

# VSM 指导下 MSDNet 轻量型融合网络建模与仿真<sup>①</sup>



邹 韵<sup>1</sup>, 王 振<sup>2</sup>

<sup>1</sup>(安徽文达信息工程学院 通识教育学院, 合肥 231201)

<sup>2</sup>(安徽文达信息工程学院 人事处, 合肥 231201)

通信作者: 邹 韵, E-mail: 2783215249@qq.com

**摘 要:** 轻量级的图像融合算法对于人眼观察和机器识别有着重要的意义. 通过对视觉显著性在红外与可见光图像融合过程中的重要性研究, 在 SDNet 融合网络的基础上, 优化并设计了一种视觉显著图 (visual saliency map, VSM) 指导下的 MSDNet 轻量型融合网络. 首先, 通过对 SDNet 的结构以及通道数进行了缩减以加快训练及推理速度, 并利用结构参数化与反参数化技术增强轻量化模型的学习能力; 然后, 针对模型的训练, 使用了基于显著值图 VSM 指导的损失函数, 实现模型的自监督训练; 最后, 在训练结束时, 将图像重建分支删除, 根据卷积参数融合得到最终的轻量化模型. 实验表明, 该轻量化网络能够在保证图像融合质量的基础上, 极大地提高了融合的速率, 使其在移动端的移植成为可能.

**关键词:** 图像融合; 红外图像; 可见光图像; 轻量型网络; 显著损失函数

引用格式: 邹韵, 王振. VSM 指导下 MSDNet 轻量型融合网络建模与仿真. 计算机系统应用, 2024, 33(10): 133-139. <http://www.c-s-a.org.cn/1003-3254/9630.html>

## Modeling and Simulation of Light-weight Fusion Network MSDNet Guided by VSM

ZOU Yun<sup>1</sup>, WANG Zhen<sup>2</sup>

<sup>1</sup>(School of Liberal Studies, Anhui Wenda University of Information Engineering, Hefei 231201, China)

<sup>2</sup>(Personnel Department, Anhui Wenda University of Information Engineering, Hefei 231201, China)

**Abstract:** Light-weight image fusion algorithm is very important for human eye observation and machine recognition. By studying the importance of visual saliency in infrared and visible image fusion, a visual saliency map (VSM)-guided MSDNet fusion network is optimized and designed based on the SDNet fusion network. Firstly, the structure and channel numbers of SDNet are reduced to accelerate training and inference speed, and the learning ability of the light-weight model is enhanced by structural parameterization and reverse parameterization techniques. Then, for model training, the loss function guided by VSM is used to achieve model self-supervised training. Finally, at the end of the training, the image reconstruction branch is deleted. So the final light-weight model is obtained by the fusion of convolution parameters. Experiments show that the light-weight network can not only ensure image fusion quality but greatly improve the speed, making its porting in mobile terminals possible.

**Key words:** image fusion; infrared image; visible image; light-weight network; significant loss function

图像融合技术是当前实现图像增强任务的重要方法<sup>[1]</sup>, 其能够对多源图像进行处理使得融合图像具有不

同图像的优势. 作为其中关键的分支, 红外与可见光图像融合通过提取和组合红外图像的目标信息和可见光

① 基金项目: 校级自然科学基金一般项目 (XZR2023B05); 安徽省高等学校科学研究重点项目 (2023AH052821); 安徽省质量工程重点项目 (2022jyxm626)

收稿时间: 2024-03-12; 修改时间: 2024-04-10; 采用时间: 2024-05-06; csa 在线出版时间: 2024-08-21

CNKI 网络首发时间: 2024-08-22

图像的背景细节信息<sup>[2]</sup>,得到更符合人眼观察和机器识别的图像,受到了广泛的关注。

现有图像融合方法主要分为传统方法和基于深度学习的方法<sup>[3]</sup>。由于深度学习网络具有较强的泛化能力<sup>[4]</sup>和多层特征提取能力,已经成为图像融合技术领域发展的趋势。研究人员先后提出了两类基于深度学习的融合方法:一类是非端到端的融合方法,另一类则是端到端的融合方法。

