

## 基于互信息特征分离表达的红外与可见光图像融合

王慧<sup>1,2,3</sup>, 罗晓清<sup>1,2,3\*</sup>, 张战成<sup>4</sup><sup>1</sup>江南大学人工智能与计算机学院, 江苏 无锡 214122;<sup>2</sup>江南大学先进技术研究院, 江苏 无锡 214122;<sup>3</sup>江苏省模式识别与计算智能工程实验室, 江苏 无锡 214122;<sup>4</sup>苏州科技大学电子与信息工程学院, 江苏 苏州 215000

**摘要** 针对红外与可见光图像融合方法存在的对源图像特征分离不充分、可解释性低且融合规则难以准确设计等问题, 提出基于互信息特征分离表达的红外与可见光图像融合方法, 有效分离特征的同时保留源图像的典型信息。首先, 采用互信息约束的编码网络提取特征, 最大化源图像与特征间互信息来保留源图像的特征表示, 同时通过最小化私有和公有特征的互信息来达到分离表达的目的; 其次, 特征融合阶段设计了层级特征自适应融合模块来有效融合不同层级的特征信息, 减小语义差距并调整感受野, 增强网络对特征的学习能力; 此外, 损失函数采用软加权强度损失来平衡红外与可见光特征分布; 最后, 对比实验的主客观评价结果表明, 所提方法能有效融合红外与可见光图像的重要信息, 具有良好的视觉感知。

**关键词** 图像处理; 红外与可见光图像; 互信息; 分离表达; 层级特征自适应融合; 软加权强度损失

中图分类号 TP391.41

文献标志码 A

DOI: 10.3788/LOP230855

## Infrared and Visible Image Fusion Based on Separate Expression of Mutual Information Features

Wang Hui<sup>1,2,3</sup>, Luo Xiaoqing<sup>1,2,3\*</sup>, Zhang Zhancheng<sup>4</sup><sup>1</sup>School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, Jiangsu, China;<sup>2</sup>Institute of Advanced Technology, Jiangnan University, Wuxi 214122, Jiangsu, China;<sup>3</sup>Jiangsu Laboratory of Pattern Recognition and Computational Intelligence, Wuxi 214122, Jiangsu, China;<sup>4</sup>School of Electronics and Information Engineering, Suzhou University of Science and Technology, Suzhou 215000, Jiangsu, China

**Abstract** To solve the challenges associated with the inadequate separation of source image features, low interpretability, and difficulty of designing accurate fusion rules, this paper proposes an infrared (IR) and visible image fusion method based on mutual information feature separation and representation, which effectively separates features while preserving the typical information of the source image. First, a mutual information constrained coding network is used to extract the features, maximize the mutual information between the source image and features to retain the feature representation of the source image, and minimize the mutual information of private and public features to achieve separation and representation. In addition, the loss function adopts a soft weighted intensity loss to balance the distribution of IR and visible features. Objective and subjective evaluation results of comparison experiments indicate that the proposed method can effectively fuse important information regarding IR and visible images and has good visual perception.

**Key words** image processing; infrared and visible image; mutual information; separation representation; hierarchical adaptive feature fusion; soft weighted intensity loss

## 1 引言

多模态图像融合任务旨在综合不同传感器所获得的同一场景信息, 从而提供信息丰富的综合视角, 典型

代表如红外与可见光图像融合<sup>[1]</sup>。其中, 红外传感器可以有效规避恶劣天气及光线变化导致的视觉障碍, 采集到场景的热辐射信息, 但图像空间分辨率较低且缺乏纹理信息; 而可见光传感器得到的图像具有较高

收稿日期: 2023-02-13; 修回日期: 2023-03-04; 录用日期: 2023-04-07; 网络首发日期: 2023-04-17

通信作者: \*xqluo@jiangnan.edu.cn

分辨率,包含丰富的纹理细节信息,但易受外界环境影响,降低目标的感知能力。因此,融合红外与可见光图像可以生成一幅符合人眼视觉特征且场景表达能力更优的高质量图像,这便于应用在其他计算机视觉任务如目标跟踪<sup>[2]</sup>、行为识别<sup>[3]</sup>等领域。

为实现红外与可见光图像融合,早期研究者们一般采用基于多尺度变换<sup>[4]</sup>和表示学习<sup>[5]</sup>等传统方法,旨在将源图像进行分解或映射到其他空间,然后设计不同融合策略进行融合。有效实现图像融合同时也存在计算过程复杂、融合效率低下的弊端。而基于深度学习<sup>[6]</sup>的图像融合方法利用神经网络强大的特征提取和重构能力,在一定程度上缓解传统方法所存在的问题,受到极大关注。

自编码网络作为深度学习中典型无监督模型,具有良好的特征重构性能<sup>[7]</sup>,被广泛应用于图像融合领域,如:Li等<sup>[8]</sup>结合编码网络和密集连接方式提出了DenseFuse,进行图像特征提取和重构;随后在此基础上提出的NestFuse<sup>[9]</sup>改进了融合策略和损失函数,获得的融合图像细节更加丰富;Jian等<sup>[10]</sup>提出带有残差块的对称编码解码网络SEDRFuse,分别提取源图像中间特征和补偿特征,对两种特征采用不同方式进行融合。但上述方法都只是简单使用编码器生成红外与可见光特征图,未单独关注不同传感器模态之间的私有和公有信息。

多模态图像之间存在私有与公有信息,私有信息是各个源图像独有的模态特征,如纹理和梯度细节等,而公有信息是指公共的区域结构特征。近年来,对特征进行有效分离受到研究者的关注。Luo等<sup>[11]</sup>提出了基于分离表达的图像融合通用框架IFsepR,采用包含两个私有分支和一个公共分支的联合编码器,将不同分支提取特征之间的余弦相似性最小化来获得图像上的私有和公有特征,但该方法缺乏对图像特征对应关系的考量,难以有效分离多模态图像上的私有与公有特征;Xu等<sup>[12]</sup>提出基于解纠缠表示的可见红外图像融合分解方法,在源图像成像过程中对特征进行解耦操作,但该方法将模态特征压缩为一维向量,造成空间信息的丢失,从而引起融合视觉模糊与信息扭曲。

基于上述分析,本文提出基于互信息特征分离表达的红外与可见光图像融合方法,通过互信息编码器分别提取源图像独有的判别信息以及共享的结构信息,编码网络具有4个参数不共享的分支;同时为了捕获源图像的全局感受野信息,额外添加编码分支进行浅层纹理特征提取;采用层级自适应特征融合模块对自下而上的特征递进融合,促进层级特征联合学习。此外,为探索源图像私有特征间的先验关系,在强度损失上采用软加权策略通过网络自我优化源图像特征所占权重。实验结果表明,所提方法能有效对源图像进行特征分离,充分融合多模态图像层级信息,融合图像在视觉感知和客观指标均取得较好效果。

## 2 本文方法

所提方法为端到端的图像融合网络,结构如图1所示。输入为已配准的红外图像(IR)和可见光图像(VIS),输出为融合图像(F)。网络整体流程如下:

1) 使用互信息编码网络En\*提取红外与可见光图像不同层级特征。公有特征编码器En<sub>C1</sub>和En<sub>C2</sub>提取源图像层级特征 $C_1^{\text{IR}}, C_2^{\text{IR}}, C_1^{\text{VIS}}, C_2^{\text{VIS}}$ ;私有特征编码器En<sub>R1</sub>和En<sub>R2</sub>提取源图像层级特征 $R_1^{\text{IR}}, R_2^{\text{IR}}, R_1^{\text{VIS}}, R_2^{\text{VIS}}$ ,用以捕获源图像中私有信息,图像融合网络整体训练时互信息编码部分参数冻结。

2) 使用普通卷积编码网络En充分提取全局感受野特征 $F_1$ ,将层级特征 $F_1, F'_1$ 和 $F''_1$ 级联后送入融合层。在融合层使用层级特征自适应融合模块(HAFF)融合提取层级特征;解码网络De解码融合的特征,得到融合图像。解码网络由4个 $3 \times 3$ 卷积层组成,第一个卷积层中仅使用LeakyReLU激活函数,而其他卷积层使用批标准化和LeakyReLU激活函数,最后使用Tanh激活函数来输出融合图像。

