ICJK
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2025.874404

RESEARCH ARTICLE

Salient Feature-Driven Bimodal Video Mimic Fusion Algorithm

Yanchen Meng[1], **Fengbao Yang**[1,*], **Linna Ji**[1], **Xiaoxia Wang**[1] **and Xiaoming Guo**[1]

[1] School of Information and Communication Engineering, North University of China, Taiyuan 030051, China

Abstract

In complex dynamic environments, infrared and visible video sequences exhibit highly variable and unpredictable feature distributions. Existing fusion algorithms with fixed architectures cannot adaptively respond to these dynamic feature changes, resulting in blurred fusion outcomes and the loss of critical detail information. To address this limitation, we propose a salient feature-driven mimic fusion algorithm that continuously monitors feature variations and dynamically reconfigures the fusion architecture to maintain optimized fusion performance. First, we extract amplitude and frequency attributes from infrared and visible video features and perform weighted fusion to calculate single-modality temporal features and cross-modal intra-frame difference features. Second, based on clustering statistical properties of feature distributions, we construct possibility distribution functions to quantify the degree of feature variation, design synthesis rules to derive comprehensive possibility values of feature change, and utilize significantly changing features as driving factors for subsequent mimic variant adjustments. Building upon this foundation, we establish a fusion validity evaluation function by analyzing correlation coefficients between feature changes and fusion quality metrics, and accordingly construct mapping relationships between features and mimic variants. Finally, we determine the optimal mimic variant combination by synthesizing various feature change characteristics to implement mimic fusion. Experimental evaluation demonstrates that our proposed method significantly outperforms existing approaches in adaptive fusion performance in dynamic scenes, with superior preservation of edge and texture details in the fusion results.

Keywords: mimic fusion, video fusion, infrared and visible image, possibility theory.

1 Introduction

Infrared and visible video fusion, a key technology in multimodal intelligent detection [1], effectively overcomes the physical limitations of single sensors [2] by deeply integrating rich textural details from the visible modality with thermal radiation information from the infrared modality. This effectively addresses perception failures in challenging environments such as low illumination and strong occlusion. The technology demonstrates irreplaceable value in

Academic Editor:
Shutao Li

Submitted: 09 May 2025
Accepted: 19 March 2026
Published: 17 April 2026

Vol. 3, **No.** 2, 2026.
10.62762/CJIF.2025.874404

***Corresponding author:**
✉ Fengbao Yang
yfengb@163.com

Citation
Meng, Y., Yang, F., Ji, L., Wang, X., & Guo, X. (2026). Salient Feature-Driven Bimodal Video Mimic Fusion Algorithm. *Chinese Journal of Information Fusion*, 3(2), 74–92.

applications including target recognition, military reconnaissance, situational awareness, and smart sensing for autonomous driving [3].

However, in practical detection and fusion tasks, environmental characteristics such as sudden changes in unstructured detection scenarios and interference from adverse weather cause the features of dual-modal videos to exhibit variable and unpredictable characteristics. This makes adaptive algorithms prone to imbalance in modal weight allocation, and the fusion architecture cannot adjust dynamically according to feature changes, leading to a decline in adaptability [4]. Various methods have been proposed over recent decades to address infrared and visible video fusion problems. These methods are primarily categorized as traditional methods and deep learning approaches. Traditional methods mainly include those based on multi-scale transform (MST), saliency, and sparse representation (SR). Multi-scale transform is one of the most mature and active fields in image fusion [5]. The consistency with human visual characteristics gives the fusion results good performance. For example, Zhu et al. [6] used three different fusion rules at different scales to obtain weight maps for each scale, significantly improving the performance of the fusion results in terms of human visual perception. Saliency methods design fusion rules or parameters by extracting salient regions of the image [7]; sparse representation methods design fusion rules based on the statistical characteristics of sparse coefficients [8]. These methods require manual construction or selection of fusion strategies for specific scene feature distributions. Although they perform well in static scenes, they cannot perceive scene changes to adjust the fusion architecture. Once the preset rule strategy does not match the actual requirements, it can easily cause severe fluctuations in fusion performance.

In deep learning, researchers have developed various CNN-based fusion models such as ResNet [9] and IFCNN [10]. However, these non-end-to-end approaches struggle to identify effective manual fusion rules [11] and still rely on predetermined architectures. For instance, IFCNN extracts features through two convolutional layers with shared weights, followed by manually selected fusion rules. End-to-end network models eliminate the arbitrariness of manual strategy design. Nevertheless, these models are difficult to optimize, and their architectures remain fixed after training, unable to adapt to scene variations. Moreover, deep learning methods heavily depend on learning feature distributions from large-scale training data. This leads to severe performance degradation in applications with insufficient training samples.

The fundamental issue lies in the inherent conflict between the fixed architecture of fusion methods and the dynamic nature of video sequences. Once designed, the algorithm's processing pipeline and fusion strategy remain static, incapable of dynamic adjustment according to significant feature changes. Meanwhile, the limited predefined fusion strategies cannot handle the diverse and unpredictable feature variations in video scenes, easily causing decision mismatches in complex changing environments. To address these challenges, this paper proposes a salient feature-driven mimic fusion method. We establish a complex mapping relationship between feature variations, fusion architecture, and fusion performance. Using significantly changing features as driving factors, we achieve real-time dynamic adjustment of mimic fusion variants. This provides a new approach for significantly enhancing infrared and visible video fusion quality in complex dynamic scenarios. The main contributions are as follows:

(1) We propose a dynamic perception mechanism for salient features based on possibility distribution. By constructing possibility distribution functions and designing synthesis rules, we achieve quantitative description of feature variations in complex dynamic videos, providing precise driving signals for dynamic adjustment of mimic fusion strategies.

(2) We establish a set-valued mapping relationship among features, mimic variants, and fusion performance. Through correlation analysis between feature changes and fusion quality metrics, we construct a fusion validity evaluation function, thereby building a set-valued mapping library between feature changes and optimal fusion variants to support decision-making in mimic fusion strategy selection.

(3) Extensive experiments fully demonstrate the superior fusion performance of our method. Results from multiple typical dynamic scenarios show that our approach outperforms existing mainstream methods in both subjective visual effects and objective evaluation metrics, significantly enhancing the preservation of edge and texture features.

The remainder of this paper is organized as follows: Section 2 reviews related work on infrared and visible image fusion and mimic fusion methods.

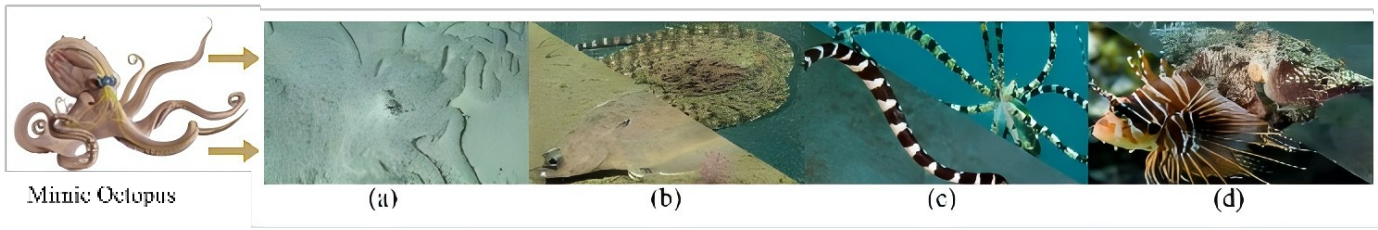

Figure 1. Mimic octopus. (a) Camouflage state of the mimic octopus, (b) Imitating a flatfish, (c) Imitating a sea snake, (d) Imitating a lionfish.

Section 3 provides a comprehensive overview of our proposed approach. Section 4 presents and analyzes comparative experimental results. Section 5 concludes the paper with discussion.

2 Related Work

2.1 Adaptive Image Fusion

To address the limitations of methods that rely on manually constructing and selecting fusion strategies based on prior knowledge - specifically their insufficient adaptability in dynamic fusion scenarios - researchers have explored numerous adaptive fusion methods.

2.1.1 Adaptive Fusion Strategy Generation Mechanisms

To enhance the environmental adaptability of algorithms, researchers began exploring dynamic generation mechanisms for fusion strategies [7]. Xu et al. [12] dynamically evaluated the importance of feature maps using a classifier and generated pixel-level fusion rules based on saliency maps. Tang et al. [13] proposed CAMF, a learnable fusion rule based on class activation mapping, which reduced reliance on manually designed rules by automatically measuring the importance of each deep feature. Liu et al. [14] learned a CNN model to dynamically generate fusion rules. Luo et al. [15] proposed IFSepR, an independent representation learning framework that designs universal fusion rules based on separated representation learning, specifically for integrating private features separated from images. While this adaptive design reduces manual intervention, its adaptability to feature distribution changes in unseen scenes is limited, making it difficult to guarantee stability in scenarios with extreme variations.

2.1.2 Dynamic Selection Mechanisms for Fusion Strategies

To overcome the limitations of single fusion strategies, many researchers have turned to dynamic selection mechanisms for fusion strategies. These methods dynamically switch fusion rules during the fusion process to cope with scene changes. As an early exploration of dynamic fusion strategies, Saeedi et al. [16] skillfully integrated pixel-based, region-based, and weighted average rules using fuzzy logic, and dynamically selected among three fusion rules based on dissimilarity measures of the source images to adaptively adjust the fusion strategy. This research fully demonstrated the potential of dynamic rule selection for improving fusion performance in complex scenes. Tong et al. [17] selected three image quality assessment parameters as inputs to a fuzzy system to obtain fusion weights, and used the classification results to guide the fusion process. Ji et al. [18] proposed a quality-based comprehensive intuitionistic possible set method, which dynamically selects the optimal fusion algorithm by constructing scoring functions for multiple fusion algorithms, effectively improving the scientific nature and accuracy of fusion strategy selection.

2.2 Mimic Fusion

The mimic octopus serves as a "transformer" in the biological world. Its core capability lies in accurately simulating up to fifteen different marine species—including lionfish, flounders, sea snakes, and sea anemones—by flexibly altering its body color, tentacle morphology, and overall appearance in response to different environmental threats. This allows it to cope with diverse survival challenges (see Figure 1). This remarkable polymorphic mimicry ability primarily relies on two core biological mechanisms:

(1) Rapid Perception and Discrimination Capability: The mimic octopus possesses a highly developed perceptual system. Through interaction with its environment, it establishes a stable mapping relationship between threat characteristics and mimicry transformations within its neural system. This enables rapid threat perception and retrieval of optimal response strategies.

(2) Powerful Learning and Imitation Capability: In harsh habitats, the mimic octopus learns

and memorizes the primary appearance and movement characteristics of organisms it imitates. Through repeated mimicry attempts and practice, it achieves optimal mimicry transformation effects. When encountering specific predators or threats subsequently, the octopus rapidly adopts the most suitable mimicry form based on its accumulated practical experience, thereby effectively deceiving or deterring the threat.

