

Supplementary Material for “DC4Former: Orientation-Stable UAV Disaster Image Segmentation via Diagonal-Complemented C4 Consistency”

Interpolation Round-Trip Diagnostic

This diagnostic measures image-level reconstruction error after applying a rotation and its inverse to the RGB test images. Right-angle rotations use exact index permutations, whereas diagonal rotations use bilinear interpolation. The resulting measurements characterize resampling differences between grid-preserving C_4 transformations and interpolation-based diagonal transformations, without evaluating model accuracy or arbitrary-angle equivariance.

All test images are resized with the same aspect-ratio-preserving rule used in the evaluation protocol before the round trip. For right-angle rotations, the image is rotated and inverse-rotated by index permutation. For diagonal rotations, the image is first rotated on an enlarged canvas with bilinear interpolation and then inversely warped back to the original resized canvas. Reconstruction metrics are computed only on the valid non-filled region, so the reported error mainly reflects interpolation and resampling effects rather than black padding.

Table S1. Image-level rotation–inverse-rotation round-trip diagnostic over the eight principal angles on the FloodNet and RescueNet test sets. MAE_{norm} is computed on normalized RGB intensities. PSNR is reported in dB. Gradient retention measures the ratio of image-gradient magnitude retained after the round trip. Valid ratio denotes the non-filled image area after rotation. Values are mean $\pm$ standard deviation over test images.

Dataset	Angle	n	$\text{MAE}_{\text{norm}} \downarrow$	PSNR $\uparrow$	Gradient retention $\uparrow$	Valid ratio
FloodNet	0°	448	0.000000 $\pm$ 0.000000	∞	1.0000 $\pm$ 0.0000	1.000000 $\pm$ 0.000000
FloodNet	45°	448	0.022769 $\pm$ 0.007582	29.65 $\pm$ 2.74	0.7306 $\pm$ 0.0393	0.983894 $\pm$ 0.000275
FloodNet	90°	448	0.000000 $\pm$ 0.000000	∞	1.0000 $\pm$ 0.0000	1.000000 $\pm$ 0.000000
FloodNet	135°	448	0.022769 $\pm$ 0.007582	29.65 $\pm$ 2.74	0.7306 $\pm$ 0.0393	0.983909 $\pm$ 0.000272
FloodNet	180°	448	0.000000 $\pm$ 0.000000	∞	1.0000 $\pm$ 0.0000	1.000000 $\pm$ 0.000000
FloodNet	225°	448	0.022769 $\pm$ 0.007582	29.65 $\pm$ 2.74	0.7306 $\pm$ 0.0393	0.983894 $\pm$ 0.000275
FloodNet	270°	448	0.000000 $\pm$ 0.000000	∞	1.0000 $\pm$ 0.0000	1.000000 $\pm$ 0.000000
FloodNet	315°	448	0.022769 $\pm$ 0.007582	29.65 $\pm$ 2.74	0.7306 $\pm$ 0.0393	0.983878 $\pm$ 0.000279
RescueNet	0°	450	0.000000 $\pm$ 0.000000	∞	1.0000 $\pm$ 0.0000	1.000000 $\pm$ 0.000000
RescueNet	45°	450	0.038587 $\pm$ 0.015497	25.01 $\pm$ 3.50	0.7619 $\pm$ 0.0272	0.983780 $\pm$ 0.000000
RescueNet	90°	450	0.000000 $\pm$ 0.000000	∞	1.0000 $\pm$ 0.0000	1.000000 $\pm$ 0.000000
RescueNet	135°	450	0.038587 $\pm$ 0.015497	25.01 $\pm$ 3.50	0.7619 $\pm$ 0.0272	0.983797 $\pm$ 0.000000
RescueNet	180°	450	0.000000 $\pm$ 0.000000	∞	1.0000 $\pm$ 0.0000	1.000000 $\pm$ 0.000000
RescueNet	225°	450	0.038587 $\pm$ 0.015497	25.01 $\pm$ 3.50	0.7619 $\pm$ 0.0272	0.983780 $\pm$ 0.000000
RescueNet	270°	450	0.000000 $\pm$ 0.000000	∞	1.0000 $\pm$ 0.0000	1.000000 $\pm$ 0.000000
RescueNet	315°	450	0.038587 $\pm$ 0.015497	25.01 $\pm$ 3.50	0.7619 $\pm$ 0.0272	0.983763 $\pm$ 0.000000

The right-angle round trips recover the original image exactly on the discrete grid. The diagonal round trips leave measurable reconstruction errors because bilinear interpolation resamples the image.

