Appendix

A. Supplementary Details

A.1. Additional Details of Task-specific Adaptation

For the details of three parallel operation branches in OAF module: (i) **HPF** branch: Conditioned by the operand $P_h \in \mathbb{R}^{\frac{N}{H\times W}\times K^2\times H\times W}$ predicted by the hyper-network $\phi_{\theta_h}(\cdot)$, this branch employs spatially-variant convolution on the input feature $X \in \mathbb{R}^{\frac{N}{H\times W}\times D\times H\times W}$ to obtain the resultant feature $\hat{X}_h \in \mathbb{R}^{\frac{N}{H\times W}\times D\times H\times W}$, where H, W, K denote the height, width and the kernel size respectively. (ii) **ADD** branch: Conditioned by the operand $P_a \in \mathbb{R}^{\frac{N}{H\times W}\times D\times H\times W}$ predicted by the hyper-network $\phi_{\theta_a}(\cdot)$, this branch performs addition on the input feature to obtain the resultant feature $\hat{X}_a \in \mathbb{R}^{\frac{N}{H\times W}\times D\times H\times W}$. (iii) **MUL** branch: Conditioned by the operand $P_m \in \mathbb{R}^{\frac{N}{H\times W}\times D\times H\times W}$ predicted by the hyper-network $\phi_{\theta_m}(\cdot)$, this branch performs Hadamard (element-wise) multiplication on the input feature to obtain the resultant feature $\hat{X}_m \in \mathbb{R}^{\frac{N}{H\times W}\times D\times H\times W}$. And all the hyper-networks used are two-layer MLPs.

A.2. Optimization Objectives

The unified objective used in Baseline is borrowed from U2Fusion [4]:

$$\begin{aligned}
\ell =& \lambda_1 \ell_{ssim} + \lambda_2 \ell_{mse}, \\
\ell_{ssim} =& w_1(1 - ssim(I_f, I_1)) + w_2(1 - ssim(I_f, I_2)), \\
\ell_{mse} =& w_1 \cdot \|I_f - I_1\|_2^2 + w_2 \|I_f - I_2\|_2^2,
\end{aligned} \tag{1}$$

where $\lambda_1 = 1, \lambda_2 = 20$, and w_1, w_2 is calculated by the information measured on VGG features.

And for Baseline-TS and the proposed TITA framework, following SwinFusion [2], the task-specific training objectives are shown as below:

$$\begin{aligned}
\ell =& \lambda_1 \ell_{ssim} + \lambda_2 \ell_{text} + \lambda_3 \ell_{int}, \\
\ell_{ssim} =& \frac{1}{2}(1 - ssim(I_f, I_1)) + \frac{1}{2}(1 - ssim(I_f, I_2)), \\
\ell_{text} =& \frac{1}{HW} \| |\nabla I_f| - \max(|\nabla I_1|, |\nabla I_2|) \|_1, \\
\ell_{int} =& \frac{1}{HW} \| I_f - M(I_1, I_2) \|_1,
\end{aligned} \tag{2}$$

where $\lambda_1 = 10, \lambda_2 = 20, \lambda_3 = 20$ are hyper-parameters, $M(\cdot)$ is task-specific element-wise aggregation operation. Specifically, $max(\cdot, \cdot)$ is employed for IVF and MFF, $mean(\cdot, \cdot)$ is applied to MEF.