Inspired by multi-mimicry and context-driven switching mechanisms, mimic fusion methods [19–27] abstract the behavioral strategy of the mimic octopus into a multi-variant reconfigurable fusion framework. By organizing candidate fusion algorithms, fusion rules, and parameter combinations into a strategy library, these methods address the performance degradation commonly observed in fixed-architecture models under complex dynamic scenes. Unlike conventional adaptive fusion approaches that adjust parameters within a fixed architecture, mimic fusion emphasizes structural adaptability. It enables dynamic architectural variation by selecting or reconstructing suitable variant combinations online according to scene characteristics and inter-modal difference features. This transition from parameter adaptation to structural adaptation has been shown to improve video fusion performance in complex dynamic environments.

However, several limitations remain. First, existing mimic fusion methods are primarily driven by intra-frame difference features between corresponding bimodal frames, while inter-frame temporal evolution in video sequences is insufficiently modeled. Consequently, dynamic information in the temporal dimension is not fully exploited. Second, conventional mimic fusion quantifies feature variation using the absolute difference of feature values. This static amplitude-based measure reflects instantaneous value discrepancies rather than the significance of the change process. In addition, most methods construct the fusion effectiveness function based on a single evaluation metric, which may overemphasize specific features and introduce bias in variant selection. To address these limitations, this study adopts feature saliency variation as the driving mechanism and develops an integrated framework that combines possibility-theory-based quantification of difference-feature variation with multi-dimensional mimic variant evaluation. Spatial and temporal variations are jointly characterized, and variant selection is determined through multi-metric evaluation, thereby reducing the bias induced by reliance on a single metric.

Beyond mimic fusion, several biomimetic fusion approaches inspired by other biological visual mechanisms have been proposed in recent years, such as infrared-visible image fusion models based on the rattlesnake visual receptive field and bimodal cells [28, 29], as well as methods based on biologically inspired empirical mode decomposition and color opponency mechanisms [30]. These approaches typically emulate specific visual pathways or receptive field structures and construct relatively fixed or weakly adaptive multi-level fusion operators or network architectures. Their adaptability is mainly reflected in parameter tuning or limited mode switching. In contrast, the present work does not replicate a particular visual structure. Instead, it abstracts the multi-mimic behavioral strategy of the mimic octopus at the strategic level, constructs a multi-layer fusion variant library, and drives variant combination selection through feature saliency variation. Structural adaptive fusion for complex scenes is thus achieved within a unified framework.

3 Methodology

The workflow of the proposed salient feature-driven bimodal video mimic fusion algorithm is illustrated in Figure 2. It primarily encompasses bimodal video inter-frame and intra-frame feature extraction, salient feature change quantification based on possibility distribution synthesis, construction of fusion validity evaluation functions, and selection and configuration of mimic variants based on set-valued mapping libraries. First, to effectively capture complex feature changes in dynamic scenes, six complementary difference features are simultaneously extracted and characterized by their amplitude and frequency attributes. These features are subsequently used to calculate single-modality inter-frame and cross-modal intra-frame difference features. Second, possibility distribution functions are constructed from clustered feature distributions to measure the likelihood of feature changes. Inter-frame and intra-frame changes are then combined by nonlinear weighted rules to produce the final saliency values. When more pronounced non-stationary variations occur in scene content or imaging conditions, the saliency measure generates a stronger driving response. This response guides the subsequent set-valued mapping library to retrieve and combine mimic variants and parameters that better match the current variation pattern.

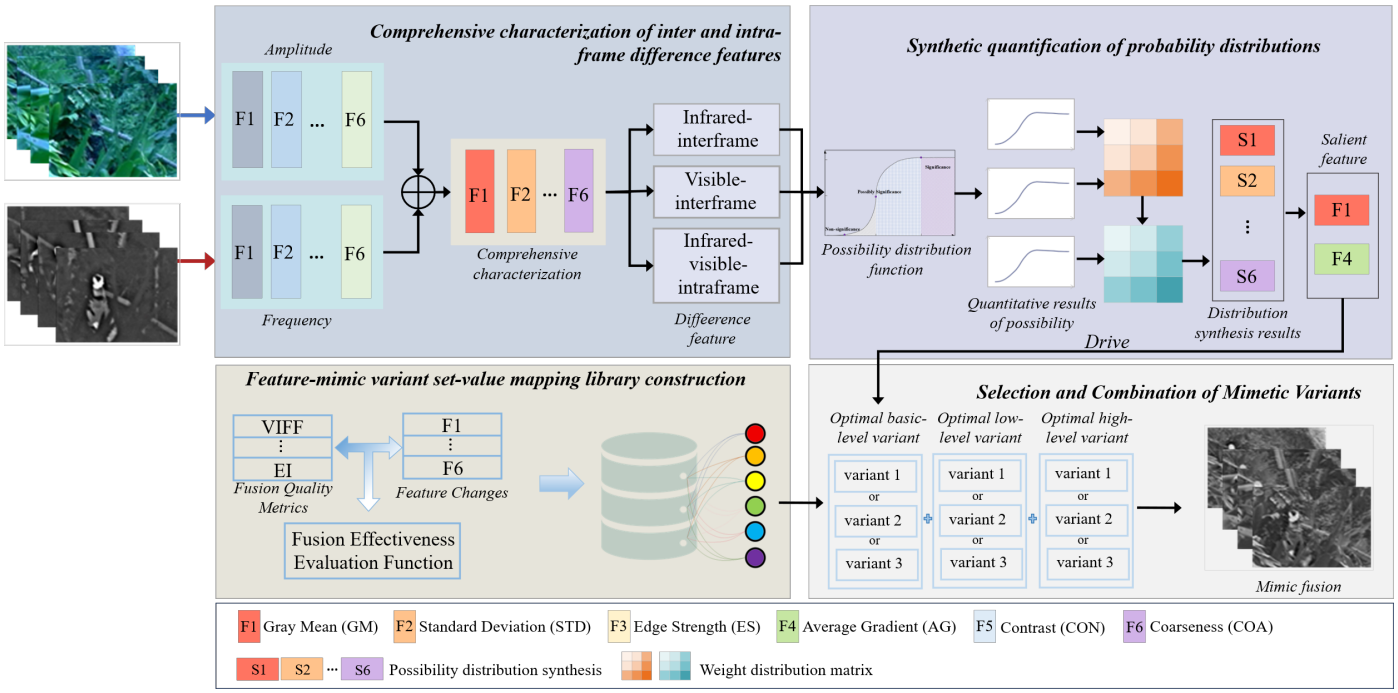

Figure 2. Workflow diagram of the mimic fusion method.

Finally, a fusion validity evaluation function is established by applying Spearman correlation analysis to characterize the association between feature changes and fusion quality metrics. Based on this function, a feature–mimic-variant set-valued mapping library is constructed to support the selection and configuration of mimic variants during the fusion process.

3.1 Bimodal Video Inter-frame and Intra-frame Feature Extraction

To achieve accurate perception of scene dynamics, this study requires the extraction of key features from infrared and visible videos that can represent their structural information and visual perception. Due to differences in the imaging mechanisms of infrared and visible video images, they exhibit significant complementarity in terms of brightness, edges, and texture details, as shown in Figure 3. Therefore, this study selects six features as the fundamental feature set $F1, F2, F3, F4, F5, F6$ [21], including gray mean (GM), standard deviation (STD), edge strength (ES), average gradient (AG), contrast (CON), and coarseness (COA). Among these, gray mean (GM, $F1$) and contrast (CON, $F5$) represent the global brightness and local dynamic range of the image, respectively, which are crucial for perceiving sudden illumination changes and the appearance of salient targets. Standard deviation (STD, $F2$) and coarseness (COA, $F6$) describe the dispersion of pixel values and the coarse texture of the image, jointly measuring the information richness of the image. Edge strength (ES, $F3$) and average gradient (AG, $F4$) focus on the local structure and detail clarity of the image, serving as core indicators for evaluating whether the fusion results preserve key contours and textures. These six types of features together form a multi-angle feature description set, providing comprehensive data support for subsequent quantification of scene changes. The specific feature characterization process is described below.

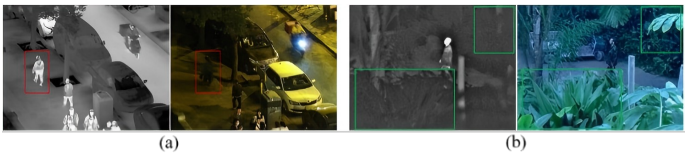

Figure 3. Infrared and visible video frames.

First, infrared and visible video frames are processed block by block without overlap using an $m \times n$ smoothing window. The feature amplitude for each block is computed as:

$$\bar{I}^r_{p,q} = f_r(I_{p,q}) \tag{1}$$

$$\bar{V}^r_{p,q} = f_r(V_{p,q}) \tag{2}$$

where $I_{p,q}$ and $V_{p,q}$ represent the (p,q)th block, f_r is the feature extraction function, $\bar{I}^r_{p,q}$ and $\bar{V}^r_{p,q}$ are the corresponding amplitudes of feature F_r for infrared and visible video frames. The cross-modal intra-frame

difference feature amplitude between infrared and visible image blocks is:

$$\Delta_{p,q}^{r} = |\bar{I}_{p,q}^{r} - \bar{V}_{p,q}^{r}| \tag{3}$$

The frequency of difference features reflects the distribution of amplitudes. In this paper, histogram statistical analysis is performed on the difference feature amplitude $\Delta_{p,q}$ of each block, the frequency of difference features in each interval is calculated as $N_{\text{space}_l}^{r}$, and the frequency weight is computed as:

$$P_{\text{space}_l}^{r} = \frac{N_{\text{space}_l}^{r}}{H \times W/m \times n} \tag{4}$$

Through weighted calculation, we obtain the comprehensive representation of amplitude and frequency for each interval of intra-frame difference features:

$$\omega_{\text{space}_l}^{r} = P_{\text{space}_l}^{r} \times X_{\text{space}_l}^{r} \tag{5}$$

where $X_{\text{space}_l}^{r}$ is the center value of feature amplitude in each interval. Finally, the cross-modal intra-frame difference feature value for the tth frame is calculated as:

$$IV\_DF^{r,t} = \sum_{l=1}^{L} \omega_{\text{space}_l}^{r} \tag{6}$$

Similarly, frequency statistical analysis is conducted on single-modality inter-frame difference features, the weighted comprehensive representation of amplitude and frequency for each interval is derived, and the single-modality inter-frame difference feature values between consecutive frames t and $t-1$ are computed as:

$$IR\_TF^{r,t} = \left| \sum_{l=1}^{L} IR_{\text{time}_l}^{r,t} - \sum_{l=1}^{L} IR_{\text{time}_l}^{r,t-1} \right| \tag{7}$$

$$VIS\_TF^{r,t} = \left| \sum_{l=1}^{L} VIS_{\text{time}_l}^{r,t} - \sum_{l=1}^{L} VIS_{\text{time}_l}^{r,t-1} \right| \tag{8}$$

3.2 Possibility Distribution Synthesis Quantification

In complex dynamic video scenes, feature changes exhibit unpredictability and uncertainty, making it difficult to assess them using a fixed and precise threshold. To address this, this section proposes a possibility distribution synthetic quantification method to model this uncertain continuous change process. This method transforms the absolute changes in feature values into the relative possibility of significant change occurring. The flowchart of its distribution synthesis module is shown in Figure 4.

