

基于语义分割的红外和可见光图像融合

周华兵^{1,2} 侯积磊^{1,2} 吴伟^{1,2} 张彦铎^{1,2} 吴云韬^{1,2} 马佳义³

¹(武汉工程大学计算机科学与工程学院 武汉 430205)

²(智能机器人湖北省重点实验室(武汉工程大学) 武汉 430205)

³(武汉大学电子信息学院 武汉 430072)

(zhouhuabing@gmail.com)

Infrared and Visible Image Fusion Based on Semantic Segmentation

Zhou Huabing^{1,2}, Hou Jilei^{1,2}, Wu Wei^{1,2}, Zhang Yanduo^{1,2}, Wu Yuntao^{1,2}, and Ma Jiayi³

¹(College of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205)

²(Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology), Wuhan 430205)

³(Electronic Information School, Wuhan University, Wuhan 430072)

Abstract Infrared images can distinguish targets from their backgrounds due to the difference in thermal radiation even in poor lighting conditions. By contrast, visible images can represent texture details with high spatial resolution. Meanwhile, both of infrared and visible images preserve corresponding semantic information. Therefore, infrared and visible image fusion should keep both radiation information of the infrared image and texture details of the visible image; additionally, it needs to reserve the semantic information of both. Semantic segmentation can transform the source images into the masks with semantic information. In this paper, an infrared and visible image fusion method is proposed based on semantic segmentation. It can overcome the shortcomings that the existing fusion methods are not specific to different regions. Considering the specific information for different regions of infrared and visible images, we design two loss functions for different regions to improve the quality of fused image under the framework of generative adversarial network. Firstly, we gain the masks of the infrared images with semantic information by semantic segmentation; then we use the masks to divide the infrared and visible images into infrared target area, infrared background area, visible target area, and visible background area. Secondly, we employ different methods to fuse the target and background area, respectively. Finally, we combine the two regions to obtain the final fused image. The experiment shows that the proposed method outperforms state-of-the-art, where our results have higher contrast in the target area and richer texture details in the background area.

Key words infrared image; visible image; image fusion; semantic segmentation; mask

摘要 红外图像即使在低光照条件下,也能根据热辐射的差异将目标与背景区分开来,而可见光图像具有高空间分辨率的纹理细节,此外,红外和可见光图像都含有相应的语义信息。因此,红外与可见光图像融合,需要既保留红外图像的辐射信息,也保留可见光图像的纹理细节,同时,也要反映出二者的语义信息。而语义分割可以将图像转换为带有语义的掩膜,提取源图像的语义信息。提出了一种基于语义

收稿日期:2020-03-31;修回日期:2020-08-20

基金项目:国家自然科学基金项目(61771353,61773295,62072350,41501505);湖北省技术创新工程项目(2019AAA045)

This work was supported by the National Natural Science Foundation of China (61771353, 61773295, 62072350, 41501505) and the Hubei Technology Innovation Project (2019AAA045).

通信作者:马佳义(jyma2010@gmail.com)

分割的红外和可见光图像融合方法,能够克服现有融合方法不能针对性地提取不同区域特有信息的缺点.使用生成式对抗神经网络,并针对源图像的不同区域设计了2种不同的损失函数,以提高融合图像的质量.首先通过语义分割得到含有红外图像目标区域语义信息的掩模,并利用掩模将红外和可见光图像分割为红外图像目标区域、红外图像背景区域、可见光图像目标区域和可见光图像背景区域;然后对目标区域和背景区域分别采用不同的损失函数得到目标区域和背景区域的融合图像;最后将2幅融合图像结合起来得到最终融合图像.实验表明,融合结果目标区域对比度更高,背景区域纹理细节更丰富,提出的方法取得了较好的融合效果.

关键词 红外图像;可见光图像;图像融合;语义分割;掩膜

中图法分类号 TP391

图像融合是一种增强技术,其目的是将不同类型传感器获取的图像结合起来,生成一幅信息丰富的图像,以便于后续处理^[1].通过不同类型传感器获取的图像一般并不是对齐的,会有一些的偏移,需要配准后才能融合^[2-4].本文主要解决精确配准条件下的图像融合问题,其在语义上具有逐像素(per pixel)的对应关系.

红外和可见光图像融合是图像融合的重要分支,红外图像可以根据辐射信息来突出目标,并且不受天气和光线的影响,可见光图像拥有较高的分辨率,图像更清晰,符合人类视觉习惯^[5-8].红外和可见光图像融合既可以保留红外图像的对比度信息,又可以保留可见光图像的高分辨率.

