness differences across the mosaic) can introduce edge artifacts that affect boundary precision. These observations suggest that further improvements could be obtained by (i) stronger domain-specific augmentation, (ii) more robust cross-spectral registration, and (iii) fusion strategies that explicitly model uncertainty or confidence at boundaries.

Finally, limitations of this study include (i) the dependency of fusion performance on the target domain (sensor characteristics, resolution, and scene structure), (ii) potential sensitivity to strong shadows, specular highlights, and radiometric drift across large mosaics, and (iii) the fact that COD model rankings can change with weed density, morphology, and crop row structure. Although we partially address domain shift by evaluating on WeedsGalore (different crop, country, platform, and multi-temporal acquisition), further multi-site and multi-season validation is required to fully characterize robustness.

Table 7
Metric evaluation results for each COD technique on our Weeds-Banana dataset, reported for the RGB baseline and the three best-ranked RGB–NIR fusion inputs (CBF, HMSD-GF, and LatLRR). Results are presented using the metric notation defined in Section 3.5, "↑ / ↓" indicates that larger or smaller is better. The top three performing results across the entire table are highlighted in red (first), blue (second), and green (third) respectively. Underlined values represent the top performance for each specific technique.

Technique	Method	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$E_\phi^{adp} \uparrow$	$E_\phi^{mean} \uparrow$	$E_\phi^{\max} \uparrow$	$F_\beta^{adp} \uparrow$	$F_\beta^{mean} \uparrow$	$F_\beta^{\max} \uparrow$
BASNet [44]	RGB	0.9562	0.5125	0.0043	0.9718	0.9828	0.9960	0.4825	0.5192	0.5371
	CBF	0.9529	0.5075	0.0045	0.9601	0.9833	0.9961	0.4701	0.5113	0.5342
	HMSD-GF	0.9599	0.5151	0.0042	0.9674	0.9843	0.9968	0.4714	0.5182	0.5416
	LatLRR	0.9339	0.4747	0.0078	0.9491	0.9698	0.9825	0.4439	0.4840	0.5033
SINet-v2 [45]	RGB	0.8761	0.3410	0.0111	0.8774	0.9491	0.9826	0.3189	0.3646	0.4002
	CBF	0.8809	0.3553	0.0105	0.8762	0.9463	0.9756	0.3221	0.3738	0.4179
	HMSD-GF	0.8857	0.3637	0.0103	0.8832	0.9455	0.9837	0.3304	0.3818	0.4266
	LatLRR	0.8852	0.3672	0.0100	0.8856	0.9437	0.9797	0.3364	0.3831	0.4292
BGNet [22]	RGB	0.7232	0.2584	0.1534	0.9716	0.8333	0.9933	0.4754	0.5019	0.5510
	CBF	0.7294	0.2566	0.1490	0.9724	0.8363	0.9934	0.4742	0.4964	0.5465
	HMSD-GF	0.7123	0.2693	0.1711	0.9725	0.8153	0.9967	0.4735	0.4996	0.5498
	LatLRR	0.7275	0.2586	0.1476	0.9466	0.8361	0.9963	0.4460	0.4951	0.5493
C^2F-Net [46]	RGB	0.6931	0.1359	0.1252	0.8462	0.8079	0.9837	0.3100	0.3598	0.4651
	CBF	0.7111	0.1755	0.1219	0.8894	0.8331	0.9927	0.3546	0.4060	0.4862
	HMSD-GF	0.7045	0.1598	0.1248	0.8901	0.8278	0.9894	0.3545	0.3985	0.4799
	LatLRR	0.6783	0.1657	0.1517	0.8830	0.8041	0.9896	0.3503	0.4048	0.4746
OCENet [47]	RGB	0.8937	0.4258	0.0079	0.9674	0.9638	0.9875	0.4238	0.4529	0.4643
	CBF	0.8895	0.4292	0.0077	0.9368	0.9615	0.9878	0.4315	0.4569	0.4686
	HMSD-GF	0.8838	0.4116	0.0081	0.9698	0.9394	0.9800	0.4259	0.4469	0.4605
	LatLRR	0.8867	0.4217	0.0077	0.9429	0.9479	0.9915	0.4398	0.4577	0.4708
EAMNet [48]	RGB	0.7475	0.1157	0.0645	0.7928	0.8579	0.9478	0.2744	0.2901	0.3522
	CBF	0.6868	0.0803	0.1103	0.7202	0.7699	0.9137	0.1970	0.1882	0.2533
	HMSD-GF	0.7467	0.1257	0.0527	0.7887	0.8522	0.9410	0.2382	0.2584	0.2998
	LatLRR	0.7719	0.1839	0.0388	0.7862	0.8620	0.9529	0.2385	0.2850	0.3436
DGNet [49]	RGB	0.8772	0.4094	0.0387	0.8876	0.9278	0.9862	0.3565	0.4219	0.4614
	CBF	0.8930	0.4231	0.0309	0.9017	0.9443	0.9889	0.3748	0.4363	0.4651
	HMSD-GF	0.9079	0.4160	0.0127	0.8759	0.9616	0.9912	0.3660	0.4299	0.4579
	LatLRR	0.8708	0.4173	0.0526	0.8971	0.9182	0.9901	0.3639	0.4285	0.4607
HitNet [20]	RGB	0.9165	0.4794	0.0159	0.9196	0.9671	0.9963	0.4381	0.4723	0.5085
	CBF	0.9263	0.4840	0.0074	0.9359	0.9748	0.9962	0.4499	0.4746	0.5142
	HMSD-GF	0.9345	0.5004	0.0084	0.9437	0.9822	0.9957	0.4744	0.4958	0.5169
	LatLRR	0.9282	0.4926	0.0123	0.9326	0.9758	0.9957	0.4662	0.4875	0.5161
PCNet [19]	RGB	0.9558	0.5294	0.0034	0.9716	0.9901	0.9971	0.4954	0.5245	0.5448
	CBF	0.9559	0.5248	0.0035	0.9696	0.9907	0.9972	0.4981	0.5242	0.5384
	HMSD-GF	0.9555	0.5274	0.0036	0.9707	0.9908	0.9971	0.5021	0.5272	0.5409
	LatLRR	0.9590	0.5278	0.0034	0.9790	0.9903	0.9967	0.5011	0.5272	0.5411

