

Supplementary Information

CAM-Threshold-Guided Weakly Supervised Lesion Segmentation and Relative Lesion-Burden Stratification of Wheat Stripe Rust Using CF2-WRD and Light-UNet

Bojian Guo, Xin LI, Temirbaeva Nazgul Ismanovna, Asel Kartanova

Supplementary Table S1. Dataset split and role in the experimental workflow.

Split	Number of images	Role in workflow
Training	27,834	Classifier training, CAM generation, pseudo-label segmentation training
Validation	3,430	Model selection and pseudo-label agreement metrics
Classifier test / burden profiling	3,430	Classifier test accuracy and descriptive relative lesion-burden profiling
Total	34,694	87 image-level categories

Supplementary Table S2. Principal CF2-WRD components and feature dimensions.

Component	Output size	Function
MobileNetV2-like branch	$1,280 \times 7 \times 7$	Global and efficient hierarchical representation
EfficientNet-B0-like branch	$1,280 \times 7 \times 7$	Texture-sensitive MBConv and channel recalibration
1×1 projections	$512 \times 7 \times 7$ per branch	Align feature dimensions
Cosine-similarity gate	2 softmax weights	Residual branch gains: $(1+w)F$
ACFFM fused feature	$512 \times 7 \times 7$	Concat $\rightarrow$ channel attention $\rightarrow 1 \times 1$ + 3×3 fusion; CAM source
Classifier head	87 logits	Image-level supervision

Supplementary Table S3. Experimental implementation settings.

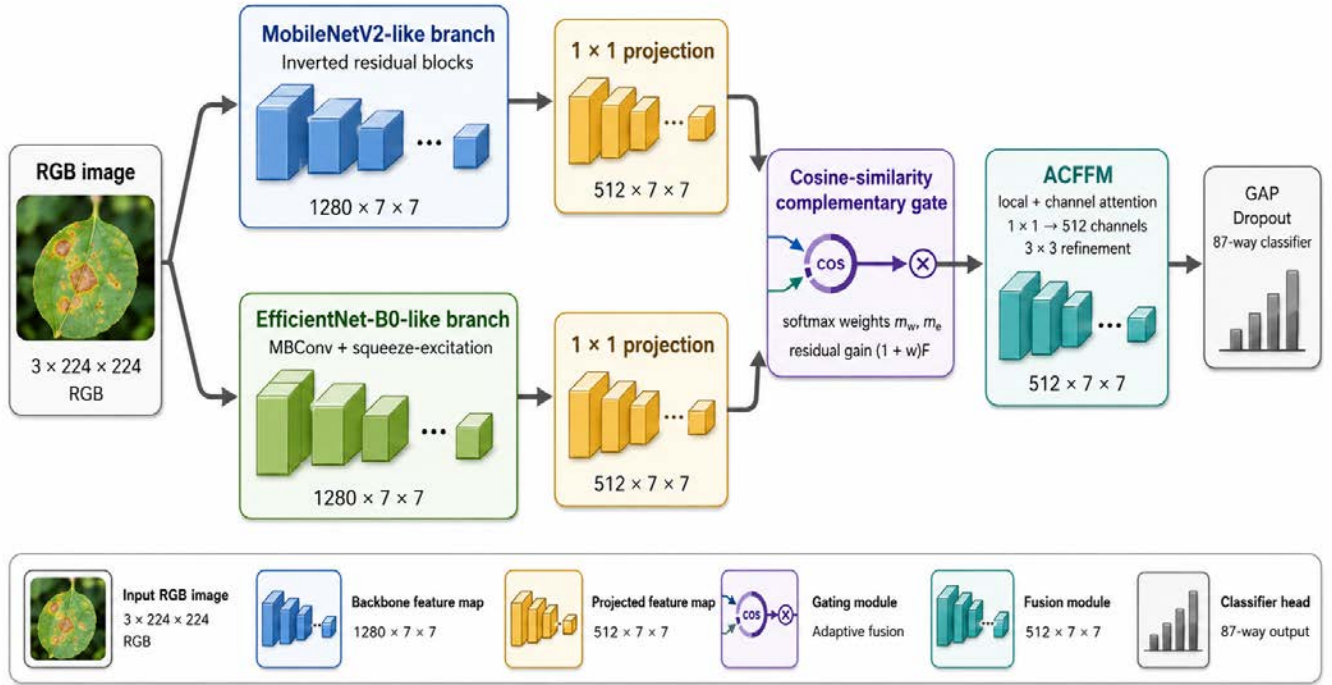
Item	Setting
Framework	PaddlePaddle 2.x; Python 3.10
Hardware	NVIDIA Tesla V100/T4; Intel Xeon; ~32 GB RAM
Input size	224×224 RGB
Segmenter	Light-UNet, base channels 32, 2 classes
Optimizer	AdamW, learning rate 1×10^{-4} , weight decay 1×10^{-4}
Batch size / epochs	16 / maximum 30; completed logs: 29, 30, 30, 28 epochs for $\tau=0.40, 0.45, 0.50, 0.55$
Loss	Cross-entropy + Dice loss
CAM feature	CF2-WRD fused_feat, $512 \times 7 \times 7$
CAM thresholds	0.40, 0.45, 0.50, 0.55
Lesion-probability threshold for relative-burden profiling	0.80
Dice smoothing constant	$\varepsilon = 1 \times 10^{-6}$
Grad-CAM target	Classifier-predicted class (argmax softmax)