A.3. Additional Details of FAMO

Considering M fusion tasks associated with M objectives $\ell_m\}_{m=1}^{M}$. In t-th iteration, the combination weights are obtained by $Z_t = Softmax(\xi_t)$, where $\xi_t \in \mathbb{R}^M$ are unconstrained logits. FAMO updates the model parameters as:

$$\theta_{t+1} = \theta_t - \alpha \sum_{m=1}^{M} \left(C_t \frac{Z_{m,t}}{\ell_{m,t}} \right) \nabla \ell_{m,t}, \tag{3}$$

where $C_t = \left(\sum_{m=1}^{M} Z_{m,t} / \ell_{m,t} \right)^{-1}$, α is the step size. And the weighting logits can be updated as:

$$\begin{aligned}
\xi_{t+1} &= \xi_t - \beta(\delta_t + \gamma \xi_t), \\
\delta_t &= \begin{bmatrix} \nabla^\top Z_{1,t} \\ \vdots \\ \nabla^\top Z_{M,t} \end{bmatrix}^\top \begin{bmatrix} \log \ell_{1,t} - \log \ell_{1,t+1} \\ \vdots \\ \log \ell_{M,t} - \log \ell_{M,t+1} \end{bmatrix}
\end{aligned} \tag{4}$$

where β is the step size, γ is the decay. By maximizing the minimum improvement rate, FAMO effectively allocates computational resources and aligns optimization objectives, ultimately improving overall performance. For detailed derivation, please refer FAMO [1].

A.4. Token Exchange

TE module [3] operates on the principle that when a uninformative token is detected, it can be replaced with a binary modal token at the corresponding position, preserving essential information while reducing noise:

$$\begin{aligned}
X_{i,1} &= X_{i,1} \odot \mathbb{I}_{\phi_{\theta_s}(X_{i,1}) \geq \gamma} + X_{i,2} \odot \mathbb{I}_{\phi_{\theta_s}(X_{i,1}) < \gamma}, \\
X_{i,2} &= X_{i,2} \odot \mathbb{I}_{\phi_{\theta_s}(X_{i,2}) \geq \gamma} + X_{i,1} \odot \mathbb{I}_{\phi_{\theta_s}(X_{i,2}) < \gamma},
\end{aligned} \tag{5}$$

where $\mathbb{I}$ is an indicator, and the threshold γ is set to 0.02 according to the paper.

B. More Results and Analysis

B.1. Ablation Study for MEF and MFF Tasks

We present the ablation study results for MEF and MFF tasks in Tab. 1 and Tab. 2. The findings on the MFF task

Table 1. The ablation study for MEF task.

TI	TA	MO	MI	FMI	Qabf	VIF
Baseline			**6.6732**	0.8953	0.6748	1.4342
Baseline-TS			5.2338	0.8976	0.7154	1.3061
✓			5.2125	0.8974	0.7148	1.3187
	✓		5.2231	0.8984	0.7227	1.3154
		✓	5.9231	0.8998	0.7039	1.4994
✓	✓		5.9976	0.8992	0.7057	1.5153
✓		✓	5.1837	0.8989	0.7215	1.3146
	✓	✓	6.1557	**0.9002**	**0.7235**	1.5274
Ours			6.2073	0.9000	0.7227	**1.5338**

Table 2. The ablation study for MFF task.

TI	TA	MO	MI	FMI	Qabf	VIF
Baseline			6.2336	0.8777	0.5768	1.6171
Baseline-TS			6.2303	0.8822	0.6455	1.5997
✓			6.2566	0.8821	0.6554	1.5931
	✓		6.2687	0.8824	0.6834	1.5963
		✓	6.2822	0.8833	0.6519	1.6223
✓	✓		6.3071	0.8836	0.6523	1.6210
✓		✓	6.3229	0.8826	0.6916	1.6050
	✓	✓	6.4689	0.8841	0.6915	1.6294
Ours			**6.5463**	**0.8847**	**0.6973**	**1.6371**

Table 3. The ablation study on task-invariant integration for MEF task.

	MI	FMI	Qabf	VIF
SF	5.2338	0.8976	0.7154	1.3061
IrSF	**5.2551**	0.8975	0.7125	1.3059
IeSF	5.2512	**0.8977**	**0.7163**	**1.3074**
TE	**5.2662**	0.8969	0.7096	**1.3155**
PA	5.2428	0.8975	0.7098	1.3056
IPA	5.2273	**0.8977**	0.7116	1.3048

align with those of the IVF task, reinforcing the same conclusions. The relatively lower impact of TI on the MEF task may be attributed to its higher reliance on global transformations, where the global operation in TA plays a more significant role.