传统的图像融合方法多是基于多尺度分解的思路.包括拉普拉斯金字塔变换(Laplacian pyramid transform, LP)^[9]、双树复小波变换(dual-tree complex wavelet transform, DTCWT)^[10]、非下采样轮廓波变换(nonsubsampled contourlet transform, NSCT)^[11]等方法.随着视觉显著性相关研究的深入,出现越来越多基于视觉显著性的图像融合方法.Zhang等人^[12]提出一种基于显著性区域提取的红外和可见光图像融合方法.首先基于显著性分析和自适应阈值算法提取出红外图像的目标区域,然后采用非下采样剪切波变换得到背景区域融合系数,完成融合任务.但这种方法得到的融合图像中目标区域直接使用了红外图像中的目标区域,丢失了可见光图像中相应区域的信息.

随着深度学习的火爆,近年来,基于深度学习的红外和可见光图像融合方法越来越多.如:Ma等人^[13]的FusionGAN使用生成式对抗神经网络来完成融合任务,通过生成器和鉴别器的对抗,来使融合图像保留更丰富的信息.Zhang等人^[14]的PMGI从图像梯度和对比度2条路径来提取图像信息,并且在同一条路径上使用特征重用,以避免由于卷积而

丢失图像信息.同时,在2条路径之间引入了路径传递块,实现了不同路径间的信息交换,保证了融合图像有更丰富的信息.但这些融合方法都是将红外和可见光图像整体放进同一个网络框架下,对源图像中的目标和背景采用同一种处理方式,没有针对性,不可避免地损失了源图像的部分信息.

红外和可见光图像具有丰富的语义信息,语义信息可以用来提取图像中的目标区域,也可以遮挡图像中的干扰区域.语义分割可以将图像转换为具有语义信息的掩膜,是计算机视觉研究的热门方向^[15].一般的分类卷积神经网络(convolutional neural network, CNN),会在网络的最后加入一些全连接层,经过softmax后就可以获得类别概率信息.但是这个概率信息是1维的,即只能标识整个图像的类别,不能标识每个像素点的类别.Long等人^[16]提出的全卷积网络(fully convolutional network, FCN)将CNN的全连接层替换成卷积层,这样就可以获得一幅2维特征图,从而成功解决了语义分割问题.Chen等人^[17]提出的Deeplabv3+采用了编码解码网络结构,逐步重构图像空间信息来更好的捕捉物体边界.并在网络结构中引入了空间金字塔池化模块(spatial pyramid pooling, SPP),利用多种比例感受野的不同分辨率特征来挖掘图像多尺度的上下文内容信息.

为了在融合任务中更具有针对性地保留源图像目标和背景区域的信息,本文基于语义分割和生成式对抗神经网络提出了一种新的红外和可见光图像融合方法,在融合任务中引入语义分割,实现了图像融合时对目标区域和背景区域采用不同的融合策略,解决了现有融合方法对源图像不同区域针对性不足的问题.本文的融合结果在目标区域保留了红外图像的对比度,在背景区域保留了可见光图像的纹理细节,图像信息更丰富,视觉效果更好.

1 融合方法

为了实现目标背景和背景区域融合时的不同偏好,本文提出一种基于语义分割的红外和可见光图像融合方法,流程图如图1所示,步骤如下:

- 1) 通过语义分割,得到带有红外图像目标区域语义信息的掩膜 I_m .
- 2) 使用掩膜 I_m 带有的语义信息来处理红外图像 I_r 和可见光图像 I_v ,得到红外图像的目标区域 I_{r1} 和背景区域 I_{r2} ,以及可见光图像的目标区域 I_{v1} 和背景区域 I_{v2} .
- 3) 将红外图像的目标区域 I_{r1} 和可见光图像的目标区域 I_{v1} 融合得到融合图像 I_{f1} .
- 4) 将红外图像的背景区域 I_{r2} 和可见光图像的背景区域 I_{v2} 融合得到融合图像 I_{f2} .
- 5) 将 I_{f1} 和 I_{f2} 融合得到最终融合图像 I_f .

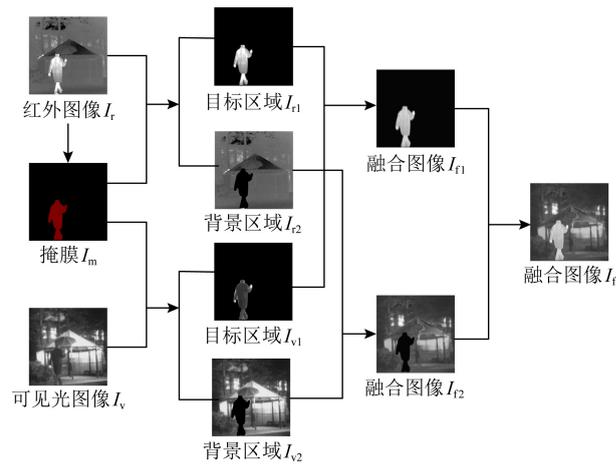


Fig. 1 Schematic of infrared and visual image fusion based on semantic segmentation

图1 基于语义分割的红外和可见光融合

2 语义分割

红外和可见光图像融合是为了融合图像目标区域保留更多的红外图像信息,背景区域保留更丰富的可见光图像信息^[18].为了达到这个目的,本文将语义分割引入图像融合,用带有语义信息的掩膜提取红外和可见光图像的目标和背景区域.