Table 8
Metric evaluation results for each COD technique on the WeedsGalore [26] dataset, reported for the RGB baseline and the three best-ranked RGB–NIR fusion inputs (FPDE, GFCE, and LatLRR). Results are presented using the metric notation defined in Section 3.5, "↑ / ↓" indicates that larger or smaller is better. The top three performing results across the entire table are highlighted in red (first), blue (second), and green (third) respectively. Underlined values represent the top performance for each specific technique.

Technique	Method	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	$M \downarrow$	$E_\phi^{adp} \uparrow$	$E_\phi^{mean} \uparrow$	$E_\phi^{\max} \uparrow$	$F_\beta^{adp} \uparrow$	$F_\beta^{mean} \uparrow$	$F_\beta^{\max} \uparrow$
BASNet [44]	RGB	0.7533	0.4428	0.0150	0.7994	0.8385	0.9253	0.3991	0.4788	0.5150
	FPDE	0.7782	0.4548	0.0138	0.8804	0.8480	0.9070	0.4357	0.4880	0.5200
	GFCE	0.7696	0.4656	0.0145	0.8144	0.8533	0.9149	0.4207	0.5042	0.5474
	LatLRR	0.7541	0.4436	0.0138	0.8530	0.8304	0.8937	0.4003	0.4777	0.5122
SINet-v2 [45]	RGB	0.6350	0.2182	0.0316	0.5266	0.7242	0.8213	0.2015	0.2436	0.2654
	FPDE	0.6465	0.2472	0.0271	0.5476	0.7556	0.8414	0.2259	0.2699	0.3025
	GFCE	0.6433	0.2441	0.0280	0.5592	0.7359	0.7925	0.2307	0.2674	0.2908
	LatLRR	0.6435	0.2428	0.0274	0.5545	0.7229	0.7774	0.2230	0.2640	0.2890
BGNet [22]	RGB	0.5854	0.1478	0.1427	0.7178	0.7016	0.8954	0.3179	0.3383	0.4114
	FPDE	0.5931	0.1558	0.1312	0.7763	0.7115	0.8457	0.3516	0.3772	0.4518
	GFCE	0.5871	0.1442	0.1228	0.7471	0.7310	0.8986	0.3554	0.3939	0.4952
	LatLRR	0.5918	0.1479	0.1397	0.7476	0.7218	0.8788	0.3390	0.3703	0.4542
C^2F-Net [46]	RGB	0.5858	0.1361	0.1080	0.5482	0.6774	0.9141	0.1957	0.2722	0.3446
	FPDE	0.5736	0.1189	0.1323	0.5137	0.6255	0.8430	0.1842	0.2446	0.3217
	GFCE	0.5941	0.1332	0.0911	0.5293	0.6798	0.8927	0.2114	0.2699	0.3545
	LatLRR	0.5640	0.1171	0.1412	0.5162	0.6208	0.8602	0.1972	0.2294	0.2973
OCENet [47]	RGB	0.6543	0.2738	0.0286	0.6216	0.7718	0.8353	0.2277	0.3033	0.3336
	FPDE	0.6869	0.3114	0.0234	0.6097	0.8065	0.8761	0.2567	0.3395	0.3676
	GFCE	0.6929	0.3214	0.0225	0.6128	0.8362	0.9177	0.2603	0.3609	0.4007
	LatLRR	0.6755	0.2971	0.0216	0.6209	0.8002	0.8820	0.2499	0.3235	0.3553
EAMNet [48]	RGB	0.5547	0.0732	0.1025	0.5112	0.5907	0.7768	0.1540	0.1541	0.2101
	FPDE	0.5112	0.0871	0.1995	0.5067	0.5447	0.7012	0.1886	0.1854	0.2388
	GFCE	0.5548	0.0647	0.1088	0.5615	0.5911	0.7416	0.1959	0.1706	0.2381
	LatLRR	0.5454	0.0830	0.1249	0.4889	0.5726	0.8004	0.1566	0.1814	0.2573
DGNet [49]	RGB	0.6208	0.1876	0.0602	0.5132	0.7000	0.8426	0.1730	0.2624	0.3061
