

基于结构-纹理分解和 VGG 深层网络的红外与可见光图像融合

杨飞燕^{1,2}, 王蒙^{1,2*}

¹昆明理工大学信息与自动化学院, 云南 昆明 650500;

²云南省人工智能重点实验室, 云南 昆明 650500

摘要 针对当前红外与可见光图像融合方法中存在低频信息未充分利用,以及高频细节易混入噪声等问题,提出一种基于结构-纹理(ST)分解和 VGG 深层网络的红外与可见光图像融合方法。首先用均值滤波将输入图像分解为高-低频子带,并引入 ST 对低频子带二次分解,采用平均法则与邻域空间频率分别对结构和纹理进行预融合;同时,将输入图像送入 VGG 网络得到多层特征映射,并用 Sigmoid 函数实现高频子带的归一化预融合;最后利用预融合的高频、低频结构和低频纹理 3 个频带进行图像融合重建。实验结果表明,提出的方法能够融合图像的深层细节特征,有效保留纹理细节并抑制噪声,且在噪声评估、结构相似性、均方误差、峰值信噪比等客观指标方面具有明显优势。

关键词 图像处理; 图像融合; 红外图像; 可见光图像; 结构-纹理分解; VGG 网络

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP212808

Infrared and Visible Image Fusion Based on Structure-Texture Decomposition and VGG Deep Networks

Yang Feiyan^{1,2}, Wang Meng^{1,2*}

¹Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, Yunnan, China;

²Key Laboratory of Artificial Intelligence in Yunnan Province, Kunming 650500, Yunnan, China

Abstract To address the problems of underutilization of low-frequency information and easy mixing of high-frequency details with noise in the current infrared and visible image fusion methods, an infrared and visible image fusion method based on structure-texture (ST) decomposition and VGG deep network is proposed. First, the input image is decomposed into high-low frequency subbands using mean filtering, and ST is introduced to re-decompose the low-frequency subbands. The structure and texture are pre-fused by absolute maximum and neighborhood spatial frequency, respectively. Subsequently, the input image is input into the VGG network to get the multi-layer feature mapping, and the Sigmoid function is used to realize the normalized pre-fusion of the high-frequency subband. Finally, the pre-fused high-frequency, low-frequency structure, and low-frequency texture are used for image fusion and reconstruction. Experimental results show that the proposed algorithm can fuse the deep detail features of images, retain texture details, and suppress noise effectively, and has significant advantages in noise assessment, structural similarity index measure, mean square error, peak signal to noise ratio, and other objective indexes.

Key words image processing; image fusion; infrared image; visible image; structure-texture decomposition; VGG network

1 引言

图像融合是一种图像增强技术,该技术有效解决了单一传感器成像能力有限,难以获得全面的空间信息等问题。红外与可见光图像各具优点,红外图像反

映热辐射信息,其具有较强的抗干扰能力,但纹理信息缺失;而可见光图像捕获光信息,具有丰富的纹理细节,但容易被恶劣环境影响。因此,红外与可见光图像融合在图像处理、监控分析等领域应用广泛。

目前,国内外学者对红外与可见光图像融合方法

收稿日期: 2021-10-26; 修回日期: 2021-11-12; 录用日期: 2021-11-30

基金项目: 国家自然科学基金(62062048)

通信作者: *vicong68@qq.com

进行了众多研究。Kumar 等^[1]利用交叉双边滤波获取细节图,并根据信息强度计算归一化权重系数。Jiang 等^[2]将源图像进行二频带分解,结合显著性特征构建权重映射进行融合。多尺度分解在红外与可见光图像融合中^[3]被实现于基于金字塔、离散小波、非下采样轮廓波、曲波的方法等。二频带分解格式在这些融合方法中已被广泛应用^[2-4],且通常采用取大法或平均加权对低频子带进行融合,这可能会导致对比度降低和边缘退化。实际上,这些尺度分解工具的本质是多次分解图像,但需要分解多少次才能将高频信息和低频信息很好地分离,这仍然是个未知数。同时,多重分解可能会导致高频和低频重建时增加复杂度,使重建结果不理想。为此,Li 等^[5]提出一种基于三频带分解和稀疏表示的融合方法,引入结构-纹理(ST)分解模型^[6],充分利用低频子带的结构与细节信息,并同步实现融合和去噪。Yin 等^[7]提出了一种将多尺度变换与字典学习相结合的融合方法,充分体现了二者融合的优势,但学习字典使得方法效率变低,耗时过长。此外,深度学习通过有效学习网络参数,能从图像中提取出显著特征并用于后续任务^[8]。这种强大的特征提取和数据表示能力已成功应用于人脸识别、图像生成等领域。Liu 等^[9]将卷积稀疏表示应用于图像融合,用于提取多层特征并生成融合图像。Wang 等^[10]提出了一种图像融合生成模型,将分析模块得到的显著特征图进行集成,通过梯度迭代反向传播生成融合图像。基于深度学习^[11]的融合方法一般能够获得较好的融合效果,但仍存在 3 个主要缺点:1) 通常需要大量特定的数据集进行训练;2) 仅能够利用神经网络最后一层的特征结果,更多有用的网络浅层特征未被综合利用;3) 将其直接应用于多模态图像融合中易于产生伪影和噪声。

本文提出一种基于 ST 分解和 VGG 深层网络的红外与可见光图像融合方法。首先将输入图像进行二频带分解,通过 ST 模型充分利用低频的结构与纹理信息;结合 VGG 深层网络提取多层特征,采用 Sigmoid 函数归一化策略实现高频子带的预融合。由于本文方法是基于预先训练的神经网络,因此可以直接有效地提取图像特征,得到更适合的权重图。同时,三频带分解能够根据特征属性有效地表示图像,充分利用低频子带的结构和纹理,提高了红外与可见光图像融合效果。

2 相关工作

2.1 ST 分解

图像通常可视为结构子带和纹理子带的叠加,为了精确分解图像结构和纹理信号, Lee 等^[6]提出的 ST 分解模型中,结构子带提取的主要目的是消除图像中的纹理信号,同时尽可能地恢复结构信号之间与输入图像的相似性^[5]。在离散信号 I 中,像素 p 的间隔梯度滤波器被定义为

$$(\nabla_{\Omega} I)_p = g_{\tilde{\sigma}}^r(I_p) - g_{\tilde{\sigma}}^l(I_p), \quad (1)$$

式中: Ω 是中心像素 p 周围的邻域; $g_{\tilde{\sigma}}^r$ 和 $g_{\tilde{\sigma}}^l$ 分别是右、左剪切一维高斯函数,具体优化公式为

$$g_{\tilde{\sigma}}^r(I_p) = \frac{1}{k_r} \sum_{n \in \Omega(p)} \omega_{\tilde{\sigma}}(n - p - 1) I_n, \quad (2)$$

$$g_{\tilde{\sigma}}^l(I_p) = \frac{1}{k_l} \sum_{n \in \Omega(p)} \omega_{\tilde{\sigma}}(p - n) I_n, \quad (3)$$

$$\omega_{\tilde{\sigma}}(x) = \begin{cases} \exp\left(-\frac{x^2}{2\tilde{\sigma}^2}\right), & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

