

基于 Tikhonov 正则化和细节重建的红外与 可见光图像融合方法

卢鑫, 杨林, 李敏, 张学武*

河海大学物联网工程学院, 江苏 常州 213022

摘要 传统的红外与可见光图像融合方法将图像分解为多个频域分量后分别融合再相加, 存在边缘模糊、对比度低等问题, 为此提出了一种基于 Tikhonov 正则化和细节重建的融合方法。首先, 利用 Tikhonov 正则化将图像分解为基本层和细节层, 针对基本层训练一种用于细节重建的生成对抗网络; 然后提取待融合图像的基本层特征, 采用主成分分析方法进行融合; 最后将基本层融合结果输入到生成对抗网络中, 重建出一幅高频信息丰富的融合图像。实验结果表明: 所提方法很好地保留了源图像中的细节信息和高亮区域, 对不同清晰度的图像具有较好的鲁棒性。

关键词 图像处理; 细节重建; Tikhonov 正则化; 生成对抗网络; 主成分分析

中图分类号 TP751

文献标志码 A

doi: 10.3788/AOS202040.0210001

Infrared and Visible Image Fusion Method Based on Tikhonov Regularization and Detail Reconstruction

Lu Xin, Yang Lin, Li Min, Zhang Xuewu*

College of Internet of Things Engineering, Hohai University, Changzhou, Jiangsu 213022, China

Abstract Traditional infrared and visible image fusion method decomposes images into several frequency components, fuses them separately, and then adds them together, resulting in problems of edge fuzziness, low contrast, and so on. The paper proposes a fusion method based on Tikhonov regularization and detail reconstruction. Firstly, images are decomposed into base layers and detail layers by Tikhonov regularization. A generative adversarial network is trained aiming at detail information reconstruction for base layers. Secondly, features of base layers to be fused are extracted, and the principal component analysis method is used for feature fusion. Finally, the fused results of base layers are input into generative network to reconstruct a fusion image with abundant high frequency information. Experimental results show that the method proposed in this paper preserves detail information and highlight areas of the source images well, with a good robustness to the images with different resolutions.

Key words image processing; detail reconstruction; Tikhonov regularization; generative adversarial network; principal component analysis

OCIS codes 100.2960; 150.0155; 300.6550; 350.2660

1 引 言

图像融合就是将包含同一场景的多张传感器图像结合起来, 获得一幅鲁棒性更好、信息更加丰富的图像^[1], 实现对图像场景和目标更加准确的描述, 以便于开展进一步的处理与研究。其中, 红外与可见光图像融合是该领域的一项重要课题, 通过红外成像传感器捕捉物体的热辐射信息, 结合可见光图像

的细节和纹理, 就可以得到一幅场景更加完整, 且更符合人类视觉感受的图像, 在遥感^[2]、医学^[3]、军事^[4]等领域都有广泛应用。

传统的红外与可见光图像融合方法包括多尺度分解法^[5]、稀疏表示法^[6], 以及基于显著性^[7]的方法等, 这些方法虽然可以有效地改进图像的融合效果, 但复杂度较高, 执行效率低。随着深度学习的兴起, 通过深度网络提取图像深层次的信息, 并利用 GPU

收稿日期: 2019-06-24; 修回日期: 2019-07-24; 录用日期: 2019-09-09

基金项目: 国家重点研发计划(2016YFC0401606)、国家自然科学基金(61671202, 61573128, 61701169)

* E-mail: lab_112@126.com

(Graphics Processing Unit)进行加速,可以重构出一幅信息丰富的图像,且算法的执行效率大大提高。Amin-Naji等^[8]在全卷积神经网络(FCN)的基础上,提出了一种基于深度学习的图像融合方法,该方法用卷积层代替了网络中的全连接层,并删除了池化层,这样可以避免图像中的细节被消除,从而得到了一幅更准确且更清晰的决策图。Li等^[9]针对深度网络只提取特征而不处理的问题,提出了一种基于深度特征和零相位成分分析(ZCA)的融合框架,将残差网络(ResNet)提取到的深度特征利用ZCA进行归一化后得到初始权值图,然后采取加权平均的策略得到了融合图像。此外,Li等^[10]还将密集连接卷积网络(DenseNet)与融合策略相结合,针对图像融合的问题设计了一套编码、解码的机制,在编码过程中利用DenseNet从源图像中获取更多的有用特征,然后利用解码器对融合后的图像进行重构。Goodfellow等^[11]利用生成的对抗网络(GAN)实现红外图像与可见光图像的融合,生成的网络保留了红外目标的亮度和可见光图像的梯度,判别网络迫使融合后的图像保留更多的细节,最终得到了一幅纹理清晰、信息丰富的图像^[12]。上述融合方法结合深度网络提取图像深层次的特征,弥补了传统算法的一些不足,具有一定的创新性,但还存在以下问题:缺少训练的过程,直接利用已有的网络模型和预训练权重进行特征提取,特征的有效性不能保证;损失函数的设定与模型预期得到的融合效果不符,缺乏说服力。

本文提出了一种结合Tikhonov正则化与细节重建的图像融合方法,针对图像通过Tikhonov正则化分解后得到的基本层,训练了一个用于将基本层还原成清晰图像的GAN,然后通过引入图像梯度损失来增强网络重建高频信息的能力。本文的主要创新点如下:将Tikhonov正则化运用到图像分解中,滤除了边缘、轮廓等梯度较大的区域,并且很好地保留了图像的主要能量和灰度分布信息,为之后的细节重建做好了准备;自定义基本层特征,基于主成分分析(PCA)的融合规则对基本层进行融合。相比于固定权重的融合方法,本文中的融合规则充分结合了红外和可见光基本层图像的灰度特性,对不同场景的适应性更好。此外,不同于传统方法将各频段分别融合后再相加的思路,本文融合结果的细节层由基本层融合图像重建得到,重建网络实现了Tikhonov正则化分解的逆过程,增强了融合图像中低频与高频信息之间的相关性。实验结果表明,本

文算法融合后的图像在主观和客观评价上均具有一定优势,视觉效果也更好。

2 基本原理

2.1 Tikhonov 正则化

Tikhonov正则化依据最小二乘准则和平滑准则,将不适定问题转换为一个近似适定问题,该方法可以应用于图像去噪、梯度计算和图像复原^[13]等问题上,具有良好的稳定性^[14]。Tsibanov等^[15]通过引入一阶导数,将该正则化方法应用于图像滤波,根据正则化函数最小化的结果,得到了分解后图像的基本层。正则化函数的表达式为

$$E(u_b) = \|u_b - u\|_{L_2}^2 + \alpha \cdot \left\| \frac{d}{dx} u_b \right\|_{L_2}^2, \quad (1)$$

式中: u 表示分解前的图像; u_b 表示图像的基本层,即滤波后的结果; $u_d = u - u_b$ 表示图像的细节层,包括图像的边缘、轮廓等信息; $\|u_b - u\|_{L_2}^2$ 为残差项,用于限制分解后的基本层与原图像的灰度分布差异; $\left\| \frac{d}{dx} u_b \right\|_{L_2}^2$ 为惩罚项,用于滤除图像中灰度跳变大的区域;正则化参数 α 用于平衡分解后图像灰度的变化以及高频分量的损失, α 越大,图像的平滑效果越明显。与传统的小波变换和低秩分解的方法相比,Tikhonov正则化对不同分辨率的图像均具有良好的分解效果,不仅具有更强的抗噪性能^[16],而且在运行效率上也有所提升。本文充分利用了Tikhonov正则化方法对大梯度的惩罚^[17],选取适当的正则化参数,在滤除高频分量时保留了图像中灰度变化相对平缓的区域,减少了微弱细节的损失,为之后的融合和重建做好了准备。

