

基于梯度方向直方图的红外与可见光融合网络的损失函数

龙雨馨, 赖文杰, 张怀元, 张鸿波, 李成世, 刘子骥*

电子科技大学光电科学与工程学院, 成都 四川 611731

摘要 基于深度学习的图像融合方法实现了良好的图像融合性能,近年来经过快速发展,被广泛应用于生物特征识别、自动驾驶和目标追踪等方面。深度学习网络在提取图像的重要纹理细节和保存重要信息等方面依然存在许多挑战。因此,提出了一种适用于红外与可见光图像融合网络的损失函数,在损失函数中引入了梯度方向直方图(HOG)损失,HOG特征可以反映图像局部的梯度方向和梯度大小,用HOG特征作损失函数可以提升网络提取图像细节信息的能力。将HOG损失与多尺度结构相似性损失相结合,用设计的损失函数训练了NestFuse、Res2Fusion和UNFusion 3个红外与可见光图像融合网络。在TNO数据集上,所提模型将融合图像的标准差(SD)分别提高2.1476%、1.2273%和1.4444%,将融合图像的视觉信息保真度(VIF)分别提高1.6529%、1.4936%和1.2902%;在RoadScene数据集上,所提模型将融合图像的SD分别提高1.0083%、1.1669%和0.7214%,将融合图像的VIF分别提高1.8093%、1.8063%和1.0406%。实验结果表明,所提损失函数可以从源图像中提取更多有效信息。

关键词 图像融合; 红外图像; 可见光图像; 梯度方向直方图; 损失函数; 深度学习

中图分类号 TP391.41

文献标志码 A

DOI: 10.3788/LOP230882

Soft Histogram of Gradients Loss: A loss Function for Optimization of the Image Fusion Networks

Long Yuxin, Lai Wenjie, Zhang Huaiyuan, Zhang Hongbo, Li Chengshi, Liu Ziji*

College of Optoelectronic Science and Engineering, University of Electronic Science and Technology, Chengdu 611731, Sichuan, China

Abstract Image fusion methods based on deep learning have achieved excellent image fusion performance and have been widely used in biometric recognition, automatic driving and target tracking. However, it is still challenging to extract important texture details and preserve information of images. Therefore, a loss function for infrared and visible image fusion networks is presented. We employ the histogram of oriented gradient (HOG) to calculate the loss function. HOG feature can reflect the direction and magnitude of local gradient in the image, and using HOG feature as the loss function can improve the ability of the network to extract image details. We combine HOG loss with multi-scale structural similarity loss, and train NestFuse, Res2Fusion and UNFusion infrared and visible image fusion networks with the designed loss function. On the TNO dataset, our model increases the standard deviation (SD) of fused images by 2.1476%, 1.2273% and 1.4444% respectively, and increases the visual information fidelity (VIF) of fused images by 1.6529%, 1.4936% and 1.2902% respectively. On the RoadScene dataset, our model increases the SD of the fused images by 1.0083%, 1.1669% and 0.7214% respectively, and increases the VIF of the fused images by 1.8093%, 1.8063% and 1.0406% respectively. The experimental results show that the proposed loss function can extract more effective information from the source image.

Key words image fusion; infrared image; visible light image; histogram of oriented gradient; loss function; deep learning

1 引言

可见光图像传感器可以记录物体的纹理信息,但

是容易受到雨、雾、光照变化等自然因素和遮挡伪装等人为因素的影响导致目标物体信息丢失。红外图像传感器可以感知视野中目标物体的热辐射信息,并且对

收稿日期: 2023-03-16; 修回日期: 2023-03-26; 录用日期: 2023-04-04; 网络首发日期: 2023-05-09

基金项目: 科技创新创业苗子工程(2023JDRC0084)

通信作者: *zjliu@uestc.edu.cn

于上述影响具有一定的鲁棒性,但是红外图像存在目标边缘纹理和背景信息缺乏、对比度低等问题。将可见光传感器和红外图像传感器所拍摄的图像进行融合,可以实现更为精准且不易受环境变化影响的目标检测、语义分割和夜视监控等。

近年来,随着计算机视觉的快速发展,基于深度学习的图像融合研究被大量发表。Prabhakar 等^[1]设计了第一个无监督的图像融合卷积神经网络(CNN)——DeepFuse;Li 等^[2]提出了基于 CNN 层和 DenseBlock 的融合网络 DenseFuse,通过在编码器中增加 DenseBlock 改善 DeepFuse 中间层信息丢失的问题;Li 等^[3]将注意力模型引入融合网络中,实现了更好的融合性能;Ma 等^[4]将生成对抗网络(GAN)引入红外与可见光融合网络设计了 FusionGAN;Cheng 等^[5]提出的 UNIFusion(unified image fusion)网络模型通过引入 Ghost 模块减少了网络的参数量,降低了网络的复杂度;Xu 等^[6]提出的 PSTLFusion(parallel scene and texture learning fusion)网络模型首次在红外与可见光融合领域引入循环神经网络(RNN)来进行纹理特征的提取。上述的融合网络大多都取得了较好的结果,但是也存在以下问题:由于缺乏融合图像的真值(ground truth),许多融合网络都以无监督的方式训练;在设计损失函数时,多数网络仅考虑结构相似性(SSIM)损失、图像梯度图损失和图像强度损失,融合结果会存在纹理信息缺乏等问题。

为了提升红外与可见光融合网络的特征提取与图像重建的能力,使融合图像有更多来自源图像的细节,本文将梯度方向直方图(HOG)作为损失函数引入红外与可见光融合网络中,用 parzen-window 方法使 HOG 特征可导从而实现 HOG 损失的反向传播。同时,结合多尺度结构相似性(MS-SSIM)损失,重新训练了 NestFuse^[3]、Res2Fusion^[7]和 UNIFusion(unified multi-scale densely connected fusion)^[8]3 个红外与可见光融合网络。在 TNO 和 RoadScene 数据集上的结果表明,设计的损失函数与网络原来的损失函数相比,在视觉上能使网络的融合图像细节更清晰,在数据上图像熵(EN)、标准差(SD)、互信息(MI)、差异相关性总和(SCD)和视觉信息保真度(VIF)几个评估指标均有提升。

2 理论基础

2.1 图像融合网络的损失函数

红外与可见光图像融合网络的损失函数,除 GAN 中固有的生成对抗损失外,比较通用的有 3 类:SSIM 损失、图像梯度图损失和图像强度损失,大部分融合网络中的损失函数都是其中 2 种或 3 种的组合。网络模型 DeepFuse^[1]、文献[9]中模型使用了多曝光融合结构相似性(MEF-SSIM)图像质量评估指标作为损失函数,实现了比图像强度的 l_1 和 l_2 等损失函数更好的性能^[1];网络模型 TC-GAN(texture condition GAN)^[10]利用梯度损失