式中: $\omega_{\tilde{\sigma}}$ 表示剪切指数加权函数; k_r 和 k_l 为归一化系数; $\tilde{\sigma}$ 是标度参数,文献^[6]建议 $\tilde{\sigma} = 3$ 。为了获得结构信息,纹理信息不应该存在于结构区域,采用式(5)将输入信号 I 的梯度细化为以下间隔梯度:

$$(\tilde{\nabla} I)_p = \begin{cases} (\nabla I)_p \cdot \omega_p, & \text{if } \text{sign}[(\nabla I)_p] = \text{sign}[(\nabla_{\Omega} I)_p] \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

式中: $(\tilde{\nabla} I)_p$ 是重标度梯度; ω_p 代表重标度权重。

$$\omega_p = \min\left(1, \frac{|(\nabla_{\Omega} I)_p| + \epsilon}{(\nabla I)_p + \epsilon}\right), \quad (6)$$

式中, ϵ 表示一个小常数,以避免数据不稳定,通常将 ϵ 设置为 10^{-4} 。对于二维图像信号,在 x 和 y 方向交替使用一维信号的区间梯度滤波操作以迭代方式收敛从而获得最终图像的结构层。

2.2 VGG-16 卷积神经网络结构

图像融合任务中的一个关键因素是获取活动水平测量,通过深度卷积神经网络的高维空间关系可以提取到更为精细的纹理、边缘等特征。

VGG 网络作为一个经典的卷积神经网络^[12],探索了网络的深度与性能之间的关系,在深度特征提取任务中功能强大,易于实现。本文提出的方法是基于 ST 分解与 VGG-16 网络相结合的融合方法, VGG-16 网络主要由 5 个卷积组构成,最后经 Softmax 输出结果^[13]。当 VGG-16 网络的输入图像大小为 $N \times N$ 时^[14],各卷积组的相关参数如表 1 所示。为获取更具

表 1 VGG-16 结构参数

Table 1 Structural parameters of VGG-16

Convolution group	Convolution	Pooling	Channel number	Output
1(1_1, 1_2)	$3 \times 3, 1$	Max, 2×2	64	$N \times N$
2(2_1, 2_2)	$3 \times 3, 1$	Max, 2×2	128	$(N/2) \times (N/2)$
3(3_1, 3_2, 3_3)	$3 \times 3, 1$	Max, 2×2	256	$(N/4) \times (N/4)$
4(4_1, 4_2, 4_3)	$3 \times 3, 1$	Max, 2×2	512	$(N/8) \times (N/8)$
5(5_1, 5_2, 5_3)	$3 \times 3, 1$	Max, 2×2	512	$(N/16) \times (N/16)$

有表达力的特征, VGG 网络将 ReLU 函数作为所有隐藏层的激活单元。

3 融合方法

本文将输入的红外与可见光图像分别记为 I_{ir} 和 I_{vis} , 融合后的图像为 I_F 。融合方法流程如图 1 所示。由于均值滤波具有高效的平滑图像能力, 因此将待融合图像通过均值滤波进行分解^[14], 表示为

$$\begin{bmatrix} I_{ir}^L \\ I_{ir}^H \end{bmatrix} = \begin{bmatrix} I_{ir} * Z \\ I_{ir} - I_{ir} * Z \end{bmatrix}, \quad (7)$$

$$\begin{bmatrix} I_{vis}^L \\ I_{vis}^H \end{bmatrix} = \begin{bmatrix} I_{vis} * Z \\ I_{vis} - I_{vis} * Z \end{bmatrix}, \quad (8)$$

式中: “*” 为滤波运算; Z 为均值滤波器, 大小为 9×9 ; $\{I_{ir}^H, I_{vis}^H\}$ 为高频子带; $\{I_{ir}^L, I_{vis}^L\}$ 为低频子带。低频子带的融合是在输入图像已被分解为高-低频两部分的基础上, 利用 ST 分解模型对低频子带进行二次分解, 得

到低频结构:

$$I_{ir}^{L,s} = \text{IGF}(I_{ir}^L, \tilde{\sigma}, \lambda), \quad (9)$$

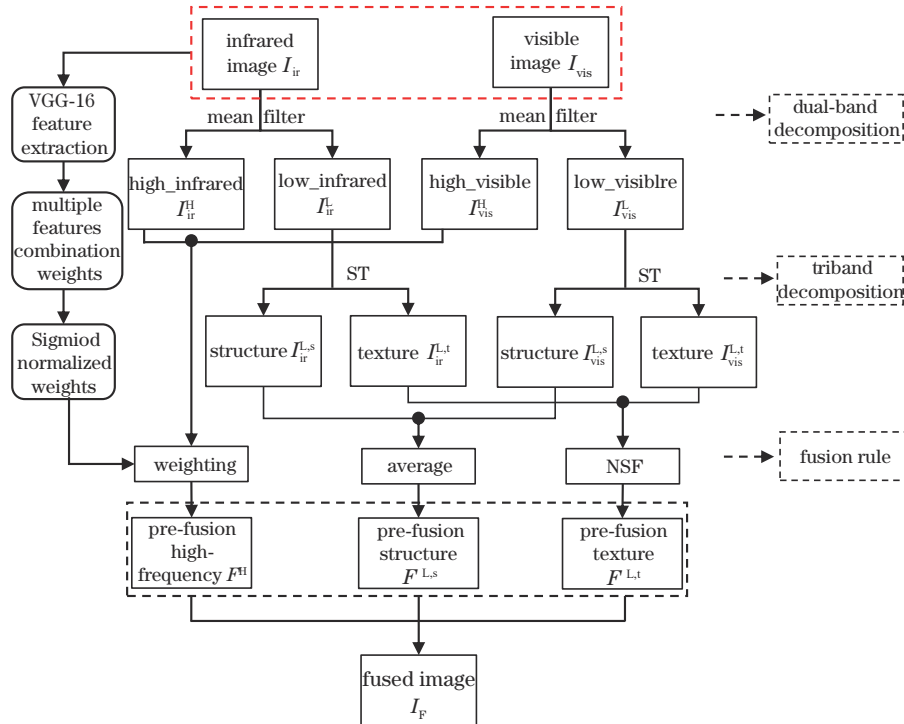
$$I_{vis}^{L,s} = \text{IGF}(I_{vis}^L, \tilde{\sigma}, \lambda), \quad (10)$$

式中: $\tilde{\sigma}$ 和 λ 分别是尺度和平滑参数; $\text{IGF}(\cdot)$ 指的是 2.1 小节中的 ST 分解操作, 具体实现细节见文献[6]。低频纹理可以由以下公式生成:

$$I_{ir}^{L,t} = I_{ir}^L - I_{ir}^{L,s}, \quad (11)$$

$$I_{vis}^{L,t} = I_{vis}^L - I_{vis}^{L,s}. \quad (12)$$

经过上述运算, 输入图像将被分解为高频、低频结构和低频纹理 3 个频带, 本文根据每一频带的特点^[5], 设计相应的融合规则进行有效融合。三频带分解结构的实例如图 2 所示, 从图中可以看出: 1) 通过均值滤波的分解, 大部分细节被保留到 (b) 列高频子带, 但 (c) 列低频子带仍然包含了一定的细节信息; 2) 经过 ST 分解后, (d) 列低频结构几乎不包含振幅信息, 局部区域通常只包含强度信息或少数大的边缘结构。



