Supplemental Material

I. Influence Analysis on Stage Number in Unfolding

Table I shows the PSNR results using varied number of unfolding stages in UDPNet. The performance of UDPNet is improved with the increasing of stage number of unfolding and saturates beyond four stages.

TABLE I
RESULTS OF VARIED NUMBER OF UNFOLDING IN UDPNET, TESTED ON REAL-MFF.

Depth	1	2	3	4	5
PSNR(dB)	39.72	39.96	40.05	40.07	40.07

II. Regarding retraining of baselines

For fair comparison, given the variability of training data in unsupervised methods, we retrained baseline models using the same data as ours, following the original work's settings and implementations. The detail results are shown in Table I in the main paper. We also report the performance difference between re-trained and original models on Real-MFF in Table II below. It can be observed that several models (FusionDN, DIFNet, U2Fusion, SwinFusion) outperform their original versions after our retraining.

TABLE II
PERFORMANCE DIFFERENCE BETWEEN RE-TRAINED AND ORIGINAL MODELS ON REAL-MFF. **POSITIVE VALUES** INDICATE **BETTER** PERFORMANCE FROM **RETRAINING**.

Metric	FusionDN	MFFGAN	DIFNet	U2Fusion	SwinFusion	SMFuse	MUFusion
PSNR	+5.7054	-1.1513	+8.9823	+2.6661	+2.4796	-0.0348	+9.4120
SSIM	+0.0873	+0.0707	+0.0711	+0.0634	+0.0020	+0.0001	+0.0281
LPIPS	+0.0431	+0.0330	+0.0601	+0.0294	+0.0089	+0.0002	+0.0382

TABLE III
QUANTITATIVE COMPARISON OF END-TO-END UNSUPERVISED MFIF METHODS ON MFI-WHU AND REAL-MFF DATASETS IN NO-REFERENCE METRICS.

MFI-WHU	NMI↑	Q_{NCIE}↑	Q_G↑	Q_M↑	Q_P↑	Q_C↑	Q_Y↑	Q_{CB}↑	MI↑	VIF↑	**ARank**↓
PMGI	0.8435	0.8244	0.5266	0.3271	0.6598	0.7838	0.8129	0.7057	6.1471	1.1092	7.70
FusionDN	0.8285	0.8238	0.5320	0.3374	0.6555	0.7675	0.8112	0.6384	6.0472	1.0956	8.40
MFF-GAN	0.8079	0.8229	0.6000	0.5808	0.6885	0.7872	0.8762	0.6187	5.8401	1.0980	7.60
DIFNet	0.8707	0.8253	0.5752	0.3473	0.6968	0.8366	0.8774	0.7451	6.2995	1.1223	5.70
U2Fusion	0.8706	0.8251	0.6157	0.4052	0.7271	0.8233	0.8877	0.7061	6.2414	1.1413	5.60
SwinFusion	0.9111	0.8281	0.6850	0.8632	0.7486	0.8453	0.9507	0.7544	6.6364	1.2662	3.50
SMFuse	**1.2143**	**0.8485**	0.7362	**2.5310**	0.7515	0.8669	0.9832	0.8202	**8.8507**	1.3743	1.60
MUFusion	0.9231	0.8285	0.6658	0.7825	0.7407	0.8507	0.9465	0.7626	6.7051	1.2524	3.50
UDPNet	1.2120	0.8481	**0.7408**	2.5168	**0.7595**	**0.8707**	**0.9895**	**0.8253**	8.8330	**1.3898**	**1.40**
Real-MFF	**NMI↑**	**Q_{NCIE}↑**	**Q_G↑**	**Q_M↑**	**Q_P↑**	**Q_C↑**	**Q_Y↑**	**Q_{CB}↑**	**MI↑**	**VIF↑**	**ARank↓**
PMGI	1.0222	0.8329	0.6078	0.5830	0.8078	0.7143	0.7280	0.5653	7.1378	1.4485	7.80
FusionDN	0.9708	0.8318	0.6219	0.5705	0.8072	0.8166	0.8402	0.6830	6.9466	1.4467	7.70
MFF-GAN	0.9210	0.8291	0.5883	0.5289	0.8257	0.7826	0.8178	0.6616	6.5189	1.4064	8.50
DIFNet	1.0806	0.8360	0.6994	0.6666	0.8750	0.9000	0.9160	0.7713	7.5874	1.5098	3.80
U2Fusion	1.0787	0.8352	0.6803	0.6504	0.8840	0.8631	0.8853	0.7329	7.4694	1.4526	4.90
SwinFusion	1.0223	0.8335	0.6628	0.8416	0.8662	0.8750	0.9162	0.7740	7.2002	1.4947	5.10
SMFuse	1.3017	0.8504	0.7432	1.9674	0.9010	0.9035	0.9560	0.8235	9.1723	1.5626	2.00
MUFusion	1.0536	0.8351	0.6709	0.8558	0.8788	0.8816	0.9245	0.7970	7.4074	1.4870	4.20
UDPNet	**1.3106**	**0.8506**	**0.7669**	**2.1700**	**0.9281**	**0.9157**	**0.9752**	**0.8426**	**9.2261**	**1.6112**	**1.00**

III. No-reference metrics results on MFI-WHU and Real-MFF datasets

In Table 1 in our main paper, we report the results in terms of full-reference metrics on MFI-WHU and Real-MFF datasets. To better analyze the performance of our method, Table III below shows the results in terms of the ten non-reference metrics used in Table II of main paper. We can see that our method still overall performs better than other methods in terms of these metrices, achieving the highest ARank.

