Task-driven Image Fusion with Learnable Fusion Loss

SUPPLEMENTARY MATERIALS

Abstract

In this document, we provide additional supplementary information for the paper "Task-driven Image Fusion with Learnable Fusion Loss". This file contains:
(I) Additional fusion comparison results of image fusion in Sec. 4.2.
(II) More downstream application results in Sec. 4.3.
(III) Class-wise results of downstream application in Sec. 4.3.
(IV) More learning results of learnable loss in Sec. 4.4.
(V) Visualization Results of Ablation Studies in Section 4.5.

1. Additional fusion results

In this section, we provide additional qualitative comparisons of fusion results, as presented in Fig. S-1.

2. Downstream application results

In this section, we provide more results on semantic segmentation and object detection, as presented in Fig. S-2 and Fig. S-3.

3. Class-wise results of downstream application

In this section, we provide detailed per-class results for downstream applications. Semantic segmentation results on the FMB and MSRS datasets are presented in Tab. S-1 and Tab. S-2, respectively, while object detection results on the M3FD and LLVIP datasets are summarized in Tab. S-3.

4. More learning results of learnable loss

The learning results of learnable loss and the corresponding fusion results on the four datasets are presented in Fig. S-4.

5. Visualization Comparison and Analysis of Ablation Studies

Fig. S-5 visualizes the ablation studies. In Exp. I, fixing w_a and w_b to 1/2 loses thermal radiation and texture information. In Exp. II, removing the gradient loss leads to poor texture and contrast. In Exp. III, the downstream task loss slightly affects fusion learning, weakening the highlights. Exp. IV alters the meta-learning process, retaining minimal infrared image information. Exp. V produces unnatural images. Both visual and quantitative results confirm the validity of our method.

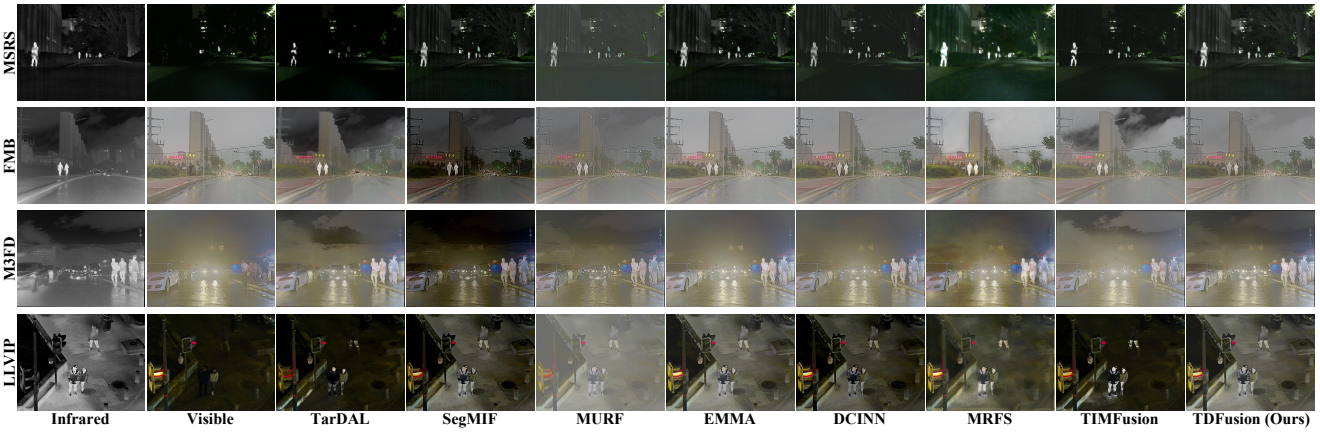

Figure S-1. Visual comparison of fusion results. The cases are "00862N" in MSRS dataset, "00123" in FMB dataset, "00421" in M3FD dataset and "200214" in LLVIP dataset.