ST: structure-texture; NSF: neighborhood spatial frequency

图 1 提出方法框架图

Fig. 1 Framework of proposed method

3.1 低频结构-纹理的融合

对于低频结构-纹理的融合, 文中首先使用平均法则来融合低频结构:

$$F^{L,s} = \alpha I_{ir}^{L,s} + \beta I_{vis}^{L,s}, \quad (13)$$

式中, α 和 β 分别表示 $I_{ir}^{L,s}$ 和 $I_{vis}^{L,s}$ 的权重值, 为了保留共同特征和减少冗余信息, 本文选择 $\alpha = \beta = 0.5$ 。邻域空间频率(NSF)能够有效地检测低频纹理层的边缘、纹理、角点等信息, 因此预融合的低频纹理层 $F^{L,t}$ 可以通过以下方式获得:

$$F^{L,t} = \max[f_{\text{NSF}}(I_{ir}^{L,t}), f_{\text{NSF}}(I_{vis}^{L,t})], \quad (14)$$

式中, $\text{NSF}(f_{\text{NSF}})$ 定义为

$$f_{\text{NSF}} = \sqrt{f_{\text{RF}}^2 + f_{\text{CF}}^2}, \quad (15)$$

$$f_{\text{RF}} = \sqrt{\frac{1}{\hat{M} \times \hat{N}} \sum_{a=1}^{\hat{M}} \sum_{b=2}^{\hat{N}} [I(a, b) - I(a, b-1)]^2}, \quad (16)$$

$$f_{\text{CF}} = \sqrt{\frac{1}{\hat{M} \times \hat{N}} \sum_{a=1}^{\hat{M}} \sum_{b=2}^{\hat{N}} [I(a, b) - I(a-1, b)]^2}, \quad (17)$$

式中: $\hat{M} \times \hat{N}$ 表示窗口大小; (a, b) 是位于窗口中的中

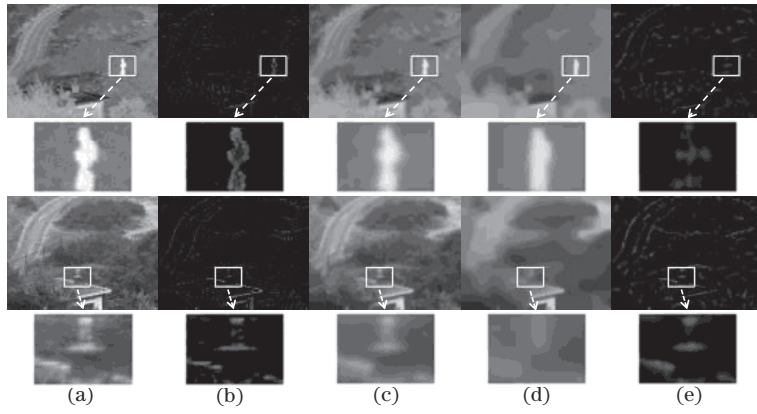


图 2 三频带分解示例图。(a)输入图像;(b)高频子带;(c)低频子带;(d)低频结构;(e)低频纹理

Fig. 2 Example diagram of triband decomposition. (a) Input images; (b) high-frequency subbands; (c) low-frequency subbands; (d) low-frequency structures; (e) low-frequency textures

心像素; f_{RF} 和 f_{CF} 分别为水平和垂直方向的空间频率。

3.2 高频子带融合

对于高频子带,先使用 VGG-16 网络提取输入图像的深层特征,然后对所提取的特征分别进行 L1 正则

化^[13]、卷积平滑^[14]、上采样和 Sigmoid 归一化策略得到最终权重图。最后将最终权重图分别作为 $\{I_{ir}^H, I_{vis}^H\}$ 的权重系数相乘得到预融合高频子带 F^H , 具体融合流程如图 3 所示。

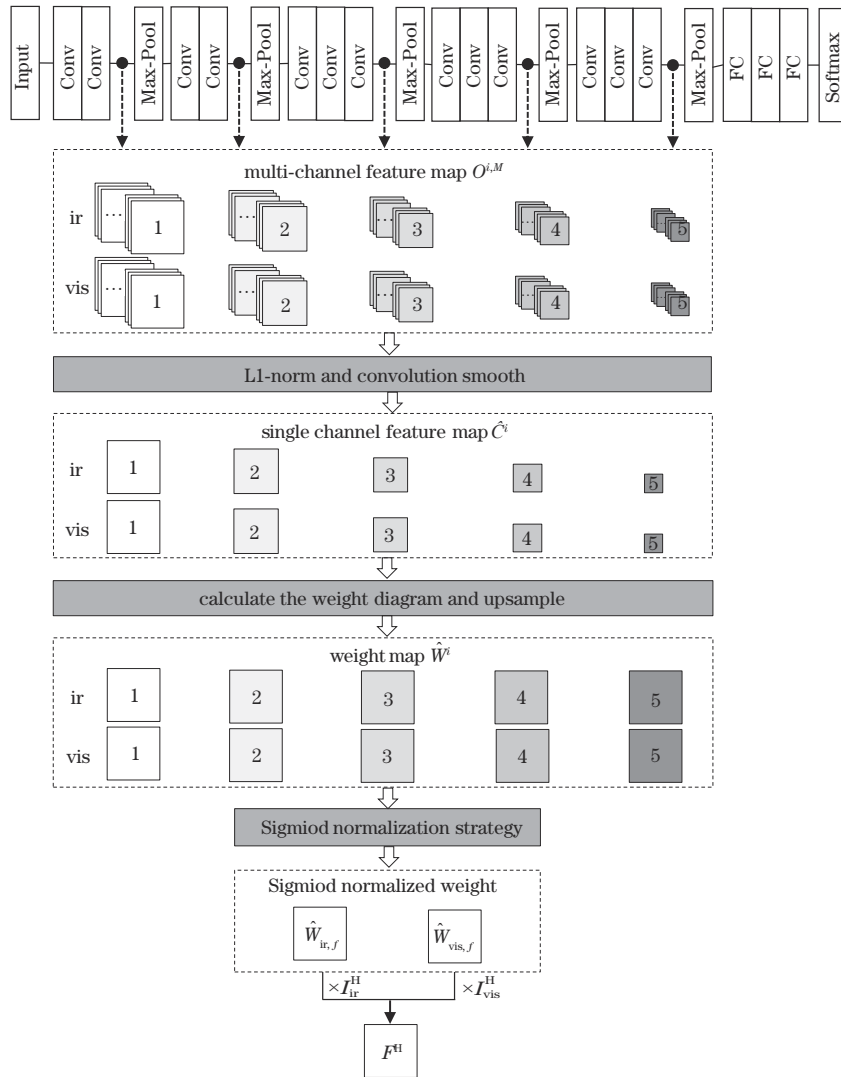


图 3 高频子带融合流程图

Fig. 3 Flow chart of high-frequency subband fusion

1) 提取多通道特征图。为了均匀全面地获得 VGG-16 网络提取到的深层特征, 本文将每个 $G_i(\cdot)$ 表示为 VGG 网络中的一个特征提取层。与文献 [13, 15] 不同的是 $i \in \{1, 2, 3, 4, 5\}$, 分别表示选取第 1、2、3、4、5 个卷积组作为特征提取层, 值得注意的是所有特征提取层均在 ReLU 运算后获得:

$$O_{ir}^{i,m} = \max[0, G_i(I_{ir}^H)], \quad (18)$$

$$O_{vis}^{i,m} = \max[0, G_i(I_{vis}^H)], \quad (19)$$

式中: $\max(0, \cdot)$ 函数表示 ReLU 操作; $O^{i,m}$ 表示高频子带经过特征提取层 $G_i(\cdot)$ 提取到的多通道特征图; m 为特征图的通道数, $m \in \{1, 2, \dots, M\}$ 。由表 1 所示的 VGG-16 网络参数可知, $O^{i,m}$ 是一个 $M = 64 \times 2^{i-1}$ 维的向量。