3) 考虑到融合图像中源图像信息占比不同,利用提取的私有特征计算源图像信息权重,通过该权重实现对强度损失软加权。其实现方式为对私有特征全局平均池化后进行级联,使用 $1 \times 1$ 卷积将通道维度压缩到2维,在通道维度使用Softmax函数得到红外与可见光信息权重 $a$ 和 $b$ 。

### 2.1 互信息编码器

目前基于自编码器与分离表达的图像融合方法一般是通过损失函数的约束来实现私有和公有特征分离,但这忽略了源图像对应空间位置特征不同的重要性导致学习到的分离表达不到位。因此,基于迁移学习的思想,使用红外与可见光TNO数据集训练文献[13]解耦模型得到互信息编码器,此训练过程分为两个阶段,包括公有特征编码器En<sub>C1</sub>和En<sub>C2</sub>,私有特征编码器En<sub>R1</sub>和En<sub>R2</sub>。利用训练好的编码器提取特征以达到分离表达的目的。

#### 2.1.1 互信息

信息论中,互信息用来度量随机变量之间的相互依赖关系。假设输入数据IR和VIS的边缘分布为 $P_{\text{IR}}$ 和 $P_{\text{VIS}}$ ,联合分布为 $P_{\text{IR}, \text{VIS}}$ 。基于Deep InfoMax启发,使用Jensen-Shannon(JS)散度代替Kullback-Leibler(KL)散度进行优化<sup>[13]</sup>,具体可表示为

$$I_{\text{IR}, \text{VIS}} = D_{\text{JS}}(P_{\text{IR}, \text{VIS}} \| P_{\text{IR}} P_{\text{VIS}}), \quad (1)$$

式中: $I_{\text{IR}, \text{VIS}}$ 为IR和VIS之间的互信息; $D_{\text{JS}}$ 为使用JS散度估计变量依赖并解耦。

对IR和VIS的互信息进行估计,具体可表示为

$$I_{\text{IR}, \text{VIS}} = E_{P_{\text{IR}, \text{VIS}}} \left\{ -\lg \left[ 1 + \exp(-T_{\text{IR}, \text{VIS}}) \right] \right\} - E_{P_{\text{IR}} P_{\text{VIS}}} \left\{ \lg \left[ 1 + \exp(T_{\text{IR}, \text{VIS}}) \right] \right\}, \quad (2)$$

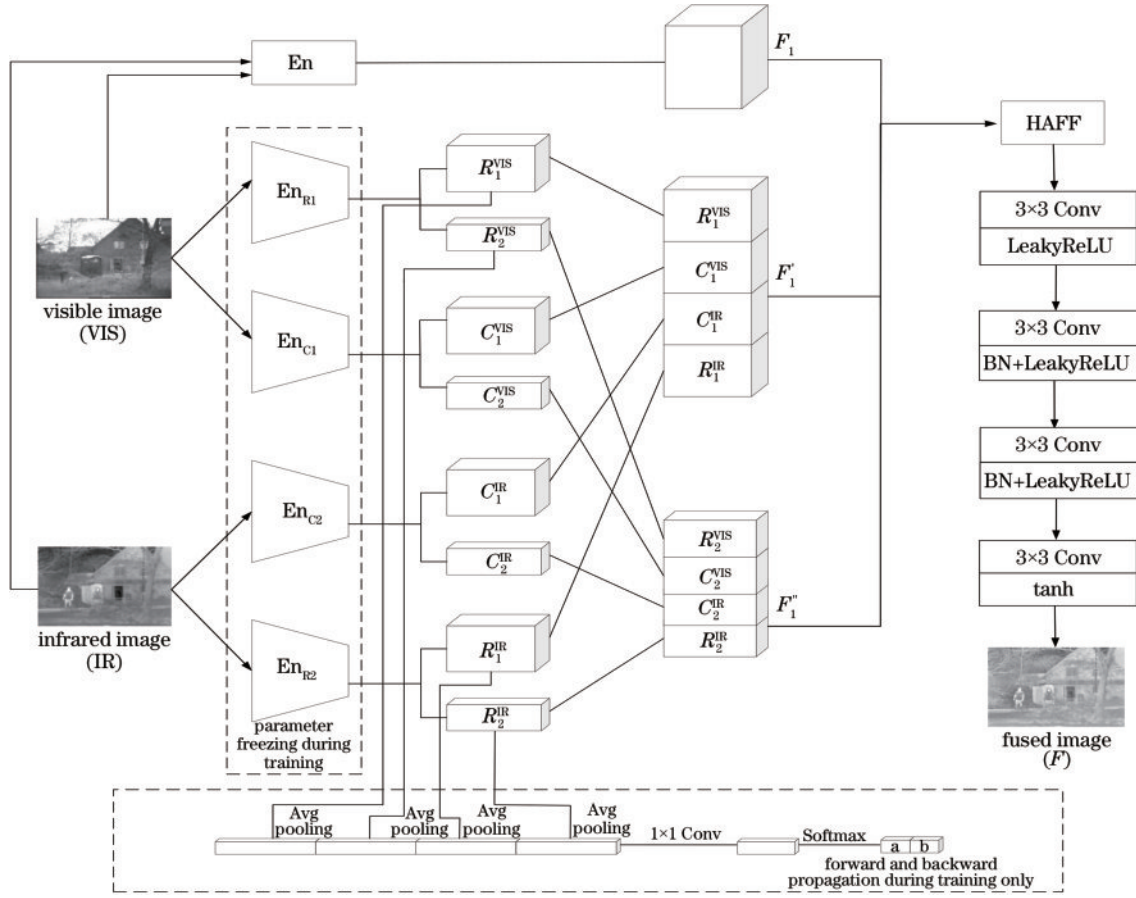


图1 融合网络整体框架

Fig. 1 Overall framework for fusion network

式中： $T$ 为Deep InfoMax中使用的深度统计网络。

$$L_{\theta, \phi}(f_{IR}, f_{VIS}) = I_{IR, VIS}, \quad (3)$$

式中： $L_{\theta, \phi}(f_{IR}, f_{VIS})$ 为优化网络的目标损失， $\theta$ 为全局统计网络 $T$ 的参数， $\phi$ 为局部统计网络 $T$ 的参数。

文献[13]中考虑到图像可以有两种互信息，一种是隐空间变量和输入图像间的全局互信息，另一种是隐空间变量和输入图像提取的特征图间的局部互信息。以IR为例， $Z$ 是参数为 $\phi$ 的神经网络提取的隐空间表示， $En_{\cdot}$ 为提取特征的编码器，则全局互信息 $L_{\theta, \phi}^{global}(f_{IR}, Z)$ 、局部互信息 $L_{\phi, \phi}^{local}(f_{IR}, Z)$ 可表示为

$$L_{\theta, \phi}^{global}(f_{IR}, Z) = I_{\theta}(f_{IR}, Z), \quad (4)$$

$$L_{\phi, \phi}^{local}(f_{IR}, Z) = I_{\phi}[En_{\cdot}(f_{IR}), Z], \quad (5)$$

式中： $I_{\theta}(f_{IR}, Z)$ 为红外图像与隐空间特征之间的互信息； $En_{\cdot}(f_{IR})$ 为编码器提取的红外图像特征。

### 2.1.2 公有特征学习

对输入的源图像对 $(f_{IR}, f_{VIS})$ 而言，假设 $E_{IR}$ 和 $E_{VIS}$ 为输入图像的隐空间表示，其具体可表示为

$$E_{IR} = (C_{IR}, R_{IR}), \quad (6)$$

$$E_{VIS} = (C_{VIS}, R_{VIS}), \quad (7)$$

式中： $C_{IR}$ 为红外图像的公有特征； $R_{IR}$ 为红外图像的私有特征； $C_{VIS}$ 为可见光图像的公有特征； $R_{VIS}$ 为可见光

