

基于显著性特征的可见光和红外图像融合算法

吴 韩¹, 王聪聪^{1*}, 陈兰兰²

¹新疆理工职业大学人工智能学院, 新疆 喀什

²新疆理工职业大学通识学院, 新疆 喀什

收稿日期: 2026年2月25日; 录用日期: 2026年3月11日; 发布日期: 2026年3月19日

摘 要

针对可见光与红外图像在成像机理与信息表达方面存在差异, 以及传统融合方法在低照度条件下易出现结构信息不完整、细节表达不足和视觉质量下降等问题, 提出一种基于显著性特征的可见光与红外图像融合算法。该方法通过构建互补特征分解模型, 将源图像分离为主体结构特征与细节纹理特征, 在主体结构特征中使用联合建模像素灰度显著性与多尺度梯度显著性, 并引入全变分约束对融合结果进行优化, 以提高结构保持能力和整体视觉一致性; 在细节纹理特征中使用参数自适应脉冲耦合神经网络(PAPCNN), 根据特征统计特性自动调节网络参数, 实现边缘与细节信息的自适应增强。基于LLVIP和M3FD数据集的对比实验结果表明, 所提方法在信息熵、特征互信息和视觉信息保真度等评价指标上均优于多种典型融合方法, 能够有效提升融合图像的信息表达能力和视觉感知质量。

关键词

可见光红外融合, 视觉显著性, 特征分解, 全变分, 自适应PCNN

A Saliency Feature-Based Visible-Infrared Image Fusion Method

Han Wu¹, Congcong Wang^{1*}, Lanlan Chen²

¹College of Artificial Intelligence, Xinjiang Vocational University of Technology, Kashgar Xinjiang

²College of General Education, Xinjiang Vocational University of Technology, Kashgar Xinjiang

Received: February 25, 2026; accepted: March 11, 2026; published: March 19, 2026

Abstract

Visible and infrared images exhibit significant differences in imaging mechanisms and information representation. Under low-illumination conditions, traditional fusion methods often suffer from

*通讯作者。

文章引用: 吴韩, 王聪聪, 陈兰兰. 基于显著性特征的可见光和红外图像融合算法[J]. 人工智能与机器人研究, 2026, 15(2): 559-569. DOI: 10.12677/airr.2026.152054

incomplete structural information, insufficient detail preservation, and degraded visual quality. To address these issues, a saliency-based visible and infrared image fusion method is proposed. The proposed method first constructs a complementary feature decomposition model to separate the source images into main structural features and detail texture features. For the structural features, a joint saliency model integrating pixel-intensity saliency and multi-scale gradient saliency is established, and total variation (TV) regularization is introduced to optimize the fusion results, thereby enhancing structural preservation and improving overall visual consistency. For the detail texture features, a parameter-adaptive pulse coupled neural network (PAPCNN) is employed, in which network parameters are automatically adjusted according to the statistical characteristics of the features to achieve adaptive enhancement of edge and fine-detail information. Comparative experiments conducted on the LLVIP and M3FD datasets demonstrate that the proposed method outperforms several representative fusion approaches in terms of information entropy, feature mutual information, and visual information fidelity, effectively improving the information representation capability and perceptual quality of the fused images.

Keywords

Visible-Infrared Image Fusion, Visual Saliency, Feature Decomposition, Total Variation, Adaptive PCNN

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

可见光与红外图像融合旨在充分利用不同成像模态之间的互补信息，将可见光图像中丰富的纹理与背景细节与红外图像中突出的目标热特征进行有效整合，从而获得信息量更丰富、视觉效果更优的融合结果[1]。该技术在夜间监控、目标检测、无人驾驶及安防预警等领域具有重要的应用价值[2]。然而，由于两类传感器的成像机理存在显著差异，可见光图像易受光照条件影响，在弱光或复杂环境下对比度明显下降；而红外图像虽然能够突出热目标，但通常存在纹理信息不足和噪声较大的问题[3]。因此，如何在突出红外显著目标的同时有效保留可见光细节信息，成为图像融合研究中的关键问题之一。

近年来，针对可见光与红外图像融合问题，研究者提出了多种方法[4]。传统方法主要包括多尺度变换方法、稀疏表示方法以及子空间方法等[5]。多尺度变换类方法通过将图像分解至不同频带并分别设计融合规则，如拉普拉斯金字塔、小波变换、NSCT和NSST等[6]。该类方法具有良好的多尺度表达能力，但融合规则通常依赖人工设计，难以在全局结构保持与局部细节增强之间取得平衡。稀疏表示和低秩表示方法能够从特征表达角度实现信息融合，但计算复杂度较高，且对参数设置较为敏感[7]。近年来，基于深度学习的方法在图像融合领域取得了一定进展[8]，但其性能依赖大量训练样本，模型的泛化能力与可解释性仍存在一定局限。

另一方面，视觉显著性理论为图像融合提供了新的研究思路[9]。通过模拟人眼视觉感知机制，可以突出图像中具有重要信息的区域，从而提高融合结果的主观视觉质量。然而，现有基于显著性的融合方法多集中于单一特征构建，如仅利用灰度或梯度信息，难以全面刻画结构与细节之间的显著性差异[10]。同时，可见光与红外图像在结构信息与纹理信息方面具有不同的表达特性，若采用统一的融合策略，容易导致目标对比度不足或细节模糊等问题。