我们将红外图像和与之对应的标注图作为输入,训练语义分割网络,这里采用 Deeplabv3+网络结构.通过此网络,可以得到带有红外图像语义信息的掩膜 I_m ,然后使用掩膜 I_m 提取出红外图像目标

区域 I_{r1} 和可见光图像目标区域 I_{v1} ,如式(1)所示:

$$I_{r1} = I_m \odot I_r, I_{v1} = I_m \odot I_v. \quad (1)$$

接着,再使用掩膜 I_m 将源图像中的目标区域遮挡起来,得到红外图像背景区域 I_{r2} 和可见光图像背景区域 I_{v2} ,如式(2)所示:

$$I_{r2} = (1 - I_m) \odot I_r, I_{v2} = (1 - I_m) \odot I_v, \quad (2)$$

其中, \odot 为 Hadamard 乘积.

3 图像融合

3.1 分割图融合

由于融合图像中目标区域和背景区域需要保留的信息差别较大.本文提出的融合方法对目标区域 I_{r1} 和 I_{v1} 以及背景区域 I_{r2} 和 I_{v2} 采用不同的融合策略,以便在融合图像不同区域能更具有针对性的保留所需要的信息.

目标区域 I_{r1} 和 I_{v1} 是为了保留更多红外图像的对比度,要让融合图像 I_{f1} 更接近红外图像目标区域 I_{r1} ,网络框架如图2所示.生成器 G_1 的目标是生成融合图像 I_{f1} 去骗过鉴别器 D_1 ,鉴别器 D_1 的目标就是将生成融合图像 I_{f1} 和可见光图像目标区域 I_{v1} 区分开来,通过这种对抗过程,网络最终能得到信息丰富的融合图像 I_{f1} .

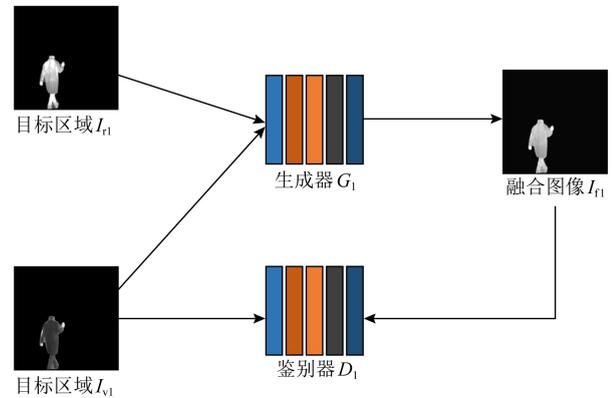


Fig. 2 Schematic of target area fusion

图2 目标区域融合网络结构图

首先将源图像目标区域 I_{r1} 和 I_{v1} 在通道维度上连接起来,一起输入到生成器 G_1 得到融合图像 I_{f1} ,通过损失函数让融合图像 I_{f1} 保留更多红外图像目标区域 I_{r1} 的信息,生成器 G_1 损失函数如式(3)所示:

$$L_{G_1} = L_{adv1} + \lambda_1 L_1, \quad (3)$$

其中, L_{G_1} 代表生成器 G_1 整体的损失, λ_1 是常数,用于平衡 2 项损失函数 L_1 和 L_{adv1} . L_{adv1} 代表生成器 G_1 和鉴别器 D_1 之间的对抗损失,如式(4)所示:

$$L_{adv1} = \frac{1}{N} \sum_{n=1}^N (D_1(I_{f1}^n) - c)^2, \quad (4)$$

其中, N 代表融合图像的数量, c 代表生成器 G_1 希望鉴别器 D_1 相信的融合图像的值。

生成器 G_1 损失函数中的第 2 项 L_1 代表目标区域内容损失, 如式(5)所示:

$$L_1 = \frac{1}{HW} (\|I_{f1} - I_{r1}\|_F^2 + \xi_1 \|\nabla I_{f1} - \nabla I_{v1}\|_F^2), \quad (5)$$