2.2 生成对抗网络

生成对抗网络(GAN)由生成器 G 和判别器 D 两个模块组成。其中,生成器将输入图像或随机噪声映射成服从真实分布的图像,判别器将生成的图像与真实图样区分开来。GAN的本质是一个“二元极小极大博弈”的问题,通过在生成器和判别器之间进行持续性的对抗学习来提高两个模型的性能,从而生成更高质量的图像。生成对抗网络的目标函数定义为

$$\min_G \{ \max_D [V(D, G)] \} = E_{x \sim P_{\text{data}}(x)} [\log_2 D(x)] + E_{z \sim P_{\text{noise}}(z)} \{ \log_2 \{ 1 - D[G(z)] \} \}, \quad (2)$$

式中: E 表示函数的期望; $P_{\text{data}}(x)$ 和 $P_{\text{noise}}(z)$ 分别为真实数据的分布与噪声分布。固定判别器的参数并引入JS散度,可以将生成器的损失函数定义为

$$J^{(G)} = -\log_2(4) + 2 \times D_{JS}(P_{data} || P_{noise})。 (3)$$

当生成数据与真实数据的分布没有重叠时,JS 散度的计算结果为 0,在网络训练过程中会产生梯度消失的现象,无法达到纳什均衡。此外,GAN 存在训练不稳定、模式易崩溃等缺陷^[18],针对这类问题,一些改进模型也随之被提出来。本文参考最小二乘 GAN(LSGAN)^[19],将生成对抗网络的目标函数由交叉熵损失替换为最小二乘损失,通过拉近生成图

像与决策边界的距离,引导图像逼近真实数据,以解决 GAN 训练不稳定以及生成质量不高等问题。

3 本文算法

针对红外图像与可见光图像的融合问题,本文参考现有的深度学习方法,设计了一个结合 Tikhonov 正则化与细节重建的图像融合方法。算法流程如图 1 所示。

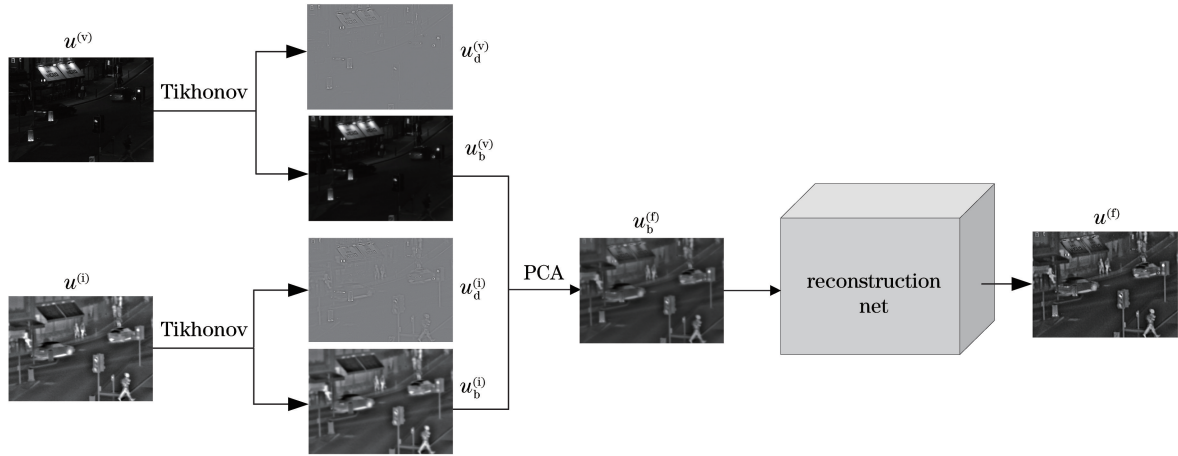


图 1 算法流程示意图

Fig. 1 Algorithm flow diagram

如图 1 所示,通过 Tikhonov 正则化方法将红外图像 $u^{(i)}$ 和可见光图像 $u^{(v)}$ 分解得到细节层 $u_d^{(v)}$ 、 $u_d^{(i)}$ 与基本层 $u_b^{(v)}$ 、 $u_b^{(i)}$,其中基本层包含了图像的大部分能量以及灰度分布信息。基于图像特征与 PCA 计算得到基本层融合权重,将两幅基本层图像加权相加后得到融合后的基本层 $u_b^{(f)}$,以 $u_b^{(f)}$ 为输入,经过细节重建网络即可生成最终的融合结果。其中,细节重建网络即为 GAN 中的生成器,由训练集图像及其分解后的基本层训练得到,网络反映了由基本层到分解前清晰图像的映射关系,实现了 Tikhonov 正则化的逆过程。

3.1 基本层融合规则

基本层反映了图像在大尺度上的变化,传统的 PCA 融合方法是基于基本层的强度的,而由于红外基本层的强度远大于可见光图像的基本层,因此计算出的融合权重会偏向于红外图像,这样就损失了可见光基本层的灰度信息,造成细节上的丢失。本文权衡了红外与可见光基本层的灰度分布特性,基于图像子块的能量和空间频率计算 PCA 融合权重,能量高的区域显著性较强,空间频率高的区域信息更加丰富。将基本层分割成 $m \times n$ 个子块,设子块的大小为 $w \times h$,则每一块的能量 E 与空间频

率 F_s 为

$$E = \frac{1}{w \times h} \sum_{i=1}^w \sum_{j=1}^h f^2(i, j), \quad (4)$$

$$F_s = \sqrt{F_R^2 + F_C^2}, \quad (5)$$

其中,

$$F_R = \frac{1}{w \times h} \sqrt{\sum_{i=1}^w \sum_{j=2}^h [f(i, j) - f(i, j-1)]^2}, \quad (6)$$

$$F_C = \frac{1}{w \times h} \sqrt{\sum_{i=2}^w \sum_{j=1}^h [f(i, j) - f(i-1, j)]^2}, \quad (7)$$

式中: $f(i, j)$ 为子块中位于 (i, j) 处的像素值; F_R 表示图像在行方向上的空间频率; F_C 表示列方向上的空间频率。将各图像子块的特征表示为 $(\sqrt{E}, k \times F_s)$,则每幅图像可以用一个长度为 $2 \times m \times n$ 的向量 l 表示, $l = [\sqrt{E_1}, \sqrt{E_2}, \dots, \sqrt{E_{m \times n}}, k \times F_{s,1}, k \times F_{s,2}, \dots, k \times F_{s,m \times n}]^T$,其中 k 为空间频率相对于能量的权重,经过实验得出 k 设为 1.5 时融合效果最好。

记红外图像基本层的特征为 l_i ,可见光图像基本层的特征为 l_v ,利用 PCA 降维的方法构建特征之间的关系。PCA 通过分析变量之间的相关性找

出变量中的主要成分,实现对原始信息的降维,将红外与可见光特征通过降维合并成单个向量,得到的合并权重即可作为基本层的融合权重,具体步骤如下。

1) 将 I_i 和 I_v 作为矩阵 A 的列向量,计算 A 的协方差矩阵 C :

$$C = \begin{bmatrix} \text{Cov}(I_i, I_i) & \text{Cov}(I_i, I_v) \\ \text{Cov}(I_v, I_i) & \text{Cov}(I_v, I_v) \end{bmatrix}; \quad (8)$$

2) 计算 C 的特征值 λ_1, λ_2 与特征向量 ϕ_1, ϕ_2 , 找出其中最大的特征值对应的特征向量 ϕ , 记 $\phi = [\phi_1, \phi_2]^T$, ϕ_1, ϕ_2 分别为向量中的数值;

3) 归一化 ϕ 的值,得到新的图像融合的权重 ω_1 和 ω_2 :

$$\omega_1 = \frac{\phi_1}{\phi_1 + \phi_2}, \omega_2 = \frac{\phi_2}{\phi_1 + \phi_2}; \quad (9)$$