2) 获取单通道特征图。对得到的多通道特征图以 L1 正则化进行压缩, 获取初始的单通道特征图 $\{C_{ir}^i, C_{vis}^i\}$, 计算如下:

$$C_{ir}^i(x, y) = \|O_{ir}^{i,m}(x, y)\|_1, \quad (20)$$

$$C_{vis}^i(x, y) = \|O_{vis}^{i,m}(x, y)\|_1, \quad (21)$$

式中, $O^{i,m}(x, y)$ 表示 $O^{i,m}$ 在 (x, y) 处的特征映射值。为了获得预融合高频子带更优的局部连续表示, 文中使用卷积核 \mathbf{R} 进行平滑 [14] 得到单通道特征图 $\{\hat{C}_{ir}^i, \hat{C}_{vis}^i\}$ 。

$$\mathbf{R} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}. \quad (22)$$

3) 计算权重图。在 $\{\hat{C}_{ir}^i, \hat{C}_{vis}^i\}$ 的基础上计算初始的权重图 $\{W_{ir}^i, W_{vis}^i\}$, 公式如下:

$$W_{ir}^i(x, y) = \frac{\hat{C}_{ir}^i(x, y)}{\hat{C}_{ir}^i(x, y) + \hat{C}_{vis}^i(x, y)}, \quad (23)$$

$$W_{vis}^i(x, y) = \frac{\hat{C}_{vis}^i(x, y)}{\hat{C}_{ir}^i(x, y) + \hat{C}_{vis}^i(x, y)}. \quad (24)$$

根据图像融合的尺寸一致原则, 记 $\{W_{ir}^i, W_{vis}^i\}$ 经 VGG-

16 网络得到的宽为 (u_i, v_i) , 源图像的宽度为 (U, V) , 对初始权重图 $\{W_{ir}^i, W_{vis}^i\}$ 进行上采样, 公式如下:

$$(u_i, v_i) = (U, V) \times \frac{1}{2^{i-1}}, \quad (25)$$

得到的 $\hat{W}_{ir}^i(x, y)$ 和 $\hat{W}_{vis}^i(x, y)$ 均为 5 维向量, 其表示为第 i 组权重图上采样后在 (x, y) 处的权值。

4) Sigmoid 平均归一化加权融合。Sigmoid 是一个可以将函数值范围压缩到 $[0, 1]$ 且幅度不变的非线性函数, 广泛用于神经激活和特征归一化。因此本文将权重图 $\{\hat{W}_{ir}^i, \hat{W}_{vis}^i\}$ 经平均权重后进行 Sigmoid 归一化, 得到最终权值映射 $\{\hat{W}_{ir,f}, \hat{W}_{vis,f}\}$, 公式如下:

$$d_{ir} = \frac{\sum_{i=1}^5 \hat{W}_{ir}^i(x, y)}{5}, \quad (26)$$

$$d_{vis} = \frac{\sum_{i=1}^5 \hat{W}_{vis}^i(x, y)}{5}, \quad (27)$$

$$\hat{W}_{ir,f}(x, y) = \frac{1}{1 + e^{-d_{ir}}}, \quad (28)$$

$$\hat{W}_{vis,f}(x, y) = \frac{1}{1 + e^{-d_{vis}}}, \quad (29)$$

$$F^H = \hat{W}_{ir,f} \times I_{ir}^H + \hat{W}_{vis,f} \times I_{vis}^H. \quad (30)$$

最后, 将最终权重图 $\{\hat{W}_{ir,f}, \hat{W}_{vis,f}\}$ 分别作为 I_{ir}^H 和 I_{vis}^H 权重系数融合得到预融合高频子带 F^H 。

3.3 融合图像重建

由新获得的 3 个预融合子带进行重建, 得到最终的融合图像 I_F , 公式如下:

$$F(x, y) = F^{L,s}(x, y) + F^{L,t}(x, y) + F^H(x, y), \quad (31)$$

式中: $F^{L,s}(x, y)$ 、 $F^{L,t}(x, y)$ 和 $F^H(x, y)$ 分别表示预融合的高频、低频结构和低频纹理在 (x, y) 处的值; $F(x, y)$ 表示最终融合图像在 (x, y) 处的像素值 [14]。本文提出方法的具体实现如表 2 所示。

表 2 提出方法的具体实现

Table 2 Concrete implementation of proposed method

Input: infrared and visible image I_{ir} and I_{vis} . Output: fused image I_F
Step 1: I_{ir} and I_{vis} are decomposed by means of mean filtering
Step 2: Low-frequency subbands I_{ir}^L and I_{vis}^L are decomposed by ST decomposition to obtain low-frequency structure $F^{L,s}$ and low-frequency texture $F^{L,t}$
Step 3: Low-frequency structure-texture fusion
Step 3.1: Low-frequency structures $F^{L,s}$ are fused using average rule
Step 3.2: Low frequency textures $F^{L,t}$ are fused using NSF rules
Step 4: High-frequency subband fusion
Step 4.1: Input image are fed into VGG-16 network to extract multi-channel feature map
Step 4.2: L1 regularization and convolution smoothing are performed on multi-channel feature map to obtain single channel feature map
Step 4.3: Single channel feature map is up sampled and corresponding weight map is calculated
Step 4.4: Five-dimensional weight map is normalized by sigmoid average normalization strategy to obtain normalized weight to guide fusion of high-frequency subbands, and pre-fused high-frequency image F^H is obtained
Step 5: Final fusion image I_F is reconstructed from three pre-fusion subbands ($F^{L,s}, F^{L,t}, F^H$)

4 实验结果及分析

4.1 实验设置

为了验证本文方法的可行性,选择公开数据集 TNO 中的 21 组图像进行实验。本文选取其中 5 组分别命名为“Camp”、“Street”、“Gate”、“Car”和“House”的可见光图像进行说明,并对人的主观感知和客观测量都进行了说明和分析。此外,将本文方法与 7 种对比方法进行实验比较,分别为:CBF^[1]、JSR^[16]、JSRSD^[17]、DRTV^[18]、CVT-SR^[19]、GTF^[20]和 MISF^[21],实验中的对比方法参数取值均为相关原文献中的默认值。本文实验硬件平台为 Windows 10 操作系统,AMD Ryzen 5 2600X Six-Core 处理器,12 GB 内存,编

程软件为 Matlab2019b,MatConvNet 为 1.0-beta19 版本,C++ 编译环境为 VS2019。为了降低复杂度,文中式(16)中设置 $\hat{N} = 43$ ^[5]。

