Computer Science and Application 计算机科学与应用, 2023, 13(12), 2562-2574
Published Online December 2023 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2023.1312255

Hans汉斯

一种基于多层卷积稀疏网络的红外与可见光图像融合方法

王静静，王少坤，吕梦莎

电科云(北京)科技有限公司，北京

收稿日期：2023年11月26日；录用日期：2023年12月22日；发布日期：2023年12月30日

摘　要

红外图像和可见光图像融合广泛应用于夜视、监视、军事等领域。融合任务的重点在于将可见光和红外光图像中的互补信息整合起来并消除多余信息。此外，大多数融合任务是在低光环境下进行的，如何保持融合结果的照明信息值得研究。为了解决存在的问题，首先，我们设计了一个多级特征模块来融合多源信息。与传统网络的并行层融合策略不同，我们提出了一种并行层和深度层相结合的融合策略。其次，我们在特征提取网络中增加了注意力计算，以提高特征提取网络的性能。第三，为了使融合图像具有良好的照明信息，我们设计了区域照明保留模块，提高了低光环境下融合算法的性能。大量实验证明了所提出的方法具有出色的性能，并且在低光环境下表现更好。此外，所提出的算法在多模式物体检测方面也显示出巨大潜力。

关键词

图像融合，红外与可见光，卷积网络

An Infrared and Visible Image Fusion Method Based on Multilayer Convolutional Sparse Network

Jingjing Wang, Shaokun Wang, Mengsha Lv

Dianke Cloud (Beijing) Technology Company, Beijing

Received: Nov. 26th, 2023; accepted: Dec. 22nd, 2023; published: Dec. 30th, 2023

Abstract

Infrared image and visible light image fusion is widely used in night vision, surveillance, military

文章引用: 王静静, 王少坤, 吕梦莎. 一种基于多层卷积稀疏网络的红外与可见光图像融合方法[J]. 计算机科学与应用, 2023, 13(12): 2562-2574. DOI: 10.12677/csa.2023.1312255

and other fields. The focus of the fusion task is to integrate complementary information in visible and infrared light images and eliminate redundant information. In addition, most of the fusion tasks are performed in the harsh environment of low light, and it is worth studying how to maintain the lighting information of the fusion results. In order to solve the problems existing, firstly, we design a multi-level feature module to fusion multi-source information. Different from the parallel layer fusion strategy of the traditional network, we proposed a fusion strategy that combined parallel layers and depth layers. Secondly, we add attention computing to the feature extraction network to improve the performance of the feature extraction network. Thirdly, in order to make the fusion image have good illumination information, we design the area illumination retention module, improving the performance of the fusion algorithm in low-light environments. A large number of experiments show that the proposed method has excellent performance and will perform better in low-light environments. In addition, the proposed algorithm also shows great potential in multi modal-object detection.

Keywords

Image Fusion, Infrared and Visible Light, Convolutional Networks

1. 引言

在复杂环境下，单个摄像机获得的图像信息是有限的。难以支持对象检测和语义分割等后续任务[1]。图像融合技术的出现解决了这个问题。其中，红外图像和可见光图像的融合是最常用的[2]。

在过去几十年中，人们使用统计方法设计了各种图像融合算法，包括基于多尺度分解的方法[3]，基于稀疏表示的方法[4]和基于显著性的方法[5]。这些方法都具有一个共同特点，它们将图像分解成多个层次，并为不同的层次设计不同的融合规则。融合结果受到分解方法的限制。为了实现良好的融合结果，必须设计极其复杂的分解方法，这对实时处理构成了挑战。此外，手动设计的分解方法没有良好的鲁棒性。近年来，随着深度学习的发展，提出了许多基于神经网络的图像融合算法。基于神经网络的融合算法可以分为基于卷积神经网络的方法[6] [7] [8]和基于对抗神经网络的方法[9] [10]。基于卷积神经网络的算法利用神经网络的特征提取能力提取特征，融合来自多个源图像的特征，并设计损失函数重建融合特征。基于对抗神经网络的方法使用生成器和判别器，通过两者之间的对抗学习获取融合图像。

2. 现存问题

目前，基于神经网络的图像融合算法可以取得良好的结果，但仍存在一些问题：

(1) 多源图像的特征提取网络彼此独立，导致自然融合结果较差。例如，在图 1 中的 SDNet [11]和 FuionGAN [12]的红框部分，两种模式的信息在特征提取过程中没有融合，导致融合结果中存在接缝感。

(2) 普通卷积可能会导致信息丢失。例如，图 1 中的 FusionGAN 使用普通卷积，导致模糊和对比度较差的融合结果。

(3) 没有考虑到照明信息，融合结果的亮度非常低。例如，图 1 中的三种方法都努力保持纹理和细节。然而，它们都没有考虑到融合过程中亮度的降低，我们可以在红框中清楚地看到这一点。GFF [13]方法的天空部分更像是可见光和红外图像的平均值，而 SDNet 和 FusionGAN 错误地保留了红外的天空部分，

