Uncertainty-Aware Spatial-Frequency Registration and Fusion for Infrared and Visible Images

Xingyuan Li[1], Haoyuan Xu[2], Xingyue Zhu[2], Jun Ma[2], Zhiying Jiang[3], Jinyuan Liu[2], Yang Zou[4*]
[1]School of Computer Science, Zhejiang University
[2]School of Software Technology & DUT-RU International School of ISE, Dalian University of Technology
[3]College of Information Science and Technology, Dalian Maritime University
[4]School of Computer Science, Northwestern Polytechnical University
xingyuan_lxy@163.com, archerv2@mail.nwpu.edu.cn

Abstract

*Corresponding author. Infrared and Visible Image Fusion (IVIF) has shown promise in visual tasks under challenging environments, but fusion under unregistered conditions faces inherent misalignments. Current studies to solve them either predict the deformation parameters coarse-to-fine (i.e., coarse registration and fine registration) or estimate the deformation fields in multi-scales for registration. Though straightforward, they overlook the cumulative errors in registration, which contaminate the fusion stage and severely deteriorate the resulting images. We introduce the **S**patial-**F**requency **R**egistration and **F**usion (**SFRF**) framework, which incorporates uncertainty estimation and infrared thermal radiation distribution consistency into a unified pipeline to handle the error accumulation for robust registration and fusion across both spatial and frequency domains. Specifically, SFRF constructs a Multi-scale Iterative Registration (MIR) framework that iteratively refines the deformation field across scales, leveraging uncertainty estimation at each stage to mitigate error accumulation and enhance alignment accuracy dynamically. To ensure the accurate alignment of infrared thermal distributions during registration, thermal radiation distribution consistency is employed as a frequency-domain supervisory signal, promoting global consistency in the frequency domain. Based on the spatial-frequency alignment, SFRF further adopts a Dual-branch Spatial-Frequency Fusion (DSFF) module, which incorporates spatial geometric features and frequency distribution information to reconstruct visually appealing images. SFRF achieves impressive performance across diverse datasets. Code is available at github.com/xhhaoyan/SFRF.

*Corresponding author

1 Introduction

Infrared and Visible Image Fusion (IVIF) is fundamental to integrating valuable information from different modalities in low-level vision [Ma *et al.*, 2019; Bai *et al.*, 2025b]. Traditional IVIF methods decompose images into multiple levels to extract information at different scales, typically relying on handcrafted feature extraction techniques and weighting rules. More recently, deep learning-based methods [Guan *et al.*, 2026; Li *et al.*, 2023; Xu *et al.*, 2023; Liu *et al.*, 2024b] have demonstrated notable potential for downstream tasks such as object detection [Liu *et al.*, 2026; Wang *et al.*, 2025; Liu *et al.*, 2022] and segmentation [Kirillov *et al.*, 2023; Liu *et al.*, 2023; Zhao *et al.*, 2026], transforming the fusion process into a task-oriented approach.

These approaches produce visually appealing results and hold unique industrial value for visual tasks in severe environments [Xiao *et al.*, 2026; Zou *et al.*, 2026]. Yet, preserving the underlying image details and maintaining the integrity of modality-specific information is more challenging than it seems. The first challenge lies in handling infrared-specific spectral characteristics. As shown in Figure 1, most existing methods [Liu *et al.*, 2024a; Li *et al.*, 2024a; Wang *et al.*, 2024; Xiao and Xiong, 2025; Li *et al.*, 2025; Wang *et al.*, 2022; Wang and Zhang, 2020] are confined to spatial registration techniques, which fundamentally do not differ from visible image registration. This spatial-only focus neglects the unique spectral properties of infrared images, such as thermal radiation distributions, which are critical for achieving effective infrared and visible image fusion [Zhao *et al.*, 2023; Bai *et al.*, 2025a]. As a result, this oversight often degrades the fusion quality, especially in scenarios where preserving modality-specific information is essential.

The second challenge lies in the inherent uncertainty in estimating the deformation field during registration, whether in coarse-to-fine (i.e., coarse and fine registration) [Xu *et al.*, 2023] or single-stage multi-scale registration frameworks [Wang *et al.*, 2024; Wang *et al.*, 2022; Huang *et al.*, 2022]. Due to the complex nature of image registration, including multi-scale spatial deformations and cross-modal disparities, predicting a precise deformation field is inherently uncertain [Decarlo and Metaxas, 2000; Brox and Malik, 2010]. This uncertainty is further exacerbated by the

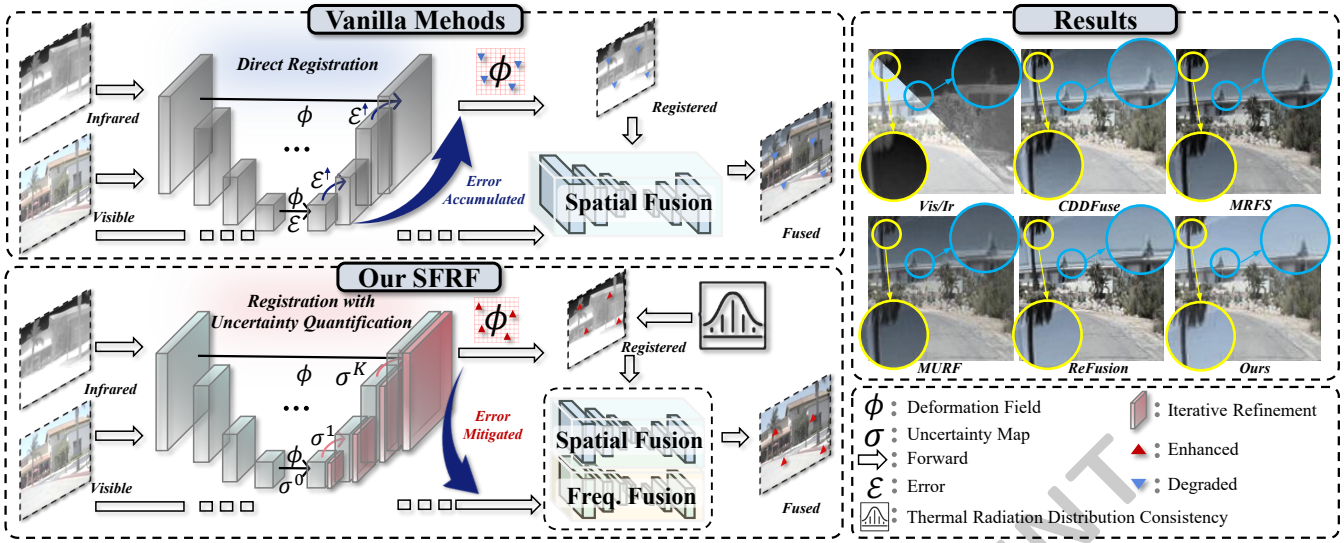

Figure 1: Vanilla methods often register the image directly, failing to account for the accumulated errors at coarse scales and severely degrading registration and fusion performance. They also neglect the specific frequency characteristics of infrared, which further diminishes the quality of the fusion. In contrast, our method addresses error accumulation and maintains thermal radiation distribution consistency through an uncertainty-aware spatial-frequency registration and fusion framework, ensuring more robust and accurate results.

single-pass prediction strategies commonly used in registration frameworks, which propagate and amplify errors.