图像的私有特征。

通过公有特征编码器 $En_{C1}$ 和 $En_{C2}$ 提取源图像特征 $C_{IR}$ 和 $C_{VIS}$ 。为了强制网络只学习红外与可见光图像之间的公共信息，对提取的特征与源图像进行交换公有特征来计算交叉互信息，具体可表示为

$$L_{MI}^C = \alpha_C [L_{\theta_1, \phi_2}^{global}(f_{IR}, C_{VIS}) + L_{\theta_2, \phi_1}^{global}(f_{VIS}, C_{IR})] + \beta_C [L_{\phi_1, \phi_2}^{local}(f_{IR}, C_{VIS}) + L_{\phi_2, \phi_1}^{local}(f_{VIS}, C_{IR})], \quad (8)$$

式中： $\phi_1, \phi_2$ 为 $En_{C1}, En_{C2}$ 的权重； $\alpha_C, \beta_C$ 为损失系数，分别设置为0.5和1。

此外，源图像存储的是同一个场景信息，应该具有相同的公有特征 $C_{IR}$ 和 $C_{VIS}$ ，对其使用 $L_1$ 损失进行一致性约束：

$$L_1 = E_{P(C_{IR}, C_{VIS})}(|C_{IR} - C_{VIS}|), \quad (9)$$

公有编码器的损失 $L_C$ 可表示为

$$L_C = L_{MI}^C - \gamma_C L_1, \quad (10)$$

式中： $\gamma_C$ 为比例参数，设置为0.1。

### 2.1.3 私有特征学习

通过私有特征编码器 $En_{R1}$ 和 $En_{R2}$ 提取源图像特征 $R_{IR}$ 和 $R_{VIS}$ 。为了学习这些特征，对源图像和其相应的特征进行互信息最大化约束，具体可表示为

$$L_{MI}^R = \alpha_R [L_{\theta_1, \omega_1}^{global}(f_{IR}, E_{IR}) + L_{\theta_2, \omega_2}^{global}(f_{VIS}, E_{VIS})] +$$

$$\beta_R [L_{\phi_1, w_1}^{\text{local}}(f_{\text{IR}}, E_{\text{IR}}) + L_{\phi_2, w_2}^{\text{local}}(f_{\text{VIS}}, E_{\text{VIS}})], \quad (11)$$

式中:  $w_1, w_2$  为  $\text{En}_{R_1}$  和  $\text{En}_{R_2}$  的权重;  $\alpha_R, \beta_R$  为损失系数, 分别设置为 0.5 和 1。

另一方面, 私有编码器用以捕获独占信息, 所得到的特征不能包含公有特征, 在对式(11)进行互信息最大化的同时也需对私有特征和公有特征进行最小化约束, 此时若使用式(2)直接进行互信息最小化, 统计网络  $T$  将无法收敛。因此, 文献[13]引入鉴别器  $D$  使用对抗性目标损失来实现互信息最小化。鉴别器  $D$  将  $P(C_{\text{IR}}, R_{\text{IR}})$  作为假样本,  $P(C_{\text{IR}})P(R_{\text{IR}})$  作为真实样本。

$$L_{\text{adv}}^{\text{IR}} = E_{P(C_{\text{IR}})P(R_{\text{IR}})} [\lg D_{\rho_{\text{IR}}}(C_{\text{IR}}, R_{\text{IR}})] + E_{P(C_{\text{IR}}, R_{\text{IR}})} \left\{ \lg [1 - D_{\rho_{\text{IR}}}(C_{\text{IR}}, R_{\text{IR}})] \right\}, \quad (12)$$

$$L_{\text{adv}}^{\text{VIS}} = E_{P(C_{\text{VIS}})P(R_{\text{VIS}})} [\lg D_{\rho_{\text{VIS}}}(C_{\text{VIS}}, R_{\text{VIS}})] + E_{P(C_{\text{VIS}}, R_{\text{VIS}})} \left\{ \lg [1 - D_{\rho_{\text{VIS}}}(C_{\text{VIS}}, R_{\text{VIS}})] \right\}, \quad (13)$$

式中:  $L_{\text{adv}}^{\text{IR}}$  为红外图像的对抗损失;  $L_{\text{adv}}^{\text{VIS}}$  为可见光图像的对抗损失;  $D_{\rho_{\text{IR}}}$  为概率密度为  $\rho$  的红外鉴别器;  $D_{\rho_{\text{VIS}}}$

为概率密度为  $\rho$  的可见光鉴别器。

私有编码器的损失  $L_R$  可表示为

$$L_R = L_{\text{MI}}^R - \gamma_{\text{adv}} (L_{\text{adv}}^{\text{IR}} + L_{\text{adv}}^{\text{VIS}}), \quad (14)$$

式中:  $\gamma_{\text{adv}}$  为比例参数, 设置为 0.025。

#### 2.1.4 编码器网络结构及特征可视化

互信息编码器  $\text{En}^*$  网络结构如图 2(a) 所示, 由 3 个  $4 \times 4$  卷积层组成, 不同于常规  $3 \times 3$  或  $7 \times 7$  的卷积核, 考虑到  $3 \times 3$  卷积核需要多层卷积来扩大感受野, 且卷积层的输出与输入尺寸不一致, 因此在多层卷积后, 可能会丢失信息;  $7 \times 7$  卷积核感受野扩大的同时会增加计算成本, 且可能会导致过拟合。基于此分析, 使用  $4 \times 4$  卷积核能在一定程度上保证神经网络感受野的范围, 有效捕获图像的全局特征从而减少信息丢失, 还能够避免计算成本的增加。第 2 和第 3 卷积层步长为 2 实现图像的多层级编码, 提取多层级特征  $F_1'$  和  $F_1''$ 。图 2(b) 为非下采样的普通卷积编码器  $\text{En}$ , 对输入图像直接编码得到特征  $F_1$ 。

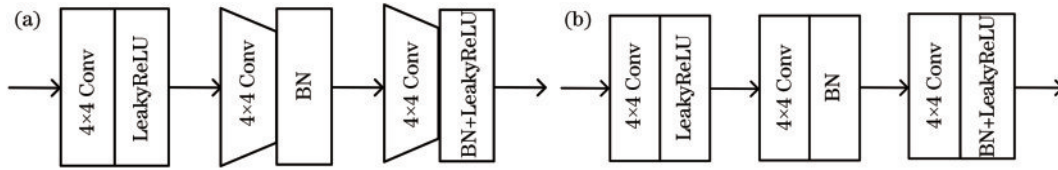


图 2 编码器结构。(a)互信息编码器;(b)普通卷积编码器

Fig. 2 Encoder construction. (a) Mutual information encoder; (b) general convolutional encoder

为验证互信息训练编码器的有效性, 将提取的源图像层级特征进行可视化分析, 如图 3 所示。图 3(b1)、图 3(b2)、图 3(d1)、图 3(d2) 为互信息编码网络提取的低层纹理信息, 图 3(c1)、图 3(c2)、图 3(e1)、图 3(e2) 为高层语义信息。由图 3(b1)、图 3(c1)、图 3(b2)、

图 3(c2) 可知, 互信息编码器提取的公有特征包含了红外图像人物信息和可见光图像房屋基本结构信息, 由图 3(d1)、图 3(d2)、图 3(e1)、图 3(e2) 可知, 提取的私有特征拥有清晰的边缘结构纹理且红外图像中的人物更加显眼。

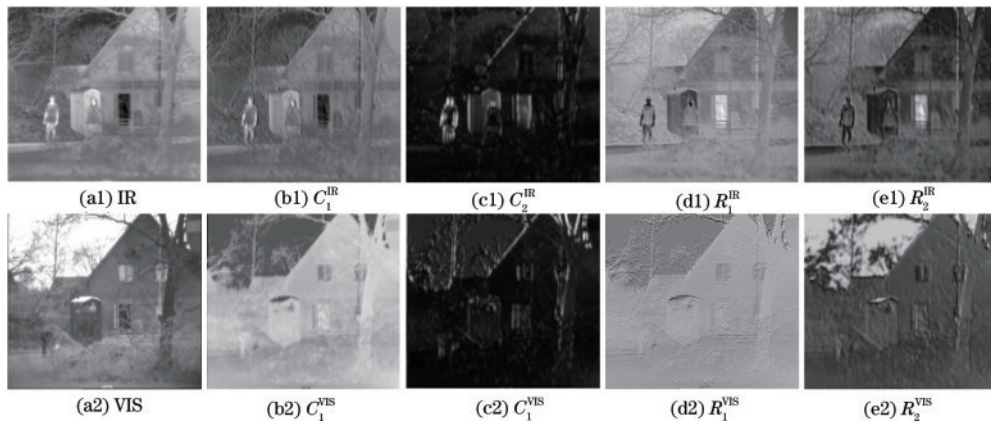


图 3 层级特征可视化。(a1) IR; (a2) VIS; (b1)  $C_1^{\text{IR}}$ ; (b2)  $C_1^{\text{VIS}}$ ; (c1)  $C_2^{\text{IR}}$ ; (c2)  $C_2^{\text{VIS}}$ ; (d1)  $R_1^{\text{IR}}$ ; (d2)  $R_1^{\text{VIS}}$ ; (e1)  $R_2^{\text{IR}}$ ; (e2)  $R_2^{\text{VIS}}$