而我们的方法成功地保留了更明亮的可见光图像的天空部分，而没有任何亮度降低。

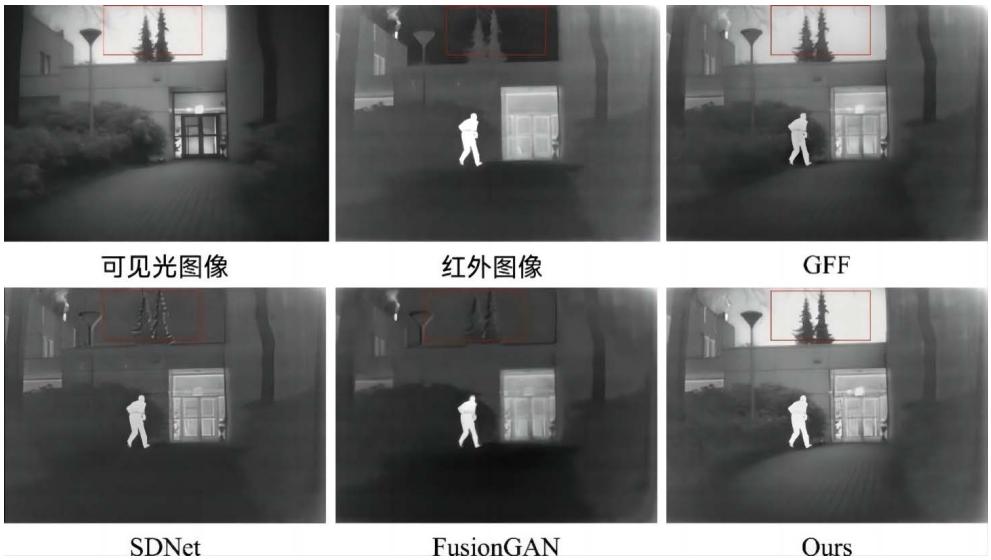


Figure 1. Comparison of typical fusion results, these are the following: visible image, infrared image, GFF, SDNet, FusionGAN, Ours
图 1. 典型融合结果对比：可见光图像、红外图像、GFF、SDNet、FusionGAN、Ours

3. 相关工作

3.1. 基于 CNN 的图像融合

随着深度学习的发展，图像融合领域出现了许多新的算法。最初，研究人员仅使用神经网络构建图像融合的权重图，而没有完全放弃传统方法。Li 等人[14]使用目标函数最小化的方法将图像分为基础层和细节层，并使用平均方法对基础层进行融合。使用 VGG 网络提取细节层的特征，并对特征图进行归一化，获得细节层的权重图进行融合。这种应用并没有充分发挥神经网络的性能。随着研究的进展，出现了许多完全基于神经网络的融合算法，通过设计不同的特征提取网络和损失函数，获得了具有自身特点的结果。Li 等人[15]设计了一个使用卷积神经网络的编码器和解码器来实现融合任务的方法，称为 DenseFuse。该方法在编码器特征中对不同的卷积层进行了不同的处理，将每一层的输出与所有后续层连接起来，并设计了同时考虑结构和像素损失的损失函数，获得了相对较好的融合结果。这种方法中向后传递浅层特征的思想启发了本文中多级融合的设计。Ma 等人[16]在融合网络的损失函数中进行了优化，并设计了 STDFusion，将图像分为显著区域和背景区域，并分别计算融合图像的显著区域和背景区域的损失，获得了更好的融合结果，但该方法的有效性依赖于分解算法的性能。通过使用传统方法，一个强大的分解算法也可以产生良好的结果。Tang 等人[17]使用类似的思路采用语义分割网络实现图像分割，为不同的区域构建不同的损失，并在特征提取网络中考虑多级特征的互补性，证明了多级特征融合的有效性。许多研究人员还尝试突破红外和可见光图像之间的障碍，进行更深层次的融合。Li 等人[18]在 DenseFuse 之后增加了特征融合，并提出了一种新的网络结构。多源图像的四个尺度的不同特征由编码器获得，输入到 RFN 模块进行融合，然后将融合的特征输入到解码器进行解码。这种方法以多层次的方式创新地结合了多源图像，融合结果展示了多层次融合的有效性。Tang 等人提出的 PIAFusion [19]也考虑了多源特征之间的关系，并设计了类似差分电路的 CMDAF 模块，用于逐步融合多源特征。融合中保留照明信息也是一个具有挑战性的问题，如何在照明过低和过高的极端条件下保持良好的照明水平需要解决。

此外，一些最新的工作利用语义信息进行图像融合，紧密联系了上游和下游任务。Xie 等人[20]提出了一个融合网络，嵌入了语义信息，并将融合任务和对齐任务组合在一个网络中进行。这种方法非常创新，推动了工程应用中图像融合技术的发展。

3.2. 基于 GAN 的图像融合

这种 GAN 方法通过设计两个相互对抗的网络来创新地利用神经网络来学习融合结果。生成器经过专门训练以生成欺骗判别器的融合图像[21] [22]。Xu [23]等人设计了一个双条件双判别器对抗神经网络，其中两个判别器的训练损失组合了可见光和红外光，其中一个判别器通过计算融合图像和可见光图像的梯度损失进行训练，另一个通过计算融合图像和红外图像的梯度损失进行训练。Ma 等人[24]设计的 DDcGAN 网络具有类似的结构。Zhang 等人[25]设计的 GAN-FM 采用了 U-Net 网络结构，在生成器的网络设计中尝试保留更多语义信息。Li 等人[26]在网络中引入了注意机制，并设计了 AttentionFGAN。Rao 等人[27]提出了一种结合语义信息的融合网络：AT-GAN，通过强度维护模块在红外图像中保持热目标信息，并通过使用语义转换模块在可见光图像中滤除噪声，从而实现了良好的融合结果。这些 GAN 的框架相似，都由一个生成器和两个判别器组成。值得指出的是，这种框架也导致了这些融合网络的复杂训练。总之，一个好的融合网络应首先考虑多个源图像的多层特征之间的融合，其次，尽可能保留多个源图像的照明信息也尤为重要，最后，一个端到端的、易于训练的网络可以更容易和快速地应用于工程中。

4. 方法

本节对本文的方法进行了全面介绍。首先，详细描述了 MLFFusion 的模块结构，然后说明了网络的损失函数，最后给出了网络的结构。

4.1. 模块结构

首先，多源图像的融合需要将不同图像的优点融合到融合图像中。红外光包含丰富的目标信息，可见光包含精细的纹理信息，将两个图像在深层次上进行融合是需要解决的问题。其次，传统的卷积网络通过卷积获得特征时会存在冗余信息，更加注重有用信息可以提高网络特征提取的质量。最后，损失的设计不能仅关注融合图像的质量，尽可能保留亮度信息不仅可以提高融合图像的对比度，还可以提高融合算法在低光和其他恶劣环境下的鲁棒性。因此，本文设计了以下三个点来解决这些问题。