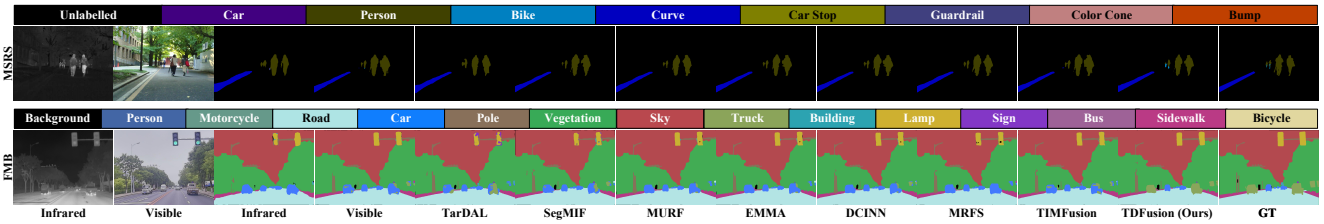

Figure S-2. Visual comparison for Semantic Segmentation. The cases are "00241D" in MSRS dataset and "01319" in FMB dataset.

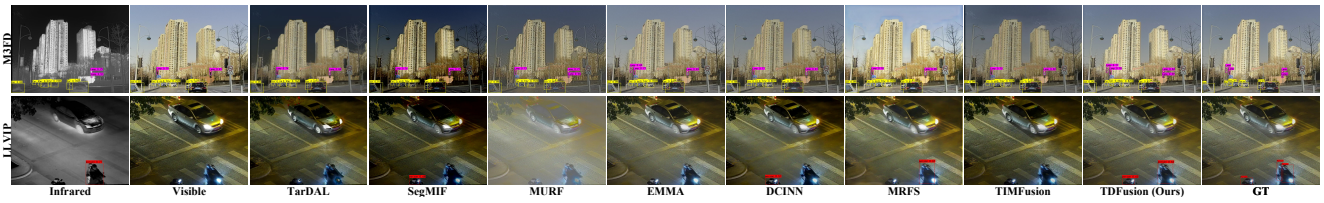

Figure S-3. Visual comparison for Object Detection. The cases are "03288" in M3FD dataset and "220219" in LLVIP dataset.

Table S-1. Quantitative semantic segmentation results of different methods on the FMB dataset. The red and blue markers represent the best and second-best values, respectively.

	Background		Road		Sidewalk		Building		Lamp		Sign		Vegetation		Sky		Person		Car		Truck		Bus		Motorcycle		Pole			
Methods	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	mAcc	mIoU
Infrared	21.28	18.20	96.57	87.07	43.33	36.87	89.89	80.62	23.23	22.07	66.07	60.55	90.74	81.25	97.44	93.05	74.24	66.45	92.64	76.02	19.96	19.34	19.10	18.75	48.33	36.52	41.03	30.91	58.85	51.98
Visible	26.09	21.47	96.22	87.38	47.46	42.59	90.19	80.61	31.50	29.48	82.31	70.52	92.58	84.70	97.25	93.49	64.79	56.26	93.05	80.56	30.87	29.26	50.05	48.36	50.47	41.45	58.90	45.29	65.12	57.96
TarDAL	24.88	19.72	96.85	86.63	42.24	37.72	90.25	80.81	29.92	28.24	84.98	72.38	93.06	85.48	97.67	94.11	71.19	64.91	93.21	80.23	17.69	17.13	18.20	18.02	60.70	43.59	59.14	45.65	62.86	55.33
SegMIF	28.64	22.83	96.79	87.91	45.24	40.97	91.11	82.19	33.53	30.15	84.27	72.55	93.10	85.68	97.61	94.37	75.73	67.23	93.16	79.80	43.62	40.50	28.13	27.63	53.10	40.06	59.50	45.87	65.97	58.41
MURF	27.57	22.19	96.85	88.20	47.44	42.82	92.55	82.27	31.50	29.28	82.48	69.82	93.10	86.30	97.12	94.30	72.60	65.83	93.69	80.10	23.31	22.41	23.64	23.26	54.37	43.63	61.21	47.10	64.10	56.96
EMMA	25.14	21.37	95.81	87.54	46.42	41.77	91.05	81.86	31.70	30.39	77.77	69.11	93.47	84.79	97.05	93.78	72.39	65.01	93.87	78.34	11.78	11.36	36.88	36.38	43.46	40.62	57.49	45.58	62.45	56.28
DCINN	26.83	23.02	96.48	88.18	43.94	39.76	91.84	81.33	29.72	27.85	79.94	70.38	93.15	85.32	97.08	93.80	71.78	64.91	93.86	78.77	8.61	8.57	19.06	18.77	45.13	40.77	57.76	45.89	61.09	54.81
MRFS	25.78	22.17	96.42	87.86	49.81	44.35	91.50	81.47	22.36	21.51	79.69	70.18	93.47	85.60	97.24	94.06	70.45	63.66	93.61	80.14	8.94	8.68	30.68	30.03	48.28	43.81	58.76	46.43	61.93	55.71
TIMFusion	24.92	21.27	95.09	87.20	49.29	44.60	89.86	80.34	31.60	29.12	80.27	70.25	93.54	84.39	97.25	93.37	69.89	62.97	93.25	81.12	17.89	17.21	40.77	39.59	50.42	44.23	57.83	45.69	63.70	57.24
TDFusion	26.53	21.58	96.07	88.72	52.57	47.26	92.07	82.78	37.56	35.15	83.30	73.31	93.29	86.25	97.59	94.08	73.37	65.87	93.91	80.73	30.99	30.00	45.32	44.32	59.75	50.14	58.02	46.79	67.17	60.50