First, based on the intra-frame difference features $IV_{DF^{r,t}}$ and inter-frame difference features $IR_{TF^{r,t}}$ and $VIS_{TF}^{r,t}$ extracted from historical frames, the segmentation points of the possibility distribution function are calculated through K-medoids clustering. Statistical analysis of historical data distribution reveals that the difference feature data distribution is right-skewed, with data mainly concentrated on the left side of the mean. Directly using Euclidean distance would likely amplify the influence of extreme values, causing the clustering centers to deviate from the main body of data. Therefore, Manhattan distance is adopted to mitigate the impact of outliers:

$$J = \sum_{j=1}^{K} \sum_{i \in S_j} |x_i - C_j| \tag{9}$$

where $x_i \in \{IR_{TF}^{r,t}, VIS_{TF}^{r,t}\}, C_j$ is the jth centroid, S_j is the sample set of cluster j, and $K = 3$ is the number of clusters. For all features, the segmentation points B_k^r are calculated as:

$$B_k^r = \frac{c_k^{r+1} + c_{k+1}^{r+1}}{2},\ k = 1, 2 \tag{10}$$

where $C^r = \{C_1^r, C_2^r, C_3^r\}$ represents the cluster centroids. Then, based on the cluster centroids, we design the following possibility distribution function:

$$\begin{aligned} \Pi(F^{r,t}) &= [F^{r,t}, C_1^r, C_2^r, C_3^r; 1] \\ &= \begin{cases} 0, & F^{r,t} \le C_1^r \\ \mu_1^r (F^{r,t} - C_1^r)^2, & C_1^r < F^{r,t} \le C_2^r \\ 1 + \mu_2^r (F^{r,t} - C_3^r)^2, & C_2^r < F^{r,t} \le C_3^r \\ 1, & F^{r,t} > C_3^r \end{cases} \end{aligned} \tag{11}$$

The function maps feature variation inputs to saliency possibility values through three centroids. Regions where the feature variation is lower than μ_1 correspond to saliency possibility values close to 0, whereas regions exceeding μ_3 correspond to values close to 1. The interval between μ_1 and μ_3 represents a fuzzy transition region of saliency, in which the centroid μ_2 defines the point of maximum uncertainty with a possibility value of 0.5. Based on the three saliency levels established through clustering and

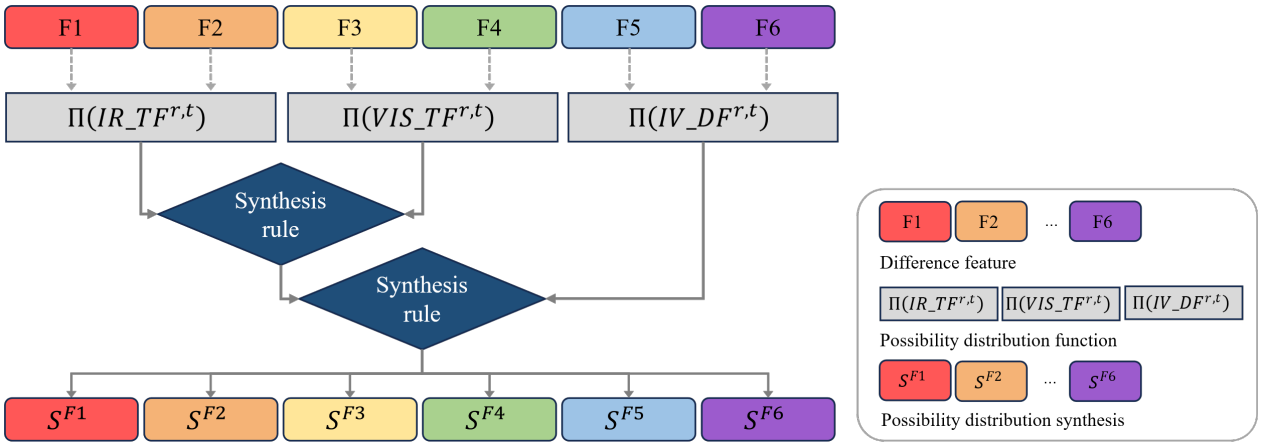

Figure 4. Schematic diagram of possibility distribution synthetic quantification.

possibility distribution analysis, a 3×3 saliency weight distribution matrix M is constructed, as illustrated in Figure 5(a) and (b). This matrix covers all possible saliency combinations and provides weighting coefficients for subsequent distribution synthesis. The weight distribution matrix is shown in Figure 9(a). The elements a in matrix M are determined according to the proportions of different cross-modal saliency combinations, including low-low ($\frac{1}{9}$), low-medium or medium-low ($\frac{3}{9}$), medium-medium ($\frac{4}{9}$), low-high or high–low ($\frac{6}{9}$), medium–high or high–medium ($\frac{8}{9}$), and high–high (1).

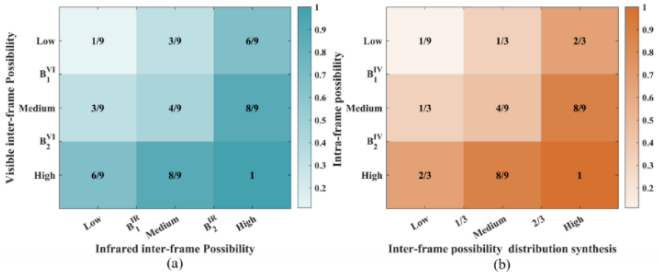

Figure 5. Weight distribution matrix.

To comprehensively evaluate the overall scene variation, we design a possibility distribution synthesis rule based on a weight distribution matrix and nonlinear weighting. This rule generates a unified comprehensive feature change saliency metric $S^{(r,t)}$ by performing weighted fusion of the possibility values from intra-frame difference features and inter-frame difference features. The inter-frame possibility distributions are synthesized as shown in Eq (12), where α represents the weight corresponding to the salient combination region in the value distribution matrix:

$$\begin{aligned}\Pi_{IRVI}^{r,t} = &\max\left(\Pi\left(IR_{TF}^{r,t}\right), \Pi\left(VI_{TF}^{r,t}\right)\right)\max\left(a, 1-\alpha\right) \\ &+ \min\left(\Pi\left(VI_{TF}^{r,t}\right), \Pi\left(IR_{TF}^{r,t}\right)\right)\min\left(a, 1-\alpha\right)\end{aligned} \tag{12}$$

Similarly, based on the weight distribution matrix in Figure 5(b), the intra-frame possibility distribution $\Pi\left(IV_{DF}^{r,t}\right)$ is synthesized again with the inter-frame synthetic possibility distribution $\Pi_{IR-VI}^{r,t}$ according to the synthesis rule, yielding the final possibility distribution synthesis result $S^{(r,t)}$. This comprehensively quantifies the saliency of changes for each feature.

3.3 Construction of Fusion Validity Evaluation Function

The Spearman correlation coefficient, as shown in Eq (13), is used to measure the nonlinear correlation between two variables. To build an effective fusion validity evaluation function for adaptively configuring fusion variants according to salient feature changes, Spearman correlation analysis is first employed to quantify the association between feature changes and fusion quality metrics. The evaluation function is then constructed from the correlation coefficients between each feature and the corresponding metrics.

$$\rho_{Spearman} = 1 - \frac{6\sum_{t=1}^{T} d_t^2}{T(T^2-1)} \tag{13}$$

where $d_t = R(S^{r,t}) - R(m_{i,t})$ is the difference in ranks between feature saliency $S^{r,t}$ and metric $m_{i,t}$ in the tth sample. Multiple experiments were conducted on several infrared and visible video datasets and the Spearman correlation coefficient $\rho_{i,r,k}$ was calculated

for each feature-metric pair. To screen out quality evaluation metrics strongly correlated with various feature changes, significance testing is performed on their correlation values. Based on the central limit theorem, assuming $\rho_{i,r,k}$ follows a normal distribution, its standard error is shown in Eq (14). Using a t-distribution with a degree of freedom of $I-1$ and a confidence level of 90% ($\alpha = 0.1$) as the upper bound of the confidence interval, as shown in Eq (15), metrics with correlation above this confidence level are selected as strongly correlated indicators, as shown in Table 1.

$$SE_r = \frac{\sigma_r = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I}(\rho_{i,r,k} - \mu_r)^2}}{\sqrt{I}} \tag{14}$$

$$upper = \mu_r + t_{\alpha/2,I-1} \cdot SE_r \tag{15}$$

Based on the derived correlation coefficients, the fusion validity evaluation function for each feature is formulated, as shown in equation (16).

$$\phi_r = \sum_{m_i \in M_{strong,r}} \frac{\rho_{i,r,k}}{\sum_{m_j \in M_{strong,r}} \rho_{i,r,k}} \cdot m_i \tag{16}$$

3.4 Construction of Set-Valued Mapping Library

The mimic high-level variants, low-level variants, and base-level variants included in this experiment are constructed as follows:

The high-level variant set includes 7 classic multi-scale fusion algorithms, specifically Wavelet Packet Transform (WPT) [31], Stationary Wavelet Transform (SWT) [32], Non-Subsampled Shearlet Transform (NSST) [33], Non-Subsampled Contourlet Transform (NSCT) [34], Laplace Pyramid Transform (LP) [35], Dual-Tree Complex Wavelet Transform (DTCWT) [36], and Curvelet Transform (CVT) [37]. These algorithms have been widely recognized in the field of image fusion and can provide reliable benchmarks and comparisons for this research.

The low-level variants primarily consist of fusion rules for high and low frequencies, which can be combined arbitrarily according to different scenes during the fusion process. Low-frequency fusion rules include Simple Adaptive Weighted average (SAW), Maximum Coefficient (MC), Measure of Energy based (ME), and Window Weighted Average (WWA); high-frequency fusion rules include Absolute Maximum Coefficient (AMC), Maximum Coefficient (MC), Measure of Energy based (ME), Window Gradient based (WG), Window Standard Deviation based (WSD), and Sum-Modified-Laplacian based (SML).

The base-level variant set refers to the decomposition levels, internal fusion parameters, etc., of the fusion algorithm, all set according to the algorithm itself.