4.1.1. MLFF 模块

MLFF (多级特征融合)模块旨在实现多级特征融合。卷积神经网络在网络深度加深时可以更详细地提取特征，但深层网络结构可能导致某些特征丢失，因此需要进行跨层特征补充。在图像融合任务中，仅在特征提取网络末端进行融合无法充分融合跨模态信息，因此在网络特征提取的图中进行渐进式融合也是必要的。因此，MLFF 模块的定义如下：

$$F_{ir}, F_{vi} = \mathrm{MLFF}\left(H_{ir}, H_{vi}, L_{ir}, L_{vi}\right)$$

MLFF 模块包含 4 个输入。表示浅层红外特征和浅层可见特征，H_{ir} 和 H_{vi} 表示更深的红外特征和更深的可见特征，F_{ir} 和 F_{vi} 多源和多水平融合红外特征和可见光特征的表示。MLFF 模块的结构如图 2 所示。

更具体地说，MLFF 模块首先在两个浅层特征之间的通道上建立连接，以完成浅层多源特征的融合，定义如下：

$$F_{low} = concat\left(F_{ir}, F_{vi}\right)$$

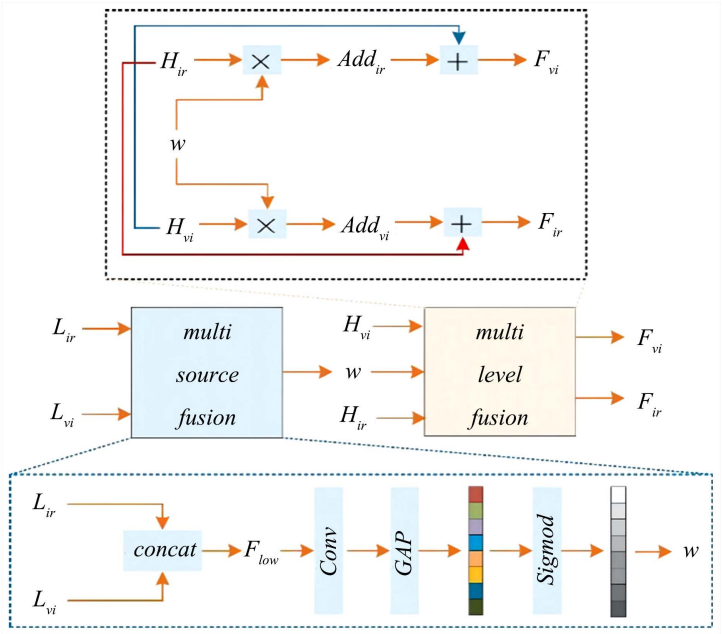


Figure 2. MLFF module schematic, the top box indicates multi-layer information fusion, the middle box indicates the MLFF module schematic and the bottom figure indicates multi-source information fusion
图 2. MLFF 模块示意图，上框表示多层信息融合，中框表示 MLFF 模块示意图，下图表示多源信息融合

这个公式表示通道之间的逐通道连接，而 1 表示融合的浅层特征。该公式表示浅层多源信息的融合以获取融合的浅层特征。然而，此时获取的浅层融合特征与深层特征的比例不匹配，因此需要进行 1 × 1 卷积来调整浅层融合特征的比例，并且可以对调整后的融合浅层特征进行池化和激活，以获取权重向量来指导深层特征的融合。该过程的定义如下：

$$w = s\left(g\left(conv\left(F_{low}\right)\right)\right)$$

其中 $conv(\cdot)$ 表示 1 × 1 卷积操作以平衡浅层和深层特征的规模，$g(\cdot)$ 表示从融合的浅层特征中提取信息的全局平均池化，$s(\cdot)$ 表示一个 sigmod 函数，用于将特征映射的值限制在[0-1]和输出之间 w 表示得到的权重向量，可以将其与深层特征逐个元素相乘，以获得融合的深层特征。定义如下：

$$F_{ir} = H_{ir} \oplus add_{ir} = H_{ir} \oplus \left(H_{vi} \otimes w\right)$$
$$F_{vi} = H_{vi} \oplus add_{vi} = H_{vi} \oplus \left(H_{ir} \otimes w\right)$$

其中，$\otimes$ 是逐通道乘法运算，$\oplus$ 是逐通道求和，H_{ir} 和 H_{vi} 是深层特征，并且 add_{ir}，add_{vi} 是从浅层特征中提取的深层互补信息，并且 F_{ir} 和 F_{vi} 是 MLFF 模块之后的多级、多源融合红外/可见光特征。此步骤类似于自注意力计算，其中浅层信息用于指导深度多模态信息融合。

神经网络可以通过堆叠卷积层不断提取特征，我们将靠近输入的特征定义为浅层特征，而靠近输出的特征定义为深层特征，见图 3。具体而言，在 MLFF (多级特征融合)模块中，有四个输入，分别是可见光浅层特征和红外光浅层特征(接近网络输入部分)，以及红外深层特征和可见光深层特征(接近网络输出部分)。随着网络的不断加深，提取的特征特性也不同。例如，在图 3 中所示，浅层特征是靠近输入的特征，包含更多关于图像的详细信息；深层特征是靠近输出的特征，不再具有详细的特征，而是包含更抽象的语义特征。