Fig. 3 Hierarchical feature visualisation. (a1) IR; (a2) VIS; (b1)  $C_1^{\text{IR}}$ ; (b2)  $C_1^{\text{VIS}}$ ; (c1)  $C_2^{\text{IR}}$ ; (c2)  $C_2^{\text{VIS}}$ ; (d1)  $R_1^{\text{IR}}$ ; (d2)  $R_1^{\text{VIS}}$ ; (e1)  $R_2^{\text{IR}}$ ; (e2)  $R_2^{\text{VIS}}$

## 2.2 层级特征自适应融合模块

不同层级特征图所包含的图像信息不同, 为了有效利用不同层级特征中有价值的信息, 基于金字塔特征融合策略 ASFF<sup>[14]</sup>, 提出了 HAFF, 进行自下而上的

特征递进融合, 以捕获不同层级多样性特征来弥合编码器、解码器之间的语义差距, 增强网络的信息表达能力。

HAFF 通过对不同层级特征分配自适应的权重来

融合层级特征图,如图 4 所示,主要分为两步:权重学习和特征融合。首先,将编码网络得到的特征图  $F_1'$  和  $F_1''$  分别上采样至  $F_1$  相同尺度得到  $F_2$  和  $F_3$ ,然后使用  $3 \times 3$  卷积将  $F_1$ 、 $F_2$  和  $F_3$  的通道维度统一至 256,接着将  $F_1$ 、 $F_2$  和  $F_3$  级联后使用  $3 \times 3$  卷积进行权重学习得到  $H \times W \times 3$  的特征权重,最后使用 Softmax 函数在通道维度对特征权重归一化,得到权重  $\alpha$ 、 $\beta$  和  $\gamma$ 。利用学习到的权重对每层特征进行加权,再将加权后的特征相加,得到最终的融合特征  $F'$ , $F'$  可表示为

$$F' = \alpha F_1 + \beta F_2 + \gamma F_3. \quad (15)$$

为更加清晰地展示融合模块自适应参数学习内容,将特征  $F_1$ 、 $F_2$  和  $F_3$  与其对应权重以及融合后特征

$F'$  进行可视化,如图 5 所示。 $F_1$  和  $\alpha$  分别为全局感受野特征及对应权重,由图 5(c)、图 5(d) 可知, $F_1$  保留了红外图像基本人物信息和汽车轮廓信息及可见光图像房屋结构信息, $\alpha$  凸显源图像基本信息; $F_2$  和  $\beta$  分别为浅层纹理特征及对应权重,由图 5(e)、图 5(f) 可知, $F_2$  较  $F_1$  凸显源图像的典型特征,如红外图像的人物和可见光图像的房屋, $\beta$  较  $\alpha$  更加关注全局信息; $F_3$  和  $\gamma$  分别为深层语义特征及对应权重,由图 5(g)、图 5(h) 可知, $F_3$  较  $F_2$  的特征更加全面,网络更多关注内容结构的完整性, $\gamma$  与  $\beta$  通过网络所学习的大体相似; $F'$  为融合特征图,由图 5(i) 可知, $F'$  保留了浅层特征的细节纹理,也包含深层特征的语义信息。

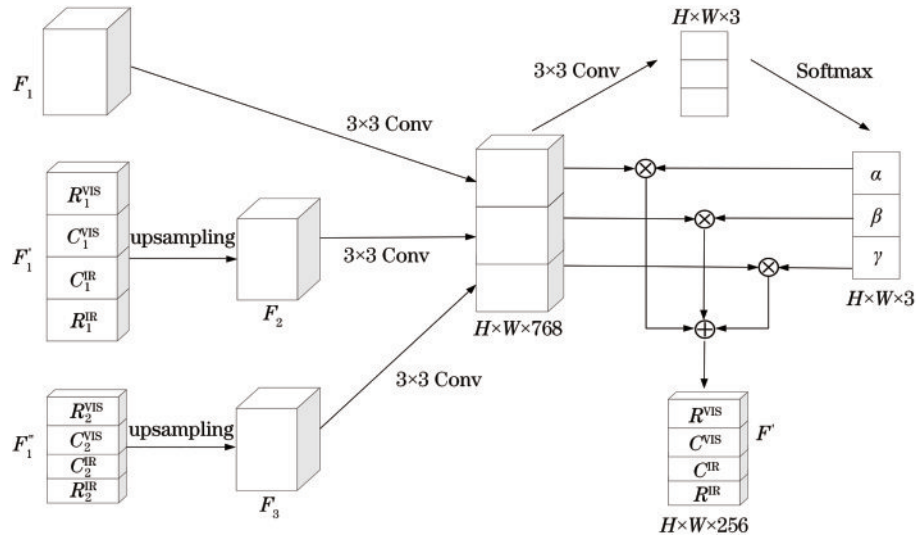


图 4 HAFF 模块结构

Fig. 4 HAFF module structure



图 5 HAFF 模块特征及参数可视化。(a) IR; (b) VIS; (c)  $F_1$ ; (d)  $\alpha$ ; (e)  $F_2$ ; (f)  $\beta$ ; (g)  $F_3$ ; (h)  $\gamma$ ; (i)  $F'$

Fig. 5 Visualisation of HAFF module features and parameters. (a) IR; (b) VIS; (c)  $F_1$ ; (d)  $\alpha$ ; (e)  $F_2$ ; (f)  $\beta$ ; (g)  $F_3$ ; (h)  $\gamma$ ; (i)  $F'$

### 2.3 损失函数设计

损失函数  $L$  采用软加权强度损失和联合梯度损失, $L$  可表示为

$$L = L_{M\_Int} + \lambda L_{grad}, \quad (16)$$

式中: $L_{M\_Int}$  为软加权强度损失; $L_{grad}$  为联合梯度损失; $\lambda$

为平衡损失间比例参数。

#### 2.3.1 软加权强度损失

传统的强度损失  $L_{Int}$  通过源图像和融合图像的欧氏距离衡量强度差, $L_{Int}$  可表示为

$$L_{\text{int}} = \frac{1}{H \times W} \left( \|F - f_{\text{IR}}\|_2 + \omega \|F - f_{\text{VIS}}\|_2 \right), \quad (17)$$

式中:  $H$ 、 $W$  分别为高度和宽度;  $\|\cdot\|_2$  为 Frobenius 范数, 权重参数  $\omega$  控制融合图像中红外图像和可见光图像的信息比, 权重倾向于某一源图像, 会相对削弱融合图像中的另一源图像特征。因此, 提出一种软加权强度损失  $L_{M,\text{int}}$ , 利用私有互信息编码器所提取的特征, 通过神经网络学习红外和可见光图像权重, 实现对强度损失软加权,  $L_{M,\text{int}}$  可表示为

$$L_{M,\text{int}} = \frac{1}{H \times W} \left( a \|F - f_{\text{IR}}\|_2 + b \|F - f_{\text{VIS}}\|_2 \right), \quad (18)$$

式中:  $a$  为网络学习的红外图像权重;  $b$  为网络学习的可见光图像权重。

### 2.3.2 联合梯度损失

联合梯度损失  $L_{\text{grad}}$  通过对比源图像的梯度, 取其梯度较大项作为损失,  $L_{\text{grad}}$  可表示为

$$L_{\text{grad}} = \frac{1}{H \times W} \left\| \nabla F - \max(\nabla f_{\text{IR}}, \nabla f_{\text{VIS}}) \right\|_2, \quad (19)$$