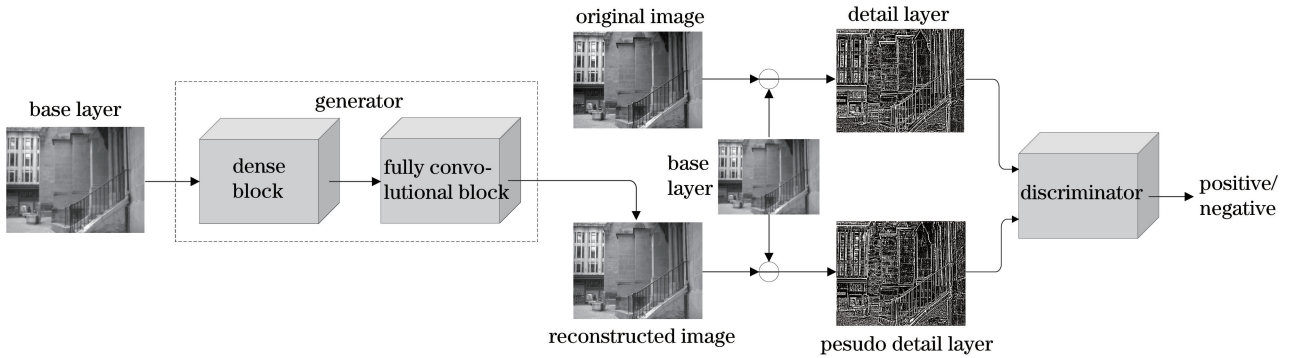


图 2 图像重建模型的框架

Fig. 2 Framework of image reconstruction model

3.2.1 生成网络的结构

生成网络主要由密集卷积模块(Dense Block)^[20]和全卷积模块(Fully Convolutional Block)组成,输入

4) 根据计算得到的权重对红外基本层与可见光基本层进行融合,得到基本层融合结果 $u_b^{(f)}$:

$$u_b^{(f)} = \omega_1 u_b^{(i)} + \omega_2 u_b^{(v)}. \quad (10)$$

3.2 细节重建

针对 Tikhonov 正则化方法分解后的基本层图像,本文设计了一种用于细节重建的生成对抗网络,网络的输入输出图像均映射至 $[-1, 1]$ 之内,模型的框架如图 2 所示。其中,生成网络用于细节重建,将图像分解得到的基本层输入网络,网络提取基本层的深层特征并重建边缘等高频信息,输出细节重建后的图像。将生成图像减去基本层图像得到伪细节层图像,然后将其与真实的细节层一起输入判别网络,在判别网络中实现这两幅细节图像的分类,在与生成网络的对抗过程中增强生成网络重建细节层的能力。

Tikhonov 正则化方法分解后的基本层,得到的输出结果为重建后的图像,它恢复了分解损失的边缘和细节信息。生成网络的结构如图 3 所示。

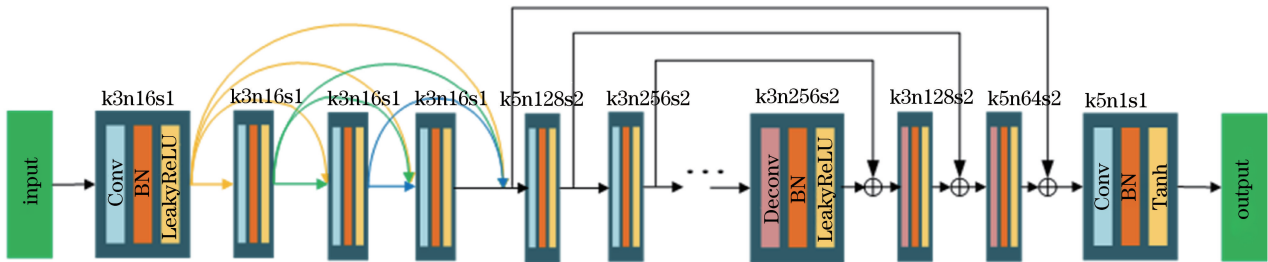


图 3 生成网络的结构

Fig. 3 Generative network structure

图 3 中, $k3n16s1$ 表示单个卷积层的属性,即该卷积层的感受野边长为 3,深度为 16,卷积过程中的步长为 1。密集卷积模块的每一层都将前面所有层的输出作为输入,建立层与层之间的密集连接,得到每一层的输出为

$$X_n = H_n([X_0, X_1, \dots, X_{n-1}]), \quad (11)$$

式中: X_n 表示第 n 层的输出; H_n 表示第 n 层网络的非线性变换,包括卷积、批归一化(BN, Batch Normalization)和带泄露整流函数(LeakyReLU)运算; $[]$ 表示拼接操作,通过对输出进行通道合并来实现特征复用,可在减轻梯度消失的同时减少网络参数,提升了运算效率,并且可以提取高质量的语义

信息。本文并非直接利用全卷积模块对输入图像进行采样操作,而是通过密集卷积模块对输入进行了初步的编码,得到一张高维特征图,为之后的采样和进一步提取深层次的信息做准备。

全卷积模块通过卷积与池化形成高维特征,再经过反卷积还原得到高分辨率的低维特征图,将上采样和下采样过程中对应的特征进行合并,从而提取图像的高级语义特征和细节信息。特征合并采用跳跃连接(Skip Connections)^[21]的方式,将上、下采样中对应尺寸和维度的特征图通过相加的方式实现特征的复用,便于恢复图像的细节信息,减轻梯度消失的问题。此外,针对计算量增加与存储资源占用的问题,在高维的层之间使用 1×1 的卷积核进行降维,产生一个瓶颈层(Bottleneck),不仅有效地减少

了参数的数目,还在保持模型效果的同时加速了网络的收敛。全卷积模块的参数信息如表 1 所示,其中:Kernel size 为卷积核的大小;Stride 表示卷积步幅,本文利用步长为 2 的卷积代替池化层(pooling)实现下采样;Padding 为卷积时图像边界填充的像素数;Output padding 用来控制上采样输出的尺寸,使上采样与下采样对应维度的特征图大小保持一致;BN 用于加速学习和收敛,减小层与层之间的相关度,增强模型的鲁棒性。模型中最后一层使用双曲正切(Tanh)激活函数,用于生成重建后的结果,其余层使用 LeakyReLU 函数,DCGAN (Deep Convolutional Generative Adversarial Networks)表明 LeakyReLU 对生成高分辨率的图片具有更好的效果。

表 1 全卷积模块的参数信息

Table 1 Parameter information of fully convolutional block

Name	Kernel size	Stride	Padding	Output padding	Output size	BN
Input	-	-	-	-	$320 \times 320 \times 64$	-
Conv1	$128 \times 5 \times 5$	2	2	-	$160 \times 160 \times 128$	✓
Conv2	$256 \times 3 \times 3$	2	1	-	$80 \times 80 \times 256$	✓
Conv3	$512 \times 3 \times 3$	2	1	-	$40 \times 40 \times 512$	✓
Conv4	$512 \times 3 \times 3$	2	1	-	$20 \times 20 \times 512$	✓
Conv5	$1024 \times 1 \times 1$	2	-	-	$10 \times 10 \times 1024$	✓
DeConv1	$512 \times 1 \times 1$	2	-	1	$20 \times 20 \times 512$	✓
Add(Conv4+DeConv1)	-	-	-	-	$20 \times 20 \times 512$	-
DeConv2	$512 \times 1 \times 1$	2	1	1	$40 \times 40 \times 512$	✓
Add(Conv3+DeConv2)	-	-	-	-	$40 \times 40 \times 512$	-
DeConv3	$256 \times 3 \times 3$	2	1	1	$80 \times 80 \times 256$	✓
Add(Conv2+DeConv3)	-	-	-	-	$80 \times 80 \times 256$	-
DeConv4	$128 \times 3 \times 3$	2	1	1	$160 \times 160 \times 128$	✓
Add(Conv1+DeConv4)	-	-	-	-	$160 \times 160 \times 128$	-
DeConv5	$64 \times 5 \times 5$	2	2	1	$320 \times 320 \times 64$	✓
Add(Input+DeConv5)	-	-	-	-	$320 \times 320 \times 64$	-
Output	$1 \times 5 \times 5$	1	2	-	$320 \times 320 \times 1$	-