Table S-2. Quantitative semantic segmentation results of different methods on the MSRS dataset. The red and blue markers represent the best and second-best values, respectively.

	Unlabelled		Car		Person		Bike		Curve		Car_stop		Guardrail		Color_cone		Bump			
Methods	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	Acc	IoU	mAcc	mIoU
Infrared	98.94	98.02	92.25	86.56	84.77	70.41	80.73	68.24	72.96	56.55	82.41	62.22	83.46	59.89	74.13	52.42	79.43	71.10	83.23	69.49
Visible	99.17	98.22	93.48	88.87	77.51	63.27	83.34	69.83	70.29	58.55	83.97	73.39	87.23	74.02	75.84	63.84	80.09	73.81	83.44	73.76
TarDAL	99.09	98.09	93.00	87.47	81.67	67.09	78.98	67.68	68.93	54.04	82.12	70.41	85.24	68.57	72.92	59.58	75.40	69.21	81.93	71.35
SegMIF	99.06	98.31	93.96	89.03	86.60	71.27	83.63	70.54	80.51	58.88	82.40	71.84	84.70	68.86	78.61	63.89	82.11	75.67	85.73	74.25
MURF	99.18	98.34	93.21	88.78	85.35	71.50	81.57	70.47	77.76	61.11	83.52	71.77	87.83	68.08	76.07	62.68	80.77	73.99	85.03	74.08
EMMA	99.15	98.37	93.63	89.32	85.79	71.15	83.09	70.58	78.61	60.71	81.12	70.37	91.44	69.59	79.65	64.75	81.45	75.50	85.99	74.48
DCINN	99.21	98.34	93.36	89.08	85.91	70.94	81.30	69.19	71.26	59.31	84.19	73.27	83.66	72.21	78.54	64.02	79.61	72.76	84.11	74.35
MRFS	99.16	98.32	93.46	88.80	85.88	70.55	82.41	70.41	72.67	58.76	85.51	73.85	87.05	74.66	76.99	62.80	79.73	72.37	84.76	74.50
TIMFusion	99.18	98.30	93.68	89.19	83.61	68.68	81.58	69.64	73.53	59.55	85.04	71.94	84.11	69.48	75.45	63.69	76.82	71.75	83.67	73.58
TDFusion	99.17	98.38	94.61	89.75	85.49	71.76	82.46	70.36	78.26	59.77	86.59	72.40	87.61	75.06	79.41	64.61	80.71	73.73	86.04	75.09

Table S-3. Quantitative object detection results of different methods on the M3FD and LLVIP datasets. The red and blue markers represent the best and second-best values, respectively.