	FPDE	0.6177	0.1929	0.0802	0.5553	0.7026	0.8682	0.2007	0.2846	0.3404
	GFCE	0.6340	0.2122	0.0714	0.5642	0.7005	0.8083	0.2137	0.2845	0.3282
	LatLRR	0.6297	0.2097	0.0723	0.5518	0.7043	0.8615	0.2080	0.2736	0.3200
HitNet [20]	RGB	0.6514	0.2781	0.0433	0.6130	0.7567	0.8255	0.2771	0.3278	0.3707
	FPDE	0.6749	0.3258	0.0228	0.6443	0.8020	0.8925	0.2916	0.3506	0.3828
	GFCE	0.6694	0.3248	0.0242	0.6316	0.8113	0.8480	0.2850	0.3642	0.4002
	LatLRR	0.6433	0.2765	0.0330	0.5508	0.6917	0.8864	0.2480	0.3053	0.3732
PCNet [19]	RGB	0.7328	0.4087	0.0158	0.8583	0.8447	0.8719	0.4212	0.4358	0.4460
	FPDE	0.7260	0.4137	0.0164	0.8281	0.8357	0.9112	0.4123	0.4433	0.4568
	GFCE	0.7385	0.4210	0.0161	0.8247	0.8140	0.8481	0.4227	0.4496	0.4607
	LatLRR	0.7006	0.3443	0.0170	0.8076	0.7365	0.8382	0.3838	0.3708	0.3830

6. Conclusions

This paper presented a practical UAV-based framework for early detection of camouflaged weeds in agricultural production environments by integrating RGB-NIR sensing, image fusion, and state-of-the-art camouflaged object detection (COD) models. Experiments on two real-world crop scenarios–a commercial banana plantation (Weeds-Banana) and a maize field (WeedsGalore)–show that RGB-NIR fusion consistently improves weed segmentation compared with RGB-only inputs, supporting its applicability in precision agriculture.

Results further indicate that the optimal fusion-model combination is crop- and scene-dependent: PCNet-LatLRR performs best in banana plantations, while BASNet-FPDE provides the most stable performance in maize fields. Overall, the proposed approach constitutes a scalable and transferable smart sensing solution for on-farm weed monitoring, enabling early intervention and supporting site-specific and sustainable weed management strategies. Future work will address broader validation, additional spectral bands, and computational optimization toward real-time UAV-based operation and closed-loop precision weed control. Finally, coupling segmentation outputs with selective spraying or mechanical weeding systems would move toward a fully automated weed management pipeline.

CRediT authorship contribution statement

Henry O. Velesaca: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization; **Andrea Mero:** Writing – original draft, Visualization, Investigation; **Hector Villegas:** Writing – original draft, Visualization, Resources, Data curation; **Angel D. Sappa:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Ethics statement

Not applicable: This manuscript does not include human or animal research.

Data availability

Data will be made available on request.