B.2. Analysis on Task-invariant Integration for MEF and MFF tasks

We conduct ablation studies to evaluate the effectiveness of IeSF and IPA in Task-invariant Integration for MEF and MFF tasks. As shown in Tab. 3 and Tab. 4, by replacing the IeSF with IrSF or replacing IPA with TE, we observe drops in overall performance, demonstrating the necessity of both IeSF and IPA modules.

Table 4. The ablation study on task-invariant integration for MFF task.

	MI	FMI	Qabf	VIF
SF	6.2303	0.8822	0.6455	1.5997
IrSF	6.1816	0.8818	0.6415	1.5916
IeSF	**6.2373**	**0.8822**	**0.6516**	**1.6019**
TE	6.2386	0.8817	0.6467	1.5952
PA	**6.2835**	**0.8825**	0.6461	1.6088
IPA	6.2690	0.8824	**0.6541**	**1.6091**

Table 5. The ablation study on task-specific adaptation for MEF task.

	MI	FMI	Qabf	VIF
W/o HPF	6.0651	0.8999	**0.7247**	1.4957
W/o ADD	5.8250	0.8995	0.7215	1.4760
W/o MUL	6.1305	0.8984	0.7099	1.5164
W/o DW	4.8277	0.8987	0.6894	1.2815
Ours	**6.2073**	**0.9000**	0.7227	**1.5338**

Table 6. The ablation study on task-specific adaptation for MFF task.

	MI	FMI	Qabf	VIF
W/o HPF	6.2898	0.8833	0.6801	1.6039
W/o ADD	6.3395	0.8845	0.6831	**1.6385**
W/o MUL	6.3512	0.8831	0.6663	1.6148
W/o DW	6.2616	0.8840	0.6862	1.6245
Ours	**6.5463**	**0.8847**	**0.6973**	1.6371

Table 7. The visualization of the dynamic weights on OAF block.

HPF	ADD	MUL	HPF	ADD	MUL
	visible			infrared	
0.0291	0.2456	0.2054	0.0243	0.2226	0.2729
	over-exposed			under-exposed	
0.0286	0.1963	0.3054	0.0313	0.2575	0.1809
	far-focused			near-focused	
0.0159	0.2480	0.2323	0.0133	0.2231	0.2675

B.3. Analysis on Task-specific Adaptation for MEF and MFF tasks

We conduct ablation studies to evaluate the effectiveness of three branches in Task-adaptive Adaptation for MEF and MFF tasks. As shown in Tab. 5 and Tab. 6, while all branches contribute to the overall performance improvement, their contributions differ. For example, the MFF task relies more heavily on the HPF branch, as clarity is directly related to high-frequency details.

B.4. Visualization of TA

The weights for each branch of OAF for three fusion tasks is visualized in Tab. 7. It can be observed that the weight assigned to the HPF branch is relatively small, as high-frequency details comprise only a minor portion of the overall information. Nevertheless, as demonstrated in the ablation study on task-specific adaptation, these high-frequency components play a critical role in boosting performance.

References

[1] Bo Liu, Yihao Feng, Peter Stone, and Qiang Liu. Famo: Fast adaptive multitask optimization. *Advances in Neural Information Processing Systems*, 36, 2024. 1

[2] Jiayi Ma, Linfeng Tang, Fan Fan, Jun Huang, Xiaoguang Mei, and Yong Ma. Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA Journal of Automatica Sinica*, 9(7):1200–1217, 2022. 1

[3] Yikai Wang, Xinghao Chen, Lele Cao, Wenbing Huang, Fuchun Sun, and Yunhe Wang. Multimodal token fusion for vision transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12186–12195, 2022. 1

[4] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):502–518, 2020. 1