使生成的图像具有更多的细节;网络模型 DenseFuse^[2]、MFEIF(multi-scale feature ensemble and an edge-attention guidance for image fusion)^[11]、NestFuse^[3]、Res2Fusion^[7]、SEDRFuse(symmetric encoder-decoder with residual block network)^[12]、UFADFuse(unity fusion attention model for multifocus image fusion)^[13]、UNFusion^[8]以及文献[14]、文献[15]的设计都是将 SSIM 损失与图像强度损失结合, Ma 等^[16]的研究说明这两种损失函数结合比单使用 SSIM 损失生成的图像细节更丰富;网络模型 FusionGAN^[4]、AMFNet(attention-guided generative adversarial network)^[17]、GAN-FM(GAN with full-scale skip connection and dual Markovian discriminators)^[18]、ERNet(extraction-and-reconstruction network)^[19]、STDFusionNet(salient target detection fusion network)^[20]和 GIDGAN(gradient and intensity discriminator GAN)^[21]中包括了图像的强度损失和图像的梯度损失;文献[22]和网络模型 IPLF(image pair learning fusion network)^[23]中利用 SSIM 损失和图像梯度来训练网络;文献[24]和网络模型 VIFGAN(GAN-based visible and infrared image fusion)^[25]、MGANFuse(multi-source image fusion based modified GAN)^[26]、UNIFusion^[5]中的损失函数使用了 3 种损失函数的组合。

一些网络模型也会根据网络结构或实际需求设计自己的损失函数。Xu 等^[27]设计的网络模型 FusionDN 和 U2Fusion^[28]中,为了使其能够满足不同的融合任务,需要进行持续学习,设计的连续学习损失(L_{ewc})可以避免学习新任务时网络参数产生的灾难性遗忘问题。Xu 等^[29]设计的 DRF(disentangled representation for visible and infrared image fusion)网络针对分别用场景编码器(scene encoder)和属性编码器(attribute encoder)提取到的场景信息、传感器模态信息和重建图像设计了场景特征一致性损失(scene feature consistency loss)、属性分布损失(attribute distribution loss)、自重建损失(self-reconstruction loss)和域变换损失(domain-translation loss),独特的源图像信息分解方式可以缓解不同类的源图像信息提取不当的问题。DenseNet^[30]中除了梯度损失和加权的强度损失外,还使用了正则化损失(orthogonalization loss)对网络的权重进行自动正则化。DIDFuse(deep image decomposition fusion network)^[31]的编码器网络分别提取了红外和可见光图像的背景和细节特征,为了使红外和可见光的背景特征差距小、细节特征差距大,网络中设计了图像分解损失函数。DIF-Net(deep image fusion network)^[32]利用图像的结构张量建立了损失函数。DRSNFuse(deep residual shrinkage fusion network)^[33]在解码器中加入了伪纹理损失来抑制梯度损失中产生的伪纹理。Yang 等^[34]设计的基于 DICFB(dualchannel information cross fusion block)的融合网络利用显著性损失来突出红外和可见光图像的显著性信息。MDFN(mask deep

fusion network)^[35]生成了红外图像的 mask,在计算内容损失时利用 mask 与源图像和融合图像的乘积来保存有价值的信息。RFN(residual fusion network)-Nest^[36]在训练中间的融合层时,为了保留更多的显著特征设计了特征增强损失函数。DRF^[29]、DIDFuse^[31]和 MDFN^[35]等网络使用了针对其特殊的网络结构或者数据设计的专用损失函数,所以它们的损失函数不具备通用性。 L_{ewc} ^[27-28]、正则化损失^[30]等不是以图像融合为目的而设计的。DIF-Net^[32]、DICFB^[34]、RFN-Nest^[36]等网络中设计的损失函数更适用于利用红外与可见光图像训练的网络,但目前配准的红外与可见光数据集中的图片数量较少,训练前需要做数据增强处理。DRSNFuse^[33]中的伪纹理损失是针对网络中的梯度损失引入的伪纹理而设计的。本文所设计的损失函数在红外与可见光图像融合网络上有通用性,而且能使网络在图像融合中提取更多细节信息,使融合图像在视觉上的表现更好。

2.2 HOG

HOG 将图像分成大小相同的块,在块中统计梯度大小以及梯度方向作为 HOG 特征。在计算机视觉领域,HOG 一般用于目标检测。Dalal 和 Triggs^[37]首次利用 HOG 特征来进行行人检测;Carcagni 等^[38]利用 HOG 来识别面部表情特征;Sugiarto 等^[39]通过 HOG 识别木材种类。在其他方面,Wei 等^[40]将 HOG 特征作为目标特征用于掩蔽特征预测(MaskFeat),也得到了很好的效果。

也有利用 HOG 特征来实现图像融合的方法:Zong 等^[41]用滑动窗口把图像分成许多个块,利用 HOG 特征将这些块分类后进行字典学习等操作;在 FILM. LrTL(fusing multifocus images using low-rank transform learning)方法中,Maggu 等^[42]也利用 HOG 特征对获得的源图像的块分类。

本文利用 HOG 作为损失函数来训练红外与可见光融合网络。HOG 不可导,无法实现梯度的反向传播,因此用高斯函数来建模 HOG 中梯度幅值落在梯度角度中的概率,将离散的 HOG 变得可导。

3 损失函数设计

3.1 图像融合网络的损失函数

所提损失函数 L_{total} 表示为

$$L_{total} = \lambda_1 L_{MS-SSIM} + \lambda_2 L_{HOG}, \quad (1)$$

式中: $L_{MS-SSIM}$ 为 MS-SSIM 损失; L_{HOG} 为 HOG 损失; λ_1 和 λ_2 为控制权衡的参数。

3.2 结构相似性损失

$L_{MS-SSIM}$ ^[43] 表示为

$$L_{MS-SSIM} = 1 - O_{MS-SSIM}(O, I), \quad (2)$$

式中: O 和 I 分别为输出图像和源图像; $O_{MS-SSIM}(\cdot, \cdot)$ 表示 MS-SSIM 运算。

MS-SSIM 运算是多尺度版本的 SSIM 运算,将图

像通过低通滤波下采样 M 次后计算其在第 i 个尺度上的对比度相似度 $c_i^{\beta_i}(O, I)$ 、结构相似度 $s_i^{\gamma_i}(O, I)$ 和第 M 个尺度上的亮度相似度 $l_M^{\alpha_M}(O, I)$,并将这些值相乘。 $O_{MS-SSIM}(\cdot, \cdot)$ 表示为

$$O_{MS-SSIM}(\cdot, \cdot) = [l_M^{\alpha_M}(O, I)] \cdot \prod_{i=1}^M \{ [c_i^{\beta_i}(O, I)] \cdot [s_i^{\gamma_i}(O, I)] \}, \quad (3)$$