IV. Robustness or limitation analysis in cases where focus-defocus boundaries are highly complex.

Figure 1 shows two examples from the MFFW dataset to analyze the performance in dealing with highly complex focus-defocus boundaries. As shown in the first example, when the difference between the foreground and the background is not large, our method succeeds in handing their boundaries with effective artifacts suppresion, compared to other unsupervised methods. However, when the difference is very large, as shown in the second example, our method may produce undesired artifacts (e.g., a noticeable white streaky border). Note that in this case, other methods cannot get satisfactory results, either. How to effective to handle such cases remains an open question for unsupervised DL methods.

Fig. 1. Examples of successful boundary processing on the MFFW dataset (Row #1-#2) and examples of failed boundary processing (Row #3-#4).

V. Supplementary Visual Results

Fig. 2 shows the results on the samples with GTs. UDPNet has clearer details on the edges of the billboards in 1st sample, the lettering of the street signs in 2nd sample, the spines of the cactus in 3rd sample, and the lettering on the bicycle seats in 4th sample. Fig. 3 shows the best and worst results on the MFI-WHU dataset. The best result is nearly identical to the GT, while the worst one has deviations in some fuzzy boundary places.

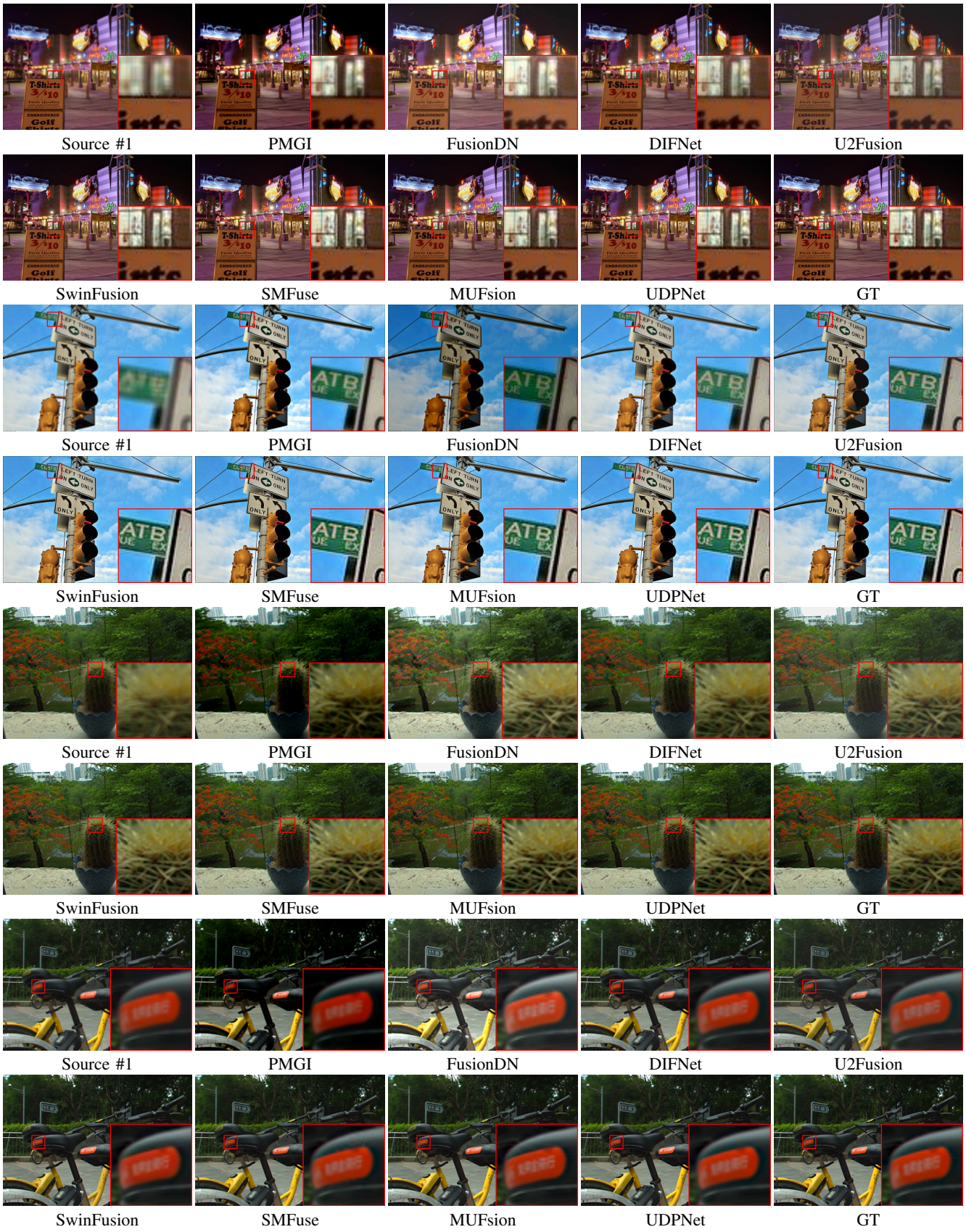

Fig. 2. Fusion results by different methods on samples from MFI-WHU (Row #1-#4) and Real-MFF (Row #5-#8).

Fig. 3. Best (upper row) and worst (bottom row) results in terms of SSIM on MFI-WHU dataset.