Sensitivity to the p_{SDA} Probability

The main comparative experiments use $p_{\text{SDA}} = 0.3$ as the shared probability held fixed for FloodNet and RescueNet. Table S2 reports seed-0 sensitivity over $p_{\text{SDA}} \in \{0.1, 0.3, 0.5\}$. The two closer settings, 0.1 and 0.3, are also evaluated with three seeds to estimate cross-seed variability; 0.5 remains a seed-0 reference because the coarse comparison provides no consistent cross-dataset advantage.

Table S2. Sensitivity of DC4Former to the stochastic diagonal augmentation probability p_{SDA} . Panel (a) reports the seed-0 diagnostic comparison over $p_{\text{SDA}} \in \{0.1, 0.3, 0.5\}$. Panel (b) reports the follow-up three-seed comparison between the two lower-probability settings.

(a) Seed-0 diagnostic comparison.										
Dataset	p_{SDA}	0°	Orth.	Diag.	All-8	Worst-8	Std.-8	Mean-24	Worst-24	Std.-24
FloodNet	0.1	77.68	77.68	78.02	77.85	77.67	0.20	77.73	77.23	0.26
FloodNet	0.3	77.22	77.15	77.37	77.26	76.97	0.30	77.07	76.13	0.45
FloodNet	0.5	75.90	76.00	76.77	76.39	75.90	0.46	76.24	75.40	0.56
RescueNet	0.1	67.24	67.34	67.99	67.66	67.24	0.41	65.30	63.26	1.76
RescueNet	0.3	68.10	68.02	68.06	68.04	67.66	0.28	65.27	62.71	2.03
RescueNet	0.5	66.86	66.86	68.35	67.60	66.85	0.77	65.45	63.17	1.67

(b) Follow-up three-seed comparison.											
Dataset	p_{SDA}	Seeds	0°	Orth.	Diag.	All-8	Worst-8	Std.-8	Mean-24	Worst-24	Std.-24
FloodNet	0.1	3	77.07 ± 0.80	77.07 ± 0.82	77.15 ± 0.82	77.11 ± 0.80	76.76 ± 0.82	0.23 ± 0.10	77.14 ± 0.73	76.51 ± 0.71	0.33 ± 0.07
FloodNet	0.3	3	76.98 ± 0.44	77.00 ± 0.38	77.39 ± 0.02	77.20 ± 0.19	76.83 ± 0.30	0.34 ± 0.11	77.19 ± 0.25	76.50 ± 0.47	0.39 ± 0.10
RescueNet	0.1	3	67.40 ± 0.71	67.41 ± 0.60	67.71 ± 0.55	67.56 ± 0.56	67.25 ± 0.47	0.28 ± 0.13	65.30 ± 0.27	63.33 ± 0.74	1.68 ± 0.59
RescueNet	0.3	3	67.82 ± 0.31	67.81 ± 0.26	68.21 ± 0.18	68.01 ± 0.04	67.66 ± 0.18	0.33 ± 0.11	65.40 ± 0.69	63.04 ± 1.51	1.91 ± 0.50

The response to p_{SDA} is non-monotonic and dataset-dependent, and 0.5 provides no consistent cross-dataset advantage in the seed-0 comparison. Across three seeds, 0.1 and 0.3 remain close on FloodNet, whereas 0.3 gives higher values for most RescueNet accuracy summaries. These results document sensitivity to the shared setting without establishing a universally optimal probability.