此外，在细节信息融合过程中，传统的取大值或加权平均策略容易放大噪声或削弱真实纹理。脉冲

耦合神经网络(Pulse Coupled Neural Network, PCNN)由于具有良好的空间同步特性,在图像处理领域得到了广泛应用[11],但其参数设置通常依赖经验,适应性不足,从而限制了其在复杂场景中的应用效果。

针对上述问题,本文提出一种基于显著性特征的可见光与红外图像融合算法。首先,通过互补特征分解将源图像分离为主体结构特征与细节纹理特征,并在分解过程中引入局部相对衰减率以抑制红外噪声。其次,在主体结构层面构建像素灰度显著性与多尺度梯度显著性联合模型,并结合全变分约束实现结构信息的优化融合,从而提高图像整体对比度与视觉一致性。然后,在细节纹理层面引入参数自适应脉冲耦合神经网络(PAPCNN),根据图像统计特性自动调整网络参数,实现边缘与纹理信息的自适应增强。最后,在公开数据集上进行实验验证,结果表明,所提方法在信息保持能力和视觉质量方面均优于多种典型融合方法。

本文的主要贡献如下:

(1) 提出一种基于互补特征分解的融合框架,实现结构信息与纹理信息的分离处理,并有效抑制红外噪声干扰;

(2) 引入参数自适应 PAPCNN 进行细节纹理融合,根据图像统计特性自动调节网络参数,提高边缘与纹理信息的保持能力。

2. 相关理论与方法基础

可见光与红外图像融合的关键在于充分挖掘不同模态之间的互补信息,以实现结构细节与目标显著性的协同表达。为构建合理有效的融合框架,有必要从图像成像特性、视觉显著性机制以及神经网络模型基础等方面进行理论分析。本节对相关理论进行简要阐述,为后续方法设计提供依据。

2.1. 可见光与红外图像特性分析

可见光与红外图像在成像机理和信息表达上存在显著差异。可见光图像基于反射光成像,能够提供丰富的纹理细节和背景结构信息,但在弱光或复杂环境下易出现对比度下降及噪声增强等问题;红外图像基于目标热辐射成像,不受环境光照影响,能够稳定突出高温目标,但通常存在纹理信息匮乏、背景层次较弱以及传感器噪声干扰等不足[12]。从信息分布角度来看,可见光图像更侧重细节与结构表达,而红外图像更强调目标显著性。若采用统一融合策略,容易导致细节信息丢失或目标对比度不足。因此,有必要对源图像进行结构与纹理分离,并针对不同特征设计差异化融合规则,以充分发挥多模态信息的互补优势。

2.2. 视觉显著性理论

视觉显著性反映人眼对图像中重要区域的优先关注特性,通常与区域对比度和结构变化强度密切相关[13]。灰度显著性通过像素与其邻域或全局均值之间的差异刻画区域对比度,有利于突出目标区域;梯度显著性则反映局部变化强度,是描述边缘与结构信息的重要特征。单一显著性难以全面表征图像内容,将灰度信息与多尺度梯度特征进行联合建模,有助于获得更加稳定且鲁棒的显著性表达。此外,为避免融合过程中权重分布不连续或产生伪影,引入全变分(Total Variation, TV)约束对融合结果进行空间正则化处理,可在保持边缘信息的同时提高结构一致性和视觉平滑性[14]。

2.3. NSST 与 PCNN 模型基础

非下采样剪切波变换(Non-Subsampled Shearlet Transform, NSST)是一种具有多尺度、多方向和平移不变特性的图像表示方法,能够有效刻画图像的边缘结构与各向异性特征。相较于传统小波或金字塔分解方法,NSST 在方向选择性与结构保持方面具有优势,因此被广泛应用于多模态图像融合领域[15]。通过

NSST 分解, 源图像可表示为低频子带与多尺度多方向高频子带, 其中低频分量反映整体亮度与结构信息, 高频分量则包含边缘与纹理等细节特征, 为分层融合提供了良好的特征基础。在高频融合过程中, 脉冲耦合神经网络(PCNN)凭借空间邻域耦合与同步激发机制, 能够利用神经元放电行为为表征区域显著程度, 从而构建融合权重。然而, 传统单通道 PCNN 对多模态信息利用不足, 且参数依赖经验设定, 易影响融合稳定性。为此, 本文在细节纹理层引入 PAPCNN 模型, 以可见光与红外图像对应的 NSST 高频系数作为双输入, 通过联合激发机制实现多源信息的协同表达, 并依据放电次数生成融合权重, 在增强边缘与纹理信息的同时抑制噪声干扰。结合 NSST 的多尺度方向表达能力与 PAPCNN 的空间同步特性, 可实现细节信息的稳定增强, 并为后续分层融合框架奠定理论基础。

3. 融合算法设计与实现

3.1. 模型总体框架

针对可见光与红外图像在结构信息与纹理信息表达上的差异性, 本文提出一种基于显著性特征的分层融合方法。该方法以互补特征分解为基础, 将源图像分解为主体结构特征(Main Structure Feature, MSF)和细节纹理特征(Detail Texture Feature, DTF), 并围绕不同层次特征构建差异化融合策略, 实现结构保持与细节增强的协同优化。算法整体流程如图 1 所示。