式中:参数 α_M 、 β_i 和 γ_i 分别用于调整亮度相似度、对比度相似度和结构相似度的相对重要性,它们的值与 HVS(human visual system)灵敏度有关,可由文献[43]中的视觉实验得出; l 、 c 和 s 分别为输出图像 O 和源图像 I 在亮度、对比度和结构方面的相似性,表示为

$$l(O, I) = \frac{2\mu_O\mu_I + c_1}{\mu_O^2 + \mu_I^2 + c_1}, \quad (4)$$

$$c(O, I) = \frac{2\sigma_O\sigma_I + c_2}{\sigma_O^2 + \sigma_I^2 + c_2}, \quad (5)$$

$$s(O, I) = \frac{\sigma_{OI} + c_3}{\sigma_O\sigma_I + c_3}, \quad (6)$$

式中: μ_O 和 μ_I 分别为输出图像 O 和源图像 I 的均值; σ_O 和 σ_I 分别为 O 和 I 的方差; σ_{OI} 为 O 和 I 的协方差; c_1 、 c_2 和 c_3 为常数。

3.3 HOG 损失

HOG 将图像划分成同样大小的小方块,并统计每个小方块里的梯度大小与梯度方向。首先计算图像在水平和垂直方向的梯度值 g_x 和 g_y :

$$g_x(x, y) = f(x+1, y) - f(x-1, y), \quad (7)$$

$$g_y(x, y) = f(x, y+1) - f(x, y-1), \quad (8)$$

式中, x 和 y 为像素点的横、纵坐标。然后计算梯度的幅度 g_{mag} 和角度 g_{ang} :

$$g_{mag}(x, y) = \sqrt{[g_x(x, y)]^2 + [g_y(x, y)]^2}, \quad (9)$$

$$g_{ang}(x, y) = \arctan[g_y(x, y)/g_x(x, y)]. \quad (10)$$

HOG 的统计是将梯度幅度图和梯度角度图分别划分成 $N \times N$ (本实验设置为 8×8) 的小单元(称为 cell),梯度方向将 $0^\circ \sim 180^\circ$ 均匀分成数量为 B (本实验设置为 9) 的角度范围(称为 bin),在每个单元内统计像素点的梯度幅度分别在这些角度范围中的占比,得到每个单元内的 HOG。在这样的 HOG 中,梯度幅度落在这些角度范围中的概率是一个 δ 函数,表示为

$$\delta(x, y) = \begin{cases} 1, & g_{ang}(x, y) \in b_{bin}(i) \\ 0, & g_{ang}(x, y) \notin b_{bin}(i) \end{cases}, \quad i = 1, 2, \dots, B, \quad (11)$$

式中: $b_{bin}(i)$ 为第 i 个角度范围;单元数为 a 、角度范围数为 b 的 HOG 就是 δ 函数与梯度幅值相乘后求和,表示为

$$H_{HOG}(a, b) = \sum_{(x, y) \in a} g_{mag}(x, y) \delta(x, y), \quad (12)$$

所以 HOG 离散且不可微,不能满足损失函数反向传播的条件。

parzen-window^[44]可以将离散分布连续化,在文献[45-46]中用于对互信息的连续联合HOG进行估计。利用parzen-window方法对离散的HOG特征进行连续化处理。将每个单元中的梯度方向分为宽度为180/B(记为S)的角度范围,数量为B。利用高斯函数对δ函数进行近似,计算单元中的梯度角度 $g_{ang}(x,y)$ 对每个角度范围的贡献:

$$W_i(x,y) = \frac{1}{S\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \times \left\{ \frac{[g_{ang}(x,y) - n_i]}{S} \right\}^2 \right\}, \quad (13)$$

式中, n_i 为第*i*个角度范围的中心角度。计算梯度角度 $g_{ang}(x,y)$ 落在第*i*个角度范围中的概率:

$$P_i(x,y) = W_i(x,y) / \left[\sum_{j=1}^B W_j(x,y) \right]. \quad (14)$$

所以第*a*个单元中第*b*个角度范围的HOG可表示为

$$H_{HOG}(a,b) = \sum_{(x,y) \in a} g_{mag}(x,y) P_b(x,y), \quad (15)$$

得到其对应的HOG的连续HOG估计:

$$L_{HOG} = \frac{1}{M} \| H_{HOG}(O) - H_{HOG}(I) \|_1, \quad (16)$$

式中, M 是图像中角度范围的总数。

4 实验分析

为验证所设计的损失函数是否能有效提高网络的性能,用其分别训练 NestFuse^[3]、Res2Fusion^[7]和 UNFusion^[8] 3个网络的模型,3个网络的结构如图1所示。这3个网络的训练集都是MS-COCO(microsoft common objects in context)^[47]可见光数据集,训练后与网络的原模型在相同的数据集进行测试并对比。测试