3.2.2 判别网络的结构

判别网络为 6 层卷积神经网络,其前 5 层为卷积层,使用带步幅的卷积(strided convolution)代替池化层实现下采样;最后 1 层为全连接层,使用 Sigmoid 激活函数进行二分类。当输入为 Tikhonov 正则化分解得到的细节层时,网络的输出为 1;当输入为生成图像与基本层差分得到的伪细节图像时,输出为 0。判别网络的结构如图 4 所示,其中 H 和 W 分别为图像的高和宽。

3.2.3 网络训练

本文通过自定义的损失函数引导网络参数的学习,生成网络的损失 L_G 共有两个部分,包括生成器和判别器的对抗损失 $V(G)$ 和内容损失 L_{content} 。其中,对抗损失用于引导网络生成符合真实数据分布的结果,内容损失用于衡量重建后的图像与原图像的距离。损失函数的定义为

$$L_G = V(G) + \lambda L_{\text{content}}, \quad (12)$$

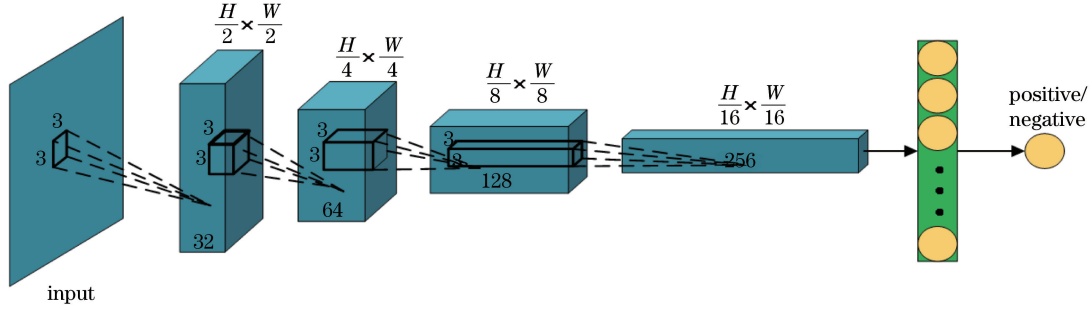


图 4 判别网络的结构

Fig. 4 Discriminant network structure

$$V(G) = \frac{1}{N} \sum_{n=1}^N [D_{\theta_D}(u_n^{(\text{gen},d)}) - c]^2, \quad (13)$$

$$L_{\text{content}} = \frac{1}{N} \sum_{n=1}^N \frac{1}{W \times H} (\|u_n^{(\text{gen})} - u_n^{(\text{lab})}\|^2 + |\nabla u_n^{(\text{gen})} - \nabla u_n^{(\text{lab})}|), \quad (14)$$

$$\nabla u = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \times u + \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \times u, \quad (15)$$

式中: c 为正样本的标签值 1; $u_n^{(\text{gen})}$ 和 $u_n^{(\text{lab})}$ 分别表示第 n 幅图像基本层重建后的结果以及其对应的原图像; $u_n^{(\text{gen},d)}$ 表示重建图像减去基本层后得到的伪细节图像; N 为每个 batch 的样本数量。本文利用 Sobel 算子计算图像的近似梯度,并在损失函数中引入图像梯度损失,使重建后的图像尽可能地还原

边缘和轮廓信息。 λ 设置为 1×10^3 ,用于衡量内容损失相对于对抗损失所占的权重。

本文使用 LSGAN 的最小二乘损失^[19]衡量分类结果和标签的距离,引导与真实图像差异较大的生成图像向决策边界靠近,优化训练的结果,并且使训练过程更加稳定。判别网络的损失函数定义为

$$L_D = \frac{1}{N} \sum_{n=1}^N [D_{\theta_D}(u_n^{(\text{lab},d)}) - a]^2 + \frac{1}{N} \sum_{n=1}^N [D_{\theta_D}(u_n^{(\text{gen},d)}) - b]^2, \quad (16)$$

式中: a 和 b 分别为正、负样本的标签值 1 和 0; $D_{\theta_D}(u_n^{(\text{lab},d)})$ 和 $D_{\theta_D}(u_n^{(\text{gen},d)})$ 分别为判别网络对第 n 组真实细节层和伪细节层的分类结果。

本文先训练判别网络,再固定判别网络的权重训练生成网络,根据自定义的函数计算损失,并利用 Adam 算法优化梯度。网络训练流程如图 5 所示。

Algorithm 1: GAN for detail reconstruction. We use default values of $k=1$, $n=4$;

Input: Dataset U

Output: Trained generator G and discriminator D

for number of training iterations **do**

Decompose images $\{u_1, u_2, \dots, u_n\}$ and obtain $\{u_1^{(b)}, u_2^{(b)}, \dots, u_n^{(b)}\}$, $\{u_1^{(d)}, u_2^{(d)}, \dots, u_n^{(d)}\}$

for k steps **do**

Reconstruct detail information of base layer by G and obtain $\{u_1^{(\text{gen})}, u_2^{(\text{gen})}, \dots, u_n^{(\text{gen})}\}$

Calculate negative detail layer $\{u_1^{(\text{gen},d)}, u_2^{(\text{gen},d)}, \dots, u_n^{(\text{gen},d)}\}$

Calculate discriminator loss L_D with samples $\{u_1^{(d)}, u_2^{(d)}, \dots, u_n^{(d)}\}$ and $\{u_1^{(\text{gen},d)}, u_2^{(\text{gen},d)}, \dots, u_n^{(\text{gen},d)}\}$

Update parameters of D by $\theta_D = \text{Adam}(\nabla_{\theta_D} L_D, \theta_D)$

end for

Reconstruct detail information of base layer by G and obtain $\{u_1^{(\text{gen})}, u_2^{(\text{gen})}, \dots, u_n^{(\text{gen})}\}$

Calculate generator loss L_G with samples $\{u_1, u_2, \dots, u_n\}$ and $\{u_1^{(\text{gen})}, u_2^{(\text{gen})}, \dots, u_n^{(\text{gen})}\}$

Update parameters of G by $\theta_G = \text{Adam}(\nabla_{\theta_G} L_G, \theta_G)$

Start a new iteration if all samples of U have been extracted

end for

图 5 网络训练流程

Fig. 5 Network training process

4 分析与讨论

本文中网络的训练采用的是 Laquerre 等提供的 RGB-NIR 数据集^[22],包括 477 张可见光图像和 477 张近红外图像,裁剪后得到 13720 张 320×320 的训练样本,测试集为 TNO 图像融合数据库,用于验证网络细节重建的能力并得到最终的融合结果。本文的实验环境为 Intel(R)Core(TM)i7-8700 CPU @ 3.20 GHz 主频/16.0 GB 内存,使用 GPU 为 NVIDIA GeForce GTX 1060,6 GB 显存。Adam 优化器的学习率设置为 2×10^{-4} ,动量设置为 0.9,网络迭代终止次数为 20,计算基本层特征时子块的大小设置为 16×16 。

本文选取了图像金字塔、引导滤波等几种图像融合时常用的分解方法,并将其与 Tikhonov 正则化进行比较,用于说明 Tikhonov 正则化的分解效果在本文算法中的优势。另外选取了 4 种结合深度网络的图像融合方法——密集连接卷积网络(DenseNet)^[10]、残差网络(ResNet)^[9]、Visual Geometry Group (VGG)^[16]和生成对抗网络(GAN)^[12],以及 3 种现