As a response, we introduce the **S**patial-**F**requency **R**egistration and **F**usion (SFRF) framework to address error accumulation while preserving thermal radiation distribution consistency. SFRF quantifies uncertainty at each scale, mitigating penalties in uncertain areas and discouraging excessively large uncertainties through a Multi-scale Iterative Registration (MIR) framework, which helps reduce accumulated errors during registration. To preserve infrared-specific characteristics, SFRF incorporates thermal radiation distribution consistency, guiding infrared image registration for improved fusion results. Additionally, SFRF employs a Dual-branch Spatial-Frequency Fusion (DSFF) module to capture spatial geometric features and frequency distribution information, thereby reconstructing visually appealing images. Our contributions are as follows:

- We propose **S**patial-**F**requency **R**egistration and **F**usion (**SFRF**) framework to address the misalignments in fusion under unregistered conditions.
- In the registration phase, we introduce the Multi-scale Iterative Registration (MIR) framework to mitigate accumulated spatial-domain errors and adopt thermal-radiation distribution consistency as a frequency-domain supervisory signal, ensuring precise alignment of infrared thermal distributions and promoting global frequency-domain consistency.
- In the fusion phase, we present the Dual-branch Spatial-Frequency Fusion (DSFF) module to reconstruct visually appealing and perceptually coherent images by integrating spatial geometric features and frequency distribution information.

2 Related Work

Multi-Modal Image Registration. Traditional multi-modal image registration methods often fall into intensity-based, feature-based, or transformation-based categories. While intensity-based approaches compare pixel intensities across modalities [Liu *et al.*, 2025b], feature-based methods rely on extracting distinctive features, and transformation-based methods estimate geometric transformations. Deep learning-based methods focus on optimizing the parameters for predicted affine transformations or deformation fields. For instance, Arar et al.[Arar *et al.*, 2020] introduced a geometrically preserving image transformation network. Ye et al.[Ye *et al.*, 2022] stacked deep networks to map image pairs to transformation parameters for improved accuracy. Nevertheless, multi-modal registration typically involves multi-scale deformation field estimation. Varying texture or lighting across scales introduces uncertainty in affine parameters [Decarlo and Metaxas, 2000; Li *et al.*, 2024c], which can amplify errors through iterative optimization. Moreover, most methods focus on spatial alignment alone, neglecting the spectral distribution in infrared images, which is crucial for high-quality fusion results.

Multi-Modal Image Fusion. Autoencoder-based fusion methods learn low-dimensional representations to fuse different modalities [Ren *et al.*, 2021], while CNN-based approaches typically involve either feature extraction with weight generation [Liu *et al.*, 2025a] or end-to-end learning [Hou *et al.*, 2020]. GAN-based methods implicitly integrate feature extraction, feature fusion, and image reconstruction [Li *et al.*, 2020; Li *et al.*, 2026; Xiao *et al.*, 2025]. Many existing methods overlook critical amplitude–phase differences. For instance, Hu et al. [Hu *et al.*, 2024] treated the amplitude and phase of infrared and visible images in a unified manner, simply concatenating them for convolutional fusion. Huang

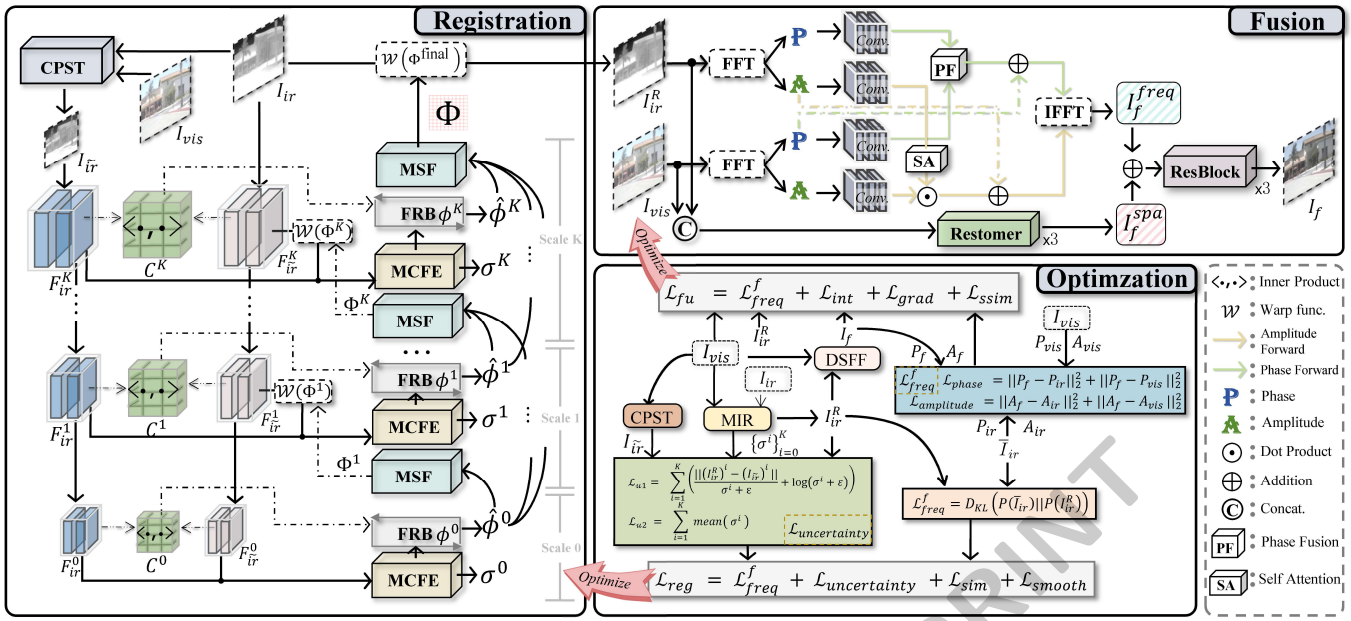

Figure 2: Overview of SFRF.

et al.[Huang *et al.*, 2022] proposed a recurrent correction network with a deformation module and attention mechanism for slight misalignments.

3 Method

As shown in Figure 2, given a pair of unaligned infrared and visible images I_{ir} and I_{vis}, respectively, our proposed **S**patial-**F**requency **R**egistration and **F**usion (**SFRF**) framework incorporates uncertainty estimation and infrared thermal radiation distribution consistency into a unified pipeline to handle the error accumulation for robust registration and fusion across both spatial and frequency domains.