The process of constructing the set-valued mapping between difference features and various mimic variants is divided into three stages, progressively determining the optimal configuration of high-level, low-level, and base-level variants corresponding to different difference features. In the high-level variant selection stage, the low-level variants and base-level variants are maintained constant, and the fusion algorithm with optimal fusion effect is selected for each difference feature through experiments. For the selection of low-level variants, the fusion algorithm and fusion parameters are fixed, permutations and combinations of low-frequency rules and high-frequency rules are performed, and the optimal fusion rules for each difference feature are determined. Finally, in the base-level variant selection stage, internal fusion parameters are modified based on the selected fusion algorithm and fusion rules to determine the optimal fusion parameter configuration. This process ultimately yields a ranking of the merits of various variants under different features and, in conjunction with confidence levels, determines the best mimic variants for each feature.

4 Experiments

This section details the overall workflow of the proposed method and analyzes the experimental results, including datasets, evaluation metrics, comparison methods, details of possibility distribution function construction, set-valued mapping library construction, algorithm performance verification, and fusion result analysis.

4.1 Datasets

Three distinct datasets are employed to validate the efficacy of the proposed method. The first dataset is the Bristol Eden Project Multi-Sensor Data Set (BEPMS) [38], which comprises sequences of a soldier traversing a jungle environment, with camera motion tracking the subject amid complex dynamic scene variations. The image resolution is 576 × 480. The second dataset is the Object Tracking and Classification in and Beyond the Visible Spectrum(OTCBVS) [39],

Table 1. Strongly correlated indicators for features.

Feature	M_{strong}, m_i	Correlation Coefficient, $\rho_{i,r,k}$
GM	RMSE, EN, FMI_pixel	0.47, 0.34, 0.35
STD	SD, EPI, EN, VIFF	0.50, 0.49, 0.46, 0.40
ES	AG, EI, FMI_w, Q_{abf}	0.46, 0.45, 0.42, 0.37
AG	AG, EI, FMI_w, SSIM	0.46, 0.45, 0.40, 0.37
CON	SD, EPI, EN, VIFF	0.52, 0.51, 0.45, 0.43
COA	FMI_pixel, VIFF, EPI, EN	0.33, 0.32, 0.32, 0.28

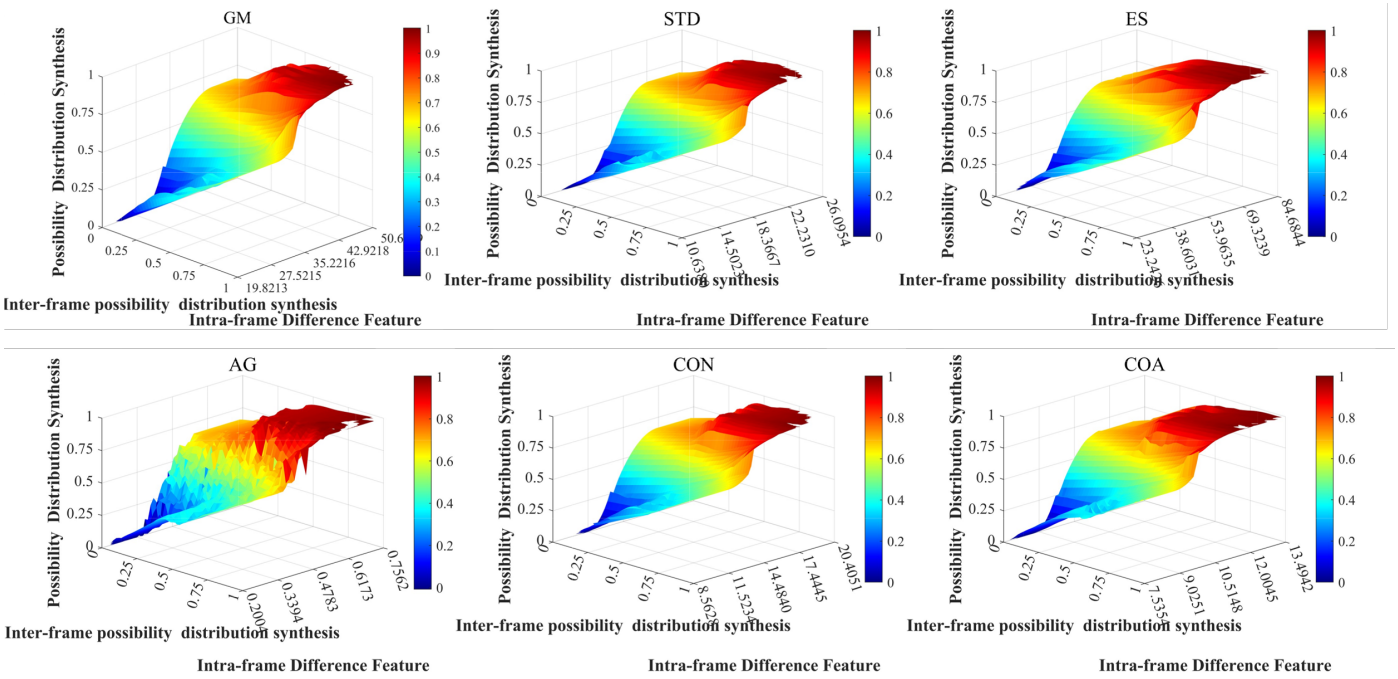

Figure 6. Possibility distribution synthesis results for various features in the BEPMS dataset.

which encompasses multiple scenarios such as rapid target movement and background interference. The image resolution is 320 × 240. The third dataset is the widely used TNO image dataset [40], which incorporates diverse outdoor natural scenes with multispectral images, including human and military scenarios. There is no fixed resolution for this dataset.

4.2 Evaluation Metrics

Performance evaluation is conducted using six quantitative metrics, including EN [41], RMSE [42], AG [43], SD [44], SF [45], and VIFF [46]. These metrics quantitatively assess information richness, edge and texture detail preservation, and visual fidelity of the fused images. Except for RMSE, all other metrics are positive indicators, wherein higher values correlate with superior fusion performance.

4.3 Case Analysis

Using the BEPMS dataset as an exemplar, the construction process is systematically delineated. First, to accurately capture the feature distribution of the entire image, the video sequence undergoes frame-by-frame processing without overlap. To balance the compound features and errors introduced by blocks that are too large or too small, block dimensions are established at $m = n = 16$, and Eqs (1) and (2) are applied to extract the amplitude of each feature. For cross-modal intra-frame difference features, difference feature amplitude calculation is executed using Eq (3), frequency distribution is obtained based on histogram statistics, and the weighted comprehensive representation is derived according to Eq (5). Finally, the cross-modal intra-frame difference feature value $\text{IV\_DF}^{r,t}$ is determined based on Eq (6). For single-modality inter-frame difference features, amplitude variations between consecutive frames are computed, similarly based on histogram statistics and weighted calculations to obtain the comprehensive representation of amplitude and frequency, and the single-modality inter-frame difference feature

values $IR_{TF^{rt}}$, $VIS_{TF^{rt}}$ are derived according to Eqs (7) and (8). Second, for different data distribution characteristics, K-medoids clustering is implemented to obtain cluster centroids C^r. The possibility distribution function $\Pi(F^{rt})$ is formulated for each feature according to Eq (11), the inter-frame and intra-frame possibility distribution functions are synthesized based on the weight distribution matrix, and the final possibility distribution synthesis values are determined, as shown in Figure 6.

Table 2 illustrates the possibility distribution synthesis results for various features in three frames (100th, 200th, and 300th), and identifies features with synthesis results exceeding the predefined threshold value (1/2) as significant change features, denoted by bold red typography in each row. Subsequently,

Table 2. Possibility distribution synthesis values for various features.

Frame	Possibility distribution synthetic values					
	GM	STD	ES	AG	CON	COA
100	0.3560	0.4653	**0.6046**	0.3478	**0.6494**	0.1045
200	0.3867	**0.5734**	**0.5236**	0.3888	0.1788	0.4007
300	0.4520	0.3711	0.0026	0.1963	0.3892	**1**

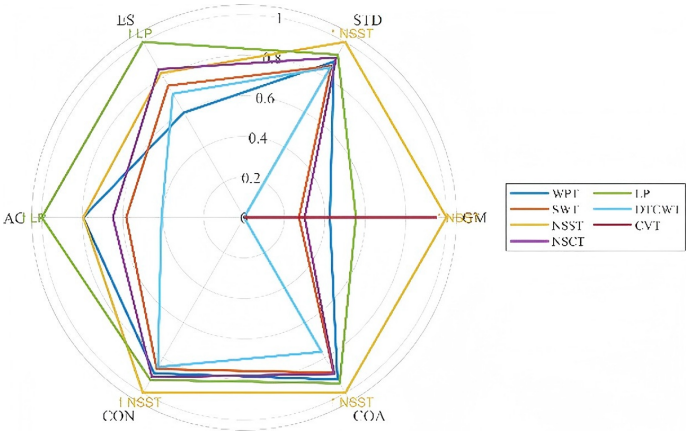

Figure 7. Mapping relationship between features and high-level variants.

utilizing the strongly correlated indicator sets $M_{\text{strong},r}$ determined in Section 3.3, evaluation scores are computed for various mimic variants according to the fusion validity evaluation function E_r in Eq (16) to establish a comprehensive set-valued mapping library between features and variants. Due to space limitations, only partial results are presented. Figures 7, 8 and 9 illustrate the mapping relationships between features and high-level variants, low-level variants under the NSST algorithm, and internal fusion parameters, respectively. The optimal mimic variant configuration for each video frame is determined through analysis of set-valued mapping relationships between significant change features and their corresponding optimal mimic variants, low-level variants, and base-level variants. As exemplified in the case of the 100th frame, the significant change features are ES and CON. According to the established mapping, the corresponding optimal high-level variants are NSST and LP. When the high-level variant is NSST, the optimal high and low-frequency combinations in the low-level variants include MC-ME, MC-WSD, ME-ME, ME-WSD, and the corresponding optimal base-level variants are pyrex and pyr.

To further validate the dynamic decision-making capability of this mapping library in video sequences and visually demonstrate the core adaptive mechanism of mimic fusion, we selected a consecutive 50-frame sequence from frame 102 to 151 for visualization analysis. This sequence completely demonstrates the entire process of mimic fusion strategy evolution with dynamic scenes. Figure 10(a) presents the temporal variations in saliency across six differential features, clearly revealing the dynamic characteristics of scene content. For instance, the gray mean feature value shows a significant change at frame 119, indicating a sudden mutation in brightness information within the video. This feature variation directly drives responsive adjustments in high-level fusion strategies. As shown in Figure 10(b), the significant change in gray mean feature F1 directly triggers the switching of its optimal high-level variant from NSST to CVT. Figure 10(c) focuses on the low-level fusion rule selection under the NSST variant, demonstrating how the system achieves precise adaptation between high-frequency and low-frequency fusion rules based on feature variations within a fixed high-level algorithm framework. These visualization results fully verify that the proposed mimic fusion algorithm does not employ a fixed architecture. Instead, it establishes a dynamic adaptive strategy that progresses from feature perception to variation-driven decision fusion, thereby ensuring continuous optimization of fusion performance in dynamic scenarios.