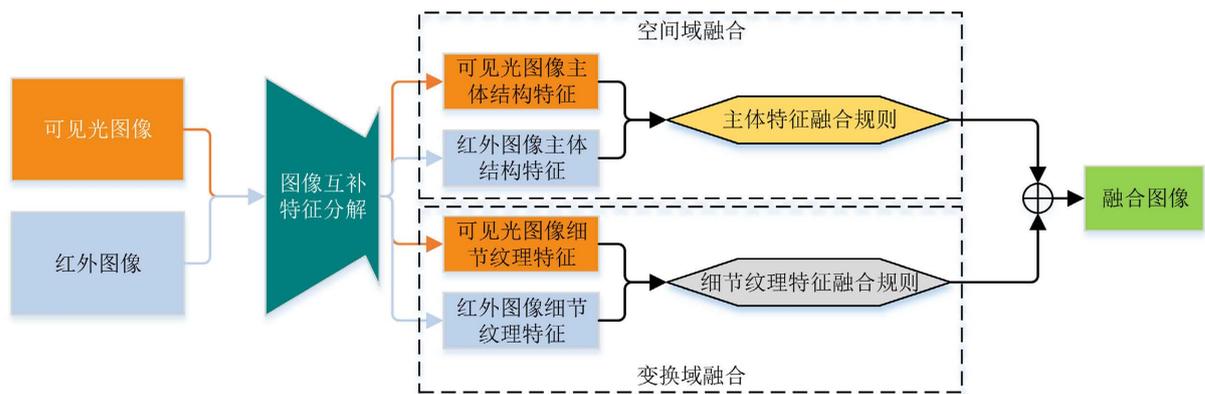


Figure 1. Overall flowchart of the algorithm

图 1. 算法整体流程图

在输入可见光图像与红外图像后, 首先进行必要的预处理, 并基于局部相对衰减率构建互补特征分解模型, 将源图像表示为 MSF 与 DTF 两部分。其中, MSF 主要描述图像的整体亮度分布与目标轮廓结构, DTF 则包含边缘、高频纹理及局部细节信息。分解过程中通过约束红外图像的局部变化特性, 实现随机噪声的有效抑制, 从而为后续融合提供稳定的特征基础。

在主体结构层面, 构建显著性驱动融合机制。通过联合建模像素灰度显著性与多尺度梯度显著性, 获得反映结构重要程度的权重信息; 在此基础上引入全变分(TV)约束, 对融合结果进行空间正则化处理, 以增强权重分布的连续性, 避免突变引起的结构失真与伪影, 从而保证融合图像的结构完整性与视觉一致性。

对于细节纹理层, 采用参数自适应脉冲耦合神经网络(PAPCNN)进行融合。两源图像的 DTF 作为网络输入, 首先使用非下采样剪切波变换(Non-subsampling shearlet transform, NSST)分解为低频系数和多方向的高频子带, 根据子带统计特性自适应设置模型参数, 通过神经元迭代激发获取放电次数, 并据此构建融合权重, 实现纹理与边缘信息的自适应增强, 同时抑制噪声区域的过度响应。最后利用融合后的低

频系数和高频子带进行重构，得到融合后的变换域图像。

最终，对融合后的 MSF 与 DTF 进行加权融合，得到完整融合图像。整体分层融合框架能够充分发挥可见光图像的细节表达优势，同时保持红外图像的目标显著性，从而获得信息量更丰富、视觉一致性更优的融合结果。

3.2. 图像互补特征分解

为实现结构信息与细节信息的有效分离，本文基于局部全变分(Local Total Variation, LTV)构建互补特征分解模型。该方法通过分析像素在滤波前后的局部变化程度，刻画区域的结构稳定性与纹理活跃程度，从而将源图像自适应分解为主体结构特征和细节纹理特征。

(1) 建立局部指标划分像素属性

图像中不同区域的空间变化特性存在显著差异。平滑区域或目标主体内部通常具有较小的梯度变化，而边缘与纹理区域则表现出较大的局部波动。为刻画像素的局部结构特性，首先定义图局部全变分响应为：

$$LTV_{\sigma}(f)(x) = L_{\sigma} \times |\nabla f|(x) \quad (1)$$

其中， $f(x)$ 为源图像， L_{σ} 表示尺度为 σ 的平滑算子， ∇f 为图像梯度算子，式(1)反映了像素邻域内的局部变化强度。

在此基础上，通过比较图像在滤波前后的局部全变分变化情况，构建像素属性判别指标，其表达式为：

$$\lambda(x) = \frac{LTV_{\sigma}(f)(x) - LTV_{\sigma}(L_{\sigma} \times f)(x)}{LTV_{\sigma}(f)(x)} \quad (2)$$

该指标描述了局部结构在平滑处理后的相对衰减程度。当 $\lambda(x)$ 较小时，说明该区域变化稳定，属于结构主导区域；当 $\lambda(x)$ 较大时，则表明该区域包含丰富的边缘或纹理信息，可判定为细节主导区域。