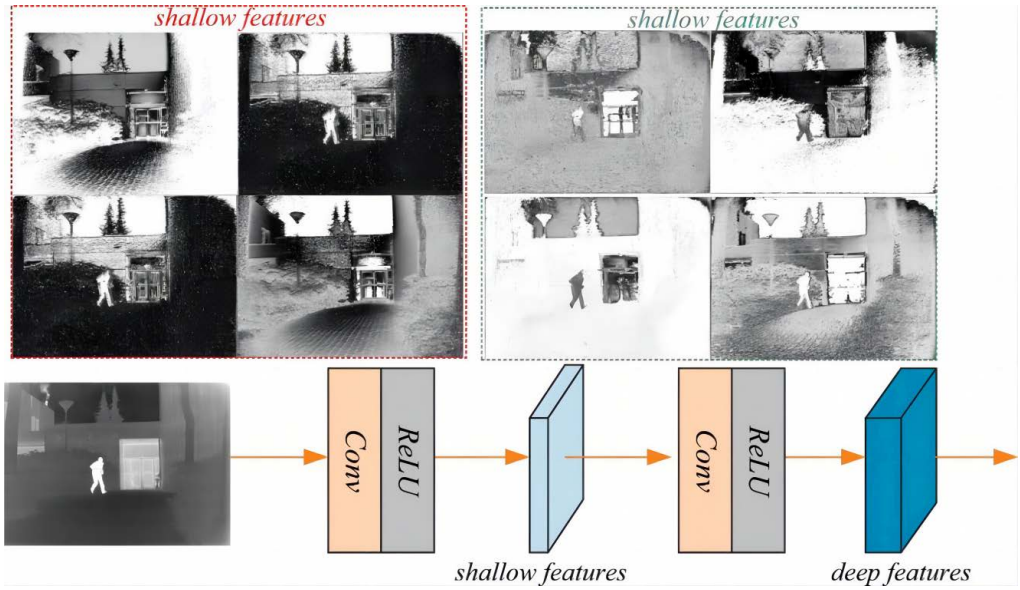


Figure 3. Definition of shallow and deep features and their visualisation
图 3. 浅层和深层的定义及其可视化

由于不同层次的特征具有不同的特点，我们需要融合多层特征，这就是为什么我们连接浅层和深层特征的原因。由于来自不同来源的特征之间存在巨大差异，我们需要分阶段融合特征。由于浅层特征包含更多的详细信息，全局池化后得到的向量的方差更大，估计的权重可以具有更大的方差分布，因此我们使用浅层特征来估计权重，这意味着我们可以将深层特征与浅层特征关联起来，并获得具有较大方差的权重用于融合，避免了“平均”情况下的融合。

4.1.2. SAconv 模块

这个传统的卷积对每个通道都赋予了相同的重要性，这是不合理的。SENet [28]提出了一种具有自我注意机制的网络，本文将这种方法应用于编码部分，并设计了编码部分的基本单元，即 SAconv 模块。图 4 显示了 SAconv 模块的结构，首先卷积输入张量，然后全局平均池化卷积结果以获得一个向量，然后对该向量进行激活，最后获得反映卷积后不同特征层重要性的向量，从而获得输出结果。

具体来说，首先对输入进行卷积以获得多个特征层，然后计算卷积结果上不同通道的不同权重：

$$\text{output} = w \otimes x = s\left(ex\left(conv\left(\text{input}\right)\right)\right) \otimes conv\left(\text{input}\right)$$

其中 x 表示输入卷积的结果，$conv(\cdot)$ 表示卷积核，$ex(\cdot)$ 表示激励[29]，$s(\cdot)$ 表示 sigmod 函数和 w 表示权重矢量。

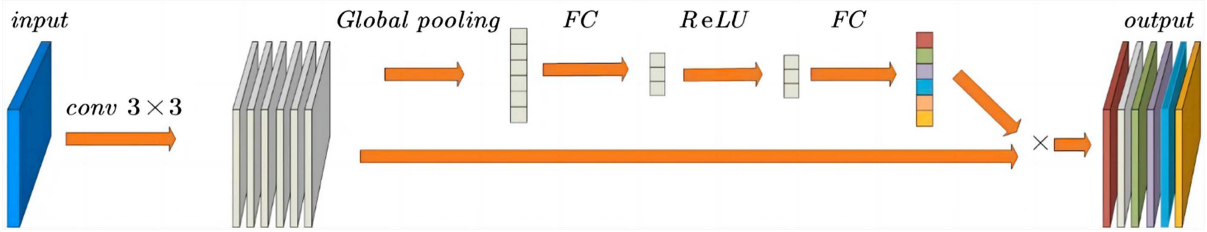


Figure 4. SAconv module structure diagram, where the input features are subjected to convolution and self-attention operations to obtain the enhanced features
图 4. SAconv 模块结构图 SAconv 模块结构图，输入特征经过卷积和自注意运算后得到增强特征

4.1.3. 区域照度模块

自然界中存在复杂的照明，在夜间等低光照条件下，图像的对比度通常较低。因此，如何在融合图像中保存光线是一个需要解决的问题。因此，本文使用二元分类网络来获取图像属于昼夜的概率。图 5 显示了二元分类网络的结构，表示为 $I(\cdot)$。这样，对于每幅图像，可以得到它属于白天的概率和属于夜晚的概率，计算公式如下：

$$\{P_d, P_n\} = I(\text{input})$$

其中，P_d 表示输入图像属于白天的概率，P_n 表示输入图像属于夜晚的概率，并且 input 表示输入分类网络的图像。

对两个概率进行归一化，结果是图像的照明水平，计算如下，输出 w 是照明水平。

$$w = \frac{P_d}{P_d + P_n}$$

4.2. 损失函数

本节将详细介绍二元分类网络的损失函数和图像融合网络的损失函数，同时还会介绍光感知模块的应用。

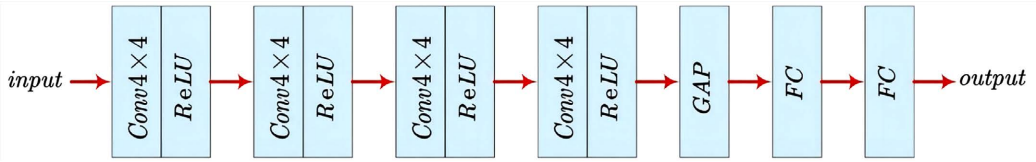


Figure 5. Binary classification network structure. The first 4 layers are convolutional, the convolutional kernel size is 4, the post convolutional is ReLU activation function, GAP denotes global average pooling and FC denotes fully connected Layer