3.1 Multi-scale Iterative Registration (MIR)

As shown in Figure 2 (Registration), given the input distorted infrared image I_{ir} and the pseudo-infrared image $I_{\tilde{ir}}$ generated from the visible image I_{vis} through a frozen CPST [Wang *et al.*, 2022] module, an encoder comprising LK-CNN blocks [Ding *et al.*, 2024] is employed to extract multi-scale feature representations of both I_{ir} and $I_{\tilde{ir}}$. For each image I, the encoder outputs K levels of feature maps from the coarsest scale $i = 0$ to the finest scale $i = K$: $\mathbf{F}_{ir} = \{F^0_{ir}, F^1_{ir}, ..., F^K_{ir}\}$, and $\mathbf{F}_{\tilde{ir}} = \{F^0_{\tilde{ir}} F^1_{\tilde{ir}}, ..., F^K_{\tilde{ir}}\}$, where F^i_{ir} and $F^i_{\tilde{ir}}$ represent the feature maps at scale i. To capture the feature-level similarity between I_{ir} and $I_{\tilde{ir}}$, we construct a correlation map $C^i \in \mathbb{R}^{B\times 1\times H\times W}$ at each scale as $C^i = \frac{F^i_{ir}\cdot(F^i_{\tilde{ir}})^T}{\|F^i_{ir}\|\|F^i_{\tilde{ir}}\|}$, for $i \in \{0, ..., K\}$, where each spatial location (y, x) in C^i encodes the mean similarity of $\left(F^i_{ir}\right)_{(y,x)}$ against all positions in $F^i_{\tilde{ir}}$.

Monte Carlo sampling-based field estimation (MCFE). At the coarsest scale ($i = 0$), an initial coarse deformation field is predicted via a Monte Carlo sampling-based field estimation (MCFE) module that incorporates dropout to enable uncertainty quantification. Let $\mathcal{D}$ denote the set of dropout-enabled forward passes through the MCFE module, and for each forward pass $n \in \{1, \ldots, N\}$, the predicted deformation field ϕ^0_n is given by $\phi^0_n = \mathcal{F}\left(F^0_{ir}, F^0_{\tilde{ir}}; \mathcal{D}_n\right)$, where $\mathcal{F}$ is the MCFE module applied to the feature maps F^0_{ir} and $F^0_{\tilde{ir}}$ at scale 0, with $\mathcal{D}_n$ denoting the n-th stochastic dropout mask applied during the forward pass. The coarse deformation field ϕ^0 is then computed as the mean of the N predictions (N = 10 in our experiments), and the uncertainty map σ^0 is estimated based on the variance, as $\phi^0 = \frac{1}{N}\sum_{n=1}^{N}\phi^0_n$ and $\sigma^0 = \frac{1}{N}\sum_{n=1}^{N}\left(\phi^0_n - \phi^0\right)^2$.

GRU-based field refinement block (FRB). Due to the complex nature of infrared-visible image registration, predicting a precise deformation field is inherently uncertain [Brox and Malik, 2010] . This uncertainty is further exacerbated by the single-pass prediction, leading to the propagation and amplification of errors. MIR iteratively refines the initial deformation field at each scale via a field refinement block (FRB) using a GRU-based update mechanism guided by the corresponding correlation map C to mitigate this error amplification. Define the FRB update operator $\mathcal{G}$ by $\mathcal{G}(\phi, h) = \mathrm{Conv}\Big(\mathrm{GRU}\Big(\mathrm{concat}(\phi, F, C), h\Big)\Big)$, where $\mathrm{concat}(\cdot)$ represents channel-wise concatenation, $\mathrm{GRU}(\cdot,\cdot)$ is the GRU cell that maps the concatenated input and previous hidden state h to an updated hidden state, and $\mathrm{Conv}(\cdot)$ is a 3×3 convolution operation that projects the hidden state into the deformation field space. The iterative refinement process over T steps is then expressed as:

$$\phi_{(t+1)} = \phi_{(t)} + \mathcal{G}\Big(\phi_{(t)}, h_{(t)}\Big), \quad t = 0, \ldots, T-1, \quad (1)$$

where the $h_{(t)}$ denotes the hidden state at iteration t (initialized with zeros), and the summation represents a residual accumulation of the GRU updates (T = 3 in our experiments). The initial deformation field ϕ^0 is then iteratively refined through the FRB to progressively enhance the alignment, resulting in the refined deformation field at the coarsest scale as $\hat{\phi}^0 = \phi^0_{(T)}$, where $\phi^0_{(T)}$ denotes the deformation field after T refinement iterations.

Multi-scale field fusion (MSF). MIR utilizes the Multi-scale Field Fusion (MSF) to mitigate the error accumulation, which refines hierarchical deformation fields by ensuring spatial consistency and uncertainty-adaptive weighting before propagating transformations to finer scales.

Given a refined deformation field $\hat{\phi}^k$ at scale k, $k < K$, we first upsample all previously estimated deformation fields $\{\hat{\phi}^i\}_{i=0}^{k-1}$ to the resolution of scale k by $\widetilde{\phi}^{(i\to k)} = \mathrm{Up}\big(\hat{\phi}^i,\, s_k\big), \forall i < k$, where s_k is the appropriate upsampling factor, and $\mathrm{Up}(\cdot)$ represents bilinear interpolation. However, due to potential spatial misalignment across scales, directly combining these deformation fields can lead to accumulated registration errors. To alleviate this, MSF embodies a correlation-based alignment step to compute a correlation map that measures the similarity between each upsampled deformation field and $\hat{\phi}^k$, formally $\mathcal{R}^{(i,k)} = \mathrm{Corr}\Big(\widetilde{\phi}^{(i\to k)},\, \hat{\phi}^k\Big)$, where $\mathrm{Corr}(\cdot,\cdot)$ is a pixel-wise normalized correlation function that estimates feature alignment quality. A small network (i.e., a 3×3 convolution + sigmoid activation) then modulates $\widetilde{\phi}^{(i\to k)}$ as:

$$\widehat{\phi}^{(i\to k)} = \widetilde{\phi}^{(i\to k)} \odot \sigma\big(\mathcal{R}^{(i,k)}\big), \tag{2}$$

where $\sigma(\cdot)$ is the sigmoid function, and $\odot$ denotes elementwise multiplication. This spatial adaptation step ensures that $\widehat{\phi}^{(i\to k)}$ better matches the finer-scale field $\hat{\phi}^k$, preventing naive field summation from amplifying spatial misalignments.