4.4 Experimental Results Analysis

This section evaluates the rationality and effectiveness of the proposed method through comparative experiments on the BEPMS and OTCBVS datasets, using NSST and CVT as benchmarks. In these experiments, NSST and CVT are incorporated into the proposed mimicry fusion framework as high-level

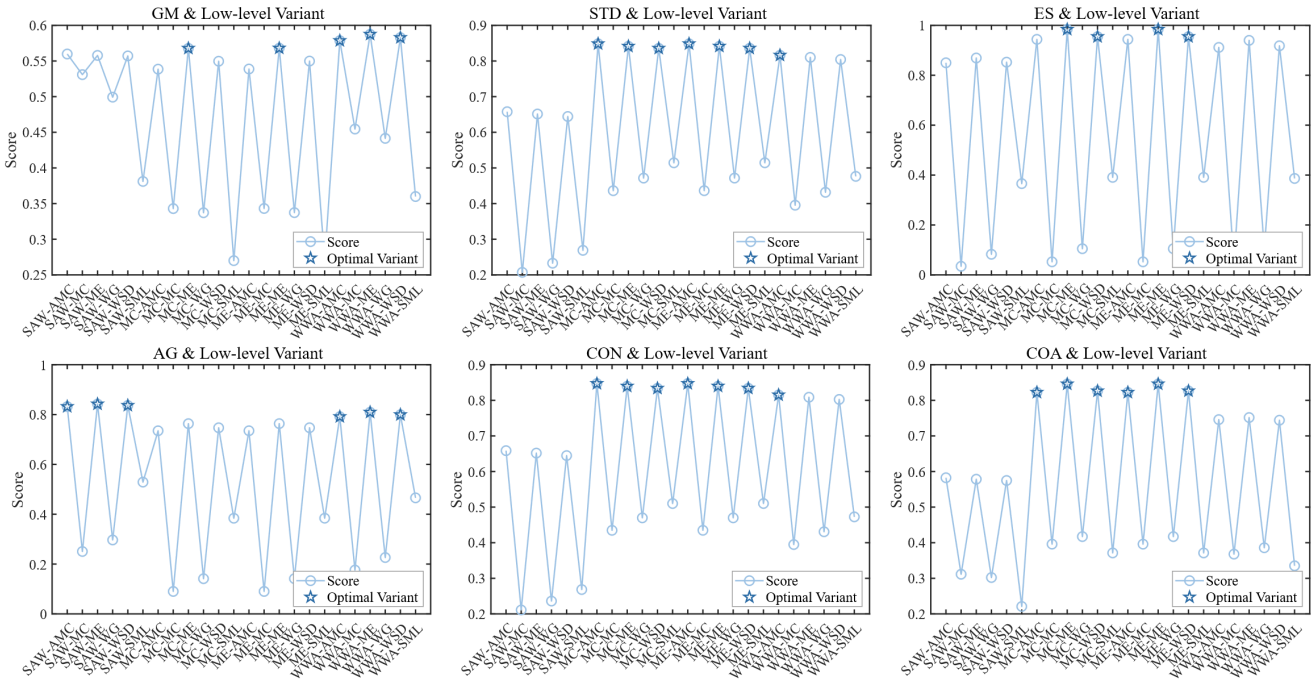

Figure 8. Mapping relationship between features and low-level variants (using NSST as an example).

Table 3. Quantitative evaluation results on the BEPMS dataset.

Frame	Methods	Evaluation Metrics					
		EN	AG	SF	VIFF	RMSE	SD
100	NSST	6.8658	**11.2250**	14.4931	**0.5211**	0.0379	30.5516
	Ours	**7.0579**	11.1983	**14.5493**	0.5208	**0.0303**	**36.7505**
200	NSST	6.8782	**11.7722**	15.1625	0.5140	0.0354	31.0492
	Ours	**7.0826**	11.7591	**15.2396**	**0.5276**	**0.0283**	**37.9438**
300	NSST	6.7606	10.2395	12.9502	0.5499	0.0327	31.3801
	Ours	**7.0532**	**10.3371**	**13.1850**	**0.6347**	**0.0247**	**45.2187**
400	NSST	6.5994	10.9826	13.9841	0.5253	0.0383	24.2054
	Ours	**7.0103**	**11.0814**	**14.1201**	**0.6007**	**0.0288**	**33.3883**
500	NSST	6.8747	11.0209	13.8391	0.5410	0.0362	29.3587
	Ours	**7.1591**	**11.0450**	**13.9676**	**0.5837**	**0.0282**	**37.5937**
Performance Improvement %		4.0745	0.3276	0.8979	8.1467	-22.1607	30.2634

Table 4. Quantitative evaluation results on the OTCBVS dataset.

Frame	Methods	Evaluation Metrics					
		EN	AG	SF	VIFF	RMSE	SD
60	CVT	7.0819	9.8006	16.4572	**0.2797**	0.0219	35.4712
	Ours	**7.5049**	**17.2926**	**29.7847**	0.2090	**0.0155**	**52.5423**
80	CVT	7.0852	9.8350	16.4929	**0.2742**	0.0220	35.4360
	Ours	**7.5021**	**17.3981**	**29.9007**	0.2064	**0.0156**	**52.3395**
100	CVT	7.0897	9.8479	16.4790	**0.2734**	0.0221	35.4731
	Ours	**7.5002**	**17.4373**	**29.7922**	0.2055	**0.0157**	**52.2141**
120	CVT	7.0876	9.8938	16.6078	**0.2715**	0.0221	35.5720
	Ours	**7.4978**	**17.4438**	**29.9514**	0.2000	**0.0157**	**52.2020**
140	CVT	7.0701	9.8498	16.5526	**0.2720**	0.0220	35.2079
	Ours	**7.4888**	**17.3544**	**29.8827**	0.2027	**0.0156**	**52.0520**
Performance Improvement %		5.8735	76.5834	80.7819	-25.3465	-29.0909	47.5215

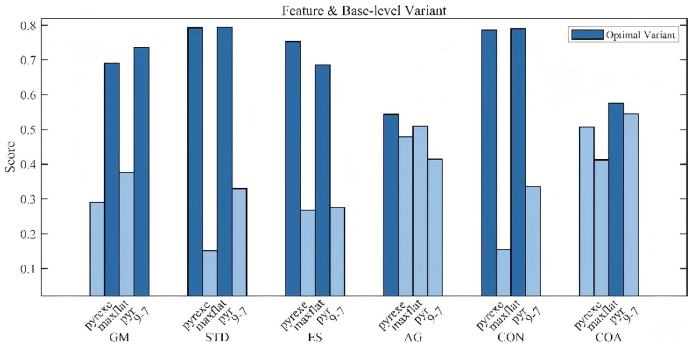

Figure 9. Mapping relationship between features and base-level variants (using NSST as an example).

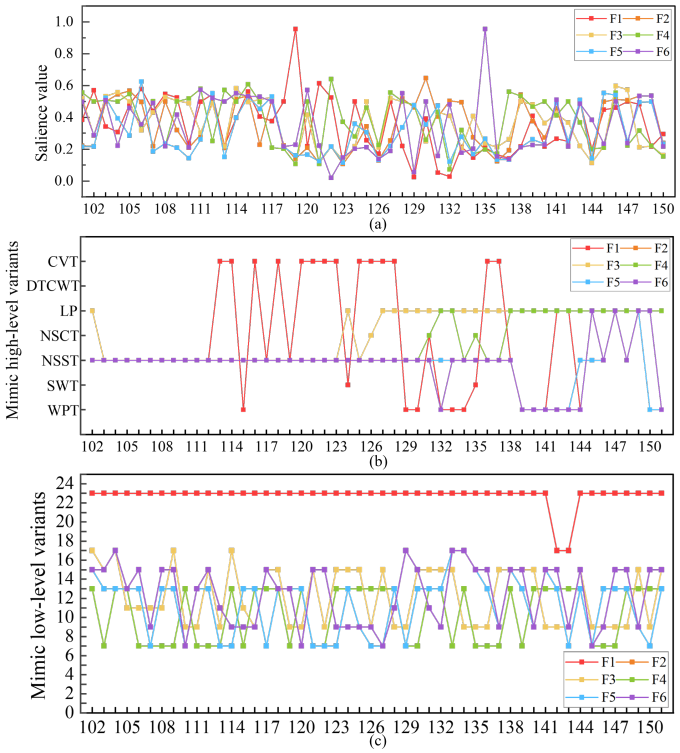

Figure 10. Visualization results of the mimic fusion dynamic decision-making process.

algorithm variants. The comparison is designed to validate the effectiveness of the framework by contrasting their native fusion strategies with their performance when dynamically optimized by the mimicry mapping library.

From the BEPMS dataset, five frames (the 100th, 200th, 300th, 400th, and 500th) were selected for analysis. Figure 11 presents a comparison of the fusion results between our method and the NSST method. Here, the proposed method operates by selecting NSST as the high-level variant while dynamically configuring optimal low-level rules and parameters through the mimicry mapping library. Our method demonstrates superior performance in preserving dynamically varying feature information. The fused images exhibit enhanced clarity and more realistic rendering of edges and textures, alongside a significant improvement in the visibility of infrared thermal targets. Quantitative results in Table 3 confirm this advantage. Our method shows significantly higher overall improvement rates across all six evaluation metrics compared to the native NSST strategy. Specifically, it achieves an SD increase of 30.2634% and an RMSE reduction of 22.1607%. These substantial improvements validate the capability of the mimicry mapping library to intelligently select superior variant combinations and, more importantly, demonstrate the enhanced adaptive performance achieved by the structural reorganization within the mimicry fusion framework. A critical test occurs when the target person moves completely into the jungle, causing a substantial scene change. In this scenario, our method maintains stable fusion quality, whereas the NSST strategy exhibits clear limitations in adapting to such changes. Further experiments on the OTCBVS dataset with CVT as the baseline (Figure 12 and Table 4) corroborate the robust feature preservation capability and the same architectural advantages of our method in dynamic scenes.

4.4.1 *Fusion Results on the BEPMS Video Dataset*

This section presents qualitative and quantitative comparisons between the proposed mimic fusion method and classical fusion algorithms, as well as three representative deep learning methods (U2Fusion [47], IFCNN [10], and FusionGAN [48]). Figure 13 shows qualitative comparison results for selected frames. The results demonstrate that when both targets and background undergo dynamic changes in the scene, our method not only effectively highlights infrared thermal target information but also finely restores texture details of leaves in the jungle background. In comparison, while the fusion results of U2Fusion and IFCNN can clearly reconstruct leaf textures, their ability to highlight infrared thermal targets is limited. This limitation becomes particularly evident when targets are severely obscured, as the target-background contrast decreases significantly. FusionGAN effectively enhances infrared thermal target information but loses substantial visible light texture details. Other single traditional fusion algorithms exhibit varying degrees of information loss in edges and texture details.