式中:  $\nabla$  为梯度算子;  $\max$  为取最大值。

## 3 实验结果与分析

为验证所提方法的有效性, 在 TNO 及 RoadScene

数据集上进行主客观实验对比, 同时与 7 种经典红外和可见光图像融合方法进行比较分析, 其中包括一种传统方法和 6 种基于深度学习的方法, 分别为 GTF<sup>[15]</sup>、GANMcC<sup>[16]</sup>、GAN-FM<sup>[17]</sup>、SDNet<sup>[18]</sup>、DenseFuse<sup>[8]</sup>、DRF<sup>[12]</sup>、IFSepR<sup>[11]</sup>。

主观上通过分析比较视觉效果, 客观上选取 8 种指标: 标准差 (SD)、熵 (EN)、清晰度 (DF)、边缘强度 (EI)、平均梯度 (AG)、空间频率 (SF)、基于相位一致性指标 (Qp)、互信息 (MI)。定量衡量各融合方法在保持纹理细节和图像融合质量等方面的性能。以上客观指标值越大, 表明融合图像质量越好。

训练阶段选取 TNO 数据集上 10 对图像, 裁剪成  $120 \times 120$  大小的图像块, 得到 10160 对图像送入网络。优化器设置为 Adam, 批处理大小为 8, 迭代轮数为 10, 学习率为 0.0001。测试阶段选取 TNO 中 43 对以及 RoadScene 中 221 对图像进行实验。实验使用 PyTorch 1.9 框架在 Python 3.7 环境下完成。硬件环境为 Ubuntu 16, CPU Intel(R) Core i7-6850k, 内存为 64 G, 实验参数设置为  $\lambda = 0.7$ 。

### 3.1 TNO 数据集实验结果及分析

图 6 为各方法在 TNO 数据集部分图像上的融合结果。融合图像部分细节信息已框出。其中, GTF 方法

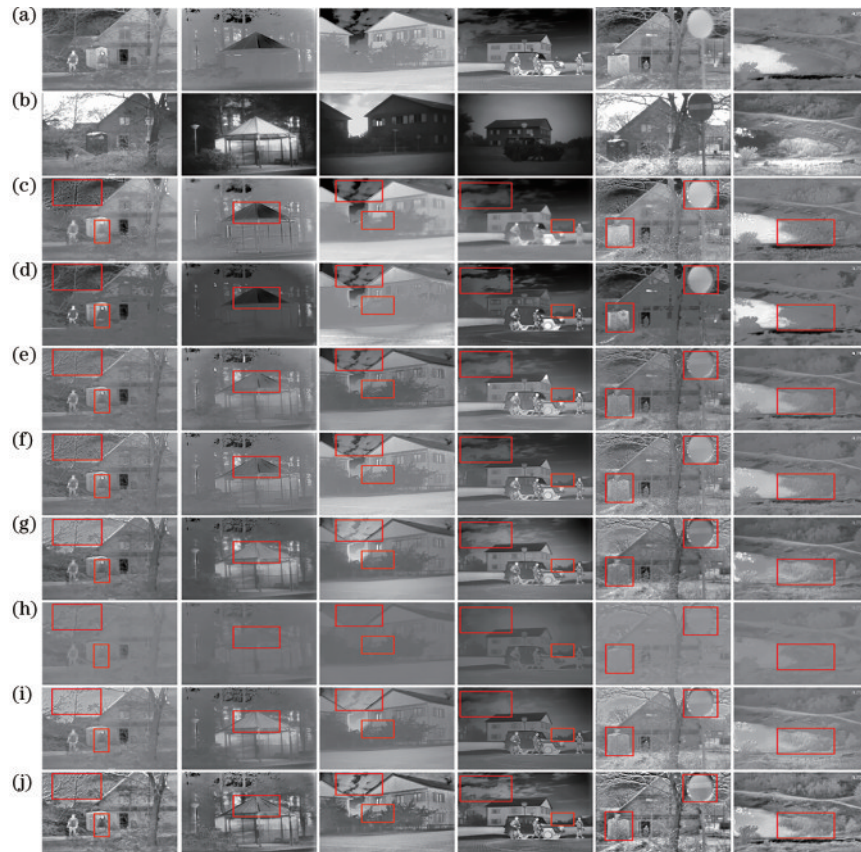


图 6 TNO 数据集对比实验结果 (a) IR; (b) VIS; (c) GTF; (d) GANMcC; (e) GAN-FM; (f) SDNet; (g) Densefuse; (h) DRF; (i) IFSepR; (j) ours

Fig. 6 Experimental results comparing the TNO dataset. (a) IR; (b) VIS; (c) GTF; (d) GANMcC; (e) GAN-FM; (f) SDNet; (g) Densefuse; (h) DRF; (i) IFSepR; (j) ours

是一种使用 VIS 图像中丰富的纹理来增强 IR 图像的方法,由图 6(c)可知,GTF 方法在所标区域内存在人工引入噪声,保留红外目标显著性的同时细节丢失严重;GANMcC 通过多分布联合估计实现源图像特征分布,由图 6(d)可知,融合图像云层边缘阴影过渡不自然,整体图像偏暗;GAN-FM 采用双判别器结构分别引入红外与可见光图像的特征信息,保留了源图像的部分显著特征,但所得图像趋于平滑,视觉效果不佳;SDNet 是一种用于实时图像融合的多功能压缩分解网络模型,由图 6(f)可知,融合图像对比度不高,且部分图像存在伪影;Densefuse 基于残差网络的图像融合方法,能够充分利用不同层次图像特征,但由于没有对源图像特征进行分离,对比所提方法图像整体对比度不高;DRF 是一种从源图像成像过程出发对特征进行解耦操作的融合算法,由图 6(h)可知,融合图像整体亮度较低,所

标云层、树木等纹理信息没有得以保留,视觉效果较差;IFSepR 是一种基于分离表达的图像融合通用框架,融合图像中保留了红外辐射信息,但树木等背景纹理信息有所丢失。对比可得,所提方法融合的图像细节纹理清晰且背景信息丰富,具有较好的视觉效果。

为验证所提方法的有效性,将各方法在 TNO 数据集上进行客观评价分析。由表 1 可知,所提方法在 8 个指标上均保持最优值。且在 SD、EI 和 AG 指标上有良好表现,较次优值提升了 4.1407、13.1772 和 2.0312。这表明软加权强度损失对图像特征进行约束,与联合梯度损失约束相互补充,使得融合图像具有清晰的边缘和较高的对比度。此外,较高的 MI 表明互信息编码器是有效的,融合图像能充分保留源图像的重要特征分布。

表 1 各方法在 TNO 数据集 43 对图像上的客观评价指标  
Table 1 Objective evaluation metrics for each method on the TNO dataset 43 pairs of images

Method	SD	EN	DF	EI	AG	SF	Qp	MI
GTF	39.3766	6.7700	3.5374	27.5437	2.7666	7.0316	0.1954	13.5402
GANMcC	30.4876	6.2167	2.2788	20.1246	1.9334	4.6448	0.1363	12.4336
GAN-FM	28.6088	6.5363	3.5750	27.3276	2.7304	6.9313	0.1945	13.0727
SDNet	33.0535	6.7042	4.7285	39.9012	3.9359	9.3848	0.2762	13.4086
Densefuse	35.6407	6.8817	3.6372	30.7780	3.0116	7.1147	0.3039	13.7635
DRF	9.7089	5.0372	0.7828	7.8756	0.7220	1.6082	0.1127	10.0745
IFSepR	26.9453	6.4900	3.3446	28.0275	2.7778	7.1973	0.3295	12.9799
Ours	43.6229	7.1650	5.7459	53.2266	5.0586	10.7925	0.4995	14.3300

### 3.2 RoadSence 数据集实验结果

为了进一步验证所提方法的有效性和泛化能力,在 RoadSence 红外与可见光数据集上进行对比实验。首先进行主观效果对比,选取部分典型图像融合结果如图 7 所示。GTF 方法整体融合效果较差,融合图像存在大量伪影;GANMcC 方法有效保留了红外辐射信息,但融合图像细节纹理清晰度低,城堡和字母只能看见整体结构,细节信息丢失;GAN-FM 生成的融合图像对比所提方法,树木和城堡纹理细节缺失;SDNet 和 DenseFuse 融合图像较为清晰但整体对比度不高,没有突出树木等边缘纹理细节;DRF 和 IFSepR 所得图像趋于平滑,视觉效果较差。对比可得,所提方法有效保留了源图像重要信息,融合图像内容更加清晰且对比度更高,视觉效果较好。