Beyond spatial alignment, MSF leverages uncertainty estimation to assign reliability-aware weights to each scale's deformation field. Given an uncertainty map σ^i associated with each $\hat{\phi}^i$, we define a confidence-based weighting scheme as $\omega^{(i)} = \frac{\exp\left(-\beta\, \sigma^i \odot \mathcal{R}^{(i,k)}\right)}{\sum_{j=0}^{k} \exp\left(-\beta\, \sigma^j \odot \mathcal{R}^{(j,k)}\right)}$, where $\beta > 0$ is a hyperparameter controlling the confidence scaling. This weighting scheme suppresses unreliable deformation fields while retaining informative ones. The aligned and confidence-weighted multi-scale flows are then fused as:

$$\widetilde{\Phi}^k = \sum_{i=0}^{k-1} \omega^{(i)}\, \widehat{\phi}^{(i\to k)} + \omega^{(k)}\, \hat{\phi}^k, \tag{3}$$

where $\widetilde{\Phi}^k$ denotes the fused deformation field at scale k. Finally, for additional regularization, we apply a smoothing operation using velocity field integration as $\Phi^k = \mathrm{VecInt}\big(\widetilde{\Phi}^k,\, n\big)$, where $\mathrm{VecInt}(\cdot, n)$ denotes an n-step scaling-and-squaring integration.

The fused deformation field Φ^k serves as a guidance transformation, applied to warp the infrared feature map at the next scale F_{ir}^{k+1} as $F_{ir}^{k+1} = \mathcal{T}\big(F_{ir}^{k+1},\ \Phi^k\big)$, where $\mathcal{T}(\cdot,\cdot)$ denotes spatial transformation using the fused field. This pre-alignment step mitigates large-scale errors, enabling finer-scale registration to focus on local refinements. Then, an initial flow ϕ^{k+1} and an uncertainty map σ^{k+1} are again predicted via the MCFE module, followed by iterative refinement through the FRB. This yields a refined field $\hat{\phi}^{k+1}$. The same MSF procedure then fuses $\hat{\phi}^{k+1}$ with all coarser fields $\{\hat{\phi}^0, \ldots, \hat{\phi}^k\}$, producing Φ^{k+1}. By iterating from coarse ($k = 0$) to fine ($k = K$) scales in this manner, MIR effectively reduces error accumulation and achieves a robust multi-scale registration. The final deformation field Φ^{final} fuses all fields $\{\hat{\phi}^0, \ldots, \hat{\phi}^K\}$ and warps the distorted infrared image I_{ir} into the registered infrared image I_{ir}^R.

Loss function for registration. Given the registered image I_{ir}^R and its undistorted ground truth $\overline{I}_{ir}$, we compute their 2D Fourier transforms $F_R = \mathcal{F}(I_{ir}^R)$ and $F_{GT} = \mathcal{F}(\overline{I}_{ir})$, and extract the magnitude spectra $M_R = |F_R|$ and $M_{GT} = |F_{GT}|$. These are normalized into probability distributions by $P(I) = \frac{A(I)}{\sum_{u,v} A(I)(u,v)}$, where $A(I)$ denotes the amplitude of I in the frequency domain. The frequency consistency loss $\mathcal{L}_{freq}^r$ is computed by measuring the frequency-domain similarity between the normalized distributions using the Kullback-Leibler (KL) divergence, formally:

$$\begin{aligned}\mathcal{L}_{freq}^r &= D_{KL}\Big(P\big(\overline{I}_{ir}\big) \,\Big\|\, P\big(I_{ir}^R\big)\Big) \\ &= \sum_{u,v} P\big(\overline{I}_{ir}\big)(u,v)\, \log\left(\frac{P\big(\overline{I}_{ir}\big)(u,v)}{P\big(I_{ir}^R\big)(u,v) + \varepsilon}\right),\end{aligned} \tag{4}$$

where $\varepsilon > 0$ is a small constant to avoid division by zero.

In the spatial domain, the proposed MIR leverages an uncertainty loss to handle ambiguous regions robustly. Recall that the Monte Carlo sampling-based field estimation (MCFE) module produces a field ϕ^i and an uncertainty map σ^i at each scale i, for $i \in \{0, 1, ..., K\}$. MIR utilizes these multi-scale uncertainty maps to incorporate two complementary terms that mitigate penalties in uncertain areas while discouraging excessively large uncertainties. First, an uncertainty-weighted reconstruction term:

$$\mathcal{L}_{u1} = \sum_{i=1}^{K} \Big(\frac{\big\| (I_{ir}^R)^i - (I_{\widetilde{ir}})^i \big\|}{\sigma^i + \varepsilon} \;+\; \log(\sigma^i + \varepsilon) \Big), \tag{5}$$

down weights residuals in low-confidence regions while penalizing excessively large σ^i. Second, an uncertainty regularization term $\mathcal{L}_{u2} = \sum_{i=1}^{K} \mathrm{mean}\big(\sigma^i\big)$, prevents inflating σ^i globally. The combined uncertainty loss $\mathcal{L}_{uncertainty}$ is computed by the sum of $\mathcal{L}_{u1}$ and $\mathcal{L}_{u2}$. By adaptively reducing penalties in uncertain areas and imposing global constraints on uncertainty, MIR achieves robust alignment in ambiguous regions without resorting to trivially large uncertainty everywhere. The uncertainty loss $\mathcal{L}_{uncertainty}$ together with the similarity loss $\mathcal{L}_{sim}$ and the smooth loss $\mathcal{L}_{smooth}$ forms the spatial consistency loss $\mathcal{L}_{spa}^r$. The total registration loss $\mathcal{L}_{reg}$ is as $\mathcal{L}_{reg} = \mathcal{L}_{freq}^r + \mathcal{L}_{spa}^r$.

3.2 Dual-branch Spatial-Frequency Fusion

After registration, SFRF employs a Dual-branch Spatial-Frequency Fusion (DSFF) to fuse the infrared and visible images by leveraging their complementary spatial and frequency information, as shown in Figure 2 (Fusion).

Frequency domain fusion. DSFF first apply the Fast Fourier Transform (FFT) to the registered infrared image I_{ir}^R and the visible image I_{vis}, obtaining their respective amplitude (P) and phase (A) components as $P_{ir}, A_{ir} = \mathrm{FFT}(I_{ir}^R)$, $P_{vis}, A_{vis} = \mathrm{FFT}(I_{vis})$. For each component, two 1×1 convolutional layers are applied to extract features, denoted as F_{ir}^P, F_{vis}^P, F_{ir}^A, F_{vis}^A.

To fuse the *phase* components, the DSFF block adopts a phase fusion module (PF) equipped with a heat-aware adaptive weighting mechanism. Let PF generate an attention map $a \in [0, 1]$ that adaptively selects between infrared and visible phase features, ensuring thermal information dominates in high-temperature regions while visible phase cues prevail in background areas. The final phase fusion P_f is then computed via:

$$P_f = a * F_{ir}^P + (1 - a) * F_{vis}^P + P_{vis}. \tag{6}$$

Similarly, *amplitude* fusion is achieved by a self-attention (SA) mechanism applied to the extracted amplitude features F_{ir}^A and F_{vis}^A. Let $\mathrm{SA}(\cdot)$ compute attention weights that highlight critical thermal structures in F_{ir}^A. We then obtain the fused amplitude A_f via

$$A_f = \mathrm{SA}\big(F_{ir}^A\big) \odot F_{vis}^A + A_{ir}. \tag{7}$$

This amplitude fusion mechanism highlights critical thermal structures, emphasizing regions with strong infrared responses. Finally, the fused frequency-domain image I_f^{freq} is reconstructed through the inverse Fourier transform as $I_f^{freq} = \mathrm{IFFT}(P_f, A_f)$, which integrates thermal saliency from I_{ir}^R while retaining the structural integrity of I_{vis} in the frequency domain.