图 5. 二元分类网络结构。前 4 层为卷积层，卷积核大小为 4，后卷积层为 ReLU 激活函数，GAP 表示全局平均池化，FC 表示全连接层

4.2.1. 分类网络

图 5 中的网络结构是一个经典的二元分类网络，使用交叉熵损失作为网络的损失函数，定义如下。其中 z 表示输入图像的标签，0 表示夜晚，1 表示白天，以及 y 是网络输出的结果，$\sigma(\cdot)$ 表示 softmax 函数。

$$L = -z\log\sigma(y) - (1-z)\log(1-\sigma(y))$$

4.2.2. 融合网络

融合图像需要保留多源图像中的显著区域，这些显著区域具有较大的梯度值。因此，保持融合图像中显著区域的任务可以被转化为获取多源图像的较大梯度的任务。我们可以定义显著损失如下：

$$L_{\text{detial}} = \frac{1}{HW}\left\|\nabla I_f\right| - \max\left(\left|\nabla I_{ir}\right|, \left|\nabla I_{vi}\right|\right)\Big\|_1$$

其中，∇ 表示图像的渐变，I_{ir} 表示红外图像，I_{vi} 表示可见图像和 I_f 表示融合图像，H，W 表示图像的高度和宽度。该公式表示融合图像的每个区域选择多源图像中梯度值较大的部分。

其次，融合图像需要在多模态图像中包含一个平滑的背景区域，我们可以通过对像素值差异进行约束来保留这一区域，定义背景损失如下：

$$L_{\text{base}} = \frac{1}{HW}\left\|I_f\right| - \max\left(\left|I_{ir}\right|, \left|I_{vi}\right|\right)\Big\|_1$$

接下来，需要计算融合网络的光照损失。每个图像的不同区域具有不同的光照水平，对整个图像计算光照损失是不合理的。因此，图像被分割成块，并分别计算每个块的光照水平，定义如下：

$$\{i_1, i_2, \cdots, i_n\} = \mathbb{I}$$

其中，$\{i_1, i_2, \cdots, i_n\}$ 表示通过分割获得的每个小图像块，n 表示通过分割获得的小图像块的数量，并且 $\mathbb{I}$ 表示输入图像。

对于每个小块的图像分割，需要先获取小块的照明强度，然后再计算损失。可以通过二分类网络得到光强度序列：

$$\{w_1, w_2, \cdots, w_n\} = \mathbb{W}$$

在计算损失时，直接使用分类网络得到的概率进行损失计算显然是不合理的。照明依赖于图像中要呈现的像素的组合，因此，像素损失定义如下：

$$L_b^a = \frac{1}{HW}\|a - b\|_1$$

其中，L_b^a 表示图像的像素损失 a 和图像 b 照度损失函数可以定义如下：

$$L_{\text{ill}} = \sum w \cdot i_{i_{vi}}^{i_f} + (1-w) \cdot L_{i_{ir}}^{i_f}, w \in \mathbb{W}, i_f \in \mathbb{I}_f, i_{vi} \in \mathbb{I}_{vi}, i_{ir} \in \mathbb{I}_{vi}$$

其中，$\mathbb{I}_{vi}$ 表示可见图像，i_{vi} 表示通过分割获得的可见图像序列。$\mathbb{I}_{ir}$ 表示红外图像，i_{ir} 表示通过分割得到的红外图像序列。$\mathbb{I}_f$ 表示融合图像，i_f 表示通过分割获得的融合图像序列。$\mathbb{W}$ 表示从融合图像获得的照度矢量，w 表示对应于每个融合图像块的照度值。$i_{i_{vi}}^{i_f}$ 表示每个融合图像块和可见图像块的像素损失，以及 $L_{i_{ir}}^{i_f}$ 表示每个融合图像块和红外图像块的像素损失。值得注意的是，n 图像块是原始图像大小相等的剪切。

最后，设置权重将三个损失联系起来。图像背景区域的照明损失和背景损失都保留，我们可以设置相同的权重，最终损失函数定义如下：

$$L = \alpha\left(L_{\text{ill}} + L_{\text{base}}\right) + \beta L_{\text{detial}}$$

其中，α 表示照明和背景损失的权重，以及 β 表示梯度损失的权重。

5. 实验配置

5.1. 数据集

MSRS [30]数据集于 2022 年发布，包括 1444 对高质量红外和可见光图像。这些图像包括明亮的白天图像和不太明亮的夜间图像。因此，我们选择了 MSRS 数据集作为我们的训练数据集。

RoadScene 有 221 个对齐的可见光和红外图像对，其中包含丰富的场景，如道路、车辆、行人等。这些图像是 FLIR 视频中极具代表性的场景。我们将之作为验证数据集。

5.2. 训练网络

融合网络使用 MSRS 数据集进行训练，该数据集包含 1444 个严格对齐的可见光和红外图像对。照明网络在处理后的 MSRS 数据集上进行训练。我们选取了 50 张光照良好的日间可见光图像和 50 张照明不佳的夜间可见光图像，共 100 张图像，然后通过裁剪将其扩展为 6400 张图像。我们将照明良好的白天图像标记为“白天”类别，将照明不佳的夜间图像标记为“夜间”类别，然后使用这 6400 张图像来训练我们的二元分类网络。值得注意的是，我们的二元分类网络的目的是结合损失函数保留融合图像的亮度信息，因此我们在训练图像的选择中避免了混淆“白天但低光”和“夜间但高亮度”的混淆。对于融合网

络，我们使用从二元分类网络中选出的 100 张可见光图像和相应的红外图像，通过裁剪将这 100 对图像扩展为 6400 对图像，然后使用这 6400 对图像来训练融合网络。

代码是使用 Pytorch-GPU 实现的，二元分类网络首先使用交叉熵损失进行训练。然后使用本文提出的损耗对融合网络进行训练。对于二元分类网络，我们将训练会话的批量大小设置为 128，学习率设置为 0.01，训练周期设置为 100。对于融合网络，我们将训练会话的批量大小设置为 128，学习率设置为 0.001，训练周期设置为 60，设置 α 为 5 和 β 为 50。