非端到端方法主要是通过深度学习网络对源图像进行处理,其中较为经典的是Li等<sup>[5]</sup>的VGG方法,该方法通过VGG-19网络对分解得到的图像细节部分进行多层特征提取与融合,然而随着网络加深,并且使用了简单的平均融合策略,导致融合图像中冗余信息较多,特征不够明显<sup>[6]</sup>。为此,Li等<sup>[6]</sup>继续进行研究,通过使用ResNet50网络首先对图像进行深度特征的提取,然后改善了融合策略,利用ZCA运算将深度特征投影到稀疏域,通过L1范数运算得到初始权重图,期望通过重构得到较为理想的融合图像。

然而,上述方法得到的图像视觉效果并不理想,并且由于方法中使用的网络过于庞大,且使用后还需要大量的后处理,导致十分耗时。因此,研究人员提出了一些端到端的融合方法,该类方法主要是通过搭建一套完整的融合框架,使得可以直接通过训练处理得到融合图像。典型的有RFN-Nest方法<sup>[7]</sup>,该方法设计了一种基于残差结构的残差融合网络模型,并且针对融合模型进行了自动编码器和损失函数的训练<sup>[7]</sup>,然而该方法计算量较大,训练起来难度较大。参照自编码器网络结构,Zhang等提出了一种端到端融合方法SDNet<sup>[8]</sup>,该方法首次将压缩分解的思想引入到融合中,设计了一种通用的融合框架,其训练模型中不仅有图像重建的分支,而且能够利用不同的损失函数进行灵活地训练,是当前一种较为优秀的融合框架,虽然该方法可以进行实时融合,但是网络的计算量依然较大,在当今轻量化深度学习网络越来越流行的前提下,其在移动端依旧难以实现理想的融合效率。

基于上述分析,以实现轻量化的实时融合模型为目的,本文选择了以Zhang等<sup>[8]</sup>提出的SDNet通用框架为基础,做出了如下的改进。

(1)为了进行网络性能的优化,将SDNet的结构以及通道数进行了缩减,同时通过参数结构化及反结构化技术,保证了模型的特征学习能力。

(2)为了改善精简后网络的融合效果,选择了基于显著值图VSM指导的损失函数来指导模型的训练。

本文将改进后的网络模型称为MSDNet。

## 1 MSDNet 轻量型融合网络

一种有效的深度学习融合方法不仅需要设计出理想的网络模型结构来提供图像融合的框架,而且融合的效果和训练时所用的策略有着密切的联系。

### 1.1 网络结构优化

#### 1.1.1 结构简化及通道缩减

由于SDNet网络不具备下采样操作并且存在跨级连接,导致其在推理时特征图占用的存储空间和消耗的算力很大,不适合在边缘设备上实时推理,因此本文采用的精简策略是尽量减少网络的中间特征图数量,降低SDNet模型的参数量,具体是:如图1的模型对比示意图所示,将网络结构以及通道数进行了缩减,删除了融合分支的最后一个特征提取层,并将每层的卷积核数量由16减少到8。图1中, $m@n \times n$ 代表 $m$ 个 $n \times n$ 大小的卷积核,加粗代表MSDNet卷积参数,非加粗代表SDNet卷积参数。

在MSDNet模型中,共有3个输出,分别是融合图像、红外重建图像、可见光重建图像。训练结束后,其前向推理网络需要将后半部分的图像重构去除,以提升推理效率。

#### 1.1.2 参数结构化及反结构化

为了保证精简后的网络学习能力,考虑到网络中的所有操作都是 $3 \times 3$ 的卷积操作,所以借鉴文献[9]中的参数结构化操作,如图1中的加粗卷积,将网络中所有的 $3 \times 3$ 卷积层进行扩展,都重构成 $3 \times 3$ , $1 \times 1$ 和identity(残差层)的组合,将这3种操作的特征图相加后作为下一层的输入。