有的图像融合方法——四叉树分解(QD)^[23]、视觉显著图(VSM)^[7]和隐式低秩表示(LatLRR)^[24],进行对比实验。

4.1 主观评价

图 6 和图 7 比较了不同分解方法对场景 Smoke 和 Heather 的分解效果,分析后可知:采用图像金字塔等多尺度分解方法分解得到的低频分量损失的信息较多;而双边滤波^[25]和引导滤波^[1]的特点在于保持大梯度的边缘,对梯度平缓区域的滤波效果类似于普通的线性滤波算子,这几种分解方法有各自的应用前景,但在本文中不利于之后细节层的重建;Tikhonov 正则化的优势在于其对大梯度的惩罚,而对重建难度较大的微弱细节保留的效果较好,当选取的参数 α 不同时,分解的效果也不同。从图 6 和图 7 中可以看出:当 $\alpha=2$ 时,图像平滑的程度较弱,图像中的高频分量得不到充分的过滤;当 $\alpha=8$ 时,图像中灰度跳变不大的区域变得模糊,图像细节信息损失严重,增大了细节层重建的难度。经过实验可以得出,当 α 取值为 4 时,图像分解和细节层重建的效果最好。

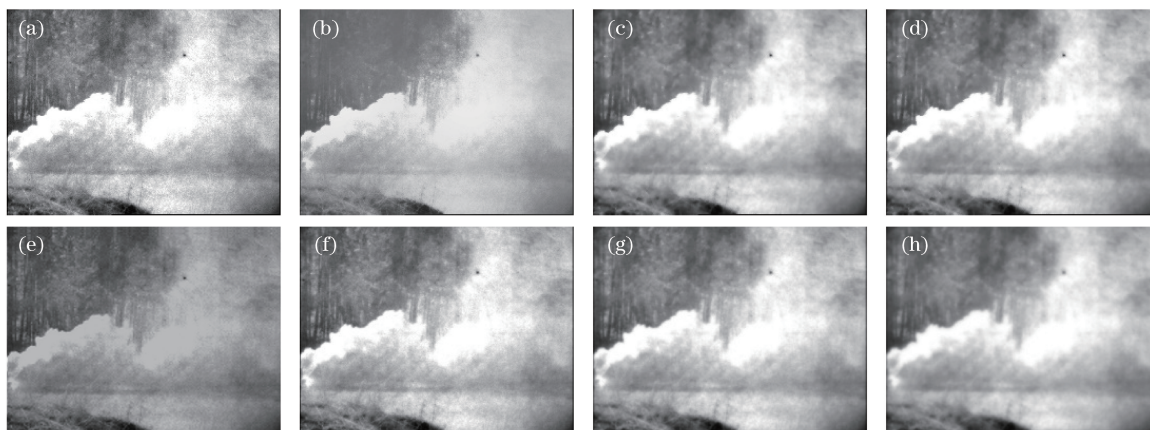


图 6 不同分解方法对 Smoke 场景的分解效果。(a)原图像;(b)双边滤波;(c)引导滤波;(d)高斯金字塔;(e)小波变换;(f) Tikhonov 正则化, $\alpha=2$;(g) Tikhonov 正则化, $\alpha=4$;(h) Tikhonov 正则化, $\alpha=8$

Fig. 6 Comparison of decomposition effects of different decomposition algorithms for "Smoke" scene. (a) Original image; (b) bilateral filtering; (c) guided filtering; (d) Gaussian pyramid; (e) wavelet transform; (f) Tikhonov regularization, $\alpha=2$; (g) Tikhonov regularization, $\alpha=4$; (h) Tikhonov regularization, $\alpha=8$

图 8~10 分别显示了不同算法对场景 Quad、Smoke 和 Nato_camp 的融合效果。由场景“Quad”的融合效果可以看出:本文算法和 ResNet 方法、VSM 方法、QD 方法均能很好地保留红外图像中的高亮区域,且 QD 方法融合后的图像中广告牌上的字符最清晰。在场景 Smoke 下,与其他算法的融合结果相比,本文算法保留了更多的细节信息,减轻了雾对丛林和战士的遮挡,但在图像重建过程中引入了

噪声。在 Nato_camp 场景中,LatLRR 和 GAN 融合后的图像的对比较低,QD 法融合得到的图像遗漏了很多可见光图像中的信息,细节保留效果最差。

使用 7×7 的高斯滤波算子对待融合图像进行模糊处理,采用不同的算法对平滑后的图像进行融合,得到了模糊图像的融合效果。由图 11 可以看出:在场景 Kaptein_1123 下,采用 VGG 方法和 LatLRR 方法融合得到的图像边缘信息损失严重,

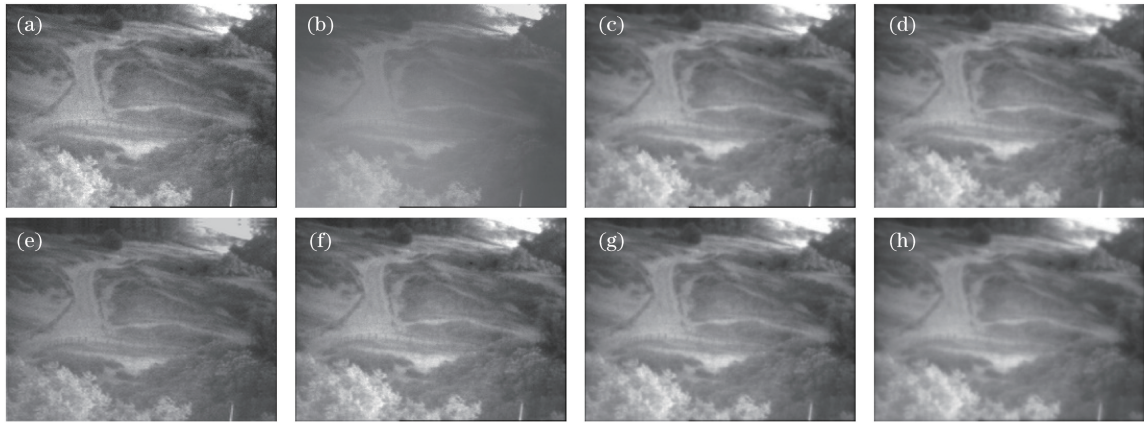


图 7 不同分解方法对 Heather 场景的分解效果。(a)原图像;(b)双边滤波;(c)引导滤波;(d)高斯金字塔;(e)小波变换;
(f) Tikhonov 正则化, $\alpha=2$;(g) Tikhonov 正则化, $\alpha=4$;(h) Tikhonov 正则化, $\alpha=8$

Fig. 7 Comparison of decomposition effects of different decomposition algorithms for “Heather” scene. (a) Original image;
(b) bilateral filtering; (c) guided filtering; (d) Gaussian pyramid; (e) wavelet transform; (f) Tikhonov regularization, $\alpha=2$; (g) Tikhonov regularization, $\alpha=4$; (h) Tikhonov regularization, $\alpha=8$