Connected-component filter	8-connectivity; retain components ≥ 80 pixels
Training-mask sample filter	Retain masks with positive area ≥ 50 pixels
DataLoader behavior	Train: shuffle=True, drop_last=True; validation: shuffle=False; checkpoint selected by maximum validation mIoU
CF2-WRD training epochs	30; best validation checkpoint selected at epoch 29
CF2-WRD optimizer	AdamW; learning rate 1×10^{-4} ; weight decay 1×10^{-4} ; fixed learning rate (no scheduler in recovered formal-training loop)
CF2-WRD classification loss	Unweighted cross-entropy
CF2-WRD batch / loader	Batch size 16; train: shuffle=True, drop_last=True; validation/test: shuffle=False
CF2-WRD train augmentation	Resize 224 $\times$ 224; random horizontal flip; random rotation $\pm 15^\circ$; tensor conversion; normalization mean [0.485, 0.456, 0.406], std [0.229, 0.224, 0.225]
CF2-WRD validation/test transform	Resize 224 $\times$ 224; tensor conversion; same normalization; no random augmentation
CF2-WRD initialization / pretraining	Newly instantiated 87-class model in recovered formal-training cell; no external pretrained state loaded before optimization
CF2-WRD checkpoint rule	Save last checkpoint every epoch; update cf2_wrd_best.pdparams when validation accuracy improves
CF2-WRD random seed	No explicit seed-setting call found in recovered formal-training notebook

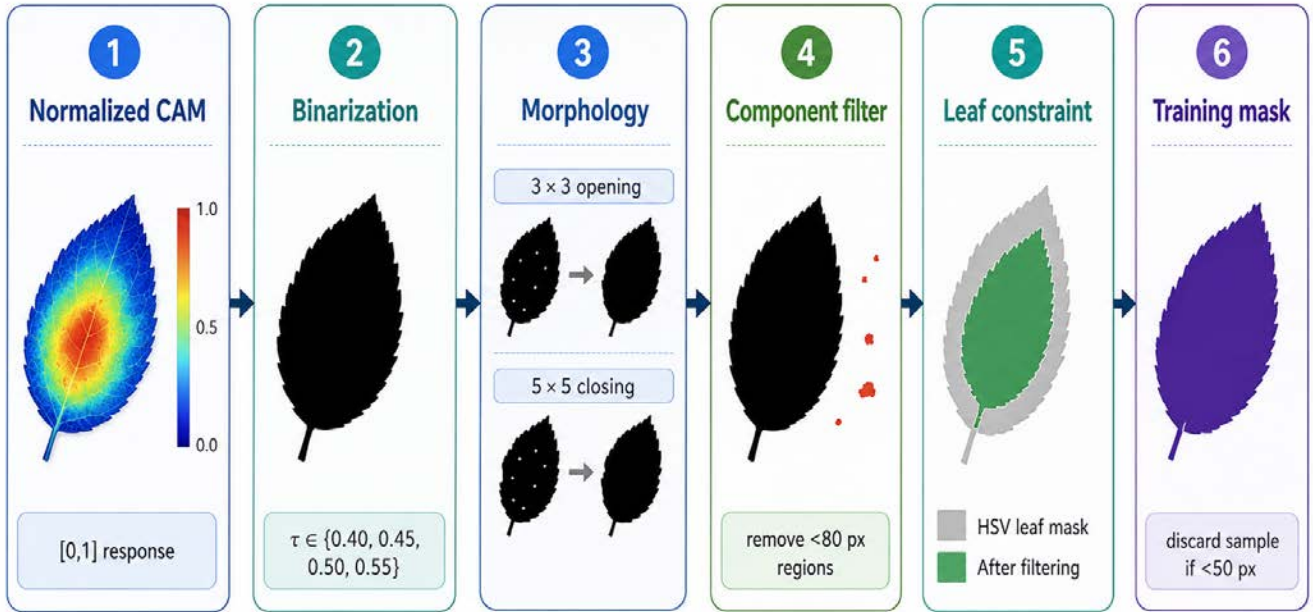
The Light-UNet training settings, Grad-CAM target selection, pseudo-label filters, Dice smoothing constant, and segmentation checkpoint-selection rule shown here were verified against the supplied ablation-training code and recovered epoch-level logs. The CF2-WRD classifier-training configuration was additionally recovered from the archived Baidu AI Studio notebook and classification train_log.csv, including the optimizer, loss, augmentation, epoch count, checkpoint-selection rule, and selected validation epoch.