(2) 原始图像与滤波图像相对衰减率的加权计算

在获得相对衰减率后，根据像素属性对原始图像与滤波图像进行自适应加权，以实现结构与纹理信息的分离。主体结构分量定义为：

$$u(x) = w(\lambda(x))L_{\sigma} \times f + (1 - w(\lambda(x)))f \quad (3)$$

细节纹理分量则表示为：

$$v(x) = f(x) - u(x) \quad (4)$$

其中， $u(x)$ 表示主体结构特征(MSF)，主要包含图像的低频结构与整体亮度信息； $v(x)$ 表示细节纹理特征(DTF)，主要反映边缘、高频纹理及局部细节。

权重函 $w(\lambda)$ 根据相对衰减率进行分段定义：

$$w(\lambda) = \begin{cases} 0 & \lambda \leq a_1 \\ (\lambda - a_1)/(a_2 - a_1) & a_1 \leq \lambda \leq a_2 \\ 1 & \lambda \geq a_2 \end{cases} \quad (5)$$

其中， a_1 和 a_2 为阈值参数，用于控制结构与纹理的分离程度。当 λ 较小时，权重趋向于原始图像以保持结构稳定性；当 λ 较大时，权重更多依赖滤波结果，以增强高频细节信息。

通过上述自适应加权策略，可获得结构稳定、噪声较低的主体结构分量以及包含丰富细节信息的纹

理分量。同时，由于平滑算子对高频随机噪声具有一定抑制作用，该分解过程能够在提取结构信息的同时降低红外图像中的噪声干扰，为后续分层融合提供可靠的特征基础。

3.3. 主体结构特征融合规则

主体结构特征(MSF)主要包含图像的整体亮度分布和目标轮廓信息，其融合质量直接影响最终图像的视觉一致性与目标显著性。为充分保留红外图像的目标响应，同时保持可见光图像的结构细节，本文从像素灰度差异和梯度变化两个方面构建结构显著性度量，并在此基础上引入全变分(TV)约束进行优化融合。

(1) 像素灰度显著性图构建

像素灰度显著性反映图像区域与整体灰度分布之间的差异程度。灰度差异越大，说明该区域越可能为视觉关注区域。设图像灰度级为 L ，像素 p 处的灰度显著性定义为

$$S_p = \sum_{i=0}^{L-1} M_i |I_p - I_i| \quad (6)$$

其中， I_p 为像素灰度值， I_i 为灰度级， M_i 表示对应灰度级的像素数量。该指标用于衡量像素与全局灰度分布之间的偏离程度。

在获得两源图像的像素显著性结果后，根据显著性差异构建像素显著性融合图，其表达式为

$$u_{PSM} = wu_{opt} + (1-w)u_{sar}, \quad w = \text{ceil}(u_{p1} - u_{p2}) \quad (7)$$

其中， u_{opt} 和 u_{sar} 分别表示可见光与红外图像的主体结构分量， w 为显著性判别权重。该策略能够在目标区域优先保留显著性更高的结构信息。

(2) 梯度变化显著性图构建

梯度信息能够有效描述图像的边缘与结构变化。为提高结构显著性评估的稳定性，本文采用多尺度梯度分析方法。

首先对 MSF 进行多尺度滤波处理，得到各尺度结构图像

$$u_j = \text{Wiener}(u^{j-1}, s^{j-i}) \quad (8)$$

并计算对应的局部梯度幅值

$$u_G^j = \sqrt{(h_x \times u_j)^2 + (h_y \times u_j)^2} \quad (9)$$

其中， $h_x = [-1 \ 0 \ 1]$, $h_y = h_x'$ 。为抑制结构外边缘和孤立噪声的影响，对梯度结果进行归一化处理。

$$u_{GN}^j = \begin{cases} \frac{u_G^j}{u^j} & 0.1 < \frac{u_G^j}{u^j} < 1 \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

随后计算多尺度相对梯度的平均差异

$$u_{GW} = \frac{\sum_{j=1}^n \text{ceil}(u_{GN2}^j - u_{GN1}^j)}{n} \quad (11)$$

根据梯度显著性结果构建结构显著性融合图

$$u_{SSM} = (2 - k_1 \cdot k_2)u_{opt} + (k_1 \cdot k_2)u_{sar} \quad (12)$$

其中权重系数定义为：

$$k_1 = \begin{cases} 1 & u_{GW} = 1 \\ 0.5 & \text{otherwise} \end{cases} \quad k_2 = \begin{cases} 1 & u_{opt} < u_{sar} \\ 0 & \text{otherwise} \end{cases}$$

该方法能够根据结构梯度强度自适应选择信息来源，从而增强目标轮廓并保持边缘连续性。

(3) 全变分约束融合

为避免显著性权重突变导致的结构不连续或伪影现象，在像素显著性图与结构显著性图的基础上，引入全变分约束模型进行优化。其能量函数定义为：

$$E(x) = \frac{1}{p} \|x - u_{PSM}\|_p^p + \lambda \frac{1}{q} \|\nabla x - \nabla u_{SSM}\|_q^q \quad (13)$$

其中，第一项保证融合结果与像素显著性保持一致，第二项通过梯度约束维持结构连续性， λ 为正则化参数。优化问题可表示为：

$$y^* = \arg \min_y \left\{ \sum_{i=1}^{mm} |y_i - (u_{PSM} - u_{SSM})| + \lambda J(y) \right\} \quad (14)$$