4.2 主观视觉评价

“Camp”的融合结果如图 4 所示。图 4(a)、4(b)为源图像,图 4(c)中的目标人物和灌木等信息包含较多噪声和伪影,视觉效果较差;图 4(d)、4(e)中的目标人物、灌木等景物存在空间“不连续”现象;图 4(f)、4(h)中目标人物不清晰,背景细节丢失严重;图 4(g)、4(i)中可见光信息较为显著,但图 4(g)中左下角灌木有泛白现象;图 4(j)为本文方法所获取的融合图像,其融合结果的灰度分布适中,整体视觉效果自然,但边缘细节过于平滑。

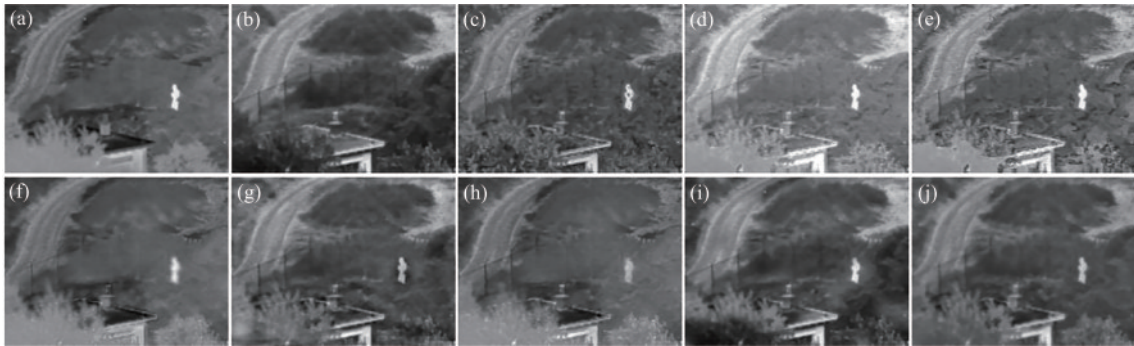


图 4 “Camp”源图像融合结果。(a)红外图像;(b)可见光图像;(c)CBF;(d)JSR;(e)JSRSD;(f)DRTV;(g)CVT-SR;(h)GTF;(i)MISF;(j)所提方法

Fig. 4 Fusion results of “Camp” source image. (a) Infrared images; (b) visible image; (c) CBF; (d) JSR; (e) JSRSD; (f) DRTV; (g) CVT-SR; (h) GTF; (i) MISF; (j) proposed method

“Street”的融合结果如图 5 所示。图 5(c)的融合结果包含较多噪声和晕影,观感不佳;图 5(d)、5(e)中的车灯、屋窗细节过于生硬,使融合图像视觉效果不协调;图 5(f)、5(h)中广告牌的可见光信息丢失,边缘模糊;图 5(g)中目标人物清晰,但图像灰度分布不均匀;图 5(i)中目标人物清晰,但左上角的窗户周围融合不完全,基本保留的都是可见光信息;图 5(j)中物体连续性好,对比度适中。

“Gate”的融合结果如图 6 所示。图 6(c)、6(f)和 6(h)中可见光信息几乎全部丢失,其中图 6(c)的天空、树木融合不协调,目标人物丢失信息且变为空心;图 6(d)、6(e)中整体亮度偏高,且存在大量人工噪声,不利于人眼观察;图 6(g)中四角位置泛白,存在背景分布灰度不均的情况,局部细节表达略有不足;图 6(i)中目标人物清晰,但天空、地面等区域明显融合不完全,造成视觉效果不协调;图 6(j)中门框的纹

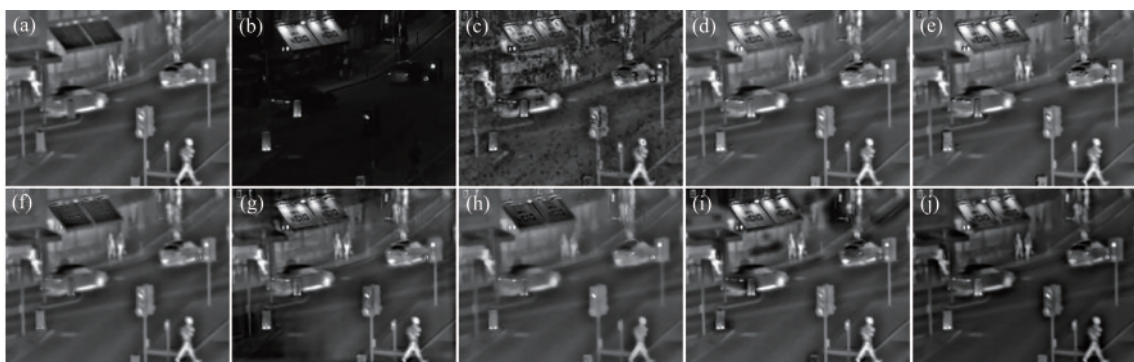


图 5 “Street”源图像融合结果。(a)红外图像;(b)可见光图像;(c)CBF;(d)JSR;(e)JSRSD;(f)DRTV;(g)CVT-SR;(h)GTF;(i)MISF;(j)所提方法

Fig. 5 Fusion results of “Street” source image. (a) Infrared images; (b) visible image; (c) CBF; (d) JSR; (e) JSRSD; (f) DRTV; (g) CVT-SR; (h) GTF; (i) MISF; (j) proposed method

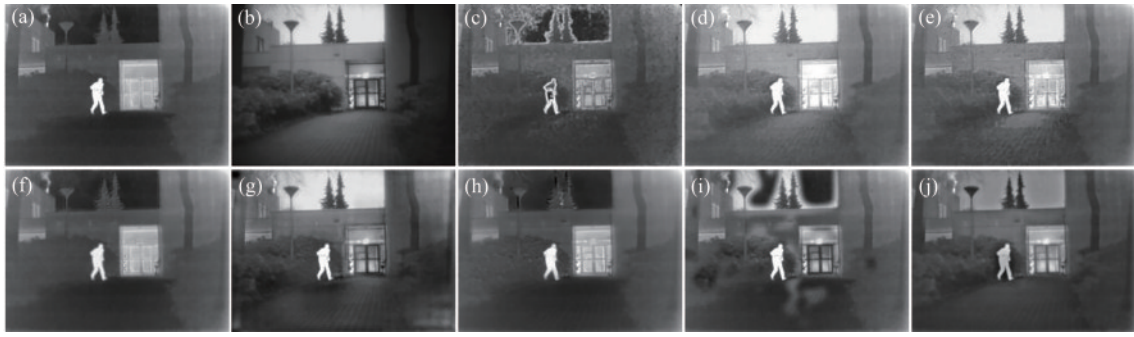


图 6 “Gate”源图像融合结果。(a)红外图像;(b)可见光图像;(c) CBF;(d) JSR;(e) JSRSD;(f) DRTV;(g) CVT-SR;(h) GTF;(i) MISF;(j)所提方法

Fig. 6 Fusion results of “Gate” source image. (a) Infrared images; (b) visible image; (c) CBF; (d) JSR; (e) JSRSD; (f) DRTV; (g) CVT-SR; (h) GTF; (i) MISF; (j) proposed method