Spatial domain fusion. In addition to frequency-domain fusion, DSFF also performs spatial fusion on the registered infrared image I_{ir}^R and the visible image I_{vis}. Specifically, we concatenate them along the channel dimension as $\mathbf{X} = \mathrm{concat}\big(I_{ir}^R,\ I_{vis}\big)$. The spatial fusion result is computed by $I_f^{spa} = \Phi_{spa}\big(\mathbf{X}\big)$, where Φ_{spa} denote the spatial fusion operator composed of three Restormer blocks [Zamir *et al.*, 2022]. Each Restormer block refines the concatenated features by exploiting long-range dependencies via self-attention, thereby balancing detail preservation (from I_{vis}) and thermal saliency (from I_{ir}^R) in the spatial domain.

Finally, the spatial fusion result I_f^{spa} and the frequency fusion result I_f^{freq} are combined and further refined by a stack of ResBlocks [He *et al.*, 2016] to produce the final fused output as $I_f^{final} = \Phi_{res}\Big(\mathrm{concat}\big(I_f^{spa},\ I_f^{freq}\big)\Big)$, where Φ_{res} denotes a sequence of four residual blocks, ensuring that the fused image retains spatially accurate details and frequency-consistent thermal cues.

Loss function for fusion. To encourage the final fused image I_f^{final} to preserve both the infrared-specific thermal cues and the visible structural fidelity in the frequency domain, we impose two separate constraints on its *phase* and *amplitude*. Let P_f, A_f be the phase and amplitude of the final fusion result I_f^{final}, while $\{P_{ir}, A_{ir}\}$ and $\{P_{vis}, A_{vis}\}$ are those of the ground-truth infrared and visible images. We define the frequency loss $\mathcal{L}_{freq}^f$ as $\mathcal{L}_{freq}^f = \mathcal{L}_{phase} + \mathcal{L}_{amplitude}$, where $\mathcal{L}_{phase} = \big\|P_f - P_{ir}\big\|_2^2 + \big\|P_f - P_{vis}\big\|_2^2$ and $\mathcal{L}_{amplitude} = \big\|A_f - A_{ir}\big\|_2^2 + \big\|A_f - A_{vis}\big\|_2^2$. This frequency loss ensures that the fuse result balances between infrared and visible to maintain thermal characteristics while preserving overall structural integrity. The total fusion loss $\mathcal{L}_{fu}$ consists of the frequency loss $\mathcal{L}_{freq}^f$ and the spatial loss $\mathcal{L}_{spa}^f$, formally $\mathcal{L}_{fu} = \mathcal{L}_{freq}^f + \mathcal{L}_{spa}^f$, where the $\mathcal{L}_{spa}^f$ comprising the intensity loss $\mathcal{L}_{int}$, gradient loss $\mathcal{L}_{grad}$ and the SSIM loss $\mathcal{L}_{ssim}$.

4 Experiments

Implementation Details. We train our model on an NVIDIA A100 GPU with Adam optimizer. The initial learning rate is set to $1e^{-3}$, employing an exponential decay strategy to refine the learning process over time. The training was executed with a patch size 128×128 and a batch size of 16.

Datasets and Evaluation Metrics. We conduct infrared and visible image registration and fusion experiments on four datasets: RoadScene [Xu *et al.*, 2020], MSRS [Tang *et al.*, 2022b], M^{3}FD [Liu *et al.*, 2022], and TNO [Toet, 2017]. Infrared images are randomly deformed to create misaligned image pairs. The RoadScene, MSRS, and M^{3}FD datasets are randomly split into 80% for training and 20% for testing, while the TNO dataset is used solely for testing due to its limited number of images. For evaluation, we use four metrics for registration: normalized cross-correlation (NCC), root mean square error (RMSE), max square error (MAE), and median square error (MEE). For fusion, we use the correlation coefficient (CC), visual information fidelity (VIF), structural similarity index (SSIM), mean gradient (MG) [Ma *et al.*, 2019], and edge intensity (EI) [Luo *et al.*, 2017] for evaluation.

4.1 Experiments on Infrared Image Registration

We compare our method with six state-of-the-art approaches, including GLU-Net [Truong *et al.*, 2020], CGRP [Wang *et al.*, 2022], SuperFusion [Tang *et al.*, 2022a], MURF [Xu *et al.*, 2023], IMF [Wang *et al.*, 2024], and BSAFusion [Li *et al.*, 2024b]. As shown in Table 1, our method achieves the best overall performance across all four test sets, and particularly dominates on M^{3}FD, yielding the highest NCC and consistently top-ranked MEE. These gains are visually confirmed in Figure 3, where our results produce darker difference maps with fewer global and local misalignments. In contrast, CGRP often fails to recover the correct global transformation, while SuperFusion struggles with local non-rigid offsets. Benefiting from multi-scale iterative refinement of the deformation field, our approach handles both global and local deformations, delivering more accurate registration.

Methods		RoadScene			MSRS			M^3FD			TNO		
		NCC↑	RMSE ↓	MEE ↓	NCC ↑	RMSE ↓	MEE ↓	NCC ↑	RMSE ↓	MEE ↓	NCC ↑	RMSE ↓	MEE ↓
GLU-Net [Truong *et al.*, 2020]	CVPR'20	0.805	6.629	16.133	0.551	6.318	10.438	0.729	6.110	9.983	0.471	**7.757**	**29.000**
CGRP [Wang *et al.*, 2022]	IJCAI'22	0.907	7.883	36.467	0.394	6.703	17.539	0.885	6.186	13.970	0.746	8.400	45.250
SuperFusion [Tang *et al.*, 2022a]	JAS'22	0.923	**6.011**	**13.267**	0.571	5.916	**7.978**	0.926	5.997	5.290	0.869	8.122	39.250
MURF [Xu *et al.*, 2023]	TPAMI'23	0.919	7.437	27.767	0.562	6.427	12.533	0.897	6.043	8.580	0.779	8.115	41.000
IMF [Wang *et al.*, 2024]	TCSVT'24	0.926	7.613	33.333	0.409	6.756	18.966	0.901	5.998	12.087	0.749	8.138	44.750
BSAFusion [Li *et al.*, 2024b]	AAAI'25	0.809	7.782	31.933	0.423	6.335	11.989	0.777	6.476	16.328	0.696	8.257	42.500
Ours	-	**0.947**	7.370	25.400	**0.573**	**5.818**	9.270	**0.948**	**5.073**	**5.043**	**0.873**	8.053	38.500

Table 1: Quantitative comparison of image registration. The best results are in bold, while the second-best results are underlined.