Declaration of competing interest

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

Acknowledgments

This work is supported in part by the Air Force Office of Scientific Research Under Award FA9550-24-1-0206; in part by Grant PID2021-128945NB-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU and Grant PID2024-162815NB-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU; in part by the ESPOL project “Advancing Camouflaged Object Detection with a cost-effective Cross-Spectral vision system (ACODCS)” (CIDIS-003-2024). The authors acknowledge the support of the Generalitat de Catalunya CERCA Program to CVC’s general activities.

References

[1] S. Ghazal, A. Munir, W.S. Qureshi, Computer vision in smart agriculture and precision farming: techniques and applications, Artif. Intell. Agric. 13 (2024) 64–83. https://doi.org/10.1016/j.aiia.2024.06.004

[2] A. Kamilaris, F. X. Prenafeta-Boldú, A review of the use of convolutional neural networks in agriculture, J. Agric. Sci. 156 (3) (2018) 312–322. https://doi.org/10.1017/S0021859618000436

[3] J.F.S. Gomes, F.R. Leta, Applications of computer vision techniques in the agriculture and food industry: a review, Eur. Food Res. Technol. 235 (2012) 989–1000.

[4] R. Punithavathi, A.D.C. Rani, K.R. Sughashini, Computer vision and deep learning-enabled weed detection model for precision agriculture, Comput. Syst. Sci. Eng. 44 (3) (2023) 2759–2774. https://doi.org/10.32604/csse.2023.027647

[5] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, L. Shao, Camouflaged object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[6] Y. Lv, J. Zhang, Y. Dai, A. Li, N. Barnes, D.-P. Fan, Toward deeper understanding of camouflaged object detection, IEEE Trans. Circuits Syst. Video Technol. 33 (7) (2023) 3462–3476.

[7] S. Lev-Yadun, A. Dafni, M.A. Flaishman, M. Inbar, I. Izhaki, G. Katzir, G. Ne’eman, Plant coloration undermines herbivorous insect camouflage, Bioessays. 26 (10) (2004) 1126–1130.

[8] S.M. Hussain, S.A. Zaidi, A. Hyder, S.M.A. Rizvi, M. Farhan, Enhanced camouflaged object detection for agricultural pest management: insights from unified benchmark dataset analysis, in: 2024 21st International Bhurban Conference on Applied Sciences and Technology (IBCAST), IEEE, 2024, pp. 273–280.

[9] A.K. Gautam, P. Preet, T.S. Rawat, P. Roy Chowdhury, L.K. Sinha, Detection of camouflaged targets in hyperspectral images, in: P.S. Sastry, J. CV, D.V.A. Raghavamurthy, S.S. Rao (Eds.), Advances in Small Satellite Technologies, Springer Singapore, Singapore, 2020, pp. 155–161.

[10] C.L. Hi, K.L. Lew, C. Zheng, T. Kurniawan, S.A. Babale, C.S. Lee, A review of camouflage object detection techniques, Int. J. Robot. Autom. Sci. 7 (1) (2025) 84–92. https://doi.org/10.33093/ijoras.2025.7.1.10

[11] M.O. Gani, S. Kuiry, A. Das, M. Nasipuri, N. Das, Multispectral object detection with deep learning, in: P. Dutta, J.K. Mandal, S. Mukhopadhyay (Eds.), Computational Intelligence in Communications and Business Analytics, Springer International Publishing, Cham, 2021, pp. 105–117.

[12] Z. Zhang, Q. Wang, Z. Li, Z. Zhou, X. Xie, H. Sun, C. Chen, M. Wang, X. Dong, Y. Wu, Z. He, B. Liu, X. Qian, J. Li, L. Wang, F. Xiu, J. Liu, W. Huang, A skin-beyond multifrequency camouflage system with self-adaptive discoloration and radar-infrared stealth, Chem. Eng. J. 494 (2024) 152867. https://doi.org/10.1016/j.cej.2024.152867

[13] S. Wang, D. Zeng, Y. Xu, G. Yang, F. Huang, L. Chen, Towards complex scenes: a deep learning-based camouflaged people detection method for snapshot multispectral images, Def. Technol. 34 (2024) 269–281. https://doi.org/10.1016/j.dt.2023.12.011

[14] H. Zhang, H. Xu, X. Tian, J. Jiang, J. Ma, Image fusion meets deep learning: a survey and perspective, Inf. Fusion. 76 (2021) 323–336.