理细节清晰,较好保留了边缘轮廓细节信息,但图像整体偏暗。

“Car”的融合结果如图 7 所示。图 7(c)中存在较多斑驳和噪声,导致视觉感观极差;图 7(d)、7(e)、7(f)

和 7(h)中纹理细节保留能力差,可见光信息丢失严重;图 7(i)中多出区域融合不完全;图 7(g)、7(j)中整体视觉效果自然,符合人眼的视觉效果。与图 7(g)相比,图 7(j)中的灰度分布更均匀。

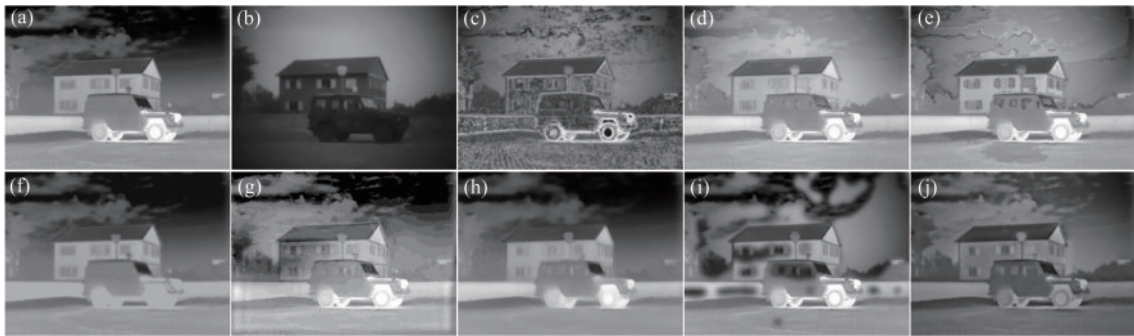


图 7 “Car”源图像融合结果。(a)红外图像;(b)可见光图像;(c) CBF;(d) JSR;(e) JSRSD;(f) DRTV;(g) CVT-SR;(h) GTF;(i) MISF;(j)所提方法

Fig. 7 Fusion results of “Car” source image. (a) Infrared images; (b) visible image; (c) CBF; (d) JSR; (e) JSRSD; (f) DRTV; (g) CVT-SR; (h) GTF; (i) MISF; (j) proposed method

“House”的融合结果如图 8 所示。图 8(c)中的天空与树木存在融合不协调,目标人物不突出的问题;图 8(d)、8(e)中整体亮度偏高,导致对比度低;图 8(f)、8(h)中纹理细节模糊,基本丢失全部的可见光信息;

图 8(g)整体得到不错的视觉效果,但图像四角位置存在背景分布灰度不均的情况;图 8(i)中天空、房屋等区域明显融合不完全;图 8(j)中整体视觉效果自然,但可见光信息表达不足,造成图像天空区域偏暗。综上主



图 8 “House”源图像融合结果。(a)红外图像;(b)可见光图像;(c) CBF;(d) JSR;(e) JSRSD;(f) DRTV;(g) CVT-SR;(h) GTF;(i) MISF;(j)所提方法

Fig. 8 Fusion results of “House” source image. (a) Infrared images; (b) visible image; (c) CBF; (d) JSR; (e) JSRSD; (f) DRTV; (g) CVT-SR; (h) GTF; (i) MISF; (j) proposed method

观分析可以得出,本文方法在红外与可见光融合图像在纹理细节保留、全局对比度和灰度分布等方面具有较好的性能。

4.3 客观性能指标评价

以上仅通过视觉效果来评价各种不同融合方法的性能,为了进一步验证本文方法的有效性,选取基于噪声评估的融合性能(NABF)、结构相似度测量(SSIM)、多尺度结构相似度测量(MS-SSIM)、相关系数(CC)、均方误差(MSE)、峰值信噪比(PSNR)等客观指标对上述算法进行评估。其中,NABF和MSE的指标数值越小,说明融合性能越好;其余4种客观指标数值越大,融合性能越好。

融合结果客观指标评价如表3所示,最优值以黑体标注。从表3中可以看出,本文方法在5组标准测试图像上取得了较好的结果,其中第二组实验中的MS-SSIM客观指标最优方法为JSR,其他客观指标的最优值均为本文方法。NABF、SSIM、CC、MSE和PSNR的提升说明了源图像的信息很好地输入到了融合图像中,体现了三频带分解与VGG网络对深度特征提取的有效性。综上所述,本文方法与7种对比方法相比泛化能力更强,得到的融合图像在较大程度上降低了噪声,保留了相对较高的结构和纹理信息,使得融合图像拥有更多、更自然的细节特征,与主观视觉所得出的结论相符。

表3 融合结果客观指标比较
Table 3 Comparison of objective indexes of fusion results

Image	Algorithm	CBF	JSR	JSRSD	DRTV	CVT-SR	GTF	MISF	Proposed
Group 1 "Camp"	NABF	0.2436	0.2319	0.3137	0.0946	0.1537	0.0754	0.0467	0.0261
	SSIM	0.6009	0.6024	0.5505	0.6802	0.6920	0.6776	0.6977	0.7452
	MS-SSIM	0.7369	0.8720	0.7530	0.6990	0.8637	0.7733	0.8334	0.8807
	CC	0.5466	0.6299	0.5618	0.4561	0.5726	0.4622	0.4978	0.6349
	MSE	0.0136	0.0478	0.0237	0.0199	0.0145	0.0192	0.0183	0.0109
	PSNR	66.7992	61.3336	64.3766	65.1374	66.5115	65.3087	65.5095	67.7416
Group 2 "Street"	NABF	0.4870	0.1804	0.1908	0.1437	0.2142	0.0803	0.0762	0.0284
	SSIM	0.4986	0.6299	0.6237	0.6245	0.6337	0.6172	0.6391	0.6741
	MS-SSIM	0.6986	0.9688	0.9471	0.9177	0.9214	0.8953	0.9124	0.9162
	CC	0.5267	0.6476	0.6172	0.4948	0.5438	0.5024	0.5280	0.6740
	MSE	0.0310	0.0462	0.0437	0.0400	0.0347	0.0381	0.0378	0.0207
	PSNR	63.2231	61.4825	61.7211	62.1136	62.7326	62.3177	62.3572	64.9705
Group 3 "Gate"	NABF	0.2554	0.2430	0.3419	0.0686	0.1729	0.0365	0.0582	0.0128
	SSIM	0.6497	0.6281	0.5665	0.6871	0.7229	0.7008	0.7042	0.7775
	MS-SSIM	0.7333	0.9037	0.8329	0.7589	0.8806	0.7979	0.8135	0.9162
	CC	0.4937	0.5862	0.5637	0.3905	0.5005	0.3910	0.4325	0.6073
	MSE	0.0331	0.0714	0.0545	0.0453	0.0401	0.0448	0.0405	0.0232
	PSNR	62.9376	59.5965	60.7645	61.5729	62.1037	61.6147	62.0529	64.478
Group 4 "Car"	NABF	0.2393	0.2692	0.3807	0.1351	0.1835	0.0771	0.0660	0.0209
	SSIM	0.6172	0.5692	0.4990	0.7117	0.6980	0.7195	0.6962	0.7643
	MS-SSIM	0.6986	0.8146	0.7746	0.7532	0.8576	0.8260	0.7998	0.8929
	CC	0.2303	0.3659	0.3336	0.2227	0.2517	0.2234	0.2460	0.3865
	MSE	0.0416	0.1115	0.0726	0.0506	0.0416	0.0508	0.0466	0.0262
	PSNR	61.9382	57.6584	59.5217	61.0862	61.9404	61.0695	61.4480	63.9461
Group 5 "House"	NABF	0.5278	0.2109	0.2801	0.1100	0.2216	0.0829	0.1595	0.0218
	SSIM	0.4575	0.6253	0.5808	0.6486	0.6537	0.6606	0.6492	0.7248
	MS-SSIM	0.5105	0.8873	0.7446	0.8055	0.8668	0.8267	0.7668	0.9105
	CC	0.2624	0.3556	0.2722	0.1519	0.2062	0.1647	0.1903	0.3794
	MSE	0.0483	0.0992	0.0845	0.0668	0.0657	0.0734	0.0659	0.0374
	PSNR	61.2878	58.1652	58.8630	59.8834	59.9580	59.4746	59.9426	62.4026