其中, H 和 W 代表输入图像的高和宽, $\|\cdot\|_F$ 表示矩阵的 Frobenius 范数, ∇ 表示梯度算子, 括号中的第 1 项用于在融合图像 I_{f1} 中保留红外图像目标区域 I_{r1} 的对比度信息, 第 2 项用于保留可见光图像目标区域 I_{v1} 的纹理细节, ξ_1 是一个常数, 用于平衡 2 项。

生成器 G_1 在没有鉴别器 D_1 时生成的融合图像会损失大量可见光图像目标区域 I_{v1} 的信息。通过生成器 G_1 和鉴别器 D_1 对抗过程, 融合图像 I_{f1} 中能加入更多可见光图像目标区域 I_{v1} 的信息, 鉴别器 D_1 的损失函数如式(6)所示:

$$L_{D1} = \frac{1}{N} \sum_{n=1}^N (D_1(I_{v1}) - a)^2 + \frac{1}{N} \sum_{n=1}^N (D_1(I_{f1}) - b)^2, \quad (6)$$

其中, a 和 b 分别代表 I_{v1} 和 I_{f1} 的标签, $D_1(I_{v1})$ 和 $D_1(I_{f1})$ 分别代表 I_{v1} 和 I_{f1} 的鉴别结果。

对于背景区域 I_{r2} 和 I_{v2} , 整体网络框架和目标区域融合网络框架一样采用了生成式对抗神经网络, 如图 3 所示。为了要让融合图像 I_{f2} 更接近可见光图像的背景区域 I_{v2} , 我们重新设计了生成器 G_2 和鉴别器 D_2 的损失函数, 生成器 G_2 的损失函数如式(7)所示:

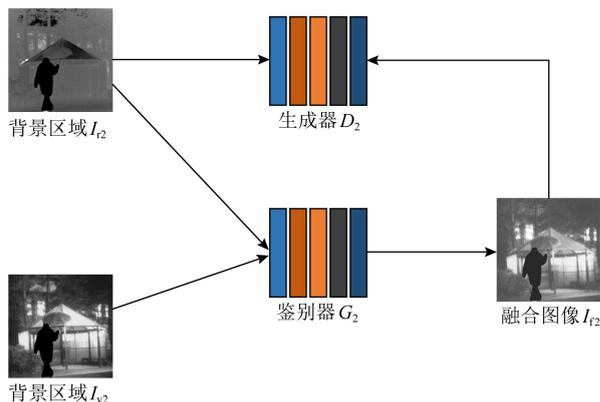


Fig. 3 Schematic of background area fusion

图 3 背景区域融合网络结构图

$$L_{G2} = L_{adv2} + \lambda_2 L_2, \quad (7)$$

其中, L_{G2} 代表生成器 G_2 整体的损失, λ_2 是常数, 用于平衡 2 项损失函数 L_2 和 L_{adv2} . L_{adv2} 代表生成器 G_2 和鉴别器 D_2 之间的对抗损失, 如式(8)所示:

$$L_{adv2} = \frac{1}{N} \sum_{n=1}^N (D_2(I_{f2}^n) - d)^2, \quad (8)$$

其中 d 代表生成器 G_2 希望鉴别器 D_2 相信的融合图像的值。

生成器 G_2 损失函数中的第 2 项 L_2 代表背景区域内容损失, 如式(9)所示:

$$L_2 = \frac{1}{HW} (\|I_{f2} - I_{r2}\|_F^2 + \xi_2 \|\nabla I_{f2} - \nabla I_{v2}\|_F^2), \quad (9)$$

其中 ξ_2 是常数, 用于平衡括号内的左右 2 项。

生成器 G_2 在没有鉴别器 D_2 时生成的融合图像会损失大量红外图像背景区域 I_{r2} 的信息。通过生成器 G_2 和鉴别器 D_2 对抗过程, 融合图像 I_{f2} 中能加入更多红外图像背景区域 I_{r2} 的信息, 鉴别器 D_2 的损失函数如式(10)所示:

$$L_{D2} = \frac{1}{N} \sum_{n=1}^N (D_2(I_{r2}) - e)^2 + \frac{1}{N} \sum_{n=1}^N (D_2(I_{f2}) - f)^2, \quad (10)$$

其中, e 和 f 分别代表 I_{r2} 和 I_{f2} 的标签, $D_2(I_{r2})$ 和 $D_2(I_{f2})$ 分别代表 I_{r2} 和 I_{f2} 的鉴别结果。