其中全变分项为：

$$J(y) = \sum_{i=1}^{mm} |\nabla_i y| = \sum_{i=1}^{mm} \sqrt{(\nabla_i^h y)^2 + (\nabla_i^v y)^2} \quad (15)$$

通过上述优化求解，可获得空间分布平滑且结构一致的主体结构融合结果，有效减少伪影并提高整体视觉质量。

3.4. 细节纹理特征融合规则

细节纹理特征(DTF)主要包含图像中的边缘信息、高频纹理及局部结构变化，是影响融合图像清晰度与细节表现能力的重要因素。近年来，将多尺度变换方法与脉冲耦合神经网络(PCNN)相结合的融合策略取得了良好效果。PCNN 具有空间同步激发特性，能够根据邻域连接关系突出纹理显著区域，在图像融合领域具有广泛应用。然而，传统 PCNN 参数依赖经验设置，模型复杂且适应性不足，容易在不同场景下产生过度或欠激发现象。

为提高模型的自适应能力，本文采用参数自适应脉冲耦合神经网络(PAPCNN)对 DTF 进行融合。该模型根据图像特征分布自动调节网络参数，使神经元响应与纹理强度相匹配，从而实现细节信息的自适应增强。

PAPCNN 神经元模型数学表达式如下：

$$F_{ij}(n) = I_{ij} \quad (16)$$

$$L_{ij}(n) = V_L \sum_{a,b} W_{ij,ab} Y_{ij,ab}(n-1) \quad (17)$$

$$U_{ij}(n) = \exp(-\alpha_f) U_{ij}(n-1) + F_{ij}(n) (1 + \beta L_{ij}(n)) \quad (18)$$

$$Y_{ij}(n) = \begin{cases} 1, & U_{ij}(n) > \theta_{ij}(n-1) \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

$$\theta_{ij}(n) = \exp(-\alpha_\theta) \theta_{ij}(n-1) + V_\theta Y_{ij}(n) \quad (20)$$

其中， n 为迭代次数； $F_{ij}(n)$ 为外部输入项； $L_{ij}(n)$ 为链接输入项； $U_{ij}(n)$ 为神经元内部活动项； $Y_{ij}(n)$ 为神经元输出脉冲； $\theta_{ij}(n)$ 为动态阈值。 $W_{ij,ab}$ 表示神经元之间的连接权系数， V_L 为连接输入的振幅； V_θ 为动态阈值的振幅； α_θ 和 α_f 为衰减系数； β 为链接强度系数。

在迭代过程中，神经元通过内部活动项与动态阈值之间的比较产生脉冲输出。纹理显著区域由于局部灰度变化剧烈，在网络中更易产生同步激发，其放电次数相对较高。本文通过统计神经元在迭代过程中的放电次数构建融合权重，实现两源图像 DTF 分量的自适应加权融合。

相比传统 PCNN 模型，PAPCNN 通过参数自适应调节机制增强了对不同场景纹理分布的适应能力，能够有效突出边缘与高频细节，同时抑制随机噪声区域的过度响应，从而提高细节层融合的稳定性与视觉清晰度。

4. 实验结果

为验证所提方法的有效性与鲁棒性，本文在 LLVIP 和 M3FD 两个公开可见光 - 红外数据集上进行实验，并与多种典型融合算法进行对比，包括 DSFusion、RFN-Nest、U2Fusion、VSFF、FCMPCA、DTCWT、DWT、IBLPCA、NSCT、PCA 和 SWT 等方法。实验从信息量、结构保持能力及视觉质量等多个方面进行定量评价。

评价指标包括：熵(EN)、边缘信息保持度(EIN)、结构相似性(SSIM)、特征互信息(FMI)、加权融合质量指数(WFQI)、边缘保持质量指数(EFQI)、视觉信息保真度(VIF)以及基于边缘的融合评价指标(Qabf)。其中，EN、FMI 和 VIF 反映图像信息丰富程度，SSIM 和 WFQI 衡量结构保持能力，EFQI 与 Qabf 用于评价边缘与细节保持性能。

4.1. LLVIP 数据集实验结果

LLVIP 数据集是一个专为低光照视觉任务设计的可见光 - 红外配对数据集，包含 16,836 对严格时空对齐的图像，采集于极暗环境(晚 6 点至 10 点)，使用 HIKVISION 双目摄像头(可见光分辨率 1920 × 1080，红外 1280 × 720)，注册后统一为 1080 × 720 分辨率以确保视场一致。该数据集通过红外图像反向映射标注可见光图像中的行人，覆盖 26 个场景的丰富行人目标(中距离中等尺寸)，支持图像融合、低光行人检测及图像转换等任务。LLVIP 在规模(16,836 对)、低光条件、时空对齐精度及行人标注密度(每对均含行人)上显著领先，尤其适用于算法在极端暗光下的鲁棒性验证。