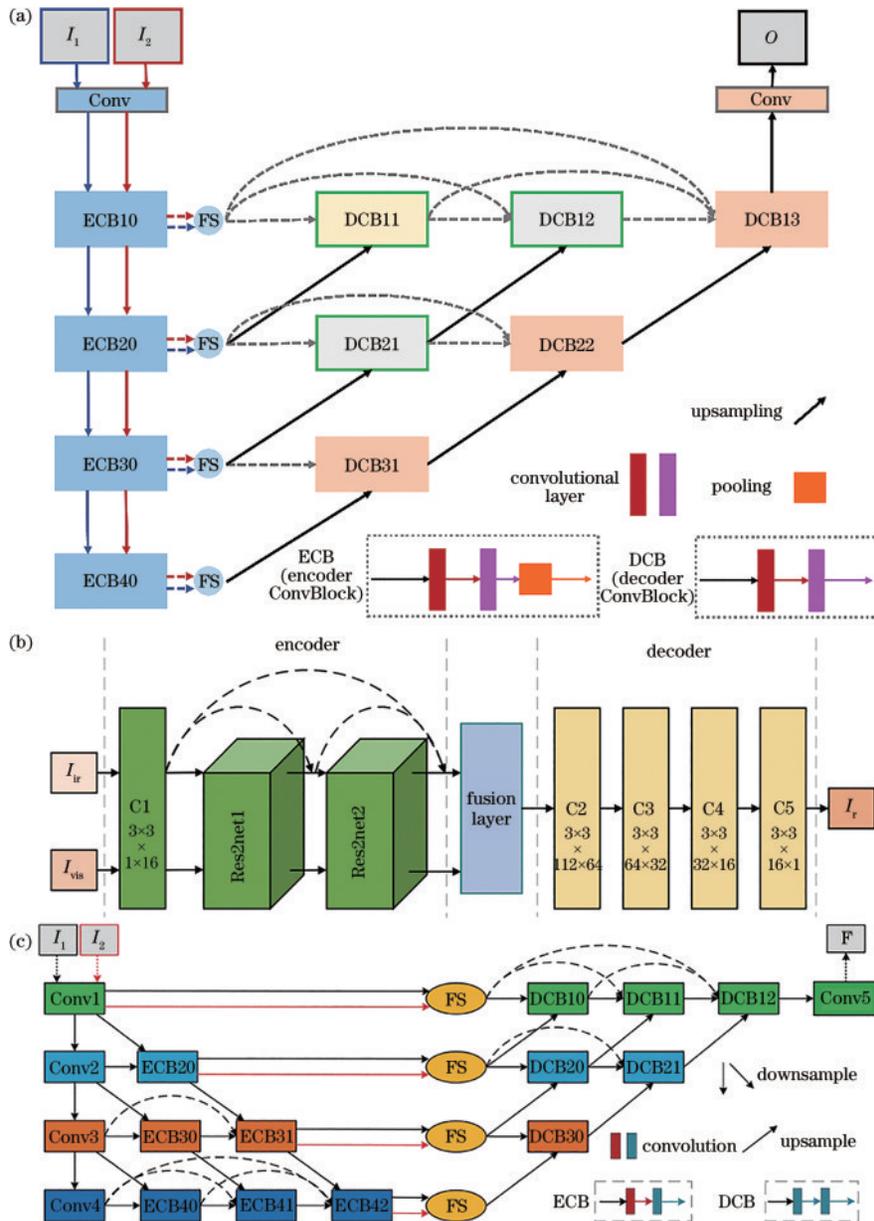


图1 训练的3个网络的结构。(a) NestFuse的结构^[3]; (b) Res2Fusion的结构^[7]; (c) UNFusion的结构^[8]

Fig. 1 Structure of three training networks. (a) Structure of NestFuse^[3]; (b) structure of Res2Fusion^[7]; (c) structure of UNFusion^[8]

实验在 TNO^[48]和 RoadScene^[27]数据集上进行。TNO 数据集包含不同场景的红外与可见光图像,从其中选取 21 对红外与可见光图像测试。RoadScene 数据集中有 221 对配准的红外与可见光图像,包含道路、车辆和行人等美国菲利尔公司(FLIR)视频中极具代表性的丰富的场景。

在实验中用所提损失函数替换网络原来的损失函数(原模型的损失函数见表 1),NestFuse^[3]和 Res2Fusion^[7]的原损失函数都是 MS-SSIM 和输出图像与源图像强度的 MSE 损失(mean-squared loss)。UNFusion^[8]的强度损失用的是最小绝对偏差损失(l_1 -norm),Zhao 等^[49]的研究结果表明在深度学习图像处理时 MS-SSIM 与 l_1 -norm 相结合的损失函数更加有效。

表 1 网络的原损失函数

Table 1 Original loss function of the network

Network	Image intensity loss	Structural similarity loss
NestFuse	MSE	MS-SSIM
Res2Fusion	MSE	MS-SSIM

实验所使用的计算机配置为 CPU Intel(R) Core (TM) i7-10700KF @ 3.80 GHz、内存大小 32 GB、GPU NVIDIA Geforce RTX3070,3 个网络均在 PyTorch 上实现。

在实验中损失函数中的参数 λ_1 除 UNFusion^[8] 设置为 100 外,均与原模型的 MS-SSIM 损失的最佳参数设置相同, λ_2 均设置为 1/9。Batch size 设置均与原模型相同。在训练次数上:UNFusion^[8] 训练了 8000 次;

NestFuse^[3] 训练了 10000 次;其他网络均与原模型训练相同次数。选择网络中结果较好的融合策略进行融合,NestFuse^[3] 的通道注意力模型选择平均算子进行池化,UNFusion^[8] 选择在 TNO 和 RoadScene 两个数据集上表现较好的 l_2 归一化注意力模型。

为了证明所提损失函数提升网络性能的能力,选择了 EN^[50]、SD^[51]、MI^[52]、SCD^[53]、MS-SSIM^[43]、离散余弦特征互信息 FMI_dct、小波特征互信息 FMI_w^[54]、基于梯度的融合性能 Q_{abf} ^[55] 和 VIF^[56] 9 种评估指标来评估融合结果的性能。其中:EN、MI、FMI_dct 和 FMI_w 是基于信息论的评估指标,可以评估图像的信息量及参考图像与源图像的相似性;SD 可以反映图像强度的分布,是基于图像特征的评估指标;SCD 和 Q_{abf} 都是基于源图像与生成图像的评估指标;MS-SSIM 是结构相似性的评估指标;VIF 是基于视觉感知的评估指标。

4.1 在 TNO 数据集上的结果

将替换损失函数后的模型与原模型测试 21 对 TNO 数据集中的图像进行对比,结果如图 2 所示。将 21 张图像的客观评价的平均值进行对比,结果如表 2 所示,将所提损失函数应用于 3 个网络,EN、SD、MI、SCD、VIF 和 Q_{abf} 几个评估指标均有提升,在 SD 和 VIF 上的提升较为明显,NestFuse^[3] 平均提升了 2.1476% 和 1.6529%,Res2Fusion^[7] 平均提升了 1.2273% 和 1.4936%,UNFusion^[8] 平均提升了 1.4444% 和 1.2902%。这些数据说明了所提损失函数训练的网络保留了源图像更多的信息,图像的失真也更小。从总体看,所提损失函数对于所测试的 3 个网络的评估指标均有提升。