接着对模型的泛化能力进行客观评价,使用总共 221 对来自 RoadScene 数据集的图像进行数据指标评价,各方法在 RoadScene 数据集上的指标结果如表 2 所示,所提方法在 8 个指标上均达到最优,且在指标 EI 和 SF 表现良好,分别超过次优值 34.6918 和 6.1823。此外,所提方法在指标 EN 上表现最优,说明融合图像的信息量丰富。TNO 和 RoadScene 两个数据集上主客观

评价的出色表现,说明所提方法能有效融合红外和可见光图像并且具有良好的泛化能力。

### 3.3 消融实验结果

采用消融实验进一步验证层级特征自适应融合模块和软加权强度损失的有效性。使用 TNO 数据集中的 10 对图像作为训练集,测试验证选取 TNO 数据集中 43 对图像。

#### 3.3.1 层级特征自适应融合

常见的融合方式有加法式、乘法式以及级联式。为验证提出的层级自适应特征融合方式的有效性,对融合方式进行消融实验。同时与基于金字塔特征融合方式 ASFF 进行对比,ASFF 的融合过程进行两次维度压缩,且第一次压缩过程是基于部分特征进行压缩,可能会造成特征信息丢失;而提出的 HAFF 只进行一次维度压缩,且是基于整体特征进行的压缩,减少信息丢失。

不同融合方式指标结果如表 3 所示。由表 3 可知,采用的特征融合方式在所有指标中均为最优值,这表明不同层级的图像特征得到有效传递,网络中图像信息得以充分利用。同时说明 HAFF 基于整体的一次特征维度压缩比 ASFF 基于部分特征的两次维度压缩

效果更好,保留了更多特征信息。

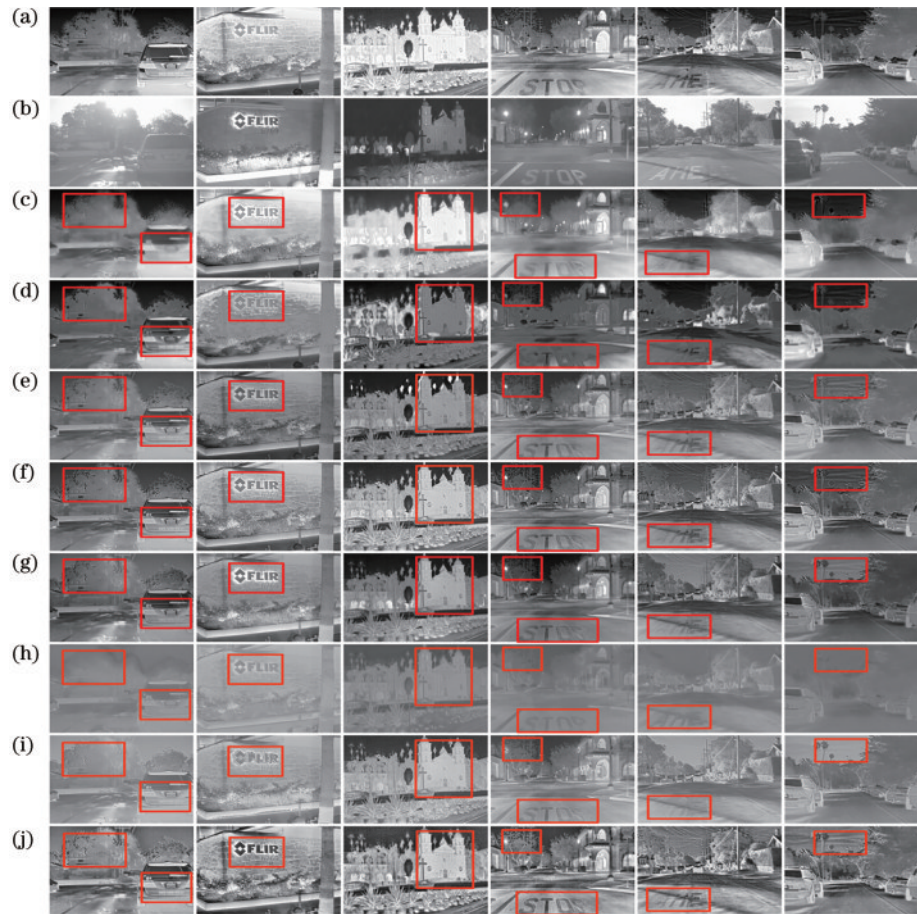


图 7 RoadScene 数据集对比实验结果。(a) IR; (b) VIS; (c) GTF; (d) GANMcC; (e) GAN-FM; (f) SDNet; (g) Densefuse; (h) DRF; (i) IFSepR; (j) ours

Fig. 7 Experimental results comparing the RoadScene dataset. (a) IR; (b) VIS; (c) GTF; (d) GANMcC; (e) GAN-FM; (f) SDNet; (g) Densefuse; (h) DRF; (i) IFSepR; (j) ours

表 2 各方法在 RoadScene 数据集 221 对图像上的客观评价指标

Table 2 Objective evaluation metrics for each method on the RoadScene dataset 221 pairs of images

Method	SD	EN	DF	EI	AG	SF	Qp	MI
GTF	53.0565	7.5013	3.9626	35.3473	3.3552	9.4492	0.2321	15.0027
GANMcC	38.6292	6.8719	3.8104	35.6541	3.3365	8.0408	0.1863	13.7437
GAN-FM	38.3201	7.0326	4.8691	41.5556	3.9687	10.4286	0.2467	14.0652
SDNet	44.9798	7.3160	7.1417	64.0914	6.0926	15.1815	0.3998	14.6320
Densefuse	42.3739	7.1708	5.3075	46.3451	4.4202	11.2749	0.3938	14.3417
DRF	17.4962	5.8409	1.3293	13.3512	1.2238	2.7764	0.0835	11.6818
IFSepR	33.4337	6.8843	5.3954	44.7533	4.4083	13.2158	0.3513	13.7686
Ours	55.2487	7.6632	10.8311	98.7832	9.3087	21.3638	0.5034	15.3265

表 3 不同融合方式在 TNO 数据集 43 对图像上的客观评价指标

Table 3 Objective evaluation metrics for different fusion methods on 43 pairs of images from the TNO dataset

Fusion	SD	EN	DF	EI	AG	SF	Qp	MI
Addition	39.1200	7.0708	5.2884	48.9269	4.6386	9.8304	0.4166	14.1415
Multiplication	37.8513	7.0431	5.1052	47.2807	4.4835	9.3957	0.3856	14.0862
Concation	42.3549	7.1600	4.8912	46.2831	4.3370	9.2844	0.2956	14.3200
ASFF	42.8398	7.1284	5.5951	51.0774	5.0215	10.0065	0.3295	14.2568
Ours	43.6229	7.1650	5.7459	53.2266	5.0586	10.7925	0.4995	14.3300



### 3.3.2 软加权强度损失

为了研究软加权强度损失对网络整体训练的优越性,对传统强度损失不同权重参数值  $\omega$  进行主观评价分析,结果如图 8 所示。由图 8(c)~图 8(m)可知,云

层部分包含明显人工噪声伪像,干扰对红外图像的有效识别。图 8(n)整体图像较为清晰且云层部分无明显伪影,这表明软加权强度损失根据源图像特征自适应分配权重,更好地处理多模态图像之间的关系。