4.4 验证与消融

1) ST分解可行性验证。在图像融合领域,基于多尺度变换^[5]的方法通常在每个分解过程中将低频子

带作为下一层的源图像,进而再将其分解为一系列高频和低频子带。因此,低频子带可以看作是一幅完整的图像,低频子带的处理一直是影响融合性能的重要

因素。图 2 可以直观地显示低频子带中不仅包含能量结构,还包含一定程度的纹理细节。为了验证所提出的三频带分解模型的有效性,将二频带分解和三频带分解的融合实验进行了比较。如图 9 所示,二频带分解采用平均法则方案对低频子带进行融合,两种分解的高频子带融合规则相同。通过观察图 9 中的(c)列和(d)列,从局部放大区域可以看出,三频带分解和二频带分解的区别主要在于细节的保留能力。三频带分解模型具有更清晰、完整的纹理和边缘特征,验证了低频结构纹理分解操作的可行性。

2) 消融实验。本文提出的方法是基于 ST 分解与 VGG 网络相结合的融合方法,上述内容已经验证了低频结构纹理分解操作的可行性。因此,本文对 VGG 网络进行了分析以确保使用 VGG 网络和 Sigmoid 归一化策略的有效性。文中共设 3 种消融方案:第一种采用绝对值最大对高频子带进行融合;第二种使用 VGG 网络和最大选择策略对高频子带进行融合^[13-15];第三种为本文方法,使用 VGG 网络和 Sigmoid 归一化策略对高频子带进行融合。3 种消融方案的低频子带

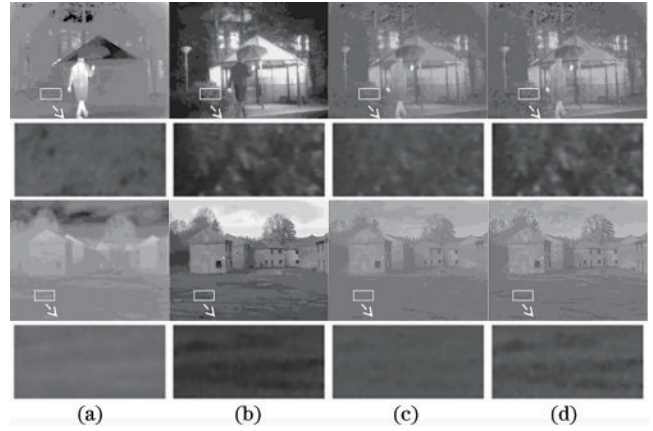


图 9 三频带分解模型验证。(a)红外图像;(b)可见光图像;(c)二频带分解;(d)三频带分解

Fig. 9 Three-band decomposition model verification. (a) Infrared images; (b) visible images; (c) dual-band decomposition; (d) triband decomposition

融合规则相同。图 10 是 21 组图像通过 3 种方案获得的融合结果图的客观指标柱状图。从图 10 中可观察到,采用 VGG 网络对高频子带进行融合能有效抑制

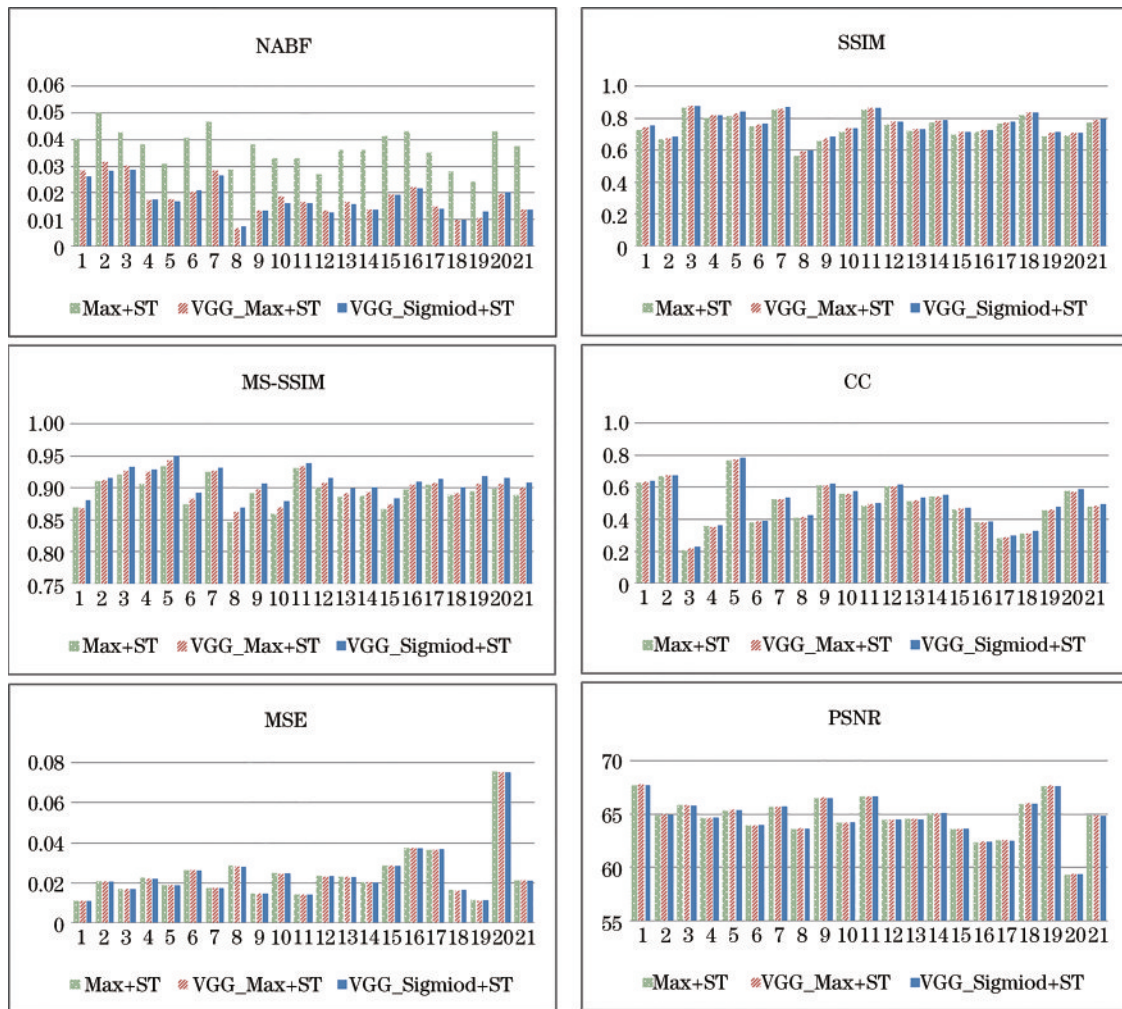


图 10 21 组融合图像客观指标柱状图

Fig. 10 Histogram of objective indexes of 21 fusion images

噪声的生成。整体来看, VGG 网络和 Sigmoid 归一化策略的使用进一步提高了融合性能, 本文方案获得的融合指标 NABF 和 MS-SSIM 性能提升较为明显, SSIM 和 CC 指标性能有较小的提升。表 4 是 21 组图像通过 3 种方案获得的融合结果图的客观指标平均