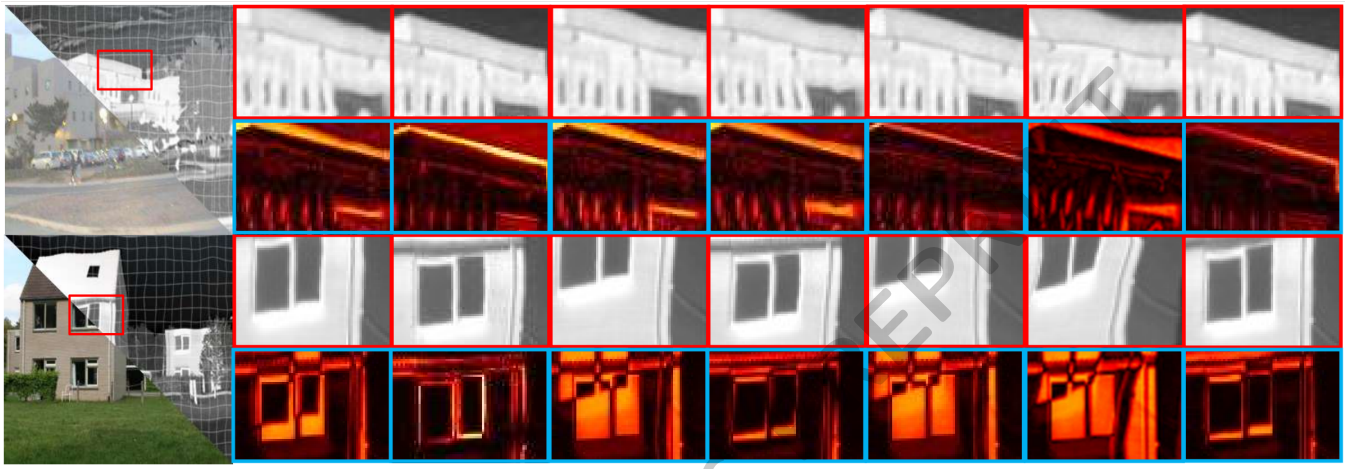

Figure 3: Qualitative results of registration. The blue boxes are the difference maps, where darker colors indicate smaller differences.

Methods		RoadScene			MSRS			M^3FD			TNO		
		CC↑	SSIM ↑	MG ↑	CC ↑	SSIM ↑	MG ↑	CC ↑	SSIM ↑	MG ↑	CC ↑	SSIM ↑	MG ↑
DIDFuse [Zhao *et al.*, 2020]	IJCAI'20	0.773	0.456	6.323	0.545	0.304	2.728	0.751	0.486	3.739	0.339	0.487	5.637
U2Fusion [Xu *et al.*, 2020]	TPAMI'20	0.752	0.464	4.975	0.565	0.577	2.278	0.777	0.674	2.891	0.351	0.532	4.388
ReCoNet [Huang *et al.*, 2022]	ECCV'22	0.771	0.538	5.087	0.586	0.384	4.546	0.794	0.656	2.690	0.297	0.554	4.588
SuperFusion [Tang *et al.*, 2022a]	JAS'22	0.785	**0.719**	5.551	0.624	0.608	3.528	0.795	**0.792**	3.412	0.320	0.602	4.111
MURF [Xu *et al.*, 2023]	TPAMI'23	0.755	0.484	5.487	0.554	0.503	2.859	0.761	0.678	3.580	0.329	0.514	4.806
CDDFusion [Zhao *et al.*, 2023]	CVPR'23	0.760	0.456	6.000	0.585	0.592	3.626	0.784	0.655	3.223	0.320	0.507	**6.100**
SegMIF [Liu *et al.*, 2023]	ICCV'23	0.761	0.547	4.512	0.600	0.506	3.889	0.783	0.696	2.836	0.325	0.528	3.762
ReFusion [Bai *et al.*, 2025b]	IJCV'24	0.735	0.468	6.464	0.584	0.643	3.413	0.763	0.739	3.528	0.292	0.546	6.076
MRFS [Zhang *et al.*, 2024]	CVPR'24	0.762	0.576	5.885	0.591	0.625	2.347	0.748	0.707	2.623	0.234	0.407	4.297
SHIP [Zheng *et al.*, 2024]	CVPR'24	0.717	0.480	6.042	0.580	0.617	3.636	0.726	0.724	3.618	0.301	0.551	5.547
BSAFusion [Li *et al.*, 2024b]	AAAI'25	0.536	0.365	5.550	0.456	0.366	3.197	0.565	0.534	3.032	0.307	0.429	4.589
Ours	-	**0.792**	0.681	**7.104**	**0.668**	**0.681**	**5.298**	**0.807**	0.760	**3.948**	**0.474**	**0.647**	5.248

Table 2: Quantitative comparison of IVIF. The best results are in bold, while the second-best results are underlined.

4.2 Experiments on IVIF

We compare our method with eleven IVIF approaches, including seven non-registration fusion methods (DIDFuse [Zhao *et al.*, 2020], U2Fusion [Xu *et al.*, 2020], CDDFuse [Zhao *et al.*, 2023], SegMif [Liu *et al.*, 2023], ReFusion [Bai *et al.*, 2025b], MRFS [Zhang *et al.*, 2024], SHIP [Zheng *et al.*, 2024]) and four registration-fusion methods (ReCoNet [Huang *et al.*, 2022], SuperFusion [Tang *et al.*, 2022a], MURF [Xu *et al.*, 2023], and BSAFusion [Li *et al.*, 2024b]). Non-registration methods take visible and deformed infrared images as input, while registration-fusion methods operate on visible and registered infrared images. As reported in Table 2, our method achieves the highest CC on all four datasets, ranking first on MSRS and consistently

Figure 4: Qualitative results of IVIF.

FRB	MSF	$\mathcal{L}_{uncertainty}$	NCC ↑	RMSE ↓	MEE ↓
	✓	✓	0.752	5.791	5.682
✓		✓	0.831	5.458	5.370
✓	✓		0.770	5.716	5.640
✓	✓	✓	0.948	5.073	5.043

Table 3: Quantitative ablation study of image registration.

$\mathcal{L}^r_{freq}$	FDF	SDF	CC ↑	SSIM ↑	MG ↑
	✓	✓	0.730	0.604	6.550
✓		✓	0.524	0.577	5.483
✓	✓		0.663	0.505	5.336
✓	✓	✓	0.792	0.681	7.104

Table 4: Quantitative ablation study of IVIF.