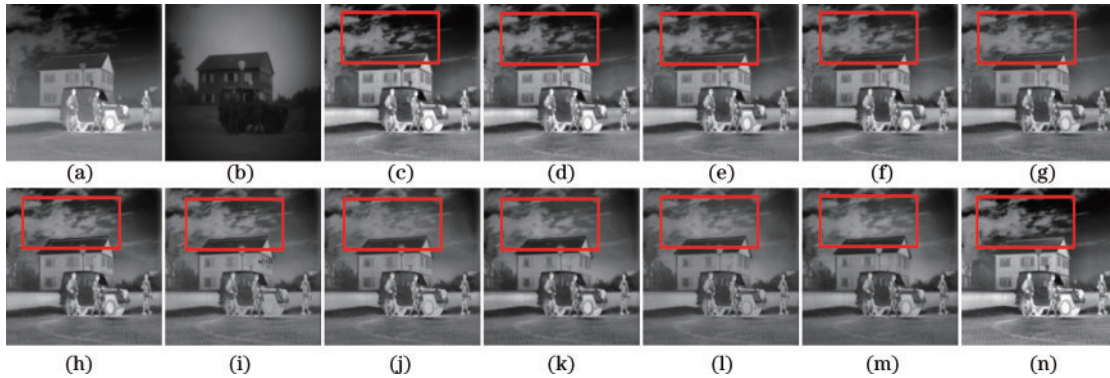


图 8 不同权重值在 TNO 数据集一对图像的融合结果。(a) IR; (b) VIS; (c)  $\omega=0$ ; (d)  $\omega=0.1$ ; (e)  $\omega=0.2$ ; (f)  $\omega=0.3$ ; (g)  $\omega=0.4$ ; (h)  $\omega=0.5$ ; (i)  $\omega=0.6$ ; (j)  $\omega=0.7$ ; (k)  $\omega=0.8$ ; (l)  $\omega=0.9$ ; (m)  $\omega=1.0$ ; (n) ours

Fig. 8 Fusion results of different weight values for a pair of images in the TNO dataset. (a) IR; (b) VIS; (c)  $\omega=0$ ; (d)  $\omega=0.1$ ; (e)  $\omega=0.2$ ; (f)  $\omega=0.3$ ; (g)  $\omega=0.4$ ; (h)  $\omega=0.5$ ; (i)  $\omega=0.6$ ; (j)  $\omega=0.7$ ; (k)  $\omega=0.8$ ; (l)  $\omega=0.9$ ; (m)  $\omega=1.0$ ; (n) ours

针对传统强度损失不同参数值  $\omega$  的客观评价指标如表 4 所示。由表 4 可知,所用的软加权强度在 6 个指标上为最优值,其余 2 个指标为次优值。当  $\omega=0$

时,指标 SD 和 EN 为最优值,表明此融合图像包含更多红外信息量,与上述采用软加权强度损失分析一致。

表 4 不同权重值在 TNO 数据集 43 对图像上的客观评价指标

Table 4 Objective evaluation metrics for different weighting values on 43 pairs of images from the TNO dataset

Method	SD	EN	DF	EI	AG	SF	Qp	MI
$\omega=0$	44.7291	7.2098	5.2815	50.0910	4.6891	10.0734	0.3338	14.3197
$\omega=0.1$	41.6797	7.1281	4.7107	44.7071	4.1901	9.0897	0.2809	14.2563
$\omega=0.2$	36.9613	7.0059	4.7370	43.4540	4.1182	8.9765	0.2711	14.0118
$\omega=0.3$	37.3071	7.0519	4.6598	43.6528	4.0920	8.8187	0.2121	14.1038
$\omega=0.4$	35.3388	6.9910	4.7573	43.8456	4.1299	8.9738	0.2201	13.9821
$\omega=0.5$	34.6621	6.9800	4.9332	44.9583	4.2583	9.2221	0.2058	13.9599
$\omega=0.6$	31.2020	6.8659	4.9018	43.3784	4.1468	9.0474	0.1835	13.7318
$\omega=0.7$	33.3493	6.9473	5.0526	45.3067	4.3051	9.3688	0.2018	13.8947
$\omega=0.8$	35.1397	7.0141	4.8359	43.7822	4.1602	8.9426	0.1962	14.0282
$\omega=0.9$	31.4273	6.8663	5.1649	45.8445	4.3726	9.5681	0.1898	13.7325
$\omega=1.0$	34.2250	6.9726	4.9425	43.7664	4.1684	9.1598	0.1991	13.9451
Ours	43.6229	7.1650	5.7459	53.2266	5.0586	10.7925	0.4995	14.3300

### 3.3.3 平衡参数 $\lambda$

网络训练时,参数  $\lambda$  用来平衡软加权强度损失与联合梯度损失在总损失中所占比例,避免某一损失值相对于另一损失值太小而对网络优化不起作用。实验中针对  $\lambda$  的不同设置来探寻软加权强度损失和联合梯度损失的最优比例。其客观评价指标值如表 5 所示。由表 5 可知,当  $\lambda=0.7$  时,所有指标均为最优值,说明此时软加权强度损失和联合梯度损失在网络训练中找到最佳权衡,得到细节丰富边缘平滑的融合图像。

### 3.4 应用场景及效率分析

#### 3.4.1 应用场景

红外与可见光图像融合技术应用场景丰富,如在军事领域进行目标识别和探测;医学领域用于疾病诊断和治疗;公共安全领域用于房屋安防监控。经上述实验表明,所提方法在融合红外与可见光真实生活场景上具有优势。

为验证所提方法在其他多模态图像如医学图像上的融合性能,选取 15 对 MRI-CT 医学图像进行融合实验。考虑到空间限制,选取一对图像进行主观显示如图 9 所示。由图 9 可知,GTF 方法保留了 CT 图像的

表 5 不同平衡参数值在 TNO 数据集 43 对图像上的客观评价指标

Table 5 Objective evaluation indicators for different balance parameter values on the TNO dataset 43 pairs of images

$\lambda$	SD	EN	DF	EI	AG	SF	Qp	MI
$\lambda = 0.1$	39.8836	6.8782	4.9477	50.3209	4.1579	9.1012	0.3946	13.7563
$\lambda = 0.2$	40.9696	6.9291	5.0301	52.4783	4.2433	9.3433	0.3980	13.8582
$\lambda = 0.3$	42.6061	6.9938	4.9322	51.9882	4.1768	10.1270	0.4042	13.9877
$\lambda = 0.4$	41.2148	7.0141	5.0890	52.2599	4.3055	9.4051	0.4031	14.0281
$\lambda = 0.5$	41.8657	7.0044	5.1368	53.0034	4.3732	9.4556	0.4057	14.0088
$\lambda = 0.6$	42.7849	6.9210	5.0168	50.3367	4.2424	10.2695	0.3902	13.8420
$\lambda = 0.7$	43.6229	7.1650	5.7459	53.2266	5.0586	10.7925	0.4995	14.3300
$\lambda = 0.8$	42.6738	6.9603	5.0696	49.1183	4.2991	10.3194	0.4150	13.9206
$\lambda = 0.9$	40.5754	6.9986	5.4492	47.2277	4.5437	9.9541	0.4081	13.9972
$\lambda = 1.0$	39.1473	7.0221	4.9707	49.8322	4.2551	9.2919	0.4082	14.0441

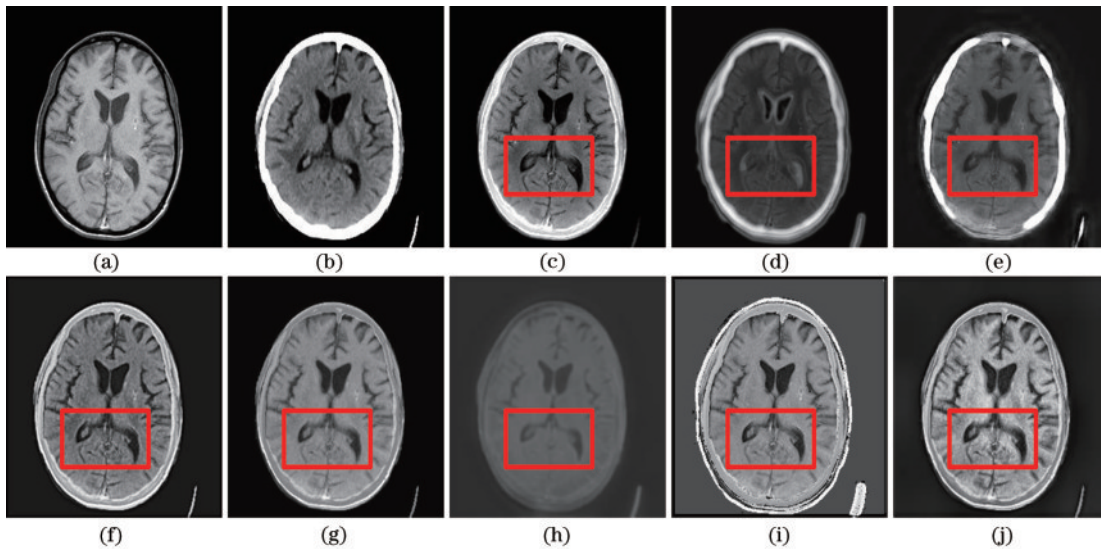


图 9 各方法在 MRI-CT 医学图像的融合结果。(a) MR-T1; (b) CT; (c) GTF; (d) GANMcC; (e) GAN-FM; (f) SDNet; (g) Densefuse; (h) DRF; (i) IFSepR; (j) ours