	M3FD Dataset														**LLVIP Dataset**		
	Bus		Car		Lamp		Motorcycle		People		Truck		mAP			Person	
Methods	AP@50	AP@75	AP@50	AP@75	AP@50	AP@75	AP@50	AP@75	AP@50	AP@75	AP@50	AP@75	AP@50	AP@75	Methods	AP@50	AP@75
Infrared	88.61	77.18	88.65	64.81	63.49	21.47	69.68	43.00	84.13	49.20	80.14	62.67	79.12	53.05	Infrared	96.03	72.07
Visible	92.29	77.40	91.53	70.73	80.86	34.09	70.79	39.22	74.34	38.49	83.48	68.98	82.21	54.82	Visible	91.78	48.66
TarDAL	88.82	73.88	91.03	70.99	78.67	33.93	75.28	40.79	82.82	49.51	82.34	69.27	83.16	56.39	TarDAL	93.79	62.71
SegMIF	89.34	77.22	91.90	71.51	79.67	36.67	73.49	45.02	83.49	49.98	83.78	69.00	83.61	58.23	SegMIF	93.95	66.45
MURF	88.57	74.95	90.17	69.07	74.48	29.35	66.94	40.89	82.65	49.08	80.65	62.02	80.58	54.22	MURF	94.24	68.04
EMMA	90.22	74.68	91.95	71.18	80.03	35.43	72.81	41.98	82.47	48.93	84.76	69.24	83.71	56.91	EMMA	94.00	66.21
DCINN	89.88	75.35	91.63	71.58	77.86	31.91	69.54	45.36	83.28	50.85	83.94	69.18	82.69	57.37	DCINN	94.92	68.34
MRFS	89.59	76.58	91.87	71.13	80.28	36.26	72.15	44.88	82.47	48.77	83.32	68.84	83.28	57.74	MRFS	93.03	67.21
TIMFusion	89.32	76.61	91.70	71.02	78.09	33.98	73.86	38.48	81.62	47.21	84.76	69.19	83.22	56.08	TIMFusion	93.76	61.33
TDFusion	92.85	78.81	93.22	73.65	82.32	38.57	78.76	46.12	85.22	50.68	85.25	70.41	86.27	59.71	TDFusion	95.00	69.18

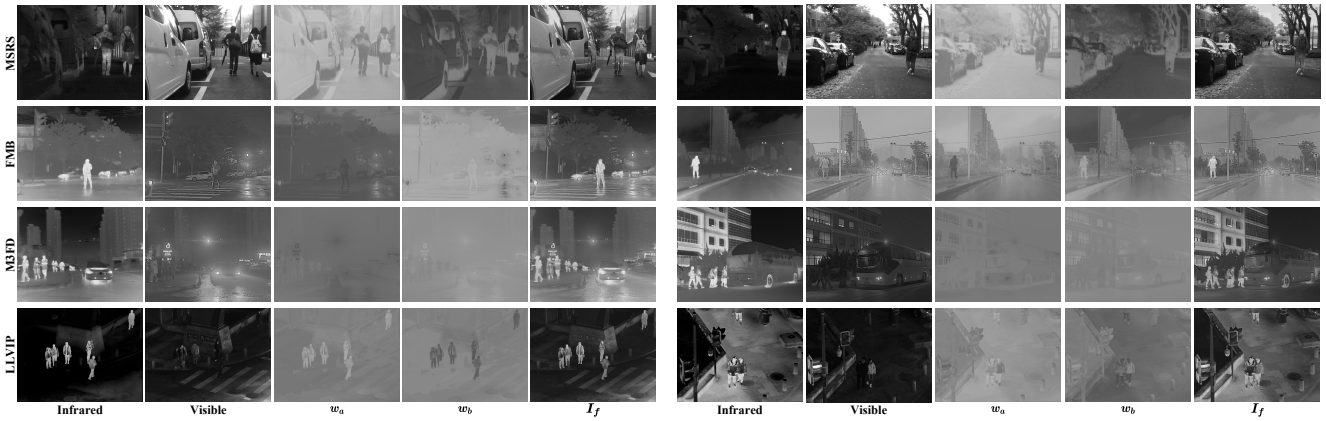

Figure S-4. Visual comparison of the results of the learnable loss function. The cases are “01389D” and ”00186D” in MSRS dataset, “00058” and ”00122” in FMB dataset, “00386” and ”03878” in M3FD dataset, “010001” and “200084” in LLVIP dataset.

Figure S-5. Visualization of ablation studies.