Quantitative comparison results in Table 5 further verify the superiority of the proposed method. Compared to other single traditional algorithms and deep learning algorithms, our method demonstrates

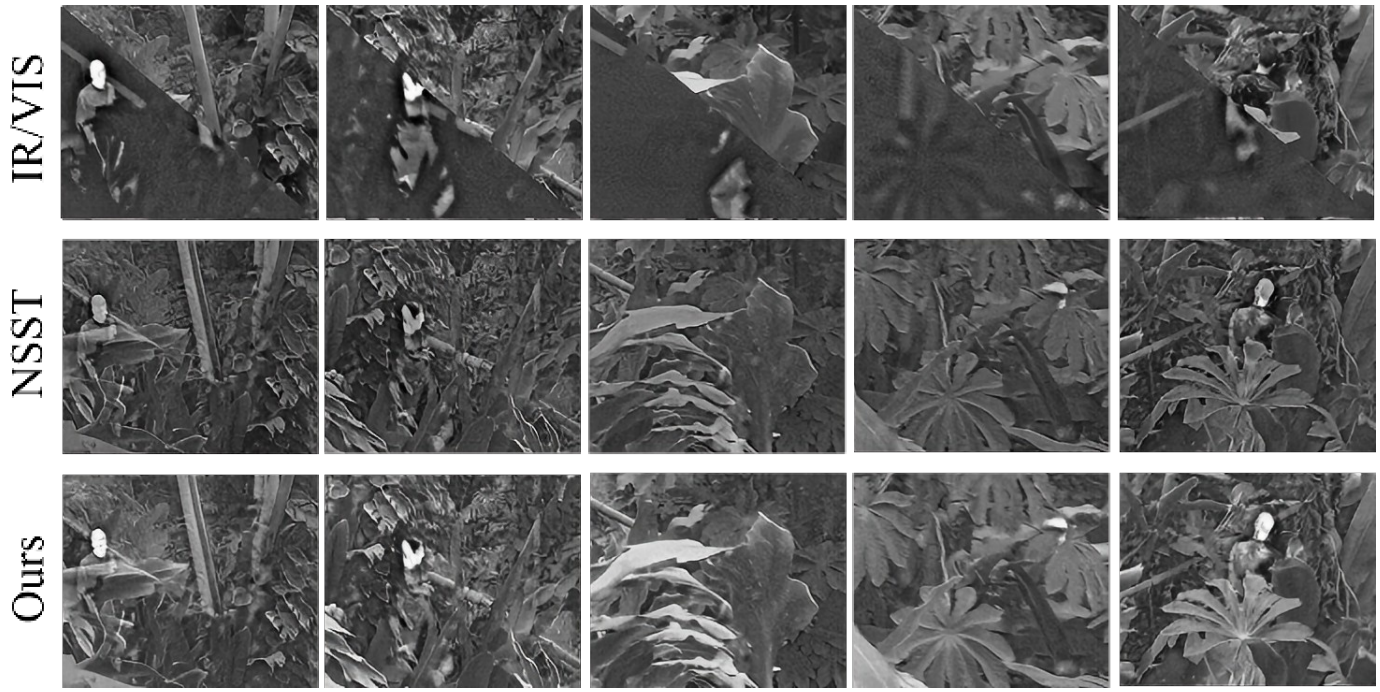

Figure 11. Qualitative comparison between the proposed method and NSST on the BEPMS dataset.

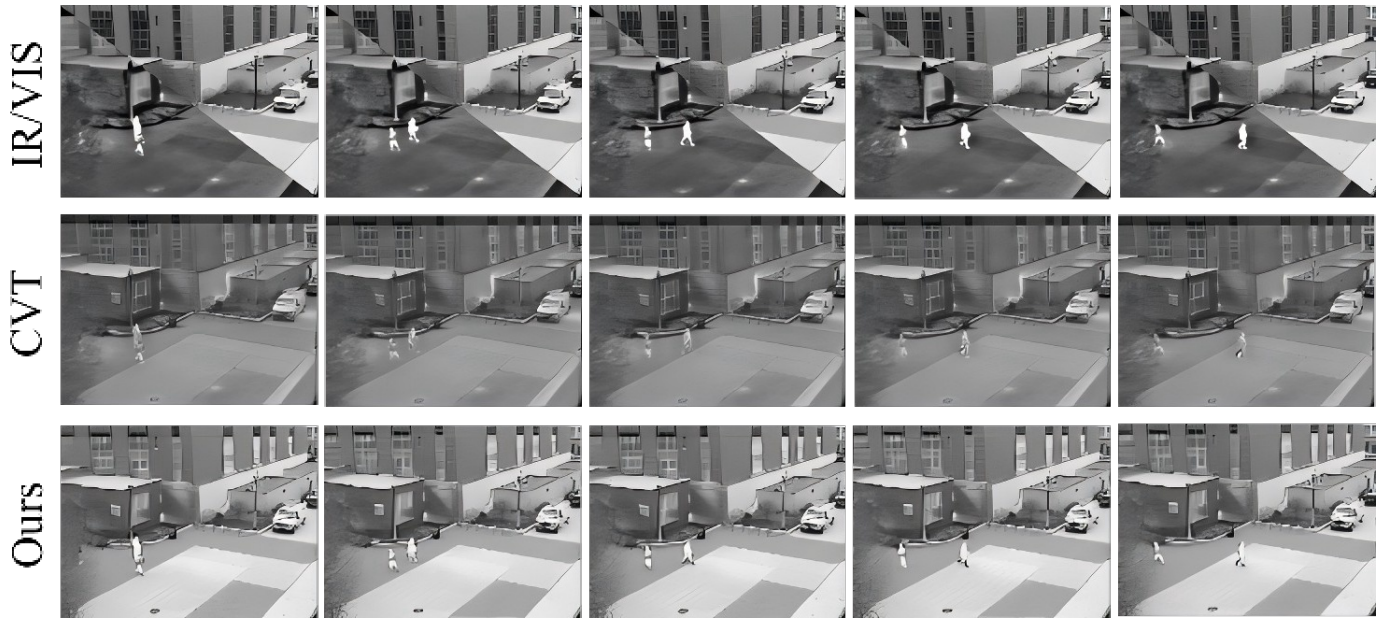

Figure 12. Qualitative comparison between the proposed method and CVT on the OTCBVS dataset.

clear advantages across all evaluation metrics. Specifically, our method maintains leading values in EN, SD, VIFF, and RMSE metrics across all test frames. These results confirm the exceptional performance of our method in detail and texture preservation, and further indicate that its fusion results contain richer information content. This advantage primarily stems from the unique dynamic variable-structure mechanism of the proposed mimic fusion and its precise feature-variant mapping relationship. In video sequences with rapidly moving targets and backgrounds, the mimic fusion method continuously perceives salient feature changes in each frame. Following the mapping logic, it rapidly identifies the optimal combination of algorithms, rules, and parameters from the fusion variant library that best suits the current content. This ensures consistently high-quality fusion output throughout the entire temporal sequence, thereby constituting the core competitive advantage of our method over traditional fixed-architecture fusion models.

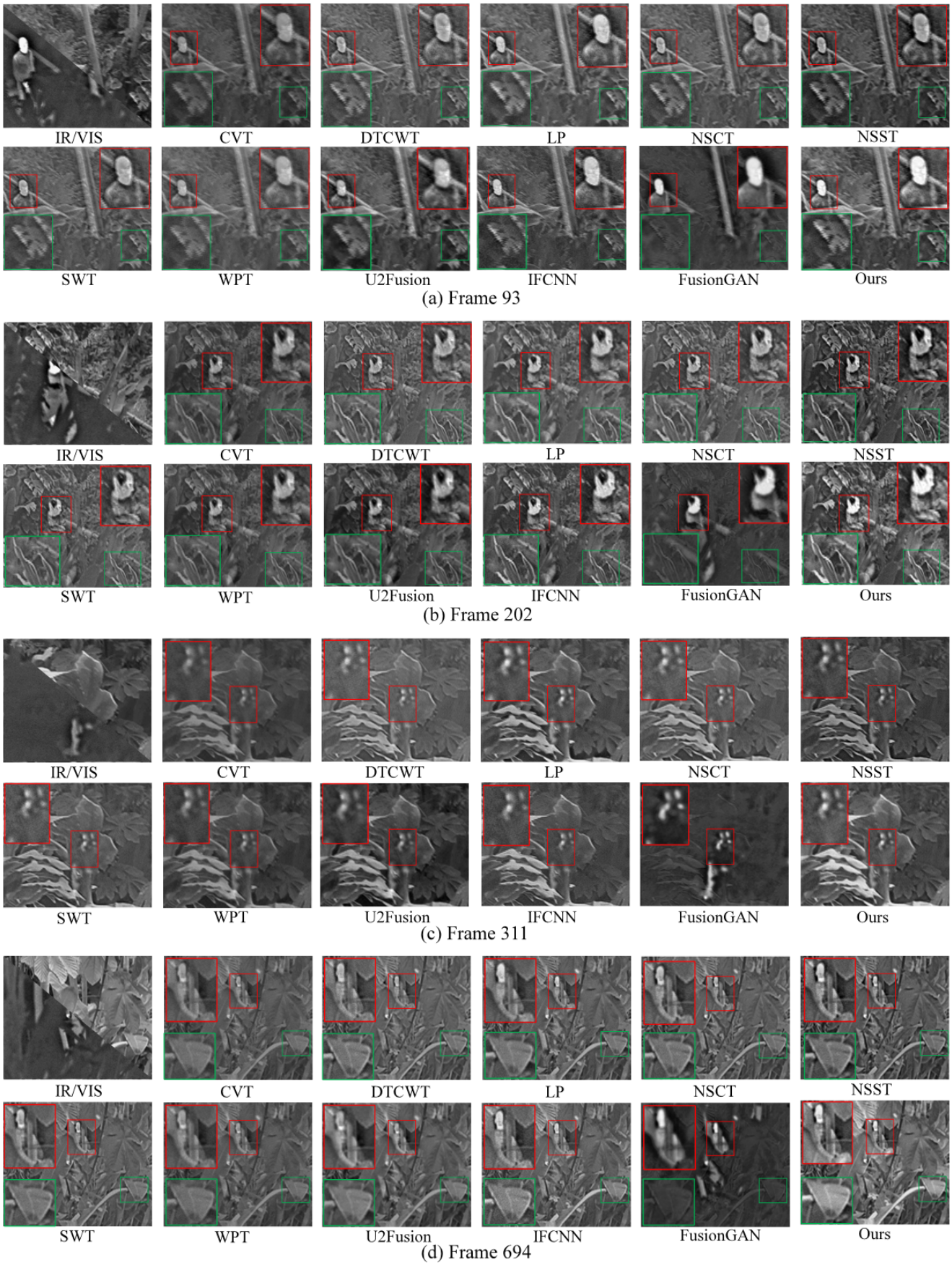

Figure 13. Qualitative comparison results on the BEPMS dataset. Red boxes highlight target person fusion details, green boxes highlight background fusion detail.