各方法在 LLVIP 数据集上的定量结果如表 1 所示。可以看出，本文方法在多个关键评价指标上均取得较优性能。其中，信息熵(EN)达到 0.6954，显著高于其他对比方法，表明融合图像包含更丰富的有效信息；特征互信息(FMI)和视觉信息保真度(VIF)分别达到 0.9478 和 0.7903，说明所提方法在源图像特征保持和视觉信息传递方面具有明显优势。同时，WFQI 和 EFQI 指标分别为 0.8222 和 0.3585，反映出融合结果在整体结构一致性和边缘细节保持方面表现良好。结合图 2 所示的主观融合效果可以进一步验证上述结论：在极低照度场景下，深度学习方法容易出现目标轮廓模糊或细节纹理缺失，而传统方法则可能引入噪声或导致对比度不足；相比之下，本文方法在突出红外目标显著性的同时，有效保留了可见光图像中的细节纹理信息，融合结果整体清晰度更高、结构更完整，暗光区域的层次感和细节表现能力均更为突出，主客观评价结果具有良好一致性。

Table 1. Experimental group design
表 1. 实验组设计

Models	DSFusion	rfn-nest	U2Fusion	VSFF	FCMPCA	DTCWT	DWT	IBLPCA	NSCT	PCA	SWT	Our
EIN	0.0866	0.0632	0.0592	0.0998	0.0487	0.0877	0.0923	0.0484	0.2005	0.0450	0.0673	0.1072
EN	0.0286	0.0988	0.1070	0.0487	0.1281	0.0623	0.3323	0.1867	0.1919	0.2699	0.0517	0.6954
SSIM	0.9978	0.9987	0.9987	0.9985	0.9988	0.9989	0.9978	0.9987	0.9985	0.9983	0.9989	0.9985

续表

FMI	0.9328	0.9172	0.9318	0.9292	0.9328	0.9332	0.9336	0.9338	0.8572	0.9414	0.9276	0.9478
WFQI	0.8125	0.6385	0.6818	0.7477	0.5322	0.8443	0.7661	0.5114	0.7106	0.3549	0.7878	0.8222
EFQI	0.3529	0.1783	0.1919	0.2483	0.1102	0.3061	0.2142	0.1003	0.0805	0.0484	0.2574	0.3585
VIF	0.4152	0.4744	0.5005	0.4169	0.4213	0.4874	0.7791	0.4280	0.3924	0.4703	0.4165	0.7903
Qabf	0.7947	0.6328	0.6642	0.7925	0.6579	0.7460	0.6598	0.6663	0.4703	0.7643	0.7102	0.7904

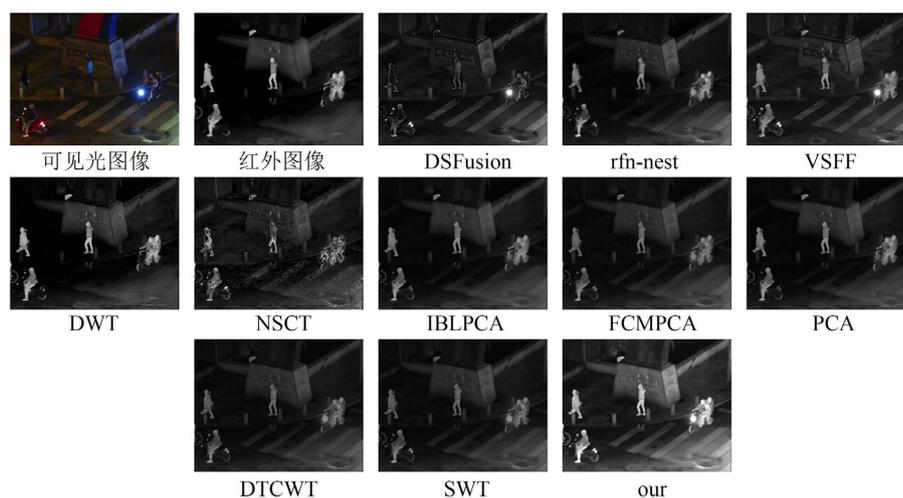


Figure 2. Experimental results on the LLVIP dataset
图 2. LLVIP 数据集实验结果图

尽管部分方法在 SSIM 指标上略高, 但差异较小, 说明本文方法在保证结构相似性的同时, 进一步提升了信息量和细节表现能力。总体而言, 在极低照度环境下, 所提方法能够有效突出红外目标并保持可见光细节, 具有较好的鲁棒性。

4.2. M3FD 数据集实验结果

M3FD 数据集: 数据采集系统包含一个双目光学相机和一个双目红外传感器, 数据采集地点包括大连理工大学校园、大连金石滩国家旅游度假区和大连金州区的主要道路。数据集共包含 8400 张图像用于融合、检测和基于融合的检测, 其中 600 张为独立场景的融合图像。图像对总数为 4200 对, 其中 300 对为独立场景的融合图像。图像格式为 24 位灰度位图(红外)和 24 位彩色位图(可见光), 图像尺寸主要为 1024×768 像素。所有图像对均经过配准, 可见光图像通过系统内部参数校准, 红外图像通过单应矩阵进行人工失真处理。数据集共标注了 34,407 个目标, 涵盖 6 种类别: 行人、汽车、公交车、摩托车、路灯和卡车。