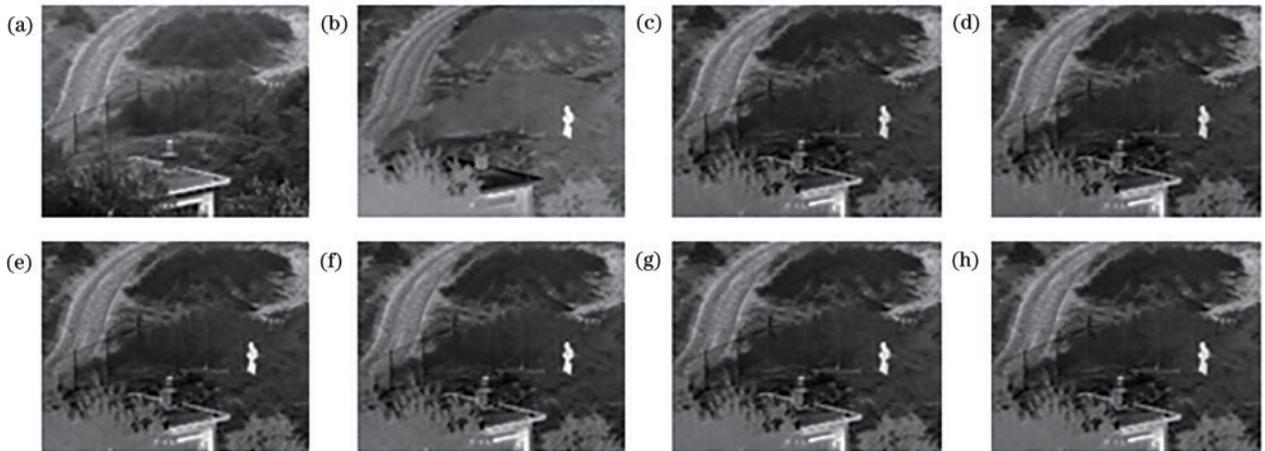


图 2 在 TNO 数据集上使用不同网络融合图像对比。(a)(b)源图像;(c) NestFuse;(d) NestFuse-HOG;(e) Res2Fusion;(f) Res2Fusion-HOG;(g) UNFusion;(h) UNFusion-HOG

Fig. 2 Image comparison fused by different networks on the TNO dataset. (a) (b) Source images; (c) NestFuse; (d) NestFuse-HOG; (e) Res2Fusion; (f) Res2Fusion-HOG; (g) UNFusion; (h) UNFusion-HOG

4.2 在 RoadScene 数据集上的结果

在 RoadScene 数据集上对比所提损失函数与其他损失函数的性能。结果如图 3 和图 4 所示,特别是在 NestFuse^[3] 网络上,利用所提损失函数训练网络,能够

提升模型在细节上的清晰度和对比度。

定量测试了这 221 张图的评估指标,结果如表 3 所示,所提损失函数对网络的提升同样体现在融合图像的 EN、SD、MI、SCD 和 VIF 指标的提升,对于 SD 和

表 2 在 TNO 数据集上的定量比较

Table 2 Quantitative comparison on the TNO dataset

Network	EN	SD	MI	SCD	MS-SSIM	FMI_dct	FMI_w	Q_{abf}	VIF
NestFuse	6.9198	82.7523	13.8397	1.7335	0.8625	0.3580	0.4371	0.4870	0.7865
NestFuse-HOG	6.9435	84.5295	13.8870	1.7491	0.8684	0.3426	0.4323	0.4892	0.7995
Res2Fusion	6.7963	77.6068	13.5927	1.7365	0.8681	0.3745	0.4373	0.4760	0.7164
Res2Fusion-HOG	6.8087	78.5593	13.6173	1.7398	0.8682	0.3721	0.4362	0.4760	0.7271
UNFusion	6.9070	83.6454	13.8139	1.7097	0.8505	0.3506	0.4380	0.4940	0.8061
UNFusion-HOG	6.9210	84.8536	13.8419	1.7131	0.8500	0.3486	0.4412	0.4964	0.8165



图 3 在 RoadScene 数据集上使用不同网络融合图像对比。(a)(b)源图像;(c) NestFuse;(d) NestFuse-HOG;(e) Res2Fusion;(f) Res2Fusion-HOG;(g) UNFusion;(h) UNFusion-HOG

Fig. 3 Image comparison fused by different networks on the RoadScene dataset. (a) (b) Source images; (c) NestFuse; (d) NestFuse-HOG; (e) Res2Fusion; (f) Res2Fusion-HOG; (g) UNFusion; (h) UNFusion-HOG

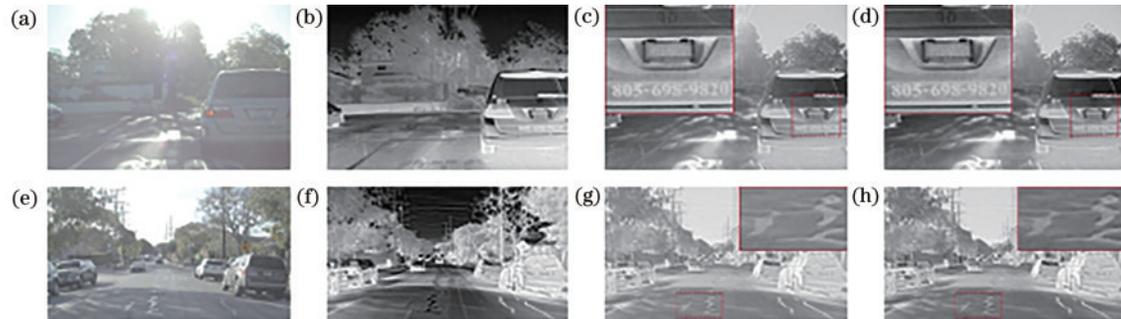


图 4 损失函数应用前后 NestFuse 网络融合图像对比。(a)(b)源图像 1;(c)源图像 1 经 NestFuse 网络融合后的图像;(d)源图像 1 经 NestFuse-HOG 网络融合后的图像;(e)(f)源图像 2;(g)源图像 2 经 NestFuse 网络融合后的图像;(h)源图像 2 经 NestFuse-HOG 网络融合后的图像

Fig. 4 Fusion image comparison before and after loss function applied in NestFuse network. (a) (b) Source image 1; (c) image of source image 1 fused by NestFuse network; (d) image of source image 1 fused by NestFuse-HOG network; (e) (f) source image 2; (g) image of source image 2 fused by NestFuse network; (h) image of source image 2 fused by NestFuse-HOG network

表 3 在 RoadScene 数据集上的定量比较

Table 3 Quantitative comparison on the RoadScene dataset

Network	EN	SD	MI	SCD	MS-SSIM	FMI_dct	FMI_w	Q_{abf}	VIF
NestFuse	7.4418	79.6857	14.8835	1.6504	0.8555	0.3446	0.4395	0.5077	0.9009
NestFuse-HOG	7.4604	80.4892	14.9209	1.6688	0.8597	0.3230	0.4332	0.5062	0.9172
Res2Fusion	7.3300	73.6833	14.6599	1.6476	0.8630	0.3697	0.4326	0.5083	0.7806
Res2Fusion-HOG	7.3469	74.5431	14.6938	1.6526	0.8616	0.3665	0.4318	0.5099	0.7947
UNFusion	7.4137	79.4455	14.8275	1.6355	0.8442	0.3386	0.4361	0.5102	0.9033
UNFusion-HOG	7.4231	80.0186	14.8462	1.6369	0.8416	0.3392	0.4406	0.5119	0.9127