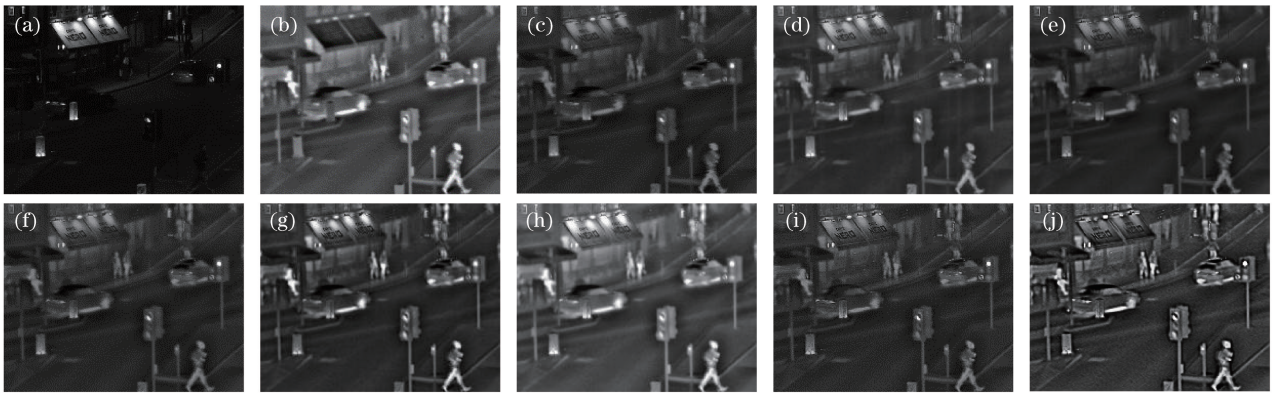


图 8 不同算法对 Quad 场景的融合效果。(a)可见光图像;(b)红外图像;(c) DenseNet; (d) LatLRR; (e) VGG;
(f) ResNet; (g) VSM; (h) QD; (i) GAN; (j)本文算法

Fig. 8 Fusion effects of different algorithms in “Quad” scene. (a) Visible image; (b) infrared image; (c) DenseNet;
(d) LatLRR; (e) VGG; (f) ResNet; (g) VSM; (h) QD; (i) GAN; (j) proposed algorithm

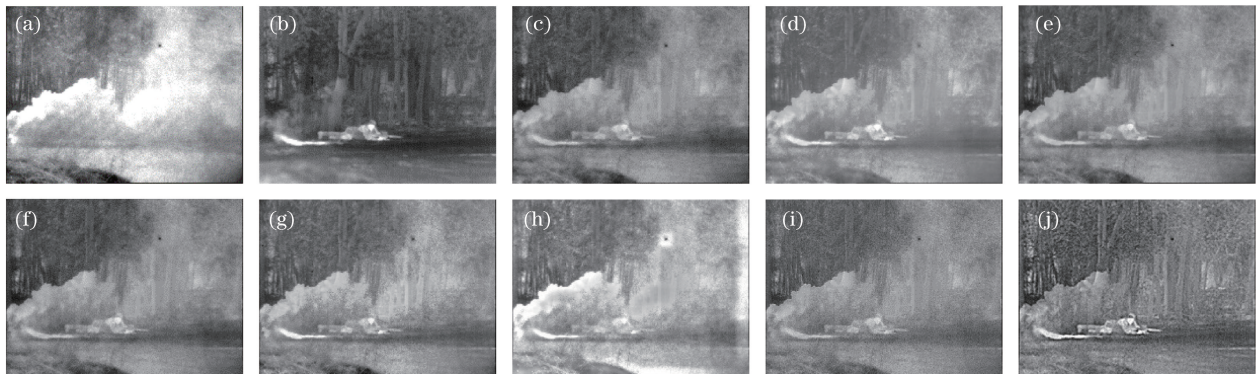


图 9 不同算法对 Smoke 场景的融合效果。(a)可见光图像;(b)红外图像;(c) DenseNet; (d) LatLRR; (e) VGG;
(f) ResNet; (g) VSM; (h) QD; (i) GAN; (j)本文算法

Fig. 9 Fusion effects of different algorithms in “Smoke” scene. (a) Visible image; (b) infrared image; (c) DenseNet;
(d) LatLRR; (e) VGG; (f) ResNet; (g) VSM; (h) QD; (i) GAN; (j) proposed algorithm

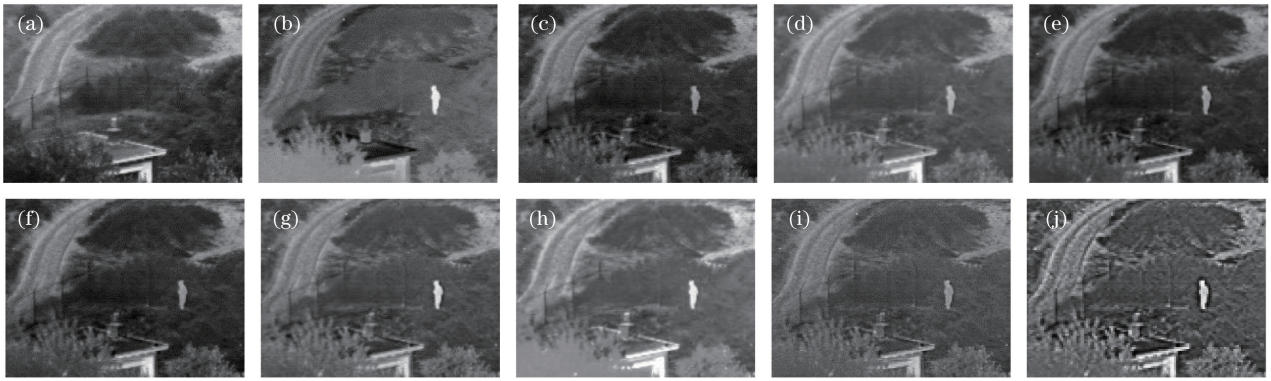


图 10 不同算法对 Nato_camp 场景的融合效果。(a)可见光图像;(b)红外图像;(c) DenseNet;(d) LatLRR;(e) VGG;(f) ResNet;(g) VSM;(h) QD;(i) GAN;(j)本文算法

Fig. 10 Fusion effects of different algorithms in “Nato_camp” scene. (a) Visible image; (b) infrared image; (c) DenseNet; (d) LatLRR; (e) VGG; (f) ResNet; (g) VSM; (h) QD; (i) GAN; (j) proposed algorithm

尤其是在路面纹理的损失上;而采用 VSM 方法和 QD 方法融合得到的图像虽然有效地突出了红外图像中的高亮区域,但 VSM 方法丢失了很多边缘信息,QD 方法存在大量的光晕;相比而言,本文算法

对路面的纹理保留得最好,且有效地突出了高亮区域。如图 12 所示,在 Heather 场景中,本文算法能够很好地恢复树叶的纹理信息,且产生的光晕较少,融合后的图像较为清晰,更符合人的视觉感受。

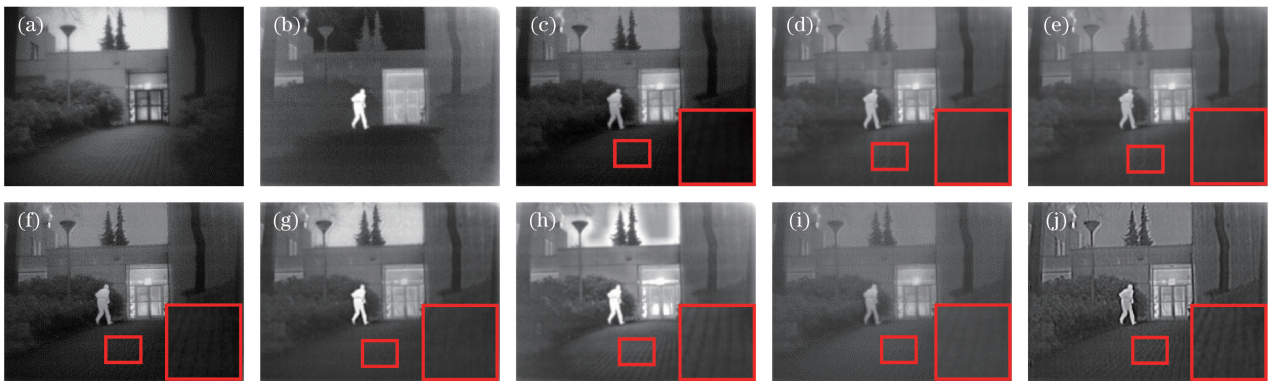


图 11 不同算法对模糊场景 Kaptein_1123 的融合效果。(a)可见光图像;(b)红外图像;(c) DenseNet;(d) LatLRR;(e) VGG;(f) ResNet;(g) VSM;(h) QD;(i) GAN;(j)本文算法