3.2 目标和背景的融合

语义分割后的图像有很多像素为 0 的区域, 这些区域虽然不带信息, 对生成器读取图像信息没有影响, 但在生成融合图像时, 生成器并不能完美识别这些区域, 会在这些区域根据学习到的风格生成像素值。为了避免这些像素影响最终融合图像, 我们先通过掩膜 I_m 带有的语义信息将融合图像 I_{f1} 中的目标部分提取出来, 同时将融合图像 I_{f2} 中的背景部分提取出来, 如式(11)(12)所示:

$$I_1 = I_m \odot I_{f1}, \quad (11)$$

$$I_2 = (1 - I_m) \odot I_{f2}. \quad (12)$$

经过处理后的图像 I_1 在目标区域以外区域像素值都为 0, I_2 在背景区域以外区域像素值都为 0, 我们直接通过简单的像素相加得到最终融合图像 I_f , 如式(13)所示:

$$I_f = I_1 + I_2. \quad (13)$$

3.3 训练过程

本文从公开的对齐数据集 TNO^① 中选取了 45 对不同场景的红外和可见光图像作为训练数据, 通过

① https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029

掩膜的语义信息将 45 对训练数据都分为红外图像目标区域、红外图像背景区域、可见光图像目标区域、可见光图像背景区域 4 部分。由于 45 对红外和可见光图像不足以训练一个好的模型,所以本文将 stride 设置为 14 来裁剪每一幅图像,裁剪后每个图像块的尺寸都是 120×120 。这样,我们可以得到 23 805 对红外和可见光图像块。

对于目标区域 I_{r1} 和 I_{v1} ,我们从训练数据中选择 32 对目标区域的红外和可见光图像块,将它们尺寸填充到 132×132 作为生成器 G_1 的输入。生成器 G_1 输出的融合图像块尺寸为 120×120 。然后,将 32 对目标区域的可见光图像块和融合图像块作为鉴别器 D_1 的输入。我们首先训练鉴别器 k 次,优化器的求解器是 Adam,然后训练生成器,直到达到最大训练迭代次数。在测试过程中,我们不重叠地裁剪测试数据,并将它们批量输入到生成器 G_1 中。然后根据裁剪的先后顺序将生成器 G_1 的结果进行拼接,得到最终的融合图像。

对于背景区域 I_{r2} 和 I_{v2} ,训练过程中对训练数据裁剪的尺寸与目标区域相同,生成器 G_2 的输入为 32 对背景区域的红外和可见光图像块,鉴别器 D_2 的输入为 32 对背景区域的红外图像块和融合图像块。在测试过程中,同样根据裁剪先后顺序将生成器 G_2 的结果进行拼接,得到最终的融合图像。

4 实验与结果分析

本文提出的方法在融合任务中引入了语义分割,为此,我们首先需要构建基于语义分割的红外和可见光图像融合数据集。其次,为了评估本文融合方法的性能,本文选取了 FusionGAN, PMGI 作为对比实验,通过主观和客观 2 方面对融合图像进行比较,所有对比实验代码都是公开的源代码,参数均是默认的参数。

4.1 数据集

本文实验所使用的红外和可见光图像来源于公开的对齐数据集 TNO。我们首先将图像尺寸统一为 450×450 ,然后挑选出红外图像的目标区域,使用 labelme 工具对目标区域进行标注,得到红外图像目标区域的标注图^①,再使用 Deeplabv3+ 网络得到带有红外图像语义信息的掩膜,通过掩膜的语义信息将红外和可见光图像分为红外图像目标区域、红

外图像背景区域、可见光图像目标区域、可见光图像背景区域 4 部分。论文发表之后,我们会公开基于语义分割的红外和可见光图像融合数据集。对比实验 FusionGAN 和 PMGI 所使用数据与本文实验数据相同,且都将图像尺寸统一为 450×450 。

4.2 主观评估

主观性能评估是基于人眼视觉系统来评价融合图像质量,因为可见光图像符合人类视觉习惯,所以红外和可见光融合图像应该在一定程度上符合人类视觉习惯。

为了验证本文融合方法的主观性能,选取了几幅图像作为评估数据,如图 4 所示,前 2 行分别是红外图像和可见光图像,第 3 行是掩膜,第 4 行是 FusionGAN 的融合结果,第 5 行是 PMGI 的融合结果,第 6 行是本文的融合结果。