VIF, NestFuse^[3] 平均提升了 1.0083% 和 1.8093%, Res2Fusion^[7] 平均提升了 1.1669% 和 1.8063%, UNFusion^[8] 平均提升了 0.7214% 和 1.0406%。 Q_{abf} 总体来说比原来的好。从 RoadScene 数据集的测试结果来看,所提损失函数对网络提取并保留图像细节信息的能力有提升,融合图像有更多来自源图像的信息。

4.3 对比实验

为了证明所提损失函数的优越性,将其中的 L_{HOG} 替换成梯度图 MSE Loss(L_{grad}) 训练网络。此外,令网络原损失函数($L_{NestFuse}$)分别加上 L_{grad} 和基于已训练

好的 VGG (visual geometry group)-19 网络计算的感知损失^[57]。对比实验用 NestFuse^[3] 网络进行,总损失为

$$L_{NestFuse-grad} = 100L_{MS-SSIM} + L_{grad}, \quad (17)$$

$$L_{NestFuse+grad} = L_{NestFuse} + L_{grad}, \quad (18)$$

$$L_{NestFuse-perceptual} = L_{NestFuse} + \sum_{i=1}^4 M_{MSE}[\phi_i(O), \phi_i(I)], \quad (19)$$

式中, ϕ_i 为通过 VGG-19 网络得到的第 i 层的特征。

在 TNO 数据集上的对比实验结果如图 5 所示。在 TNO 数据集上的定量测试结果如表 4 所示,在测试的 9 种评估指标中,所提损失函数取得了 7 种指标的最优值。

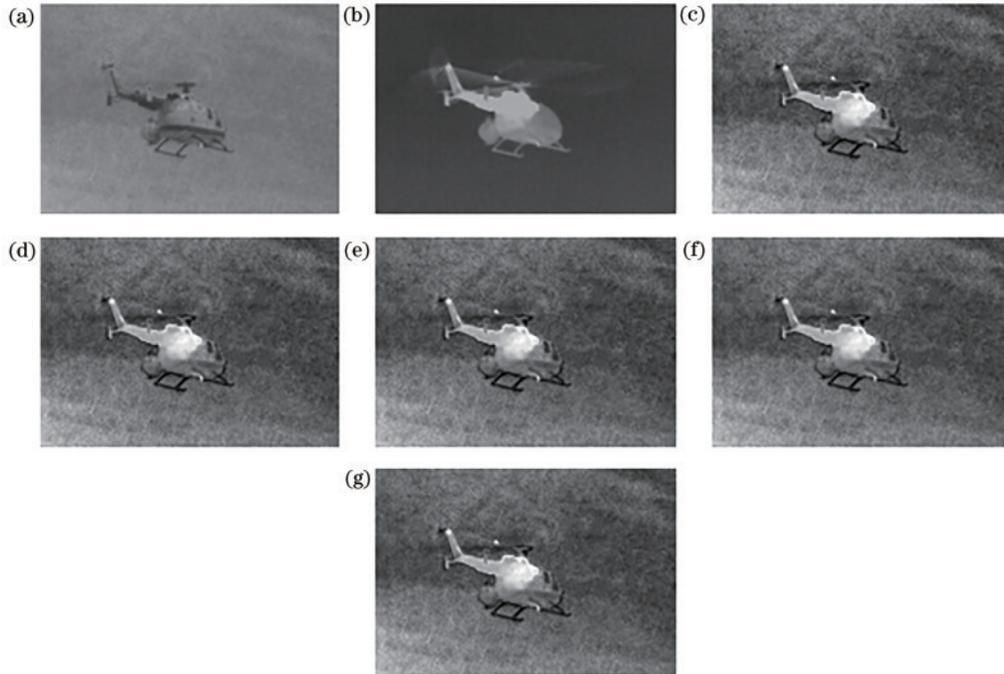


图 5 在 TNO 数据集上使用不同网络融合图像对比。(a)(b)源图像;(c) NestFuse;(d) NestFuse-grad;(e) NestFuse+grad;(f) NestFuse-perceptual;(g) NestFuse-HOG

Fig. 5 Image comparison fused by different networks on the TNO dataset. (a) (b) Source images; (c) NestFuse; (d) NestFuse-grad; (e) NestFuse+grad; (f) NestFuse-perceptual; (g) NestFuse-HOG

表 4 在 TNO 数据集上的定量比较

Table 4 Quantitative comparison on the TNO dataset

Network	EN	SD	MI	SCD	MS-SSIM	FMI_dct	FMI_w	Q_{abf}	VIF
NestFuse	6.9198	82.7523	13.8397	1.7335	0.8625	0.3580	0.4371	0.4870	0.7865
NestFuse-grad	6.9091	82.2634	13.8181	1.7390	0.8652	0.3443	0.4337	0.4868	0.7781
NestFuse+grad	6.9297	83.5418	13.8593	1.7433	0.8665	0.3489	0.4347	0.4851	0.7928
NestFuse-perceptual	6.8998	81.2038	13.7995	1.7347	0.8630	0.3500	0.4353	0.4878	0.7684
NestFuse-HOG	6.9435	84.5295	13.8870	1.7491	0.8684	0.3426	0.4323	0.4892	0.7995

在 RoadScene 数据集上的对比实验结果如图 6 所示, L_{grad} 训练的网络在对比度上的表现优于其他损失函数训练的网络。通过定量对比(表 5),所提损失函