值, 最优值以黑体标注。从表 4 可以看出, 第二种方案的各项指标均值均优于第一种方案, 验证了使用 VGG 网络提取特征的必要性。与最大选择策略相比, Sigmoid 归一化策略得到了 5 个最优均值, 验证了 Sigmoid 归一化策略具有一定的有效性。

表 4 消融实验

Table 4 Ablation experiment

Scheme	NABF	SSIM	MS-SSIM	CC	MSE	PSNR
Max+ST	0.0371	0.7477	0.8958	0.4869	0.0244	64.6999
VGG_Max+ST	0.0186	0.7629	0.9028	0.4904	0.0241	64.7542
VGG_Sigmoid+ST	0.0177	0.7659	0.9096	0.4995	0.0241	64.7526

5 结 论

采用传统方法与深度学习相结合的方式, 进行了红外与可见光图像融合研究, 并取得了良好的结果。该方法利用均值滤波高效平滑的特性将输入图像分解为高-低频子带; 引入 ST 分解模型将低频子带分解为低频结构和低频纹理, 并根据特征属性采用平均法则和 NSF 策略进行有效融合; 高频子带则利用预先训练的 VGG 深度网络得到多重特征组合权重进行 Sigmoid 归一化以获得最终权重图, 由最终权重图作为权重系数进行高频子带的预融合; 最后将三频带的预融合子图进行重建。文中对 21 组红外与可见光图像进行了实验, 并对其中 5 组图像进行了说明和分析, 结果表明: 该方法获得的融合图像不仅保留了重要纹理细节, 而且有效降低了噪声; 与常用的 7 种融合方法相比, 该方法的融合图像对比度适宜, 更适合人眼观察, 因此具有一定的实用价值和理论意义。

参 考 文 献

- [1] Kumar B K S. Image fusion based on pixel significance using cross bilateral filter[J]. Signal, Image and Video Processing, 2015, 9(5): 1193-1204.
- [2] 江兆银, 王磊. 基于显著性检测与权重映射的可见光与红外图像融合算法[J]. 电子测量与仪器学报, 2021, 35(1): 174-182.
Jiang Z Y, Wang L. Visible and infrared image fusion algorithm based on significance detection and weight mapping[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(1): 174-182.
- [3] 赵程, 黄永东. 基于滚动导向滤波和混合多尺度分解的红外与可见光图像融合方法[J]. 激光与光电子学进展, 2019, 56(14): 141007.
Zhao C, Huang Y D. Infrared and visible image fusion via rolling guidance filtering and hybrid multi-scale decomposition[J]. Laser & Optoelectronics Progress, 2019, 56(14): 141007.
- [4] Li J, Peng Y X, Song M H, et al. Image fusion based on guided filter and online robust dictionary learning[J]. Infrared Physics & Technology, 2020, 105: 103171.
- [5] Li X S, Zhou F Q, Tan H S. Joint image fusion and denoising via three-layer decomposition and sparse representation[J]. Knowledge-Based Systems, 2021, 224: 107087.
- [6] Lee H, Jeon J, Kim J, et al. Structure-texture decomposition of images with interval gradient[J]. Computer Graphics Forum, 2017, 36(6): 262-274.
- [7] Yin H T. Sparse representation with learned multiscale dictionary for image fusion[J]. Neurocomputing, 2015, 148: 600-610.
- [8] 沈瑜, 陈小朋, 苑玉彬, 等. 基于显著矩阵与神经网络的红外与可见光图像融合[J]. 激光与光电子学进展, 2020, 57(20): 201007.
Shen Y, Chen X P, Yuan Y B, et al. Infrared and visible image fusion based on significant matrix and neural network[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201007.
- [9] Liu Y, Chen X, Ward R K, et al. Image fusion with convolutional sparse representation[J]. IEEE Signal Processing Letters, 2016, 23(12): 1882-1886.
- [10] Wang M, Liu X W, Jin H P. A generative image fusion approach based on supervised deep convolution network driven by weighted gradient flow[J]. Image and Vision Computing, 2019, 86: 1-16.
- [11] 侯春萍, 王霄聪, 夏晗, 等. 基于双通路生成对抗网络的红外与可见光图像融合方法[J]. 激光与光电子学进展, 2021, 58(14): 1410024.
Hou C P, Wang X C, Xia H, et al. Infrared and visible image fusion method based on dual-channel generative adversarial network[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410024.
- [12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2021-04-05]. <https://arxiv.org/abs/1409.1556>.
- [13] Li H, Wu X J, Kittler J. Infrared and visible image fusion using a deep learning framework[C]//2018 24th International Conference on Pattern Recognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE Press, 2018: 2705-2710.
- [14] 李恒, 张黎明, 蒋美容, 等. 一种基于 ResNet152 的红外与可见光图像融合算法[J]. 激光与光电子学进展, 2020, 57(8): 081013.
Li H, Zhang L M, Jiang M R, et al. An infrared and

- visible image fusion algorithm based on ResNet152[J]. *Laser & Optoelectronics Progress*, 2020, 57(8): 081013.
- [15] 沈瑜, 陈小朋. 基于DLatLRR与VGG Net的红外与可见光图像融合[J]. *北京航空航天大学学报*, 2021, 47(6): 1105-1114.
- Shen Y, Chen X P. Infrared and visible image fusion based on latent low-rank representation decomposition and VGG Net[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2021, 47(6): 1105-1114.
- [16] Zhang Q H, Fu Y L, Li H F, et al. Dictionary learning method for joint sparse representation-based image fusion [J]. *Optical Engineering*, 2013, 52(5): 057006.
- [17] Liu C H, Qi Y, Ding W R. Infrared and visible image fusion method based on saliency detection in sparse domain[J]. *Infrared Physics & Technology*, 2017, 83: 94-102.
- [18] Du Q L, Xu H, Ma Y, et al. Fusing infrared and visible images of different resolutions via total variation model [J]. *Sensors*, 2018, 18(11): 3827.
- [19] Liu Y, Liu S P, Wang Z F. A general framework for image fusion based on multi-scale transform and sparse representation[J]. *Information Fusion*, 2015, 24: 147-164.
- [20] Ma J Y, Chen C, Li C, et al. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. *Information Fusion*, 2016, 31: 100-109.
- [21] Zhan K, Kong L W, Liu B, et al. Multimodal image seamless fusion[J]. *Journal of Electronic Imaging*, 2019, 28(2): 023027.