5.3. 验证与评价

RoadScene 数据集的场景亮度较高，因此本文提出的区域光照保留模块在该数据集中没有最大价值，但该方法在该数据集中仍有较好的性能。从表 1 中可以看出，本文中的算法在 EN 指标和 Qabf 指标上没有获得最佳结果，但它也排在前三名。Roadscreen 数据集中图像的整体亮度较高，大部分是光照充足的白天场景，甚至数据集中的部分图像也存在曝光过度的问题。因此，它削弱了区域照度保持模块在本文中的作用，甚至对于一些曝光过度的图像，区域照度保持模块也会起到相反的作用。本文中的方法仍然具有良好的性能，证明了本文算法的鲁棒性。

本文中的算法在 RoadScene 数据集下所有 5 种评价指标中都获得了最高分。其中，AG、EN 和 SF 是针对单独图像的三个评价指标，反映了融合图像的图像质量。表 1 中的信息表明，传统方法可以很好地执行，但程度不是非常高，并且 FusionGAN 方法是一种高度创新的特殊框架，但需要进一步优化网络以更好地执行融合任务。其他三种基于 CNN 的方法，DenseFuse，RFN-Nest 和 SDNet，给出了更好的融合结果。得益于区域照明信息保存模块，该方法在 RoadScene 数据集的评估指标中得分最高。Qabf、VIF、MI 反映了融合图像中多源信息的集成程度，并且由于 PIA 中差分融合模块的设计获得了更好的结果。显然，本文中的 MLFF 模块具有更强的性能。

Table 1. Comparative analysis of eleven algorithms: mean performance metrics for six evaluation indicators on the RoadScene dataset
表 1. 比较 11 种算法，RoadScene 数据集下 6 个评估指标的性能平均值

	AG	EN	SF	Qabf	VIF	MI
GFF	4.4488	7.3086	4.5608	0.3923	0.7670	3.0701
ADF	4.6268	7.0102	4.9643	0.4851	0.6951	2.7476
TIF	5.4967	7.1185	5.6856	0.4458	0.7458	2.4963
DLF	3.6297	6.8455	3.7390	0.4481	0.6836	2.8870
FusionGAN	3.3905	7.0732	3.4320	0.2579	0.5785	2.7550
PIAFusion	5.8886	7.1251	6.1980	0.4951	0.7802	3.2109
DenseFuse	3.3493	6.8242	3.3581	0.3874	0.6719	2.9224
RFN-Nest	3.3635	7.3393	3.0812	0.2982	0.7248	2.7464
SDNet	6.1086	7.3172	5.9805	0.5106	0.7703	3.2715
SuperFusion	4.4693	6.9901	4.7812	0.4502	0.7762	3.3687
SwinFusion	4.5157	7.0004	4.7743	0.4437	0.7803	3.3755
Ours	6.1476	7.3189	6.5908	0.4896	0.7824	3.5626

为了证明该算法的有效性，从 RoadScene 数据集中选择了四组图像进行分析，如图 6 所示，传统方法在处理具有复杂亮度的环境时不能产生良好的融合效果。例如，在第一组图像的结果中，包括 DLF 方

法在内的三种传统方法的融合结果对比度差，肉眼效果差，虽然 GFF 方法在清晰度方面表现不错，但在第四组图像中可以看出，这种清晰度是偶然的，并不健壮。

Figure 6. Visual comparison of our method with the 11 algorithms on the RoadScene dataset. For a clear comparison, we chose to highlight textured areas with red and green boxes

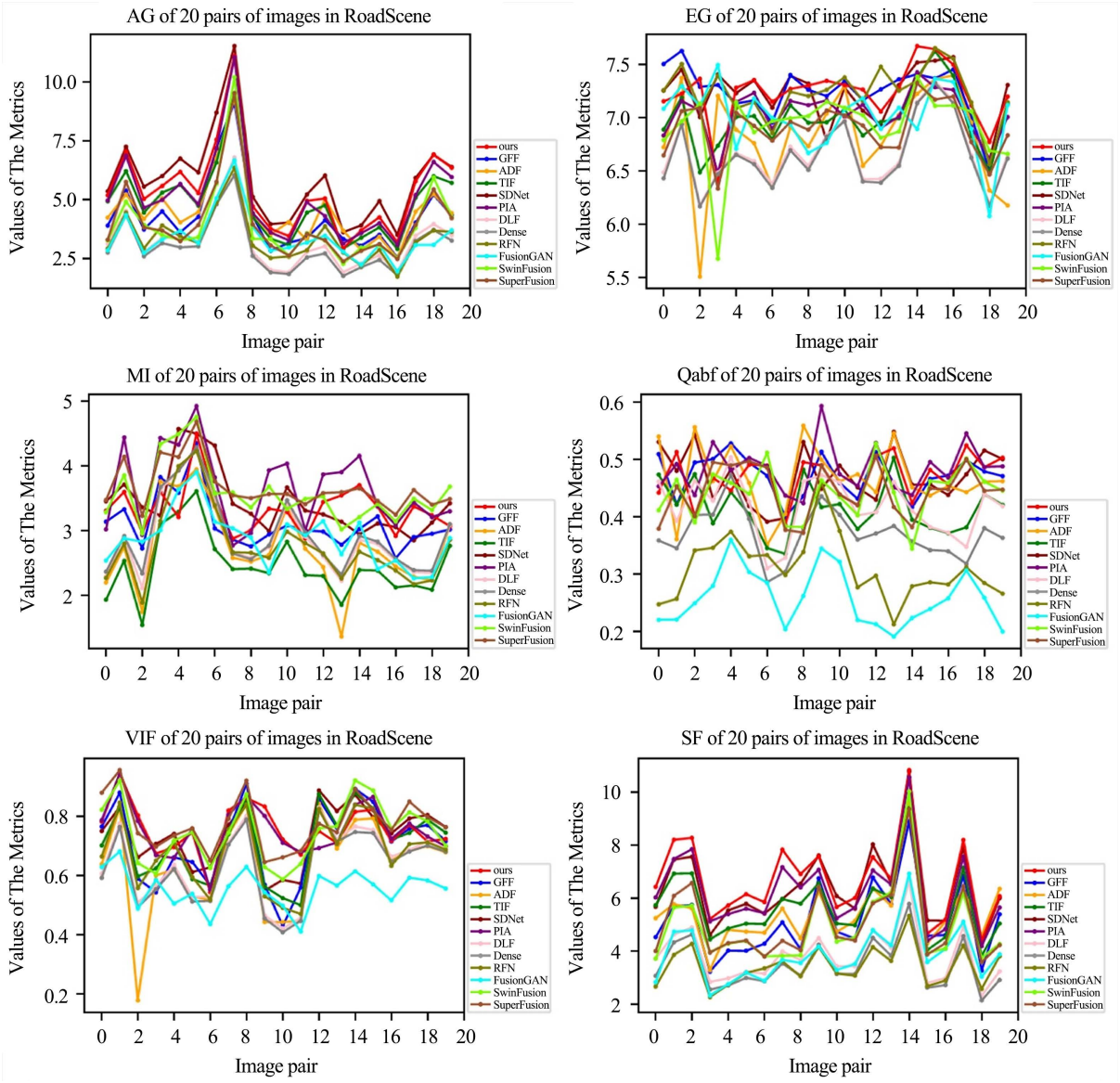
图 6. 将我们的方法与 RoadScene 数据集上的 11 种算法进行可视化比较。为了进行清晰的比较，我们选择用红色和绿色框突出显示纹理区域