Supplementary Table S4. CF2-WRD image-level classification performance.

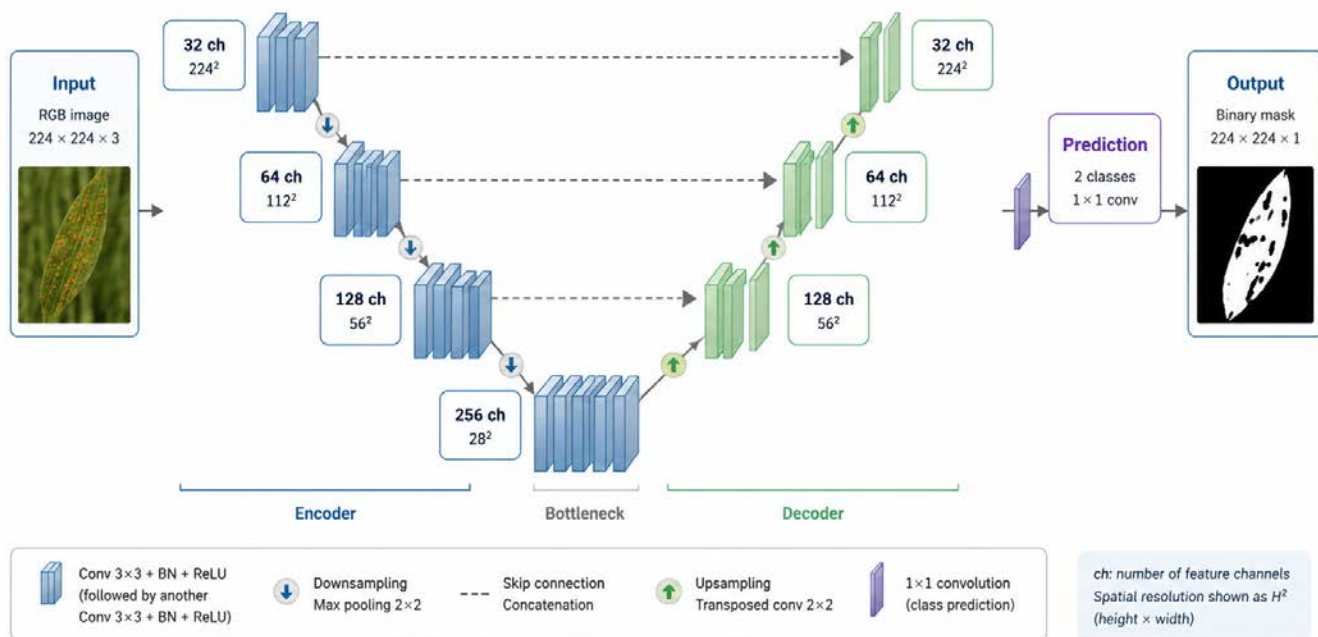
Model	Best validation accuracy	Test accuracy	Selected epoch
CF2-WRD	0.8776	0.8711	29 of 30



Supplementary Fig. S1. CF2-WRD architecture and fused feature representation used for Grad-CAM extraction. The branch-similarity descriptor is the inner product of L2-normalized pooled feature vectors (cosine similarity). A two-layer softmax gate produces branch weights that are applied as residual gains, $(1 + w_m)$ and $(1 + w_e)$, before concatenation, channel attention, and fusion.