数取得了 6 种评估指标的最优值。总的来说在 RoadScene 数据集上, L_{HOG} 训练的网络的性能优于测试的其他损失函数训练的网络。

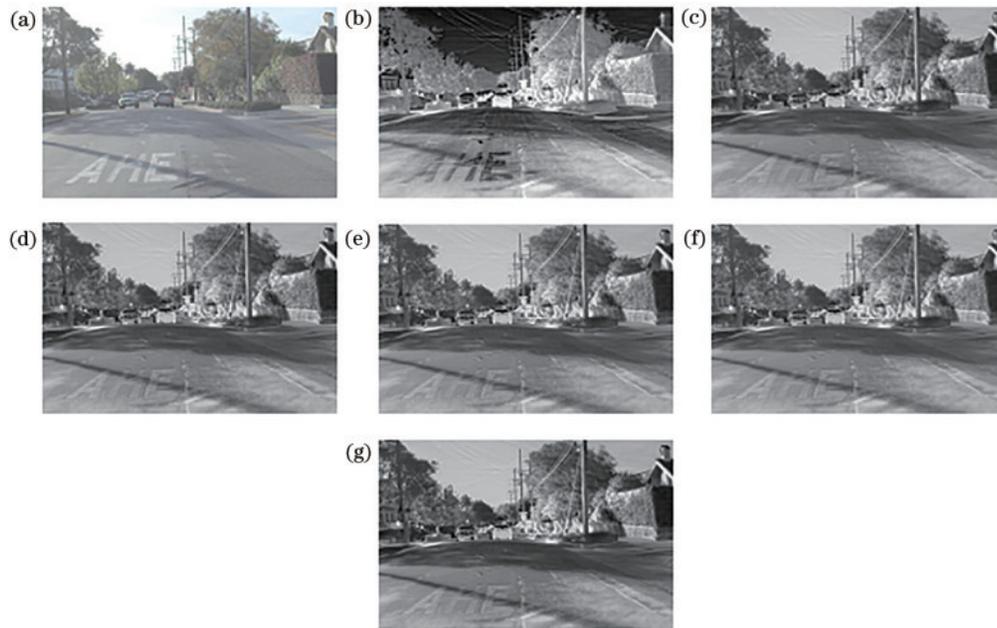


图 6 在 RoadScene 数据集上使用不同网络融合图像对比。(a)(b)源图像;(c) NestFuse;(d) NestFuse-grad;(e) NestFuse+grad;(f) NestFuse-perceptual;(g) NestFuse-HOG

Fig. 6 Image comparison fused by different networks on the RoadScene dataset. (a) (b) Source images; (c) NestFuse; (d) NestFuse-grad; (e) NestFuse+grad; (f) NestFuse-perceptual; (g) NestFuse-HOG

表 5 在 RoadScene 数据集上的定量比较

Table 5 Quantitative comparison on the RoadScene dataset

Network	EN	SD	MI	SCD	MS-SSIM	FMI _{dct}	FMI _w	Q _{abf}	VIF
NestFuse	7.4418	79.6857	14.8835	1.6504	0.8555	0.3446	0.4395	0.5077	0.9009
NestFuse-grad	7.4475	79.8023	14.8950	1.6603	0.8585	0.3270	0.4353	0.5061	0.9032
NestFuse+grad	7.4394	79.3834	14.8789	1.6566	0.8577	0.3345	0.4366	0.5056	0.8977
NestFuse-perceptual	7.4368	79.3249	14.8735	1.6522	0.8559	0.3346	0.4375	0.5076	0.8947
NestFuse-HOG	7.4604	80.4892	14.9209	1.6688	0.8597	0.3230	0.4332	0.5062	0.9172

5 结 论

针对红外与可见光图像融合网络的融合图像细节纹理保留不全等问题,提出了一种基于 HOG 的红外与可见光融合网络的损失函数,利用 parzen-window 对图像的 HOG 特征做了可微化,结合 MS-SSIM 损失使网络提取的图像信息更全面。利用这个损失函数训练 NestFuse、Res2Fusion 和 UNFusion 3 个图像融合的网络模型,实验结果表明,所提损失函数能提取更丰富的细节信息,融合图像对比度有明显提升。此外,也将 HOG 损失与其他损失函数进行了对比实验,对比实验的结果表明 HOG 损失在实验的网络上优于测试的其他损失函数。

参 考 文 献

- [1] Prabhakar K R, Srikar V S, Babu R V. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4724-4732.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Cheng C Y, Wu X J, Xu T Y, et al. UNIFusion: a lightweight unified image fusion network[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 5016614.
- [6] Xu M L, Tang L F, Zhang H, et al. Infrared and visible image fusion via parallel scene and texture learning[J]. Pattern Recognition, 2022, 132: 108929.
- [7] Wang Z S, Wu Y Y, Wang J Y, et al. Res2Fusion: infrared and visible image fusion based on dense Res2net and double nonlocal attention models[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71:

- 5005012.
- [8] Wang Z S, Wang J Y, Wu Y Y, et al. UNFusion: a unified multi-scale densely connected network for infrared and visible image fusion[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(6): 3360-3374.
- [9] 陈国洋, 吴小俊, 徐天阳. 基于深度学习的无监督红外图像与可见光图像融合算法[J]. *激光与光电子学进展*, 2022, 59(4): 0410010.
Chen G Y, Wu X J, Xu T Y. Unsupervised infrared image and visible image fusion algorithm based on deep learning[J]. *Laser & Optoelectronics Progress*, 2022, 59(4): 0410010.
- [10] Yang Y, Liu J X, Huang S Y, et al. Infrared and visible image fusion via texture conditional generative adversarial network[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(12): 4771-4783.
- [11] Liu J Y, Fan X, Jiang J, et al. Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(1): 105-119.
- [12] 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: 5002215.
- [13] Zang Y S, Zhou D M, Wang C C, et al. UFA-FUSE: a novel deep supervised and hybrid model for multifocus image fusion[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5008717.
- [14] 杨阳, 任振南, 李北辰. 联合卷积神经网络和转换器的红外与可见光图像融合[J/OL]. *激光与光电子学进展*: 1-21[2023-03-16]. <http://kns.cnki.net/kcms/detail/31.1690.TN.20221121.0940.020.html>.
Yang Y, Ren Z N, Li B C. Infrared and visible image fusion with convolutional neural network and transformer [J/OL]. *Laser & Optoelectronics Progress*: 1-21[2023-03-16]. <http://kns.cnki.net/kcms/detail/31.1690.TN.20221121.0940.020.html>.
- [15] 尹海涛, 周伟. 基于多尺度空洞卷积网络的多聚焦图像融合算法[J]. *激光与光电子学进展*, 2023, 60(2): 0210003.
Yin H T, Zhou W. Multi-scale dilated convolutional neural network based multi-focus image fusion algorithm [J]. *Laser & Optoelectronics Progress*, 2023, 60(2): 0210003.
- [16] Ma K D, Duanmu Z F, Zhu H W, et al. Deep guided learning for fast multi-exposure image fusion[J]. *IEEE Transactions on Image Processing*, 2019, 29: 2808-2819.
- [17] Wang J, Yu L, Tian S W, et al. AMFNet: an attention-guided generative adversarial network for multi-model image fusion[J]. *Biomedical Signal Processing and Control*, 2022, 78: 103990.
- [18] 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.
- [19] Su W J, Huang Y D, Li Q F, et al. Infrared and visible image fusion based on adversarial feature extraction and stable image reconstruction[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 2510214.
- [20] Ma J Y, Tang L F, Xu M L, et al. STDFusionNet: an infrared and visible image fusion network based on salient target detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5009513.
- [21] Zhou H B, Hou J L, Zhang Y D, et al. Unified gradient-and intensity-discriminator generative adversarial network for image fusion[J]. *Information Fusion*, 2022, 88: 184-201.
- [22] Fu Y, Wu X J, Durrani T. Image fusion based on generative adversarial network consistent with perception [J]. *Information Fusion*, 2021, 72: 110-125.
- [23] Zhu D P, Zhan W D, Jiang Y C, et al. IPLF: a novel image pair learning fusion network for infrared and visible image[J]. *IEEE Sensors Journal*, 2022, 22(9): 8808-8817.
- [24] Fu Y, Wu X J, Kittler J. Effective method for fusing infrared and visible images[J]. *Journal of Electronic Imaging*, 2021, 30(3): 033013.
- [25] Zhang H Z, Shen Y F, Ou Y Y, et al. A GAN-based visible and infrared image fusion algorithm[J]. *Proceedings of SPIE*, 2021, 12061: 120610Z.
- [26] Zhong Z, Yang J F. A novel pig-body multi-feature representation method based on multi-source image fusion [J]. *Measurement*, 2022, 204: 111968.
- [27] Xu H, Ma J Y, Le Z L, et al. FusionDN: a unified densely connected network for image fusion[C]//The 34th AAAI Conference on Artificial Intelligence, February 7-12, 2020, New York, USA. Menlo Park: AAAI Press, 2020: 12484-12491.
- [28] Xu H, Ma J Y, Jiang J J, et al. U2Fusion: a unified unsupervised image fusion network[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 502-518.
- [29] 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.
- [30] Jiang X H, Nie R C, Wang C C, et al. DenseNet with orthogonal kernel for infrared and visible image fusion [C]//2021 17th International Conference on Computational Intelligence and Security (CIS), November 19-22, 2021, Chengdu, China. New York: IEEE Press, 2022: 146-150.
- [31] Zhao Z X, Xu S, Zhang C X, et al. DIDFuse: deep image decomposition for infrared and visible image fusion [EB/OL]. [2022-12-09]. <https://www.ijcai.org/proceedings/2020/135>.
- [32] Jung H, Kim Y, Jang H, et al. Unsupervised deep image fusion with structure tensor representations[J]. *IEEE Transactions on Image Processing*, 2020, 29: 3845-3858.
- [33] Wang H F, Wang J Z, Xu H N, et al. DRSNFuse: deep residual shrinkage network for infrared and visible image