C4Former-Mean-SDA Diagnostic

This seed-0 control compares mean and constrained fusion with and without SDA in the full C4 operator architecture. C4Former-Mean-SDA retains uniform mean fusion and adds SDA, allowing comparison with C4Former-Mean and DC4Former.

Table S3. Seed-0 C4Former-Mean-SDA diagnostic. The comparison provides a controlled contrast between mean and constrained fusion with and without SDA and reports both eight-angle and 24-angle rotation summaries.

Dataset	Model	Fusion	SDA	0°	Orth.	Diag.	All-8	Worst-8	Std.-8	Mean-24	Worst-24	Std.-24
FloodNet	C4Former-Mean	Mean	No	76.25	76.24	72.08	74.16	71.72	2.08	73.59	71.72	1.43
FloodNet	C4Former	Constrained	No	76.53	76.53	73.09	74.81	72.97	1.72	74.33	72.97	1.17
FloodNet	C4Former-Mean-SDA	Mean	Yes	77.03	77.02	77.17	77.09	77.00	0.10	77.08	76.62	0.33
FloodNet	DC4Former	Constrained	Yes	77.22	77.15	77.37	77.26	76.97	0.30	77.07	76.13	0.45
RescueNet	C4Former-Mean	Mean	No	68.22	68.25	52.68	60.47	52.21	7.79	58.01	52.21	5.21
RescueNet	C4Former	Constrained	No	68.13	68.09	52.64	60.37	52.53	7.72	57.72	52.53	5.27
RescueNet	C4Former-Mean-SDA	Mean	Yes	65.37	65.50	66.38	65.94	65.37	0.45	63.95	61.60	1.53
RescueNet	DC4Former	Constrained	Yes	68.10	68.02	68.06	68.04	67.66	0.28	65.27	62.71	2.03

On RescueNet, adding SDA to C4Former-Mean increases the diagonal mean from 52.68% to 66.38% and reduces Std.-8 from 7.79 to 0.45. C4Former-Mean-SDA remains below DC4Former in All-8, Worst-8, Mean-24, and Worst-24, while its Std.-24 is lower in this seed-0 comparison.

Full C4 Consistency Diagnostic

This table expands the compact C4 consistency diagnostic reported in the main manuscript. The diagnostic uses the existing LRFormer, C4Former, and DC4Former checkpoints without retraining. For each checkpoint, the 0° , 90° , 180° , and 270° versions of the same test image are evaluated using the same whole-image inference path and the same fixed-canvas rotation protocol as the rotation-wise mIoU evaluation. Forward hooks record the encoder stage outputs and raw logits; rotated features, logits, and predictions are inverse-rotated to the 0° coordinate system and compared on their common centered spatial support.

Values are averaged over 90° , 180° , and 270° rotations and three random seeds. Stage-1 to Stage-4 and Logit L2 are relative L2 errors; lower values indicate stronger consistency. Prediction-consistency mIoU is the foreground mIoU between aligned predictions, and pixel disagreement is reported in percent.

Across both datasets, C4Former and DC4Former show lower aligned feature and logit discrepancies and lower pixel disagreement than LRFormer, together with higher prediction-consistency mIoU.

Table S4. Full feature-, logit-, and prediction-level C4 consistency diagnostic.

Dataset	Model	Stage-1 L2	Stage-2 L2	Stage-3 L2	Stage-4 L2	Logit L2	Pred. Cons. mIoU	Pixel Disagree.
FloodNet	LRFormer	1.054 ± 0.009	0.860 ± 0.010	0.771 ± 0.012	0.588 ± 0.008	0.274 ± 0.002	81.51 ± 0.32	7.947 ± 0.179
FloodNet	C4Former	0.637 ± 0.005	0.545 ± 0.012	0.389 ± 0.008	0.236 ± 0.001	0.151 ± 0.003	96.14 ± 0.05	0.971 ± 0.008
FloodNet	DC4Former	0.638 ± 0.013	0.549 ± 0.006	0.396 ± 0.002	0.234 ± 0.006	0.146 ± 0.002	96.26 ± 0.02	0.958 ± 0.047
RescueNet	LRFormer	1.132 ± 0.002	0.947 ± 0.005	0.763 ± 0.018	0.587 ± 0.036	0.246 ± 0.012	68.84 ± 1.92	9.227 ± 0.820
RescueNet	C4Former	0.754 ± 0.008	0.615 ± 0.002	0.404 ± 0.006	0.237 ± 0.001	0.130 ± 0.003	94.74 ± 0.12	1.227 ± 0.110
RescueNet	DC4Former	0.755 ± 0.004	0.615 ± 0.006	0.404 ± 0.003	0.236 ± 0.001	0.127 ± 0.004	94.73 ± 0.39	1.221 ± 0.129