Supplementary Fig. S2. Pseudo-label generation and filtering workflow.



Supplementary Fig. S3. Light-UNet architecture used for binary lesion segmentation.

Supplementary Table S5. Sensitivity of predicted relative lesion burden to the recorded segmentation-probability endpoints on a 300-image profiling subset.

Probability threshold	Mean S (%)	Median S (%)	Samples >30%
0.50	55.02	52.90	98.33%
0.80	29.63	31.39	53.00%

The finalized summary record retained complete descriptive statistics for the 0.50 and 0.80 endpoints. At 0.80, no sample exceeded 80% predicted burden.

Supplementary Table S6. Qualitative comparison with representative wheat stripe rust studies.

Study	Task	Supervision	Primary output	Relevance to this work
Mi et al. [13]	Severity grading	Image-level severity classes	Discrete grade	Attention-based stripe rust grading
Liu et al. [16]	Oriented detection	Bounding boxes	Lesion detections	Captures elongated rust geometry
Jiang et al. [18,19]	Severity assessment	Processed lesion percentages	Severity classes/models	Area-based severity rationale
StripeRust-Pocket [20]	Mobile severity quantification	Model-assisted pixel labels	Absolute severity estimate	Practical deployment benchmark
Present study	Weakly supervised segmentation	Image labels → CAM pseudo-labels	Lesion mask + relative burden	Low-cost classification-to-segmentation transfer

Supplementary Table S7. Consolidated experimental values reported in the manuscript.

Module	Reported value(s)	Use in manuscript
CF2-WRD classification	Validation accuracy 0.8776; test accuracy 0.8711	Supplementary Table S4 and Abstract
CAM $\tau=0.40$	Completed epochs 30; selected epoch 14; train mIoU 0.7004; valid mIoU 0.6584; Dice 0.7101; lesion IoU 0.5555; pixel accuracy 0.8238	Main Table 1
CAM $\tau=0.45$	Completed epochs 30; selected epoch 17; train mIoU 0.6963; valid mIoU 0.6490; Dice 0.6739; lesion IoU 0.5136; pixel accuracy 0.8306	Main Table 1
CAM $\tau=0.50$	Completed epochs 30; selected epoch 20; train mIoU 0.6904; valid mIoU 0.6415; Dice 0.6480; lesion IoU 0.4855; pixel accuracy 0.8349	Main Table 1
CAM $\tau=0.55$	Completed epochs 30; selected epoch 18; train mIoU 0.6756; valid mIoU 0.6312; Dice 0.6206; lesion IoU 0.4557; pixel accuracy 0.8379	Main Table 1
Lesion-probability sensitivity	$\tau=0.50$: mean 55.02%, median 52.90%, >30%=98.33%; $\tau=0.80$: mean 29.63%, median 31.39%, >30%=53%, >80%=0%	Supplementary Table S5
Relative lesion-burden quartiles	15.7567%, 27.7772%, 37.6485%; counts 858/857/857/858	Main Table 2 and Figs. 5–6
Verified implementation details	Dice smooth $\epsilon=1 \times 10^{-6}$; Grad-CAM target=predicted class; Light-UNet AdamW lr= 1×10^{-4} , weight decay= 1×10^{-4} , batch size=16; maximum 30 epochs; checkpoint selected by highest validation mIoU; CF2-WRD AdamW lr= 1×10^{-4} , weight decay= 1×10^{-4} , batch size=16, 30 epochs, unweighted cross-entropy, best checkpoint selected by validation accuracy (epoch 29)	Methods and Supplementary Table S3
Dataset / split verification	34,694 total images; 27,834 train; 3,430 validation; 3,430 test; 87 class codes; no exact image-path overlap among split lists	Methods, Supplementary Tables S1 and S8
Pseudo-label sample accounting	Retained at $\tau=0.40/0.45/0.50/0.55$: 27,803 / 27,774 / 27,744 / 27,695; excluded <50-pixel masks: 31 / 60 / 90 / 139; empty masks: 20 / 28 / 45 / 78	Methods and Supplementary Table S9

CAM-threshold results in this table were re-extracted from the recovered epoch-level training logs using maximum validation mIoU as the single checkpoint-selection rule. The recovered 27,834-image training list is consistent with the threshold-specific pseudo-label filtering records, allowing retained and excluded sample counts to be reported separately in Supplementary Table S9.