各方法在 M3FD 数据集上的定量结果如表 2 所示。可以看出, 本文方法在多个关键指标上保持了较为突出的综合性能, 其中边缘信息指标(EIN)和信息熵(EN)分别达到 0.4994 和 0.8815, 处于对比方法中的较高水平, 表明融合结果在目标轮廓表达和整体信息量保持方面具有明显优势; 同时, 视觉信息保真度(VIF)达到 0.6353, 说明所提方法能够更有效地传递源图像中的重要视觉信息。结合图 3 所示的主观融合结果可以发现, 在复杂道路与多目标场景下, 部分方法虽然在局部对比度上表现尚可, 但容易出现细节模糊或背景结构不完整的问题, 而本文方法能够在突出红外目标的同时较好地保留可见光图像中的道路结构、车辆轮廓及场景层次信息, 融合图像整体清晰度和细节表现更为均衡, 进一步验证了该方法在复

杂场景下的鲁棒性与有效性。

Table 2. Experimental group design
表 2. 实验组设计

Models	DSFusion	rfn-nest	U2Fusion	VSFF	FCMPCA	DTCWT	DWT	IBLPCA	NSCT	PCA	SWT	Our
EIN	0.3109	0.1899	0.2418	0.4080	0.0471	0.0830	0.0869	0.0472	0.2254	0.0477	0.0594	0.4994
EN	0.9396	0.6409	0.8800	0.7826	0.3001	0.2975	0.3209	0.3001	0.3786	0.3013	0.3018	0.8815
SSIM	0.9870	0.9922	0.9923	0.9852	0.9991	0.9991	0.9983	0.9991	0.9989	0.9991	0.9991	0.9982
FMI	0.8432	0.8449	0.8357	0.8419	0.9532	0.9524	0.9495	0.9533	0.8784	0.9537	0.9524	0.9444
WFQI	0.8331	0.7116	0.7753	0.8057	0.7150	0.8726	0.8007	0.7163	0.7058	0.7253	0.8302	0.8136
EFQI	0.3450	0.2990	0.2096	0.3311	0.3239	0.4458	0.3863	0.3274	0.0605	0.3397	0.3532	0.3291
VIF	0.3370	0.2965	0.4500	0.6112	0.3508	0.4654	0.4882	0.3502	0.3668	0.3459	0.3838	0.6353
Qabf	0.8121	0.6521	0.6426	0.8190	0.7071	0.7837	0.7795	0.7075	0.2951	0.7125	0.7433	0.8078

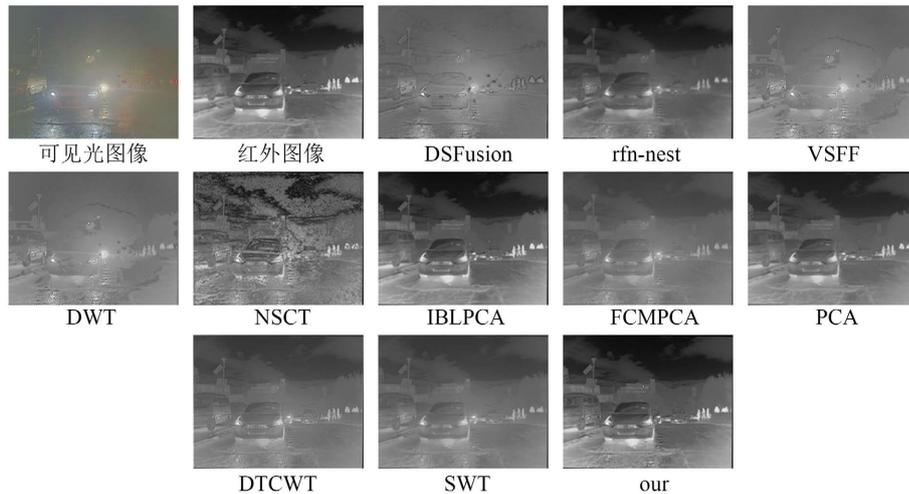


Figure 3. Experimental results on the M3FD dataset
图 3. M3FD 数据集实验结果图

在结构保持指标方面，SSIM 与 WFQI 均保持较高水平，说明融合结果在不同场景下具有良好的结构一致性与稳定性。与 LLVIP 实验结果相一致，本文方法在复杂室外环境中仍能有效突出目标区域并保持背景细节，表现出较好的泛化能力。

4.3. 结果分析

综合 LLVIP 与 M3FD 两个数据集的实验结果可以看出，本文方法在 EN、FMI 和 VIF 等信息量相关指标上均取得较优性能，表明融合图像能够有效保留源图像中的关键信息；同时，在 WFQI、EFQI 和 Qabf 等指标上保持较高水平，说明所提方法在结构保持与边缘细节表现方面具有良好效果。其性能提升主要得益于基于局部相对衰减率的互补特征分解能够实现结构与纹理信息的有效分离并抑制红外噪声，以及结构层显著性约束与纹理层 PAPCNN 自适应增强的协同作用。实验结果表明，所提方法在低光环境和复杂场景下均具有良好的稳定性与鲁棒性，能够获得信息丰富、结构清晰且视觉效果良好的融合结果。