[15] B.K. Shreyamsha Kumar, Image fusion based on pixel significance using cross bilateral filter, Signal Image Video Process. 9 (5) (2015) 1193–1204. https://doi.org/10.1007/s11760-013-0556-9

[16] Z. Zhou, M. Dong, X. Xie, Z. Gao, Fusion of infrared and visible images for night-vision context enhancement, Appl. Opt. 55 (23) (2016) 6480–6490. https://doi.org/10.1364/AO.55.006480

[17] H. Li, X.-J. Wu, Infrared and visible image fusion using latent low-rank representation, arXiv preprint arXiv:1804.08992 (2018).

[18] F. Xiao, S. Hu, Y. Shen, C. Fang, J. Huang, C. He, L. Tang, Z. Yang, X. Li, A survey of camouflaged object detection and beyond, 2024. arXiv:2408.14562

[19] J. Yang, Q. Wang, F. Zheng, P. Chen, A. Leonardis, D.-P. Fan, PlantCamo: plant camouflage detection, arXiv preprint arXiv:2410.17598 (2024).

[20] X. Hu, S. Wang, X. Qin, H. Dai, W. Ren, D. Luo, Y. Tai, L. Shao, High-resolution iterative feedback network for camouflaged object detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, 37, 2023, pp. 881–889.

[21] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, L. Shao, Pyramid vision transformer: a versatile backbone for dense prediction without convolutions, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 568–578.

[22] T. Chen, J. Xiao, X. Hu, G. Zhang, S. Wang, Boundary-guided network for camouflaged object detection, Knowl .- Based Syst. 248 (2022) 108901.

[23] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, P. Torr, Res2net: a new multi-scale backbone architecture, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2) (2019) 652–662.

[24] S. Haug, J. Ostermann, A crop/weed field image dataset for the evaluation of computer vision - based precision agriculture tasks, in: European Conference on Computer Vision, Springer, 2014, pp. 105–116.

[25] M. Krestenitis, E.K. Raptis, A.C. Kapoutsis, K. Ioannidis, E.B. Kosmatopoulos, S. Vrochidis, I. Kompatsiaris, Cofly-WeedDB: a UAV image dataset for weed detection and species identification, Data Brief. 45 (2022) 108575.

[26] E. Celikkan, T. Kunzmann, Y. Yeskaliyev, S. Itzerott, N. Klein, M. Herold, Weeds-Galore: a multispectral and multitemporal UAV-based dataset for crop and weed segmentation in agricultural maize fields, 2025. arXiv:2502.13103

[27] C. Lu, K. Gehring, S. Kopfinger, H. Bernhardt, M. Beck, S. Walther, T. Ebertseder, M. Minceva, Y. Hu, K. Yu, Weed instance segmentation from UAV orthomosaic images based on deep learning, Smart Agric. Technol. (2025) 100966.

[28] D.M. Gates, H.J. Keegan, J.C. Schleter, V.R. Weidner, Spectral properties of plants, Appl. Opt. 4 (1) (1965) 11–20. https://doi.org/10.1364/AO.4.000011

[29] E.B. Knipling, Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation, Rem. Sens. Environ. 1 (3) (1970) 155–159. https://doi.org/10.1016/S0034-4257(70)80021-9

[30] S.V. Ollinger, Sources of variability in canopy reflectance and the convergent properties of plants, New Phytol. (2011). https://doi.org/10.1111/j.1469-8137.2010.03536.x

[31] X. Zhang, P. Ye, G. Xiao, VIFB: a visible and infrared image fusion benchmark, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020.