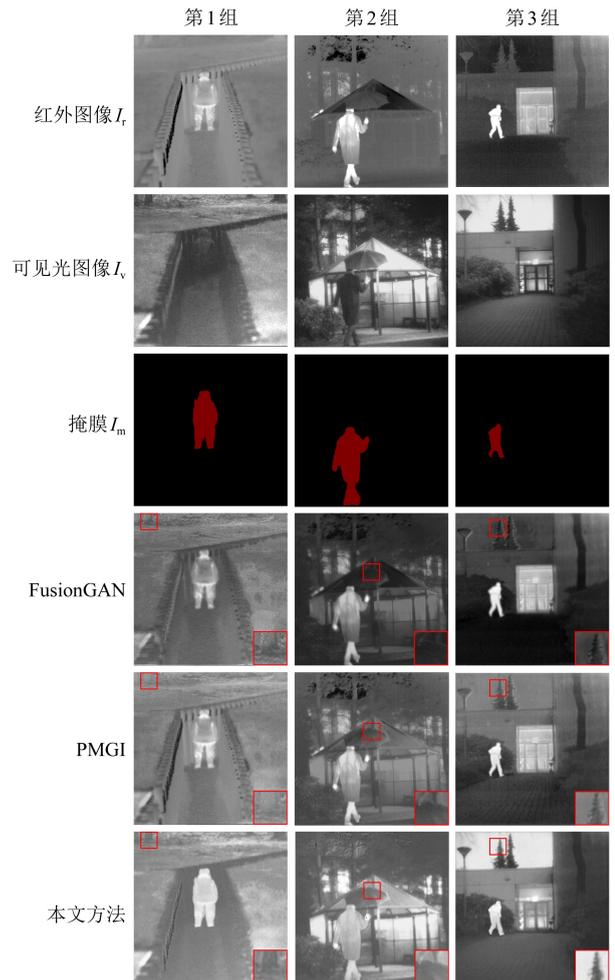


Fig. 4 Fusion results of different methods

图 4 不同方法融合结果

① <https://github.com/Jilei-Hou/FusionDataset>

本文的融合方法通过引入语义分割,实现了对目标和背景区域采用不同的融合策略,与 FusionGAN 和 PMGI 相比,3 组融合图像中,本文方法的融合结果目标区域保留的对比度信息更丰富,能更好地突出目标,有利于目标检测.背景区域本文的融合结果纹理细节保留的更好,第 1 组融合结果中,方框内树干的边界清晰,颜色保留的也更接近可见光图像,其他 2 组结果也是这样,特别是在第 3 组融合结果中表现最明显,相比与 FusionGAN 和 PMGI,我们的融合结果在树梢的细节和天空的颜色上视觉效果都更好.这说明在融合任务中引入语义信息的方法是可行的,融合图像视觉效果明显优于现有的方法.

4.3 客观评估

主观性能评估虽然能根据人类视觉系统来评估融合图像的质量,但是会受人类主观情绪的影响,为了更全面地评估融合图像的质量,本文还采用了客观性能评估.客观性能评估是依赖于数学模型的评估指标,不受人类视觉系统和主观情绪的干扰,是评价融合图像质量的重要手段,但单一的客观评估指标不能充分反映融合图像的质量,因此,本文采用了 3 种典型的客观评估指标,分别是熵(EN)、标准差(SD)和互信息(MI).

1) 熵

熵(EN)是统计图像特征的一种常用方法,融合图像的熵反映了图像从红外和可见光图像中获取的信息的多少^[19],数学定义如式(14)所示:

$$EN = - \sum_{l=0}^{L-1} p_l \lg p_l, \quad (14)$$

其中, L 表示图像灰度级, p_l 是融合图像中灰度值为 l 的标准化直方图.熵的值越大,说明融合图像中保留的源图像的信息越丰富,融合方法的性能越好.

2) 标准差

标准差(SD)反映了图像灰度值相对于灰度平均值的离散情况^[20],定义如式(15)所示:

$$SD = \sqrt{\sum_{i=1}^M \sum_{j=1}^N (F(i,j) - \mu)^2}, \quad (15)$$

其中, $F(i,j)$ 表示融合图像 F 在 (i,j) 处的像素值,融合图像 F 的尺寸为 $M \times N$, μ 表示融合图像的像素平均值.由于人类视觉系统对对比度信息很敏感,人类的注意力会被高对比度区域所吸引.因此,融合图像的标准差越大表明融合图像对比度越高,意味着融合图像的视觉效果更好.

3) 互信息

互信息(MI)是信息论中的基本概念,可以度量 2 个随机变量之间的相关性.在图像融合中互信息用来度量源图像和融合图像的相关性^[21].红外和可见光图像融合互信息定义如式(16)所示:

$$MI = MI_{r,f} + MI_{v,f}, \quad (16)$$

其中 $MI_{r,f}$ 和 $MI_{v,f}$ 分别表示红外图像和可见图像与融合图像的相关性.任意一幅源图像和融合图像之间的互信息可用定义如式(17)所示:

$$MI_{X,F} = \sum_{x,f} p_{X,F}(x,f) \log \frac{p_{X,F}(x,f)}{p_X(x)p_F(f)}, \quad (17)$$

其中, $p_X(x)$ 和 $p_F(f)$ 分别表示源图像 X 和融合图像 F 的边缘直方图. $p_{X,F}(x,f)$ 表示源图像 X 和融合图像 F 的联合直方图,互信息越大意味着融合图像与红外和可见光图像相关性越大,融合性能越好.