Table 5. Quantitative comparison results on the BEPMS dataset.

Database	Fusion Method	EN	AG	SF	VIFF	RMSE	SD
BEPMS	CVT	6.4066	7.2381	9.1972	0.3457	0.0371	22.8484
	DTCWT	6.6248	11.6728	14.9435	0.4248	0.0290	26.1341
	LP	6.7804	11.5246	14.7860	0.4751	0.0295	29.7104
	NSCT	6.6969	11.6728	15.0068	0.4757	0.0300	27.5923
	NSST	6.8410	11.6554	15.3869	0.5458	0.0344	30.5266
	SWT	6.6255	10.8338	13.8548	0.4395	0.0307	26.3897
	WPT	6.5426	8.3496	10.5931	0.3918	0.0346	25.2247
	U2Fusion	6.8401	5.5506	7.5331	0.4299	0.0342	31.0032
	IFCNN	6.7746	13.6234	17.2455	0.4990	0.0330	31.7920
	FusionGAN	5.8635	3.5708	4.9423	0.0645	0.0254	19.9498
	Ours	7.1073	11.9762	15.5977	0.5968	0.0274	39.4605

The best result is highlighted in red, the second-best results are marked in blue.

4.4.2 *Fusion Results on the TNO Image Dataset*

Additional experiments were conducted on the TNO image dataset. Qualitative results are presented in Figure 14. The proposed method enhances infrared thermal targets while preserving background texture details and contour contrasts of structures such as chimneys and street lamps. Although FusionGAN emphasizes infrared thermal targets, it results in significant loss of visible information. Other methods do not adequately balance target enhancement and background detail preservation. Quantitative results in Table 6 indicate that the proposed method achieves the highest values in EN, RMSE, and SD, and attains suboptimal values in AG and VIFF. Overall, it demonstrates competitive performance compared with other algorithms. These results indicate that the proposed method effectively preserves edge structures and texture detail.

4.5 Algorithm Complexity Analysis

This section analyzes the computational efficiency and storage requirements of the proposed algorithm by evaluating its time and space complexity. In terms of time complexity, the computational cost primarily concentrates on the feature variation perception and quantification stages. Given an input video sequence with height H and width W, the feature extraction stage requires traversing all blocks and computing six types of features, resulting in a time complexity of $O(HW)$. The possibility distribution synthesis stage involves fixed computational load, with time complexity $O(1)$. The feature-mimic variant mapping library is constructed offline, requiring only query operations during real-time fusion, also with $O(1)$ time complexity. Consequently, the overall time complexity of the perception stage is $O(HW)$. The fusion execution stage exhibits unique dynamic characteristics in time complexity. Its computational overhead directly depends on the specific variant combination selected through mimic decision-making for each frame. This stage's time complexity can be expressed as $O(F(A_l, R_l, R_h, P_l))$, where A_l, R_l, R_h and P_l represent the high-level fusion algorithm, low-frequency fusion rule, high-frequency fusion rule, and basic-level fusion parameters selected for the current frame, respectively. It should be noted that while different multi-scale transform algorithms vary in time complexity, all candidate variants originate from a verified pool of classical methods with predefined computational bounds. The instantaneous complexity per frame remains deterministic and controllable, ensuring predictable performance during actual operation.

Regarding space complexity, the algorithm's memory footprint is primarily determined by image data and feature matrices. The processing requires simultaneous storage of two consecutive frames of image pairs (i.e., both the current and previous registered infrared-visible image pairs), with a space complexity of $O(HW)$. The block feature matrices generated during feature extraction also require $O(HW)$ storage space. The historical feature window for temporal analysis occupies $O(T)$ space, where T represents the fixed number of frames used for historical analysis. The constructed feature-mimic variant mapping library requires only constant space $O(1)$. Therefore, the overall space complexity of the algorithm is $O(HW)$. This analysis demonstrates that while maintaining dynamic adaptive advantages, the algorithm's complexity characteristics align with

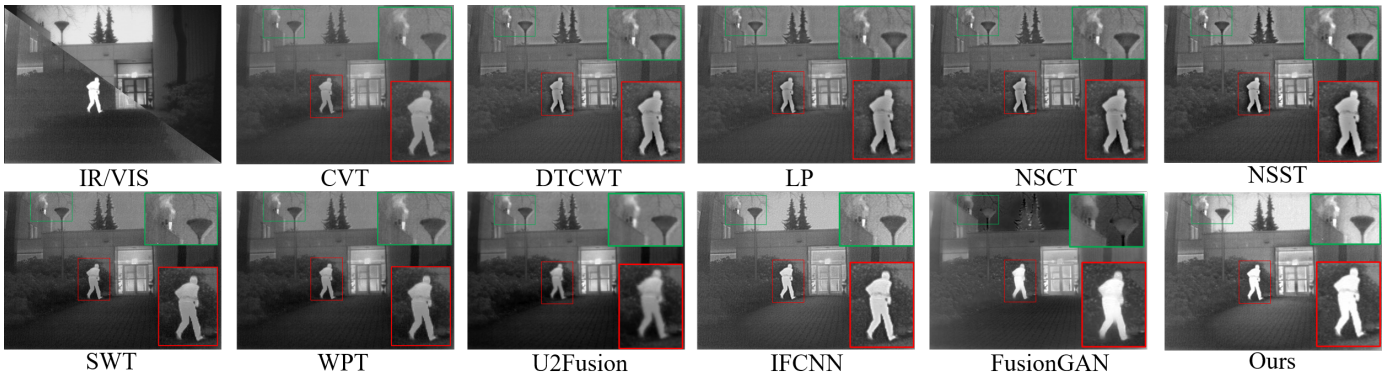

Figure 14. Qualitative comparison results on the TNO dataset.

Table 6. Quantitative comparison results on the TNO dataset.

Database	Fusion Method	EN	AG	SF	VIFF	RMSE	SD
TNO-Kaptein	CVT	6.5279	4.0053	5.2626	0.2965	0.0258	31.5343
	DTCWT	6.4726	5.5676	7.4742	0.2734	0.0254	28.4914
	LP	6.6080	5.7719	7.9535	0.1049	0.0231	32.5467
	NSCT	6.4827	5.5752	7.5320	0.3162	0.0254	28.7201
	NSST	6.7308	5.9195	8.2227	0.4026	0.0253	34.3827
	SWT	6.9154	5.9770	8.3603	0.4344	0.0231	40.2205
	WPT	6.8908	5.1608	6.7720	0.4171	0.0229	40.5046
	U2Fusion	6.9985	2.8529	4.6228	0.3918	0.0212	44.1103
	IFCNN	6.9390	7.6348	10.0022	0.3530	0.0226	42.4484
	FusionGAN	6.9525	3.5175	5.8367	0.1850	0.0185	38.1369
	Ours	7.1835	6.8238	8.2162	0.4251	0.0184	51.4504

The best result is highlighted in red, the second-best results are marked in blue.

the memory requirements of mainstream image processing algorithms, confirming its feasibility for practical system deployment.

5 Conclusion

In this study, we propose a salient feature-driven dual-modal video mimic fusion method. By utilizing salient feature variations as the driving force and adaptively adjusting mimic fusion variant combinations, our approach achieves excellent fusion performance. The method effectively overcomes the limitations of rigid structures inherent in traditional fusion models through continuous perception of scene feature changes and precise matching of optimal mimic combinations for significantly varying features. Dynamic strategy adjustment based on the fusion validity evaluation function enables our method to select the optimal variant combination tailored to different scene characteristics. This ensures both edge information integrity and texture detail fidelity, thereby providing higher-quality input for subsequent high-level vision tasks such as target detection and scene recognition. Extensive qualitative comparisons and quantitative analyses fully demonstrate that the proposed mimic fusion method possesses superior adaptive capability and maintains stable fusion performance in dynamically changing scenes. The fusion results significantly surpass those obtained by other single fusion methods.

Although the experiments focus on infrared and visible video fusion, the proposed method retains decision-level generality. The core principle is decision updating through feature variation quantification, salient-feature-driven mechanisms, and adaptive fusion variant selection. It does not depend on a specific imaging modality. By redesigning feature sets and fusion validity metrics according to different imaging mechanisms and task requirements, the framework can be extended to other domains, such as medical imaging and multispectral imaging. Future work will expand the multi-layer mimic variant library and refine the set-valued mapping between features and variants to accommodate more complex dynamic variations, broader imaging conditions, and diverse

modality combinations. This will further improve the coverage, adaptability, and robustness of the method.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 61972363 and Grant 61672472; in part by the Fundamental Research Program of Shanxi Province under Grant 202203021221104; in part by the Shanxi Provincial Postgraduate Academic Innovation Project under Grant 2025XS435; in part by the Graduate Research and Innovation Project of North University of China under Grant 2024202; in part by the Scientific and Technological Innovation Programs of Higher Education Institution Shanxi under Grant 2025L051.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Ma, J., Ma, Y., & Li, C. (2019). Infrared and visible image fusion methods and applications: A survey. *Information fusion, 45*, 153-178. [CrossRef]

[2] Tang, L., Xiang, X., Zhang, H., Gong, M., & Ma, J. (2023). DIVFusion: Darkness-free infrared and visible image fusion. *Information Fusion, 91*, 477-493. [CrossRef]

[3] Luo, Y., & Luo, Z. (2023). Infrared and visible image fusion: Methods, datasets, applications, and prospects. *Applied Sciences, 13*(19), 10891. [CrossRef]

[4] Zhang, Y., Zhang, T., Wu, C., & Tao, R. (2023). Multi-scale spatiotemporal feature fusion network for video saliency prediction. *IEEE Transactions on Multimedia, 26*, 4183-4193. [CrossRef]

[5] Chen, H., Deng, L., Chen, Z., Liu, C., Zhu, L., Dong, M., ... & Guo, C. (2024). SFCFusion: Spatial–frequency collaborative infrared and visible image fusion. *IEEE Transactions on Instrumentation and Measurement, 73*, 1-15. [CrossRef]

[6] Zhu, J., Jin, W., Li, L., Han, Z., & Wang, X. (2018). Multiscale infrared and visible image fusion using gradient domain guided image filtering. *Infrared Physics & Technology, 89*, 8-19. [CrossRef]

[7] Liu, J., Dian, R., Li, S., & Liu, H. (2023). SGFusion: A saliency guided deep-learning framework for pixel-level image fusion. *Information Fusion, 91*, 205-214. [CrossRef]

[8] Wang, C., Wu, Y., Yu, Y., & Zhao, J. Q. (2022). Joint patch clustering-based adaptive dictionary and sparse representation for multi-modality image fusion. *Machine Vision and Applications, 33*(5), 69. [CrossRef]

[9] Gao, C., Qi, D., Zhang, Y., Song, C., & Yu, Y. (2021). Infrared and visible image fusion method based on ResNet in a nonsubsampled contourlet transform domain. *IEEE Access, 9*, 91883-91895. [CrossRef]