5. 结语

针对可见光与红外图像在结构表达与纹理分布方面存在差异、传统融合方法难以兼顾目标显著性与细节保持的问题，本文提出一种基于显著性特征的可见光和红外图像融合算法。首先，基于局部相对衰减率构建互补特征分解模型，将源图像自适应分解为主体结构特征和细节纹理特征，在主体结构特征中使用联合像素灰度显著性与多尺度梯度显著性进行建模，并引入全变分约束优化融合结果，提高结构连续性与视觉一致性；最后，在细节纹理特征中采用 NSST 加参数自适应脉冲耦合神经网络实现细节信息的自适应增强。实验结果表明，所提方法在 LLVIP 和 M3FD 数据集上在信息量、视觉保真度及边缘保持等指标上均表现出较优性能，能够在低光与复杂场景条件下获得信息丰富、结构清晰且视觉效果良好的融合结果。未来将进一步研究算法的实时实现及其在目标检测等下游任务中的应用性能。

参考文献

- [1] Luo, Y. and Luo, Z. (2023) Infrared and Visible Image Fusion: Methods, Datasets, Applications, and Prospects. *Applied Sciences*, **13**, Article 10891. <https://doi.org/10.3390/app131910891>
- [2] Ma, W., Wang, K., Li, J., Yang, S.X., Li, J., Song, L., et al. (2023) Infrared and Visible Image Fusion Technology and Application: A Review. *Sensors*, **23**, Article 599. <https://doi.org/10.3390/s23020599>
- [3] Wang, Q., Jin, P., Wu, Y., Zhou, L. and Shen, T. (2025) Infrared Image Enhancement: A Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **18**, 3281-3299. <https://doi.org/10.1109/jstars.2024.3523418>
- [4] Liu, J., Wu, G., Liu, Z., Wang, D., Jiang, Z., Ma, L., et al. (2025) Infrared and Visible Image Fusion: From Data Compatibility to Task Adaption. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **47**, 2349-2369. <https://doi.org/10.1109/tpami.2024.3521416>
- [5] Zou, D. and Yang, B. (2023) Infrared and Low-Light Visible Image Fusion Based on Hybrid Multiscale Decomposition and Adaptive Light Adjustment. *Optics and Lasers in Engineering*, **160**, Article 107268. <https://doi.org/10.1016/j.optlaseng.2022.107268>
- [6] Hu, P., Wang, C., Li, D. and Zhao, X. (2023) An Improved Hybrid Multiscale Fusion Algorithm Based on NSST for Infrared-Visible Images. *The Visual Computer*, **40**, 1245-1259. <https://doi.org/10.1007/s00371-023-02844-8>
- [7] Wang, W., Zhang, J., Liu, H., Xiong, W. and Zhang, C. (2023) Joint Low-Rank and Sparse Decomposition for Infrared and Visible Image Sequence Fusion. *Infrared Physics & Technology*, **133**, Article 104828. <https://doi.org/10.1016/j.infrared.2023.104828>
- [8] Zhang, X. and Demiris, Y. (2023) Visible and Infrared Image Fusion Using Deep Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 10535-10554. <https://doi.org/10.1109/tpami.2023.3261282>
- [9] Liu, J., Dian, R., Li, S. and Liu, H. (2023) SGFusion: A Saliency Guided Deep-Learning Framework for Pixel-Level Image Fusion. *Information Fusion*, **91**, 205-214. <https://doi.org/10.1016/j.inffus.2022.09.030>
- [10] Gu, X., Xia, Y. and Zhang, J. (2024) Multimodal Medical Image Fusion Based on Interval Gradients and Convolutional Neural Networks. *BMC Medical Imaging*, **24**, Article No. 232. <https://doi.org/10.1186/s12880-024-01418-x>
- [11] Ibrahim, S.I., El-Tawel, G.S. and Makhlof, M.A. (2024) Brain Image Fusion Using the Parameter Adaptive-Pulse Coupled Neural Network (PA-PCNN) and Non-Subsampled Contourlet Transform (NSCT). *Multimedia Tools and Applications*, **83**, 27379-27409. <https://doi.org/10.1007/s11042-023-16515-2>
- [12] Yuan, D., Zhang, H., Shu, X., Liu, Q., Chang, X., He, Z., et al. (2024) Thermal Infrared Target Tracking: A Comprehensive Review. *IEEE Transactions on Instrumentation and Measurement*, **73**, 1-19. <https://doi.org/10.1109/tim.2023.3338701>
- [13] Wang, H., Lou, J., Liu, X., Tan, H., Whitaker, R. and Liu, H. (2024) SSPNet: Predicting Visual Saliency Shifts. *IEEE Transactions on Multimedia*, **26**, 4938-4949. <https://doi.org/10.1109/tmm.2023.3327886>
- [14] Zhao, J., Zhang, T., Fang, S., Gao, J., Wang, J. and Gong, M. (2026) Spatial-Spectral Texture-Preserved Total Variation: A Novel Regularization for Hyperspectral Image Denoising. *IEEE Transactions on Circuits and Systems for Video Technology*, **36**, 248-260. <https://doi.org/10.1109/tcsvt.2025.3598368>
- [15] Gupta, P. and Jain, N. (2024) Anisotropic Diffusion Filter Based Fusion of NSST Transformed Medical Images. *Biomedical Signal Processing and Control*, **90**, Article 105819. <https://doi.org/10.1016/j.bspc.2023.105819>