Fig. 9 Fusion results of each method in MRI-CT medical images. (a) MR-T1; (b) CT; (c) GTF; (d) GANMcC; (e) GAN-FM; (f) SDNet; (g) Densefuse; (h) DRF; (i) IFSepR; (j) ours

边缘信息,但 MR-T1 的边缘细节有所丢失; GANMcC、GAN-FM 和 DRF 融合结果视觉效果不佳,融合图像较为模糊,在一定程度上丢失了 MR-T1 中的亮度信息;SDNet、DenseFuse 和 IFSepR 方法获得的融合图像内部结构细节清晰,但对亮度上不如所提方

法,且 IFSepR 方法边缘似乎存在人工伪影。对比可得所提方法融合图像内容清晰且边缘信息保留更加完善,在多模态医学图像融合场景中也具备较好的效果。

各方法在 15 对医学图像上的客观评价平均值如表 6 所示,所提方法在 5 个指标上(SD、EN、DF、AG、

表 6 各方法在 15 对 MRI-CT 医学图像上的客观评价指标

Table 6 Objective evaluation index of each method on 15 pairs of MRI-CT medical images

Method	SD	EN	DF	EI	AG	SF	Qp	MI
GTF	60.6857	4.4535	6.6577	57.8738	5.6328	21.7096	0.0264	8.9071
GANMcC	48.0953	4.5493	3.9745	37.3691	3.5264	11.1650	0.0095	9.0986
GAN-FM	58.4541	5.8833	5.0845	43.8918	4.2130	18.6376	0.0186	11.7667
SDNet	62.4863	5.1234	7.9727	70.8844	6.8823	23.5017	0.0308	10.2468
Densefuse	63.0141	4.4407	5.4302	47.4444	4.5968	17.6804	0.0295	8.8815
DRF	23.5208	4.4353	0.9759	10.0041	0.9146	2.6619	0.0213	8.8707
IFSepR	54.7893	5.2338	8.7076	89.1554	7.0692	40.3196	0.0113	10.4677
Ours	63.3236	6.6326	8.7420	74.8356	7.3261	23.1368	0.0277	13.2653

MI)取得最优值,说明所提方法获得的融合图像具有清晰的细节和丰富的信息,客观评价与主观效果保持了一致性。从上述红外和可见光数据集及多模态医学图像

### 3.4.2 模型效率分析

选取参数数量和每秒传输帧数(FPS)两个指标进行模型效率分析。将所提方法与其他基于深度网络的图像融合方法进行效率对比,指标结果如表7所示。

表7 基于深度学习方法的模型效率对比

Table 7 Comparison of the efficiency of models based on deep learning methods

Parameter	GANMcC	GAN-FM	SDNet	DenseFuse	DRF	IFSepR	Ours
Parameter number / $10^6$	8.680	59.014	0.256	4.245	183.636s	1.960	8.893
FPS / (frame · s <sup>-1</sup> )	1.210	14.285	26.316	38.610	1.112	0.806	217

## 4 结 论

针对当前图像融合方法存在特征分离不充分且融合规则难以设计等问题,提出基于互信息特征分离表达的红外与可见光图像融合方法。所提方法利用互信息编码器实现私有和公有特征分离,并采用层级特征自适应融合网络来充分利用特征层级关系。此外,使用软加权强度损失进一步驱动网络学习源图像特征分布。通过公开数据集的对比和消融实验,验证了所提方法具有红外与可见光图像融合优势,融合图像质量有所提升,同时模型运行效率也取得较优结果,还可应用于多模态图像其他场景。后续工作中将考虑在提取图像显著特征的同时设计更加通用的融合规则。

### 参 考 文 献

- [1] 周晓玲, 江泽涛. 结合脉冲耦合神经网络与引导滤波的红外与可见光图像融合[J]. 光学学报, 2019, 39(11): 1110003.  
Zhou X L, Jiang Z T. Infrared and visible image fusion combining pulse-coupled neural network and guided filtering[J]. Acta Optica Sinica, 2019, 39(11): 1110003.
- [2] 梁佳明, 杨莘, 田立凡. 基于图像增强和滚动引导滤波的红外与可见光图像融合[J]. 激光与光电子学进展, 2023, 60(2): 0210006.  
Liang J M, Yang S, Tian L F. Infrared and visible image fusion based on image enhancement and rolling guidance filtering[J]. Laser & Optoelectronics Progress, 2023, 60(2): 0210006.
- [3] 卢鑫, 杨林, 李敏, 等. 基于Tikhonov正则化和细节重建的红外与可见光图像融合方法[J]. 光学学报, 2020, 40(2): 0210001.  
Lu X, Yang L, Li M, et al. Infrared and visible image fusion method based on Tikhonov regularization and detail reconstruction[J]. Acta Optica Sinica, 2020, 40(2): 0210001.
- [4] Jin H Y, Wang Y Y. A fusion method for visible and infrared images based on contrast pyramid with teaching learning based optimization[J]. Infrared Physics & Technology, 2014, 64: 134-142.
- [5] Li H, Wu X J, Kittler J. MDLatLRR: a novel decomposition method for infrared and visible image fusion[J]. IEEE Transactions on Image Processing, 2020: 4733-4746.
- [6] 杨飞燕, 王蒙. 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[J]. 激光与光电子学进展, 2023, 60(2): 0210008.  
Yang F Y, Wang M. Infrared and visible image fusion based on structure-texture decomposition and VGG deep networks[J]. Laser & Optoelectronics Progress, 2023, 60(2): 0210008.
- [7] 苗壮, 张湧, 李伟华. 基于双重对抗自编码网络的红外目标建模方法[J]. 光学学报, 2020, 40(11): 1111002.  
Miao Z, Zhang Y, Li W H. Modeling method of infrared target based on double countermeasure self-coding network[J]. Acta Optica Sinica, 2020, 40(11): 1111002.
- [8] Li H, Wu X J. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614-2623.
- [9] Li H, Wu X J, Durrani T. NestFuse: an infrared and visible image fusion architecture based on nest connection and spatial/channel attention models[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12): 9645-9656.
- [10] Jian L H, Yang X M, Liu Z, et al. SEDRFuse: a symmetric encoder-decoder with residual block network for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-15.
- [11] Luo X Q, Gao Y H, Wang A Q, et al. IFSepR: a general framework for image fusion based on separate representation learning[J]. IEEE Transactions on Multimedia, 2023, 25: 608-623.
- [12] Xu H, Wang X Y, Ma J Y. DRF: disentangled representation for visible and infrared image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 5006713.
- [13] Sanchez E H, Serrurier M, Ortner M. Learning disentangled representations via mutual information estimation[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12367: 205-221.

- [14] Liu S T, Huang D, Wang Y H. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-21) [2023-02-05]. <https://arxiv.org/abs/1911.09516>.
- [15] 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.
- [16] Ma J Y, Zhang H, Shao Z F, et al. GANMcC: a generative adversarial network with multiclassification constraints for infrared and visible image fusion[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5005014.
- [17] Zhang H, Yuan J T, Tian X, et al. GAN-FM: infrared and visible image fusion using GAN with full-scale skip connection and dual Markovian discriminators[J]. *IEEE Transactions on Computational Imaging*, 2021, 7: 1134-1147.
- [18] Zhang H, Ma J Y. SDNet: a versatile squeeze-and-decomposition network for real-time image fusion[J]. *International Journal of Computer Vision*, 2021, 129(10): 2761-2785.