Fig. 11 Fusion effects of different algorithms in blurred “Kaptein_1123” scene. (a) Visible image; (b) infrared image; (c) DenseNet; (d) LatLRR; (e) VGG; (f) ResNet; (g) VSM; (h) QD; (i) GAN; (j) proposed algorithm

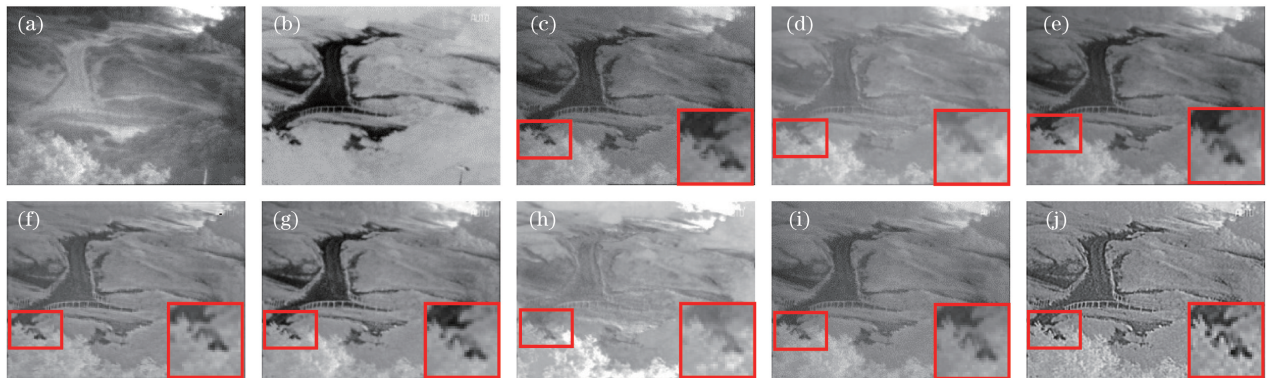


图 12 不同算法对模糊场景 Heather 的融合效果。(a)可见光图像;(b)红外图像;(c) DenseNet;(d) LatLRR;(e) VGG;(f) ResNet;(g) VSM;(h) QD;(i) GAN;(j)本文算法

Fig. 12 Fusion effects of different algorithms in blurred “Heather” scene. (a) Visible image; (b) infrared image; (c) DenseNet; (d) LatLRR; (e) VGG; (f) ResNet; (g) VSM; (h) QD; (i) GAN; (j) proposed algorithm

图 13 展示了本文算法对其他场景的融合效果, 经过分析可以得出, 本文算法充分重建了图像的细节层信息, 并且很好地保留了图像的显著区域, 对一

些微弱的边缘也起到了增强的效果, 融合的结果对不同清晰度的图像均具有较高的鲁棒性, 但在图像重建过程中也引入了轻微的噪声, 需要进一步改进。

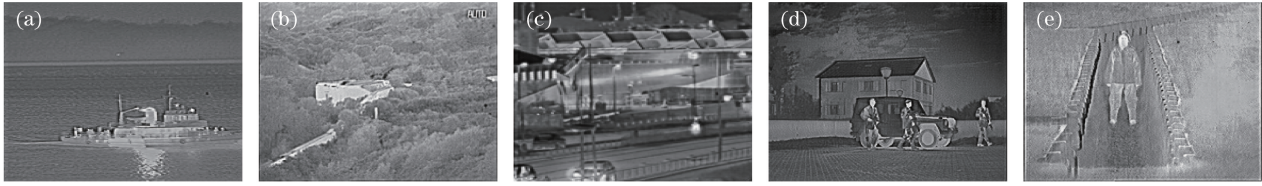


图 13 本文算法对其他场景的融合效果。(a) Steamer; (b) Bunker; (c) Street; (d) Jeep; (e) Soldier

Fig. 13 Fusion results of proposed algorithm in other scenes. (a) Steamer; (b) Bunker; (c) Street; (d) Jeep; (e) Soldier

4.2 客观评价

选取了多个评价指标对最终的融合结果进行客观评价, 主要包括图像熵(EN)、标准差(SD)、结构相似性(SSIM)、相关系数(CC)和空间频率(SF)。其中: 图像熵用于衡量图像包含的信息量; 标准差反映了图像像素值与均值的离散程度, 较高的标准差表明融合图像的质量更好; 结构相似性用于衡量融

合图像相较于红外图像、可见光图像的失真和信息损失, 主要包括协方差损失、亮度失真和对比度失真三个方面; 相关系数表明融合结果与源图像的线性相关度; 空间频率与 3.1 节中的定义一致。选取场景 Bunker、Heather、Sandpath、Jeep 的融合图像进行数据统计, 得到不同算法的评价结果, 如表 2 所示。

表 2 不同融合方法的客观评价结果

Table 2 Objective evaluation results of different fusion methods

Image	Metric	DenseNet	LatLRR	VGG	ResNet	VSM	QD	GAN	Proposed method
Bunker	EN	6.9807	6.8143	6.7277	6.8048	7.1073	7.0720	6.7043	7.1513
	SD	31.403	28.379	26.137	28.186	36.006	39.471	25.947	37.497
	SSIM	1.2790	1.1806	1.1504	1.1850	1.2137	1.0272	1.1479	1.1767
	CC	0.6270	0.6316	0.6345	0.6397	0.6223	0.5394	0.6326	0.6274
	SF	0.0180	0.0198	0.0202	0.0203	0.0213	0.0209	0.0212	0.0213
Heather	EN	6.9413	6.6036	6.8643	6.7372	7.1234	6.7995	6.7411	7.0281
	SD	32.674	26.337	30.489	28.537	38.267	31.185	30.191	37.529
	SSIM	1.0185	0.9573	1.0136	1.0044	1.0434	0.8266	0.9095	0.9770
	CC	0.5575	0.5574	0.5649	0.5680	0.5380	0.4624	0.5106	0.5545
	SF	0.0174	0.0186	0.0197	0.0194	0.0210	0.0191	0.0204	0.0212
Sandpath	EN	6.7642	6.2525	6.5866	6.5419	6.5419	6.7348	6.1159	6.7899
	SD	29.812	22.357	26.159	27.952	27.952	32.385	18.072	28.921
	SSIM	0.9022	0.8195	0.8885	0.8811	0.8811	0.8018	0.7723	0.8921
	CC	0.4780	0.4733	0.4858	0.4681	0.4681	0.4286	0.4816	0.4725
	SF	0.0268	0.0262	0.0269	0.0273	0.0273	0.0270	0.0269	0.0273
Jeep	EN	7.1496	6.5470	7.1331	6.9949	7.0240	7.2358	6.7980	7.2032
	SD	35.988	23.430	35.080	33.683	35.427	39.738	28.669	38.340
	SSIM	0.6512	0.5018	0.6460	0.6288	0.6322	0.5875	0.4970	0.6019
	CC	0.3641	0.3617	0.3650	0.3671	0.3553	0.3174	0.2818	0.3428
	SF	0.0132	0.0129	0.0151	0.0151	0.0161	0.0158	0.0152	0.0170