在训练结束之后,再通过一个反结构化操作将此结构替换成一个完全等效的 $3 \times 3$ 卷积操作,具体步骤如下。

(1) $3 \times 3$ 卷积核参不变。

(2)将 $1 \times 1$ 卷积核通过扩充,在四周补零,得到等效的 $3 \times 3$ 卷积核。

(3)identity操作相当于自身数据的复制,可以将其等效成 $1 \times 1$ 卷积操作:设特征图输入通道数为 $n$ ,则卷积核的尺寸是 $n \times n \times 1 \times 1$ ( $n$ 组 $n \times 1 \times 1$ 的向量),要使输出的特征图和输入的完全相同,那么就需要将第 $m$

( $m \leq n$ ) 组  $n \times 1 \times 1$  张量中第  $m$  个通道的值设为 1, 其余通道设为 0. 这样就将 identity 操作等效成  $1 \times 1$  卷积操作. 然后按照步骤 2, 将  $1 \times 1$  卷积再等效成  $3 \times 3$  卷积

操作.

(4) 最后利用卷积可加性原理, 将 3 组  $3 \times 3$  卷积相加, 得到合并后的  $3 \times 3$  卷积操作.

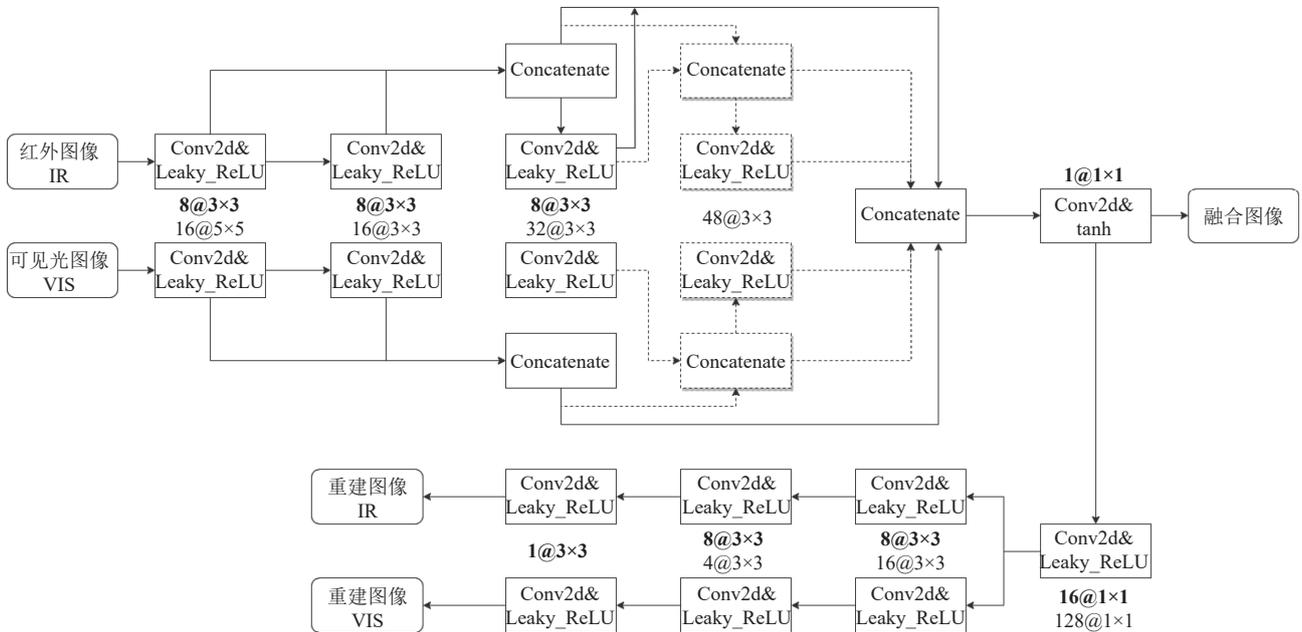


图 1 SDNet 和 MSDNet 网络的模型对比示意图

参数结构化和反结构化操作如图 2 所示.