Supplementary Table S8. Recovered 87-class dataset-code mapping and per-split image counts.

Class codes were recovered from class_to_idx.json and idx_to_class.json; counts were reconstructed from the recovered train_list.txt, valid_list.txt, and test_list.txt files. All 87 codes occurred in all three partitions. The three list files contained unique image paths and no exact path overlap between partitions. The recovered materials did not include a separate semantic codebook for the alphanumeric labels, so the dataset codes are reported exactly as stored.

Class ID	Dataset code	Training	Validation	Test	Total
0	1a	312	39	39	390
1	1b	299	37	37	373
2	1d	337	42	42	421
3	1l	313	38	38	389
4	1r	316	39	39	394
5	1u	335	41	41	417
6	2a	290	36	36	362
7	2b	305	38	38	381
8	3a	305	38	38	381
9	3b	317	39	39	395
10	4a	297	36	36	369

Class ID	Dataset code	Training	Validation	Test	Total
11	4b	360	44	44	448
12	5a	320	40	40	400
13	5b	318	39	39	396
14	5d	345	43	43	431
15	5l	360	44	44	448
16	5r	338	42	42	422
17	5u	359	44	44	447
18	6a	316	39	39	394
19	6b	305	37	37	379
20	7a	307	38	38	383
21	7b	311	38	38	387
22	8a	296	37	37	370
23	8b	301	37	37	375
24	9a	302	37	37	376
25	9b	290	36	36	362
26	9d	378	47	47	472
27	9l	347	43	43	433
28	9r	377	47	47	471
29	9u	360	45	45	450
30	10a	315	39	39	393
31	10b	289	36	36	361
32	11a	296	36	36	368
33	11b	287	35	35	357
34	12a	300	37	37	374
35	12b	294	36	36	366
36	13a	291	36	36	363
37	13b	293	36	36	365
38	13d	370	46	46	462
39	13r	353	44	44	441
40	13u	341	42	42	425
41	14a	299	37	37	373
42	14b	297	36	36	369
43	15a	292	36	36	364
44	15b	297	36	36	369
45	16a	297	37	37	371
46	16b	304	37	37	378
47	17a	275	34	34	343
48	17b	389	48	48	485
49	17d	350	43	43	436
50	17l	380	47	47	474
51	17r	352	43	43	438
52	17u	354	44	44	442
53	18a	292	36	36	364
54	18b	297	36	36	369
55	19a	288	36	36	360
56	19b	317	39	39	395
57	20a	285	35	35	355
58	20b	314	39	39	392
59	21a	288	35	35	358
60	21b	384	47	47	478
61	21d	359	44	44	447
62	21l	354	44	44	442
63	21r	360	44	44	448

Class ID	Dataset code	Training	Validation	Test	Total
64	21u	377	46	46	469
65	23a	288	35	35	358
66	23b	297	36	36	369
67	24a	408	50	50	508
68	24b	317	39	39	395
69	25a	283	35	35	353
70	25b	308	38	38	384
71	25d	352	43	43	438
72	25u	358	44	44	446
73	26a	312	39	39	390
74	26b	337	42	42	421
75	27a	287	35	35	357
76	27b	296	37	37	370
77	28a	267	33	33	333
78	28b	307	38	38	383
79	29a	310	38	38	386
80	29b	305	37	37	379
81	30a	270	33	33	336
82	30b	307	38	38	383
83	31a	310	38	38	386
84	31b	306	38	38	382
85	32a	351	43	43	437
86	32b	312	39	39	390
Total	—	27,834	3,430	3,430	34,694

Supplementary Table S9. CAM-threshold-specific pseudo-label sample accounting for the recovered 27,834-image training pool.

Masks with fewer than 50 positive pixels were excluded from Light-UNet training. Empty masks are a subset of the <50-pixel exclusions and are reported separately for traceability.

CAM τ	Initial train images	Retained masks	Excluded (<50 px)	Empty masks
0.40	27,834	27,803	31	20
0.45	27,834	27,774	60	28
0.50	27,834	27,744	90	45
0.55	27,834	27,695	139	78