对表 2 中的数据进行分析后可知: 采用 DenseNet 方法融合得到的图像与源图像的相似度最高, 这源于其在训练网络时引入了生成图像与源

图像的结构相似度作为损失函数的一部分, 从融合的结果也可以看出该方法很好地保留了源图像中的信息; QD 方法的各项指标虽然很高, 但从主观的角

度考虑,其丰富的图像信息一部分来源于融合后产生的光晕和伪影,但不能说明其真正具有很好的融合效果;本文的算法和 VSM 方法融合得到的图像,在图像熵、标准差和空间频率指标上要优于其他方法,这也证明了本文算法在重建细节层和保留显著区域方面的优势。但由于本文中的图像是由融合后的基本层重建而来,损失了源图像本身的一部分高频信息,且在重建过程中引入了噪声,因此与源图

像的相关性和相似度要略低于其他算法。此外,相比于传统的融合方法,本文算法以及 DenseNet、GAN 等算法均利用 GPU 加速运算,缩短了图像融合所需的时间,这也是深度学习相关方法所具备的一大优势。程序运行时间对比如图 14 所示,可以看出本文算法的运行速度较快。此外,由于 DenseNet 方法在图像融合时提供的是一个端到端的网络,因此该方法的效率最高。

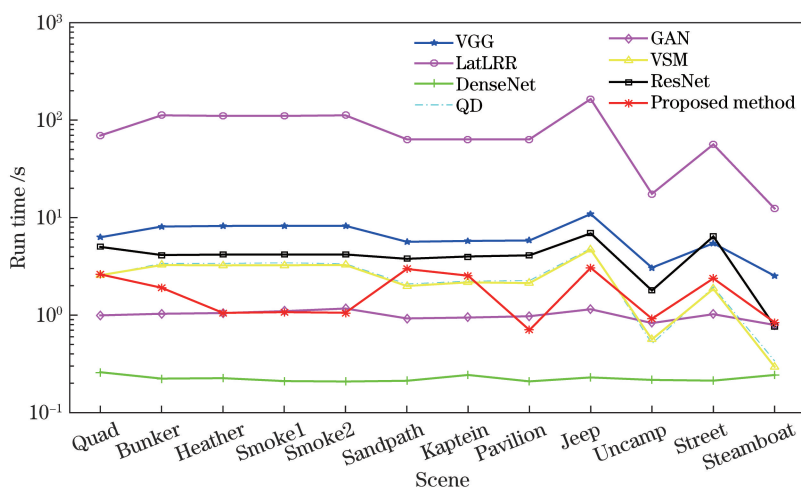


图 14 不同算法的程序运行时间对比

Fig. 14 Comparison of program running time of different algorithms

5 结 论

本文提出了一种基于 Tikhonov 正则化和细节重建的图像融合方法,首先采用 Tikhonov 正则化将红外图像与可见光图像进行分解,分解出基本层,计算基本层的特征后采用 PCA 融合得到低频融合图像;再针对基本层训练得到一个用于图像细节信息重建的网络,将融合后的低频图像输入网络后得到最终的融合结果。实验结果表明,本文提出的算法保留了源图像中的显著区域,而且重建后的图像中包含了丰富的高频信息,对不同清晰度的图像具有很好的鲁棒性,在视觉质量和运行效率上都具有一定的优势,但在细节重建过程中引入了一些噪声,如何解决这一问题并进一步提高融合质量是下一步研究的方向。

参 考 文 献

[1] Jiang Z T, Wu H, Zhou X L. Infrared and visible image fusion algorithm based on improved guided filtering and dual-channel spiking cortical model[J]. Acta Optica Sinica, 2018, 38(2): 0210002.
江泽涛, 吴辉, 周晓玲. 基于改进引导滤波和双通道

脉冲发放皮层模型的红外与可见光图像融合算法[J]. 光学学报, 2018, 38(2): 0210002.

- [2] Yu X C, Gao G Y, Xu J D, et al. Remote sensing image fusion based on sparse representation [C] // 2014 IEEE Geoscience and Remote Sensing Symposium, July 13-18, 2014, Quebec City, QC, Canada. New York: IEEE, 2014: 2858-2861.
- [3] Zhao W D, Lu H C. Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation [J]. IEEE Transactions on Instrumentation and Measurement, 2017, 66(9): 2283-2294.
- [4] Li Y S, Tao C, Tan Y H, et al. Unsupervised multilayer feature learning for satellite image scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(2): 157-161.
- [5] Rang C Z, Jia Y X, Yang Y, et al. Fusion of infrared and visible images based on multi-scale edge-preserving decomposition and sparse representation [C] // 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), October 13-15, 2018, Beijing, China. New York: IEEE, 2018: 18432125.
- [6] Yang B, Li S T. Multifocus image fusion and restoration with sparse representation [J]. IEEE

- Transactions on Instrumentation and Measurement, 2010, 59(4): 884-892.
- [7] Ma J L, Zhou Z Q, Wang B, et al. Infrared and visible image fusion based on visual saliency map and weighted least square optimization[J]. Infrared Physics & Technology, 2017, 82: 8-17.
- [8] Amin-Naji M, Aghagolzadeh A, Ezoji M. Fully convolutional networks for multi-focus image fusion [C]//2018 9th International Symposium on Telecommunications (IST), December 17-19, 2018, Tehran, Iran. New York: IEEE, 2018: 553-558.
- [9] Li H, Wu X J, Durrani T S. Infrared and visible image fusion with ResNet and zero-phase component analysis[J/OL]. (2018-12-18)[2019-06-15]. <https://arxiv.org/abs/1806.07119>.
- [10] Li H, Wu X J. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623.
- [11] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J/OL]. (2014-06-10)[2019-06-15]. <https://arxiv.org/abs/1406.2661>.
- [12] Ma J Y, Yu W, Liang P W, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. Information Fusion, 2019, 48: 11-26.
- [13] Bouhamidi A, Jbilou K. Sylvester Tikhonov-regularization methods in image restoration[J]. Journal of Computational and Applied Mathematics, 2007, 206(1): 86-98.
- [14] Zheng C L, He D D, Fei Q G. Improved digital image correlation method based on gray gradient denoised by regularization method[J]. Acta Optica Sinica, 2018, 38(8): 0812002.
郑成林, 何顶顶, 费庆国. 基于灰度梯度正则化去噪的改进数字图像相关法[J]. 光学学报, 2018, 38(8): 0812002.
- [15] Tsibanov V, Denisov A, Krylov A. Edge detection method by Tikhonov regularization[C]//Proceedings of 14 International Conference Graphicon' 2004, September 6-10, 2004, Moscow, Russia. [S.l.: s.n.], 2004: 163-165.
- [16] 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, 2018: 2705-2710.
- [17] Zhong G, Xiang S, Zhou P, et al. Spatially adaptive tensor total variation-Tikhonov model for depth image super resolution[J]. IEEE Access, 2017, 5: 13857-13867.
- [18] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs [J/OL]. (2016-06-10)[2019-06-15]. <https://arxiv.org/abs/1606.03498>.
- [19] Mao X D, Li Q, Xie H R, et al. Least squares generative adversarial networks [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2813-2821.
- [20] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2261-2269.
- [21] Mao X J, Shen C H, Yang Y B. Image restoration using convolutional auto-encoders with symmetric skip connections [J/OL]. (2016-08-30)[2019-06-15]. <https://arxiv.org/abs/1606.08921>.
- [22] Brown M, Süssstrunk S. Multi-spectral SIFT for scene category recognition [C] // CVPR 2011, June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE, 2011: 177-184.
- [23] Zhang Y, Zhang L J, Bai X Z, et al. Infrared and visual image fusion through infrared feature extraction and visual information preservation [J]. Infrared Physics & Technology, 2017, 83: 227-237.
- [24] Li H, Wu X J. Infrared and visible image fusion using Latent Low-Rank Representation [J/OL]. (2018-04-24)[2019-06-15]. <https://arxiv.org/abs/1804.08992>.
- [25] Shreyamsha Kumar B K. Image fusion based on pixel significance using cross bilateral filter [J]. Signal, Image and Video Processing, 2015, 9(5): 1193-1204.