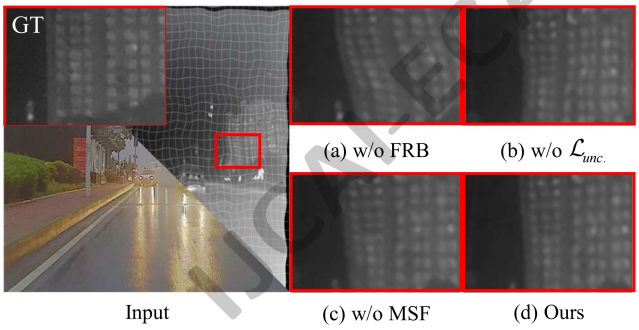

Figure 5: Qualitative ablation study of image registration.

top-two on RoadScene and M^{3}FD. These quantitative gains are further reflected in the visual results in Figure 4. Non-registration methods suffer from noticeable artifacts caused by infrared deformation, whereas registration-fusion methods still exhibit residual distortions. Our approach produces visually coherent fusion results with enhanced thermal saliency and spatial consistency, benefiting from the proposed Dual-branch Spatial-Frequency Fusion (DSFF) module.

4.3 Ablation Studies

Tables 3 and 4 show that removing any key component consistently degrades performance. For registration, discarding the GRU-based FRB notably reduces NCC/RMSE and introduces visible misalignment (Figure 5), validating the benefit of iterative refinement. Removing uncertainty modeling (either $\mathcal{L}_{uncertainty}$ or MSF) further harms stability, especially in blurry regions. For fusion, omitting $\mathcal{L}^r_{freq}$ weakens frequency-domain consistency and thermal saliency, while removing FDF or SDF degrades fusion quality by losing thermal cues or structural details, respectively (Tables 4).

5 Conclusion

We propose the Spatial-Frequency Registration and Fusion (SFRF) framework to address the misalignment and error accumulation in multi-modal image fusion. By incorporating uncertainty estimation and infrared thermal radiation distribution consistency, SFRF improves the robustness of registration and fusion across both spatial and frequency domains. Our method provides a promising solution for accurate and robust infrared and visible image fusion.

Acknowledgments

This work was partially supported by the China Postdoctoral Science Foundation (2023M730741), the National Natural Science Foundation of China (No.62302078, No.62372080), and the Key Research and Development Program of Liaoning Province under Grant (No.2023JH26/10200014).

References

[Arar *et al.*, 2020] Moab Arar, Yiftach Ginger, Dov Danon, Amit H Bermano, and Daniel Cohen-Or. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13410–13419, 2020.

[Bai *et al.*, 2025a] Haowen Bai, Jiangshe Zhang, Zixiang Zhao, Yichen Wu, Lilun Deng, Yukun Cui, Tao Feng, and Shuang Xu. Task-driven image fusion with learnable fusion loss. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 7457–7468, 2025.

[Bai *et al.*, 2025b] Haowen Bai, Zixiang Zhao, Jiangshe Zhang, Yichen Wu, Lilun Deng, Yukun Cui, Baisong Jiang, and Shuang Xu. Refusion: Learning image fusion from reconstruction with learnable loss via meta-learning. *International Journal of Computer Vision*, 133(5):2547–2567, 2025.

[Brox and Malik, 2010] Thomas Brox and Jitendra Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):500–513, 2010.

[Decarlo and Metaxas, 2000] Douglas Decarlo and Dimitris Metaxas. Optical flow constraints on deformable models with applications to face tracking. *International Journal of Computer Vision*, 38(2):99–127, 2000.

[Ding *et al.*, 2024] Xiaohan Ding, Yiyuan Zhang, Yixiao Ge, Sijie Zhao, Lin Song, Xiangyu Yue, and Ying Shan. Unireplknet: A universal perception large-kernel convnet for audio video point cloud time-series and image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5513–5524, 2024.

[Guan *et al.*, 2026] Tianwei Guan, Haozhen Wei, Yuhan Zhou, Jun Ma, Zecheng Xu, Zhiying Jiang, Jinyuan Liu, and Xingyuan Li. Domain adaptation guided infrared and visible image fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 4376–4384, 2026.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Hou *et al.*, 2020] Ruichao Hou, Dongming Zhou, Rencan Nie, Dong Liu, Lei Xiong, Yanbu Guo, and Chuanbo Yu. Vif-net: An unsupervised framework for infrared and visible image fusion. *IEEE Transactions on Computational Imaging*, 6:640–651, 2020.

[Hu *et al.*, 2024] Kun Hu, Qingle Zhang, Maoxun Yuan, and Yitian Zhang. Sfdfusion: An efficient spatial-frequency domain fusion network for infrared and visible image fusion. In *ECAI 2024*, pages 482–489. IOS Press, 2024.

[Huang *et al.*, 2022] Zhanbo Huang, Jinyuan Liu, Xin Fan, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Reconet: Recurrent correction network for fast and efficient multi-modality image fusion. In *European conference on computer Vision*, pages 539–555. Springer, 2022.

[Kirillov *et al.*, 2023] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[Li *et al.*, 2020] Jing Li, Hongtao Huo, Chang Li, Renhua Wang, and Qi Feng. Attentionfgan: Infrared and visible image fusion using attention-based generative adversarial networks. *IEEE Transactions on Multimedia*, 23:1383–1396, 2020.

[Li *et al.*, 2023] Xingyuan Li, Yang Zou, Jinyuan Liu, Zhiying Jiang, Long Ma, Xin Fan, and Risheng Liu. From text to pixels: A context-aware semantic synergy solution for infrared and visible image fusion. *arXiv preprint arXiv:2401.00421*, 2023.

[Li *et al.*, 2024a] Huafeng Li, Junyu Liu, Yafei Zhang, and Yu Liu. A deep learning framework for infrared and visible image fusion without strict registration. *International Journal of Computer Vision*, 132(5):1625–1644, 2024.

[Li *et al.*, 2024b] Huafeng Li, Dayong Su, Qing Cai, and Yafei Zhang. Bsafusion: A bidirectional stepwise feature alignment network for unaligned medical image fusion. *arXiv preprint arXiv:2412.08050*, 2024.

[Li *et al.*, 2024c] Zhuoyuan Li, Jiacheng Li, Yao Li, Li Li, Dong Liu, and Feng Wu. In-loop filtering via trained look-up tables. In *2024 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–5, 2024.

[Li *et al.*, 2025] Xingyuan Li, Zirui Wang, Yang Zou, Zhixin Chen, Jun Ma, Zhiying Jiang, Long Ma, and Jinyuan Liu. Difiisr: A diffusion model with gradient guidance for infrared image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 7534–7544, 2025.

[Li *et al.*, 2026] Xingyuan Li, Songcheng Du, Yang Zou, HaoYuan Xu, Zhiying Jiang, and Jinyuan Liu. Unifusion: A unified image fusion framework with robust representation and source-aware preservation. *arXiv preprint arXiv:2603.14214*, 2026.

[Liu *et al.*, 2022] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5802–5811, 2022.

[Liu *et al.*, 2023] Jinyuan Liu, Zhu Liu, Guanyao Wu, Long Ma, Risheng Liu, Wei Zhong, Zhongxuan Luo, and Xin Fan. Multi-interactive feature learning and a full-time multi-modality benchmark for image fusion and segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8115–8124, 2023.