- fusion[J]. *Sensors*, 2022, 22(14): 5149.
- [34] Yang Y, Kong X K, Huang S Y, et al. Infrared and visible image fusion based on multiscale network with dual-channel information cross fusion block[C]//2021 International Joint Conference on Neural Networks (IJCNN), July 18-22, 2021, Shenzhen, China. New York: IEEE Press, 2021.
- [35] Guo C X, Fan D D, Jiang Z X, et al. MDFN: Mask deep fusion network for visible and infrared image fusion without reference ground-truth[J]. *Expert Systems with Applications*, 2023, 211: 118631.
- [36] Li H, Wu X J, Kittler J. RFN-Nest: an end-to-end residual fusion network for infrared and visible images[J]. *Information Fusion*, 2021, 73: 72-86.
- [37] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. New York: IEEE Press, 2005: 886-893.
- [38] Carcagnì P, Del Coco M, Leo M, et al. Facial expression recognition and histograms of oriented gradients: a comprehensive study[J]. *SpringerPlus*, 2015, 4: 645.
- [39] Sugiarto B, Prakasa E, Wardoyo R, et al. Wood identification based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) classifier[C]//2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), November 1-2, 2017, Yogyakarta, Indonesia. New York: IEEE Press, 2018: 337-341.
- [40] Wei C, Fan H Q, Xie S N, et al. Masked feature prediction for self-supervised visual pre-training[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 14648-14658.
- [41] Zong J J, Qiu T S. Medical image fusion based on sparse representation of classified image patches[J]. *Biomedical Signal Processing and Control*, 2017, 34: 195-205.
- [42] Maggu J, Saini J K, Verma P. FILM. LrTL: FusIng MuLtiFocus IMages using low-rank transform learning [C]//2022 1st International Conference on Informatics (ICI), April 14-16, 2022, Noida, India. New York: IEEE Press, 2022: 60-65.
- [43] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]//The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, November 9-12, 2003, Pacific Grove, CA, USA. New York: IEEE Press, 2004: 1398-1402.
- [44] Duda R O, Hart P E, Stork D G. *Pattern classification* [M]. Singapore: Wiley-Interscience, 2000.
- [45] Guo C K. *Multi-modal image registration with unsupervised deep learning*[D]. Cambridge: Massachusetts Institute of Technology, 2019.
- [46] Xu R, Chen Y W, Tang S Y, et al. Parzen-window based normalized mutual information for medical image registration[J]. *IEICE-Transactions on Information and Systems*, 2008, E91-D(1): 132-144.
- [47] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. *Computer vision-ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8693: 740-755.
- [48] Toet A, Hogervorst M A. Progress in color night vision [J]. *Optical Engineering*, 2012, 51(1): 010901.
- [49] Zhao H, Gallo O, Frosio I, et al. Loss functions for image restoration with neural networks[J]. *IEEE Transactions on Computational Imaging*, 2017, 3(1): 47-57.
- [50] Roberts J W, van Aardt J, Ahmed F. Assessment of image fusion procedures using entropy, image quality, and multispectral classification[J]. *Journal of Applied Remote Sensing*, 2008, 2(1): 23522.
- [51] Rao Y J. In-fibre Bragg grating sensors[J]. *Measurement Science and Technology*, 1997, 8(4): 355.
- [52] Qu G H, Zhang D L, Yan P F. Information measure for performance of image fusion[J]. *Electronics Letters*, 2002, 38(7): 313-315.
- [53] Aslantas V, Bendes E. A new image quality metric for image fusion: the sum of the correlations of differences[J]. *AEU-International Journal of Electronics and Communications*, 2015, 69(12): 1890-1896.
- [54] Haghighat M, Razian M A. Fast-FMI: Non-reference image fusion metric[C]//2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), October 15-17, 2014, Astana, Kazakhstan. New York: IEEE Press, 2015.
- [55] Xydeas C, Petrovi V. Objective image fusion performance measure[J]. *Electronics Letters*, 2000, 36(4): 308-309.
- [56] Han Y, Cai Y Z, Cao Y, et al. A new image fusion performance metric based on visual information fidelity [J]. *Information Fusion*, 2013, 14(2): 127-135.
- [57] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[EB/OL]. (2016-03-27)[2022-11-09]. <https://arxiv.org/abs/1603.08155>.