[32] B.K. Shreyamsha Kumar, Image fusion based on pixel significance using cross bilateral filter, Signal Image Video Process. 9 (2015) 1193–1204.
[33] Z. Zhou, M. Dong, X. Xie, Z. Gao, Fusion of infrared and visible images for night-vision context enhancement, Appl. Opt. 55 (23) (2016) 6480–6490.
[34] Z. Zhou, B. Wang, S. Li, M. Dong, Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters, Inf. Fusion. 30 (2016) 15–26.
[35] D.P. Bavirisetti, R. Dhuli, Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform, IEEE Sens. J. 16 (1) (2015) 203–209.
[36] D.P. Bavirisetti, G. Xiao, J. Zhao, R. Dhuli, G. Liu, Multi-scale guided image and video fusion: a fast and efficient approach, Circuits, Syst. Signal Process. 38 (12) (2019) 5576–5605.
[37] V. Naidu, Image fusion technique using multi-resolution singular value decomposition, Def. Sci. J. 61 (5) (2011) 479.
[38] J. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, Inf. Fusion. 31 (2016) 100–109.
[39] D.P. Bavirisetti, R. Dhuli, Two-scale image fusion of visible and infrared images using saliency detection, Infrared Phys. Technol. 76 (2016) 52–64.
[40] D.P. Bavirisetti, G. Xiao, G. Liu, Multi-sensor image fusion based on fourth - order partial differential equations, in: 2017 20th International Conference on Information Fusion (Fusion), IEEE, 2017, pp. 1–9.
[41] J. Ma, Z. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, Infrared Phys. Technol. 82 (2017) 8–17.
[42] Y. Liu, X. Chen, J. Cheng, H. Peng, Z. Wang, Infrared and visible image fusion with convolutional neural networks, Int. J. Wavelets Multiresolut. Inf. Process. 16 (03) (2018) 1850018.
[43] Y. Zhang, L. Zhang, X. Bai, L. Zhang, Infrared and visual image fusion through infrared feature extraction and visual information preservation, Infrared Phys. Technol. 83 (2017) 227–237.
[44] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, M. Jagersand, BASNet: boundary-aware salient object detection, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
[45] D.-P. Fan, G.-P. Ji, M.-M. Cheng, L. Shao, Concealed object detection, IEEE Trans. Pattern Anal. Mach. Intell. 44 (10) (2021) 6024–6042.
[46] G. Chen, S.-J. Liu, Y.-J. Sun, G.-P. Ji, Y.-F. Wu, T. Zhou, Camouflaged object detection via context-aware cross-level fusion, IEEE Trans. Circuits Syst. Video Technol. 32 (10) (2022) 6981–6993.
[47] J. Liu, J. Zhang, N. Barnes, Modeling aleatoric uncertainty for camouflaged object detection, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 1445–1454.
[48] D. Sun, S. Jiang, L. Qi, Edge-aware mirror network for camouflaged object detection, in: 2023 Thinspace IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2023, pp. 2465–2470.
[49] G.-P. Ji, D.-P. Fan, Y.-C. Chou, D. Dai, A. Liniger, L. Van Gool, Deep gradient learning for efficient camouflaged object detection, Mach. Intell. Res. 20 (1) (2023) 92–108.
[50] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Conf. on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[51] M. Tan, Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.
[52] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, L. Shao, Pvt v2: improved baselines with pyramid vision transformer, Comput. Visual Media. 8 (3) (2022) 415–424.
[53] D.M. Bulanon, T.F. Burks, V. Alchanatis, Image fusion of visible and thermal images for fruit detection, Biosyst. Eng. 103 (1) (2009) 12–22.
[54] J.W. Roberts, J.A. Van Aardt, F.B. Ahmed, Assessment of image fusion procedures using entropy, image quality, and multispectral classification, J. Appl. Remote Sens. 2 (1) (2008) 023522.
[55] G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, Electron Lett. 38 (7) (2002) 313–315.
[56] P. Jagalingam, A.V. Hegde, A review of quality metrics for fused image, Aquat. Proc. 4 (2015) 133–142.
[57] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[58] G. Cui, H. Feng, Z. Xu, Q. Li, Y. Chen, Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition, Opt. Commun. 341 (2015) 199–209.
[59] B. Rajalingam, R. Priya, Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis, Int. J. Eng. Sci. Invent. 2 (Special issue) (2018) 52–60.
[60] Y.-J. Rao, In-fibre Bragg grating sensors, Meas. Sci. Technol. 8 (4) (1997) 355.
[61] C.S. Xydeas, V. Petrovic, Objective image fusion performance measure, Electron Lett. 36 (4) (2000) 308–309.
[62] Y. Chen, R.S. Blum, A new automated quality assessment algorithm for image fusion, Image Vis. Comput. 27 (10) (2009) 1421–1432.
[63] H. Chen, P.K. Varshney, A human perception inspired quality metric for image fusion based on regional information, Inf. Fusion. 8 (2) (2007) 193–207.
[64] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, A. Borji, Structure-measure: a new way to evaluate foreground maps, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4548–4557.
[65] R. Margolin, L. Zelnik-Manor, A. Tal, How to evaluate foreground maps?, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 248–255.
[66] F. Perazzi, P. Krähenbühl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: 2012 Thinspace IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 733–740.
[67] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, A. Borji, Enhanced-alignment measure for binary foreground map evaluation, arXiv preprint arXiv:1805.10421 (2018).
[68] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: 2009 Thinspace IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 1597–1604.