传统方法使用统计像素分布信息来生成融合图像，但在像素分布信息的每个部分都获得高分的复杂环境中，融合方法难以识别真正的“重要”区域，因此会产生对比度差的融合图像。

FusionGAN 方法仍然无法摆脱融合图像亮度低和目标边缘模糊的问题，如第一张和第三图中的目标所证明的那样。DenseFuse 和 RFN-Nest 在 RoadScene 数据集上不会表现优异。SDNet 性能良好，获得了高清融合结果，但 RoadScene 数据集具有良好的照明度表现。PIA 方法考虑了照明信息，但是是全局的，在光线充足的环境中，融合的图像太亮，这会影响视觉效果，反而会降低融合的质量。本文中的区域光

照保持方法不存在这个问题，融合图像目标清晰，纹理细腻。

为了进一步说明，我们在 RoadScene 数据集上选择了 20 对图像来制作图像指标的变化曲线，如图 7 所示。曲线表明，本文中的算法在 RoadScene 数据集上效果斐然。



注：在 RoadScreen 数据集中使用 EN，AG，Qabf，VIF，MI 中的 6 个指标和 11 种高级算法选择了 20 个图像对。水平坐标是图像编号，垂直坐标是图像的评价指标值。

Figure 7. Quantitative comparison of six indicators
图 7. 6 个指标的定量比较

6. 结论

1) 在本文中，我们提出了一种具有区域照明保留的多级多模态特征融合网络，缩写为 MLFFusion。首先，针对骨干网络设计 SAconv 模块，提高网络的特征提取能力。

2) 设计 MLFF 模块对不同层次、不同模式的信息进行整合，提高融合网络的信息集成能力。

3) 设计区域照度保持模块将结果与优良的照度信息融合，大大提高了融合算法在低照度地面恶劣环境下的鲁棒性。大量的实验证明了该算法的优越性。此外，该文算法在多模态目标检测任务中显示出巨大的潜力。

参考文献

[1] Zhang, J., Lei, W., Li, S., Li, Z. and Li, X. (2023) Infrared and Visible Image Fusion Withentropy-Based Adaptive Fusion Module and Mask-Guided Convolutional Neural Network. *Infrared Physics & Technology*, **131**, Article ID: 104629. https://doi.org/10.1016/j.infrared.2023.104629

[2] Ma, J., Ma, Y. and Li, C. (2019) Infrared and Visible Image Fusion Methods and Applications: A Survey. *Information Fusion*, **45**, 153-178. https://doi.org/10.1016/j.inffus.2018.02.004

[3] Ma, J. and Zhou, Y. (2020) Infrared and Visible Image Fusion via Gradientlet Filter. *Computer Vision and Image Understanding*, **197-198**, Article ID: 103016. https://doi.org/10.1016/j.cviu.2020.103016

[4] Xing, C., Wang, Z., Ouyang, Q., Dong, C. and Duan, C. (2019) Image Fusion Method Based on Spatially Masked Convolutional Sparse Represenitation. *Image and Vision Computing*, **90**, Article ID: 103806. https://doi.org/10.1016/j.imavis.2019.08.010

[5] Bavirisetti, D.P. and Dhuli, R. (2016) Two-Scale Image Fusion of Visible and Infrared Images Using Saliency Detection. *Infrared Physics & Technology*, **76**, 52-64. https://doi.org/10.1016/j.infrared.2016.01.009

[6] Li, H., Cen, Y., Liu, Y., Chen, X. and Yu, Z. (2021) Different Input Resolutions and Arbitrary Output Resolution: A Meta Learning-Based Deep Framework for Infrared and Visible Image Fusion. *IEEE Transactions on Image Processing*, **30**, 4070-4083. https://doi.org/10.1109/TIP.2021.3069339

[7] Jian, L., Yang, X., Liu, Z., Jeon, G. and Chisholm, D. (2020) SEDRFuse: A Symmetric Encoder-Decoder with Residual Block Network for Infrared and Visible Image Fusion. *IEEE Transactions on Instrumentation and Measurement*, **70**, 1-15. https://doi.org/10.1109/TIM.2020.3022438

[8] Liu, R., Liu, J., Jiang, Z., Fan, X. and Luo, Z. (2020) A Bilevel Integrated Model with Data-Driven Layer Ensemble for Multi-Modality Image Fusion. *IEEE Transactions on Image Processing*, **30**, 1261-1274. https://doi.org/10.1109/TIP.2020.3043125

[9] Yang, Y., Liu, J., Huang, S., Wan, W. and Guan, J. (2021) Infrared and Visible Image Fusion via Texture Conditional Generative Adversarial Network. *IEEE Transactions on Circuits and Systems for Video Technology*, **31**, 4771-4783. https://doi.org/10.1109/TCSVT.2021.3054584