[10] Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., & Zhang, L. (2020). IFCNN: A general image fusion framework based on convolutional neural network. *Information Fusion, 54*, 99-118. [CrossRef]

[11] Lu, R., Gao, F., Yang, X., Fan, J., & Li, D. (2023). A novel infrared and visible image fusion method based on multi-level saliency integration. *The Visual Computer, 39*(6), 2321-2335. [CrossRef]

[12] Xu, H., Zhang, H., & Ma, J. (2021). Classification saliency-based rule for visible and infrared image fusion. *IEEE Transactions on Computational Imaging, 7*, 824-836. [CrossRef]

[13] Tang, L., Chen, Z., Huang, J., & Ma, J. (2023). CAMF: An interpretable infrared and visible image fusion network based on class activation mapping. *IEEE Transactions on Multimedia, 26*, 4776-4791. [CrossRef]

[14] Liu, Y., Chen, X., Peng, H., & Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. *Information Fusion, 36*, 191-207. [CrossRef]

[15] Luo, X., Gao, Y., Wang, A., Zhang, Z., & Wu, X. J. (2021). IFSepR: A general framework for image fusion based on separate representation learning. *IEEE Transactions on Multimedia, 25*, 608-623. [CrossRef]

[16] Saeedi, J., & Faez, K. (2012). Infrared and visible image fusion using fuzzy logic and population-based optimization. *Applied Soft Computing, 12*(3), 1041-1054. [CrossRef]

[17] Tong, Y., Liu, L., Zhao, M., Chen, J., & Li, H. (2016). Adaptive fusion algorithm of heterogeneous sensor networks under different illumination conditions. *Signal Processing, 126*, 149-158. [CrossRef]

[18] Ji, L., Guo, X., & Yang, F. (2024). A fusion algorithm selection method for infrared image based on quality synthesis of intuition possible sets. *Measurement, 236*, 115163. [CrossRef]

[19] Yang, F. B. (2017). Research on theory and model of mimic fusion between infrared polarization and intensity images. *Journal of North University of China*

(*Natural Science Edition*), *38*(1), 1-8. [CrossRef]

[20] Sheng, L., Fengbao, Y., Linna, J., & Xiangdong, W. (2025). Infrared intensity and polarization image mimicry fusion based on the combination of variable elements and matrix theory. *Opto-Electronic Engineering, 45*(12), 180188-1. [CrossRef]

[21] Zhang, L., Yang, F., & Ji, L. (2017). Multi-scale fusion algorithm based on structure similarity index constraint for infrared polarization and intensity images. *IEEE Access, 5*, 24646-24655. [CrossRef]

[22] Guo, X., Yang, F., & Ji, L. (2022). MLF: A mimic layered fusion method for infrared and visible video. *Infrared Physics & Technology, 126*, 104349. [CrossRef]

[23] Guo, X., Yang, F., & Ji, L. (2023). A mimic fusion method based on difference feature association falling shadow for infrared and visible video. *Infrared Physics & Technology, 132*, 104721. [CrossRef]

[24] Hu, P., Yang, F., Wei, H., Ji, L., & Wang, X. (2019). Research on constructing difference-features to guide the fusion of dual-modal infrared images. *Infrared Physics & Technology, 102*, 102994. [CrossRef]

[25] Meng, Y., Yang, F., Xi, J., Wang, X., Ji, L., Li, B., & Guo, X. (2025). STMFuse: spatiotemporal feature saliency change-driven mimic fusion for infrared and visible video. *Optics & Laser Technology, 192*, 113640. [CrossRef]

[26] Li, W., Zhao, J., & Xiao, B. (2018). Multimodal medical image fusion by cloud model theory. *Signal, Image and Video Processing, 12*(3), 437-444. [CrossRef]

[27] Guo, X., Yang, F., & Ji, L. (2024). A Mimic Fusion Algorithm for Dual Channel Video Based on Possibility Distribution Synthesis Theory. *Chinese Journal of Information Fusion, 1*(1), 33–49. [CrossRef]

[28] Wang, Y., & Zou, H. (2024). Image Fusion Based on Bioinspired Rattlesnake Visual Mechanism Under Lighting Environments of Day and Night Two Levels. *Journal of Bionic Engineering, 21*(3), 1496-1510. [CrossRef]

[29] Wang, Y., Liu, H., Xie, W., & Wang, S. (2022). Image fusion based on the rattlesnake visual receptive field model. *Displays, 74*, 102171. [CrossRef]

[30] Sissinto, P., & Ladeji-Osias, J. (2013). Bio-empirical mode decomposition: visible and infrared fusion using biologically inspired empirical mode decomposition. *Optical Engineering, 52*(7), 073101-073101. [CrossRef]

[31] Bao, W., & Zhu, X. (2015). A novel remote sensing image fusion approach research based on HSV space and bi-orthogonal wavelet packet transform. *Journal of the Indian Society of Remote Sensing, 43*(3), 467-473. [CrossRef]

[32] Bashir, R., Junejo, R., Qadri, N. N., Fleury, M., & Qadri, M. Y. (2019). SWT and PCA image fusion methods for multi-modal imagery. *Multimedia tools and applications, 78*(2), 1235-1263. [CrossRef]

[33] Asha, C. S., Lal, S., Gurupur, V. P., & Saxena, P. P. (2019). Multi-modal medical image fusion with adaptive weighted combination of NSST bands using chaotic grey wolf optimization. *IEEE Access, 7*, 40782-40796. [CrossRef]

[34] Hu, Q., Cai, W., Xu, S., Hu, S., Wang, L., & He, X. (2024). Adaptive convolutional sparsity with sub-band correlation in the NSCT domain for MRI image fusion. *Physics in Medicine & Biology, 69*(5), 055022. [CrossRef]

[35] Dogra, A., Goyal, B., & Agrawal, S. (2018). Osseous and digital subtraction angiography image fusion via various enhancement schemes and Laplacian pyramid transformations. *Future Generation Computer Systems, 82*, 149-157. [CrossRef]

[36] Lewis, J. J., O'Callaghan, R. J., Nikolov, S. G., Bull, D. R., & Canagarajah, N. (2007). Pixel-and region-based image fusion with complex wavelets. *Information fusion, 8*(2), 119-130. [CrossRef]

[37] Zhao, X., Jin, S., Bian, G., Cui, Y., Wang, J., & Zhou, B. (2023). A curvelet-transform-based image fusion method incorporating side-scan sonar image features. *Journal of Marine Science and Engineering, 11*(7), 1291. [CrossRef]

[38] Lewis, J. J., Nikolov, S. G., Loza, A., Fernandez Canga, E., Cvejic, N., Li, J., Cardinali, A., Canagarajah, C. N., Bull, D. R., Riley, T., Hickman, D., & Smith, M. I. (2006, April 10). *The Eden Project multi-sensor data set* (Technical Report No. TR-UoB-WS-Eden-Project-Data-Set). University of Bristol & Waterfall Solutions Ltd.

[39] Ariffin, S. (2016). OTCBVS database [Data set]. Ohio State University. http://vcipl-okstate.org/pbvs/bench/

[40] Toet, A. (2014). TNO Image Fusion Dataset [Data set]. Figshare. [CrossRef]

[41] Roberts, J. W., Van Aardt, J. A., & Ahmed, F. B. (2008). Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *Journal of Applied Remote Sensing, 2*(1), 023522. [CrossRef]

[42] Huang, B., Yang, F., Yin, M., Mo, X., & Zhong, C. (2020). A review of multimodal medical image fusion techniques. *Computational and mathematical methods in medicine, 2020*(1), 8279342. [CrossRef]

[43] Cui, G., Feng, H., Xu, Z., Li, Q., & Chen, Y. (2015). Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. *Optics Communications, 341*, 199-209. [CrossRef]

[44] Rao, Y. J. (1997). In-fibre Bragg grating sensors. *Measurement science and technology, 8*(4), 355-375. [CrossRef]

[45] Eskicioglu, A. M., & Fisher, P. S. (2002). Image quality measures and their performance. *IEEE Transactions on communications, 43*(12), 2959-2965. [CrossRef]

[46] Han, Y., Cai, Y., Cao, Y., & Xu, X. (2013). A new

image fusion performance metric based on visual information fidelity. *Information fusion, 14*(2), 127-135. [CrossRef]

[47] Xu, H., Ma, J., Jiang, J., Guo, X., & Ling, H. (2020). U2Fusion: A unified unsupervised image fusion network. *IEEE transactions on pattern analysis and machine intelligence, 44*(1), 502-518. [CrossRef]

[48] Ma, J., Yu, W., Liang, P., Li, C., & Jiang, J. (2019). FusionGAN: A generative adversarial network for infrared and visible image fusion. *Information fusion, 48*, 11-26. [CrossRef]

Yanchen Meng received the B.S. degree in engineering from the North University of China, Taiyuan, Shanxi, China, in 2024, and is currently pursuing the M.S. degree in information and communication engineering at the same university. Her research interests include multimodal image fusion and image processing. (Email: mycnuc@163.com)

Fengbao Yang received his Ph.D. degree in measurement technology and instrument from North University of China, Taiyuan, China, in 2003. From 2004 to 2007, he was a post-doctoral research fellow with the Beijing Institute of Technology. He is a full professor at North University of China. He is a Fellow of the information fusion branch of the China Society of Aeronautics and Astronautics. Now, he is the leader of the Institute of Information Fusion and Recognition Technology at the North University of China. He has published approximately 100 papers in the information fusion and image processing journals and eight books on uncertain information reasoning methods and infrared technology. His current research interests include information fusion, performance detection and evaluation of complex systems, and information fusion theory of uncertainty. (Email: yfengb@163.com)

Linna Ji received her Ph.D. degree in Signal and Information Processing from North University of China, Taiyuan, China, in 2015. Her current research interests include infrared multi-source image fusion, uncertain information processing and performance detection and evaluation of complex systems. She hosted one item of the National Natural Science Foundation and has been engaged in some national and provincial projects. (Email: jlnnuc@163.com)

Xiaoxia Wang received her Ph.D. degree in Signal and Information Processing from North University of China, Taiyuan, China, in 2014. Her current research interests include uncertain information processing and decision-making, complex system evaluation, multi-source information fusion, target recognition, and ghost imaging technology. She has led or participated in more than 20 projects including those supported by the National Natural Science Foundation of China, Shanxi Province Natural Science Foundation, and Provincial Study Abroad Foundation. She has received one second prize in Shanxi Province Natural Science and Technology Award and one first prize in Shanxi Province Teaching Achievement Award. (Email: wangxiaoxia@nuc.edu.cn)

Xiaoming Guo, Lecturer at North University of China. Her main research interests include multimodal image fusion and uncertain information processing. She is currently presiding over projects such as the Shanxi Higher Education Institution Innovation Project and the Shanxi Key Laboratory Fund. (Email: Gxmnuc@163.com)