[Liu *et al.*, 2024a] Jinyuan Liu, Xingyuan Li, Zirui Wang, Zhiying Jiang, Wei Zhong, Wei Fan, and Bin Xu. Promptfusion: Harmonized semantic prompt learning for infrared and visible image fusion. *IEEE/CAA Journal of Automatica Sinica*, 2024.

[Liu *et al.*, 2024b] Jinyuan Liu, Guanyao Wu, Zhu Liu, Long Ma, Risheng Liu, and Xin Fan. Where elegance meets precision: towards a compact, automatic, and flexible framework for multi-modality image fusion and applications. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 1110–1118, 2024.

[Liu *et al.*, 2025a] Jinyuan Liu, Bowei Zhang, Qingyun Mei, Xingyuan Li, Yang Zou, Zhiying Jiang, Long Ma, Risheng Liu,

and Xin Fan. Dcevo: Discriminative cross-dimensional evolutionary learning for infrared and visible image fusion. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2226–2235, 2025.

[Liu *et al.*, 2025b] Yating Liu, Yang Zou, Xingyuan Li, Xingyue Zhu, Kaiqi Han, Zhiying Jiang, Long Ma, and Jinyuan Liu. Toward a training-free plug-and-play refinement framework for infrared and visible image registration and fusion. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 1268–1277, 2025.

[Liu *et al.*, 2026] Jinyuan Liu, Xingyuan Li, Qingyun Mei, Haoyuan Xu, Zhiying Jiang, Long Ma, Risheng Liu, and Xin Fan. Bridging human evaluation to infrared and visible image fusion. *arXiv preprint arXiv:2603.03871*, 2026.

[Luo *et al.*, 2017] Xiaoqing Luo, Zhancheng Zhang, Cuiying Zhang, and Xiaojun Wu. Multi-focus image fusion using hosvd and edge intensity. *Journal of Visual Communication and Image Representation*, 45:46–61, 2017.

[Ma *et al.*, 2019] Jiayi Ma, Yong Ma, and Chang Li. Infrared and visible image fusion methods and applications: A survey. *Information fusion*, 45:153–178, 2019.

[Ren *et al.*, 2021] Long Ren, Zhibin Pan, Jianzhong Cao, and Jiawen Liao. Infrared and visible image fusion based on variational auto-encoder and infrared feature compensation. *Infrared Physics & Technology*, 117:103839, 2021.

[Tang *et al.*, 2022a] Linfeng Tang, Yuxin Deng, Yong Ma, Jun Huang, and Jiayi Ma. Superfusion: A versatile image registration and fusion network with semantic awareness. *IEEE/CAA Journal of Automatica Sinica*, 9(12):2121–2137, 2022.

[Tang *et al.*, 2022b] Linfeng Tang, Jiteng Yuan, Hao Zhang, Xingyu Jiang, and Jiayi Ma. Piafusion: A progressive infrared and visible image fusion network based on illumination aware. *Information Fusion*, 83:79–92, 2022.

[Toet, 2017] Alexander Toet. The tno multiband image data collection. *Data in brief*, 15:249, 2017.

[Truong *et al.*, 2020] Prune Truong, Martin Danelljan, and Radu Timofte. Glu-net: Global-local universal network for dense flow and correspondences. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6258–6268, 2020.

[Wang and Zhang, 2020] Jian Wang and Miaomiao Zhang. Deepflash: An efficient network for learning-based medical image registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4444–4452, 2020.

[Wang *et al.*, 2022] Di Wang, Jinyuan Liu, Xin Fan, and Risheng Liu. Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration. *arXiv preprint arXiv:2205.11876*, 2022.

[Wang *et al.*, 2024] Di Wang, Jinyuan Liu, Long Ma, Risheng Liu, and Xin Fan. Improving misaligned multi-modality image fusion with one-stage progressive dense registration. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[Wang *et al.*, 2025] Zirui Wang, Jiayi Zhang, Tianwei Guan, Yuhan Zhou, Xingyuan Li, Minjing Dong, and Jinyuan Liu. Efficient rectified flow for image fusion. *Advances in Neural Information Processing Systems*, 2025.

[Xiao and Xiong, 2025] Zeyu Xiao and Zhiwei Xiong. Incorporating degradation estimation in light field spatial super-resolution. *Computer Vision and Image Understanding*, 252:104295, 2025.

[Xiao *et al.*, 2025] Zeyu Xiao, Zhuoyuan Li, and Wei Jia. Occlusion-embedded hybrid transformer for light field super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 8700–8708, 2025.

[Xiao *et al.*, 2026] Zeyu Xiao, Zhuoyuan Li, Yang Zhao, Yu Liu, Zhao Zhang, and Wei Jia. Learning dual modality interactions for event-based motion deblurring. *IEEE Transactions on Multimedia*, 2026.

[Xu *et al.*, 2020] 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.

[Xu *et al.*, 2023] Han Xu, Jiteng Yuan, and Jiayi Ma. Murf: Mutually reinforcing multi-modal image registration and fusion. *IEEE transactions on pattern analysis and machine intelligence*, 45(10):12148–12166, 2023.

[Ye *et al.*, 2022] Yuanxin Ye, Tengfeng Tang, Bai Zhu, Chao Yang, Bo Li, and Siyuan Hao. A multiscale framework with unsupervised learning for remote sensing image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.

[Zamir *et al.*, 2022] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022.

[Zhang *et al.*, 2024] Hao Zhang, Xuhui Zuo, Jie Jiang, Chunchao Guo, and Jiayi Ma. Mrfs: Mutually reinforcing image fusion and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26974–26983, 2024.

[Zhao *et al.*, 2020] Zixiang Zhao, Shuang Xu, Chunxia Zhang, Junmin Liu, Pengfei Li, and Jiangshe Zhang. Didfuse: Deep image decomposition for infrared and visible image fusion. *arXiv preprint arXiv:2003.09210*, 2020.

[Zhao *et al.*, 2023] Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Shuang Xu, Zudi Lin, Radu Timofte, and Luc Van Gool. Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5906–5916, 2023.

[Zhao *et al.*, 2026] Rui Zhao, Ruiqin Xiong, Dongkai Wang, Shiyu Xuan, Jian Zhang, Xiaopeng Fan, and Tiejun Huang. Spike camera optical flow estimation based on continuous spike streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4):4756–4770, 2026.

[Zheng *et al.*, 2024] Naishan Zheng, Man Zhou, Jie Huang, Junming Hou, Haoying Li, Yuan Xu, and Feng Zhao. Probing synergistic high-order interaction in infrared and visible image fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 26384–26395, 2024.

[Zou *et al.*, 2026] Yang Zou, Zhixin Chen, Zhipeng Zhang, Xingyuan Li, Long Ma, Jinyuan Liu, Peng Wang, and Yanning Zhang. Contourlet refinement gate framework for thermal spectrum distribution regularized infrared image super-resolution. *International Journal of Computer Vision*, 134(1):23, 2026.