[10] Zhou, H., Wu, W., Zhang, Y., Ma, J. and Ling, H. (2023) Semantic-Supervised Infrared and Visible Image Fusion via a Dual-Discriminator Generative Adversarial Network. *IEEE Transactions on Multimedia*, **25**, 635-648. https://doi.org/10.1109/TMM.2021.3129609

[11] Zhang, H. and Ma, J. (2021) SDNet: A Versatile Squeeze-and-Decomposition Network for Real-Time Image Fusion. *International Journal of Computer Vision*, **129**, 2761-2785. https://doi.org/10.1007/s11263-021-01501-8

[12] Ma, C. (2019) FusionGAN: A Generative Adversarial Network for Infrared and Visible Image Fusion. *Information Fusion*, **48**, 11-26. https://doi.org/10.1016/j.inffus.2018.09.004

[13] Li, S., Kang, X. and Hu, J. (2013) Image Fusion with Guided Filtering. *IEEE Transactions on Image Processing*, **22**, 2864-2875. https://doi.org/10.1109/TIP.2013.2244222

[14] Hui, L., Wu, X.J. and Kittler, J. (2018) Infrared and Vrisible Image Fusion Using a Deep Learning Framework. 2018 24*th International Conference on Pattern Recognition* (*ICPR*), Beijing, 20-24 August 2018, 2705-2710.

[15] Hui, L. and Wu, X.J. (2018) DenseFuse: A Fusion Approach to Infrared and Visible Images. *IEEE Transactions on Image Processing*, **28**, 2614-2623. https://doi.org/10.1109/TIP.2018.2887342

[16] Ma, J., Tang, L., Xu, M., Zhang, H. and Xiao, G. (2021) STDFusionNet: An Infrared and Visible Image Fusion Network Based on Salient Target Detection. *IEEE Transactions on Instrumentation and Measurement*, **70**, 1-13. https://doi.org/10.1109/TIM.2021.3075747

[17] Tang, L., Yuan, J. and Ma, J. (2022) Image Fusion in the Loop of High-Level Vision Tasks: A Semantic-Aware Real-Time Infrared and Visible Image Fusion Network. *Information Fusion*, **82**, 28-42. https://doi.org/10.1016/j.inffus.2021.12.004

[18] Li, X. (2021) RFN-Nest: An End-to-End Residual Fusion Network for Infrared and Visible Images. *Information Fusion*, **73**, 72-86. https://doi.org/10.1016/j.inffus.2021.02.023

[19] Tang, L., Yuan, J., Zhang, H., Jiang, X. and Ma, J. (2022) PIAFusion: A Progressive Infrared and Visible Image Fusion Network Based on Illumination Aware. *Information Fusion*, **83-84**, 79-92.

https://doi.org/10.1016/j.inffus.2022.03.007

[20] Xie, H., Zhang, Y., Qiu, J., Zhai, X., Liu, X., Yang, Y., Zhao, S., Luo, Y. and Zhong, J. (2023) Semantics Lead All: Towards Unified Image Registration and Fusion from a Semantic Perspective. *Information Fusion*, **98**, Article ID: 101835. https://www.sciencedirect.com/science/article/pii/S1566253523001513 https://doi.org/10.1016/j.inffus.2023.101835

[21] Tang, L., Liu, G., Xiao, G., Bavirisetti, D.P. and Zhang, X. (2022) Infrared and Visible Image Fusion Based on Guided Hybrid Model and Generative Adversarial Network. *Infrared Physics & Technology*, **120**, Article ID: 103914. https://doi.org/10.1016/j.infrared.2021.103914

[22] Liu, X., Wang, R., Huo, H., Yang, X. and Li, J. (2023) An Attention-Guided and Wavelet-Constrained Generative Adversarial Network for Infrared and Visible Image Fusion. *Infrared Physics & Technology*, **129**, Article ID: 104570. https://doi.org/10.1016/j.infrared.2023.104570

[23] Xu, H., Liang, P., Yu, W., Jiang, J. and Ma, J. (2019) Learning a Generative Model for Fusing Infrared and Visible Images via Conditional Generative Adversarial Network with Dual Discriminators. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI*-19, *International Joint Conferences on Artificial Intelligence Organization*, Macao, 10-16 August 2019, 3954-3960. https://doi.org/10.24963/ijcai.2019/549

[24] Ma, J., Xu, H., Jiang, J., Mei, X. and Zhang, X.P. (2020) DDcGAN: A Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion. *IEEE Transactions on Image Processing*, **29**, 4980-4995. https://doi.org/10.1109/TIP.2020.2977573

[25] Zhang, H., Yuan, J., Tian, X. and Ma, J. (2021) GAN-FM: Infrared and Visible Image Fusion Using GAN with Full-Scale Skipconnection and Dual Markovian Discriminators. *IEEE Transactions on Computational Imaging*, **7**, 1134-1147. https://doi.org/10.1109/TCI.2021.3119954

[26] Li, J., Huo, H.T., Li, C., Wang, R. and Feng, Q. (2020) AttentionFGAN: Infrared and Visible Image Fusion Using Attention-Based Generative Aidversarial Networks. *IEEE Transactions on Multimedia*, **23**, 1383-1396. https://doi.org/10.1109/TMM.2020.2997127

[27] Rao, Y., Wu, D., Han, M., Wang, T., Yang, Y., Lei, T., Zhou, C., Bai, H. and Xing, L. (2023) AT-GAN: A Generative Adversarial Network with Attention and Transition for Infrared and Visible Image Fusion. *Information Fusion*, **92**, 336-349. https://www.sciencedirect.com/science/article/pii/S156625352200255X https://doi.org/10.1016/j.inffus.2022.12.007

[28] Jie, H., Li, S., Gang, S. and Albanie, S. (2017) Squeeze-and-Excitation Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141.

[29] Toet, A. (2022) TNOimage Fusion Dataset. https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029

[30] Xu, H., Ma, J., Jiang, J., Guo, X. and Ling, H. (2022) U2fusion: A Unified Unsupervised Image Fusion Network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 502-518. https://doi.org/10.1109/TPAMI.2020.3012548