使用熵、标准差和互信息对 5 组图像进行客观性能评估,实验结果如表 1 所示:

Table 1 Objective Evaluation of Fusion Results

表 1 融合结果客观评估

组别	指标	FusionGAN	PMGI	本文方法
第 1 组	EN	6.6078	6.5791	6.9013
	SD	25.5816	25.6540	34.9157
	MI	1.5101	1.7313	2.7615
第 2 组	EN	6.4507	6.7182	6.9256
	SD	32.3312	32.1926	46.8908
	MI	1.6992	1.9394	2.7613
第 3 组	EN	6.8617	6.9893	6.9974
	SD	47.8476	43.2334	58.9566
	MI	2.5916	2.5041	3.3121
第 4 组	EN	6.8094	6.9273	6.9791
	SD	37.8644	33.8260	39.0462
	MI	1.8391	2.3089	2.5151
第 5 组	EN	6.8426	6.8958	6.9404
	SD	29.7839	30.3861	37.1456
	MI	1.7548	1.8977	2.4588

注:黑体数字为最优结果.

通过实验表明,相比 FusionGAN 和 PMGI,本文提出的基于语义分割的红外和可见光图像融合方法在 3 个客观评估指标上表现均为最佳.

熵和互信息有最优的值表明,本文的融合结果从红外和可见光图像中获得的总信息量最多,说明本文的融合方法确实是有效的融合方法,能保留丰富

的源图像信息.标准差有最优的值表明,本文的融合图像对比度更高,证明通过引入语义信息,针对性的对目标区域和背景区域采用不同融合方法是有效的,实现了融合图像在目标区域保留更多红外图像信息,背景区域保留更多可见光图像信息.

FusionGAN,PMGI 和本文方法的平均运行时间分别为 0.0586 s,0.0322 s,0.1118 s.因为本文提出的方法需要目标区域融合网络和背景区域融合网络 2 个不同的网络,所以本文方法的平均运行时间相比 FusionGAN 和 PMGI 有一定增长,但本文方法的融合结果在主客观上相比现有方法提升更为明显.

5 结 论

本文在原有的基于深度学习的红外和可见光图像融合方法基础上,通过语义分割引入图像语义信息,对源图像目标区域和背景区域采用不同的融合方法,以求得到质量更高的融合图像.实验表明,相比现有方法,本文提出的方法达到了预期的效果,融合图像目标区域保留了大量的对比度信息,背景区域保留丰富的纹理细节信息,在主观和客观上都有更好的融合效果.

参 考 文 献

- [1] Lin Suzhen, Zhu Xiaohong, Wang Dongjuan, et al. Multi-band image fusion based on embedded multi-scale transform [J]. Journal of Computer Research and Development, 2015, 52(4): 952-959 (in Chinese)
(蔺素珍, 朱小红, 王栋娟, 等. 基于嵌入式多尺度变换的多波段图像融合[J]. 计算机研究与发展, 2015, 52(4): 952-959)
- [2] Zhou Huabing, Ma Jiayi, Tan C C, et al. Cross-weather image alignment via latent generative model with intensity consistency [J]. IEEE Transactions on Image Processing, 2020, 29: 5216-5228
- [3] Ma Jiayi, Zhou Huabing, Zhao Ji, et al. Robust feature matching for remote sensing image registration via locally linear transforming [J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(12): 6469-6481
- [4] Ma Jiayi, Jiang Junjun, Zhou Huabing, et al. Guided locality preserving feature matching for remote sensing image registration [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(8): 4435-4447
- [5] Ma Jiayi, Liang Pengwei, Yu Wei, et al. Infrared and visible image fusion via detail preserving adversarial learning [J]. Information Fusion, 2020, 54: 85-98