Sensitivity to the Constrained Fusion Bound

The main comparative experiments use $\alpha_{\max} = 0.25$ as the shared structural cap held fixed for both datasets. This seed-0 sensitivity analysis uses the full C4Former architecture and the fixed-canvas rotation-wise evaluation protocol, with only $\alpha_{\max}$ varied.

Table S5. Seed-0 sensitivity analysis of the constrained fusion bound $\alpha_{\max}$.

$\alpha_{\max}$	FloodNet			RescueNet		
	All-8	Worst-8	Std.-8	All-8	Worst-8	Std.-8
0.00	74.16	71.72	2.08	60.47	52.21	7.79
0.10	75.19	73.54	1.54	60.58	53.20	7.04
0.25	74.81	72.97	1.72	60.37	52.53	7.72
0.50	74.29	71.85	2.06	60.74	52.95	7.42
0.75	75.20	73.47	1.50	60.50	52.72	7.46
1.00	75.37	74.14	1.02	60.35	52.15	7.85

The response is non-monotonic and differs between datasets. Larger values improve several FloodNet seed-0 metrics, whereas the same trend is not reproduced on RescueNet. These results document sensitivity around the shared $\alpha_{\max} = 0.25$ setting without establishing a universally optimal value.

Convergence Diagnostic for the Constrained Fusion Bound

This diagnostic examines whether larger constrained-fusion bounds cause optimization instability. Using the same seed-0 runs as Table S5, we extract the training logs for $\alpha_{\max} = 0.25$,

0.5, and 1.0 on FloodNet and RescueNet. The training loss is smoothed for visualization, and validation mIoU is plotted at each validation interval. All six runs completed the prescribed schedules without NaNs, loss explosion, or numerical divergence.

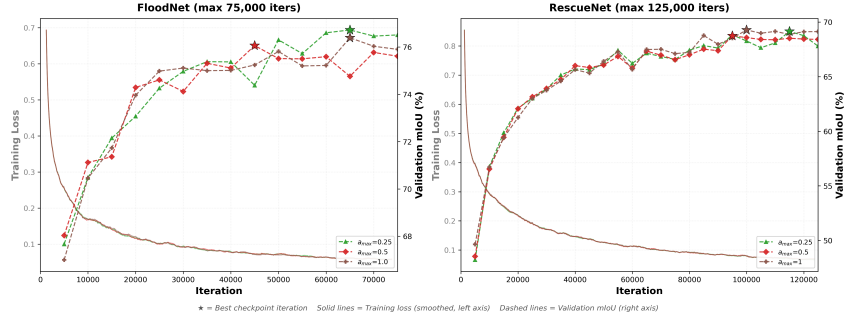


Figure S1. Training-loss and validation-mIoU trajectories for $\alpha_{\max} = 0.25, 0.5$, and 1.0 on FloodNet and RescueNet. Solid lines denote smoothed training loss on the left axis, dashed lines denote validation mIoU on the right axis, and stars mark the best validation checkpoint iteration.

Table S6. Best validation checkpoint iteration for the constrained-fusion bound diagnostic.

Dataset	$\alpha_{\max}$	Best checkpoint iteration	Best validation mIoU (%)
FloodNet	0.25	65,000	76.72
FloodNet	0.5	45,000	76.06
FloodNet	1.0	65,000	76.39
RescueNet	0.25	115,000	69.14
RescueNet	0.5	95,000	68.74
RescueNet	1.0	100,000	69.28