- [6] Ma Jiayi, Xu Han, Jiang Junjun, et al. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion [J]. IEEE Transactions on Image Processing, 2020, 29: 4980-4995
- [7] Chen Jun, Li Xuejiao, Luo Linbo, et al. Infrared and visible image fusion based on target-enhanced multiscale transform decomposition [J]. Information Sciences, 2020, 508: 64-78
- [8] Li Hui, Wu Xiaojun. Densefuse: A fusion approach to infrared and visible images [J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623
- [9] Burt P, Adelson E. The laplacian pyramid as a compact image code [J]. IEEE Transactions on Communications, 1983, 31(4): 532-540
- [10] Wang Xiaohui, Jia Jia, Cai Lianhong. Expression detail synthesis based on wavelet-based image fusion [J]. Journal of Computer Research and Development, 2013, 50(2): 387-393 (in Chinese)
(王晓慧, 贾珈, 蔡莲红. 基于小波图像融合的表情细节合成[J]. 计算机研究与发展, 2013, 50(2): 387-393)
- [11] Da Cunha A L, Zhou Jianping, Do M N. Nonsubsampled contourlet transform: Filter design and applications in denoising [C/OL] //Proc of IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2005 [2020-03-31]. <https://ieeexplore.ieee.org/abstract/document/1529859>
- [12] Zhang Baohua, Lu Xiaoqi, Pei Haiquan, et al. A fusion algorithm for infrared and visible images based on saliency analysis and non-subsampled shearlet transform [J]. Infrared Physics & Technology, 2015, 73: 286-297
- [13] Ma Jiayi, Yu Wei, Liang Pengwei, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion [J]. Information Fusion, 2019, 48: 11-26
- [14] Zhang Hao, Xu Han, Xiao Yang, et al. Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity [C] //Proc of AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 12797-12804
- [15] Zhang Ruimao, Peng Jiefeng, Wu Yang, et al. The semantic knowledge embedded deep representation learning and its applications on visual understanding [J]. Journal of Computer Research and Development, 2017, 54(6): 1251-1266 (in Chinese)
(张瑞茂, 彭杰锋, 吴恙, 等. 融合语义知识的深度表达学习及在视觉理解中的应用[J]. 计算机研究与发展, 2017, 54(6): 1251-1266)
- [16] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3431-3440
- [17] Chen L, Zhu Yukun, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2018: 801-818

- [18] Xu Han, Liang Pengwei, Yu Wei, et al. Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators [C] //Proc of Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2019: 3954-3960
- [19] Roberts J W, Aardt J V, Ahmed F B. Assessment of image fusion procedures using entropy, image quality, and multispectral classification [J]. Journal of Applied Remote Sensing, 2008, 2(1): 1-28
- [20] Ma Jiayi, Ma Yong, Li Chang. Infrared and visible image fusion methods and applications: A survey [J]. Information Fusion, 2019, 45: 153-178
- [21] Qu Guihong, Zhang Dali, Yan Pingfan. Information measure for performance of image fusion [J]. Electronics Letters, 2002, 38(7): 313-315



Zhou Huabing, born in 1983. Received his BSc and MSc degrees in computer science and technology from Wuhan Institute of Technology, Wuhan China, in 2005 and 2008, respectively, and the PhD degree in control science and engineering from Huazhong University of Science and Technology, Wuhan, China, in 2012. Currently associate professor in the School of Computer Science and Engineering at Wuhan Institute of Technology. His main research interests include computer vision, remote sensing image analysis and intelligent robot.

周华兵,1983年生,博士,副教授.主要研究方向为计算机视觉、遥感图像分析和智能机器人。



Hou Jilei, born in 1997. Master candidate. Received his BSc degree in computer science and technology from Wuhan Institute of Technology, Wuhan China, in 2019. His main research interests include computer vision and machine learning. (houjilei455@gmail.com)

侯积磊,1997年生,硕士研究生.主要研究方向为计算机视觉和机器学习。



Wu Wei, born in 1996. Master candidate. Received his BSc degree in computer science and technology from Wuhan Institute of Technology, Wuhan China, in 2019. His main research interests include computer vision and machine learning. (2026630389@qq.com)

吴伟,1996年生,硕士研究生.主要研究方向为计算机视觉和机器学习。



Zhang Yanduo, born in 1971. Received his PhD degree in the School of Astronautics in Harbin Institute of Technology, China, in 1999. Currently professor in Wuhan Institute of Technology. His main research interests include artificial intelligence, intelligent robot, and computer vision. (zhangyanduo@hotmail.com)

张彦铎,1971年生,博士,教授.主要研究方向为人工智能、智能机器人和计算机视觉。



Wu Yuntao, born in 1973. Received his PhD degree in information and communication engineering from the National Key Laboratory for Radar Signal Processing at Xidian University in 2003. Currently professor in Wuhan Institute of Technology. His main research interests include signal detection and parameter estimation. (ytwu@sina.com)

吴云韬,1973年生,博士,教授.主要研究方向为信号检测、参数估计等。



Ma Jiayi, born in 1986. Received his BSc degree in information and computing science and the PhD degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. Currently professor with the Electronic Information School, Wuhan University. His main research interests include computer vision and machine learning.

马佳义,1986年生,博士,教授.主要研究方向为计算机视觉、机器学习。