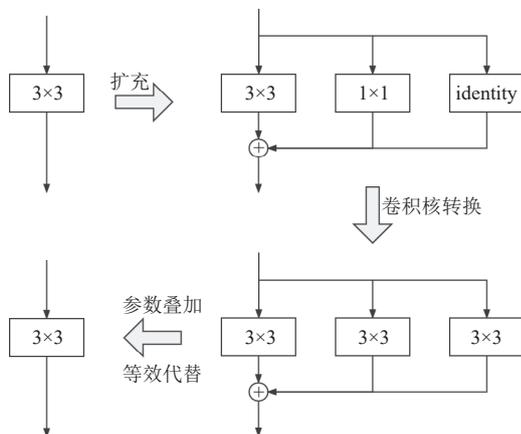


图 2 参数结构化与反结构化操作示意图

经过上述参数结构化和反结构化技术, 能大大提高网络中  $3 \times 3$  卷积的特征学习能力, 同时确保了最终的运算效率不变.

### 1.2 损失函数

像素是图像中最重要的元素, 像素的强度表示整体亮度分布, 反映了图像的对比度特征; 而像素之间的差异构成梯度, 代表了图像的纹理细节.

SDNet 方法将多模态图像融合建模为强度和梯度信息的提取和重建, 并在其后续的损失函数设计中选择了这种策略进行网络训练<sup>[8]</sup>. 而本文更加关注人眼的注意力和实际观察效果, 因此损失函数的构建主要是以 VSM 显著图指导的 VSM loss<sup>[10]</sup>, 使得输出的融合图像更加符合预期.

对于显著值图 VSM, 为了模型训练的高效和简洁性, 本文采用了 FT 显著值图求解方法<sup>[10]</sup>, FT 显著值图的公式如下:

$$S(x, y) = \|I_u - I_{\omega_{nc}}(x, y)\| \quad (1)$$

由于测试图像都是单通道图像, 因此对算法进行了简化: 首先对输入图像进行高斯滤波得到图像  $I_{gf}$ , 然后计算高斯滤波图像的整体均值  $m_g$ , 最后根据如下公式计算当前图像的显著值图  $S$ :

$$\begin{cases} t = (I_{gf} - m_g) \odot (I_{gf} - m_g) \\ S = (t - \min(t)) / (\max(t) - \min(t)) \end{cases} \quad (2)$$

其中,  $\odot$  代表图像矩阵逐像素相乘, 下同. 分别得到两种图像的显著值图  $S$  后, 按照式 (3) 计算得到两种图像的融合权重:

$$\begin{cases} score_{vis} = 0.5 + 0.5 \times (S_{vis} - S_{ir}) \\ score_{ir} = 0.5 + 0.5 \times (S_{ir} - S_{vis}) \end{cases} \quad (3)$$

根据图1, 在训练过程中, 有两个输入图像: 红外图像  $I_{ir}$  和可见光图像  $I_{vis}$ , 3 个输出图像: 融合  $I_{fused}$ 、红外重建图像  $De_{ir}$  和可见光重建图像  $De_{vis}$ 。

对于 VSM loss, 本文将其分为两部分。

(1) 首先利用原图和显著图得到的权重计算融合后的图像作为目标图像, 然后和网络输出的融合图像进行损失计算第 1 部分。

(2) 考虑到 VSM 融合时没有进行边缘的强化, 因此增加了 Laplace 梯度算子来计算原图和生成图像的梯度差异, 用来强化融合图像的边缘特征。下面是两个 VSM loss 的公式:

$$\begin{cases} l_{vsm1} = \frac{\sum_{i=1}^n ((I_{fused})_i - (score_{vis} \odot I_{vis} + score_{ir} \odot I_{ir}))_i^2}{n} \\ l_{vsm2} = \frac{\sum_{i=1}^n (\Delta(I_{fused})_i - \Delta(score_{vis} \odot I_{vis} + score_{ir} \odot I_{ir}))_i^2}{n} \end{cases} \quad (4)$$

其中,  $\Delta$  代表利用 Laplace 算子的梯度运算。

然后给出网络融合图像和源图像的直接损失公式  $l_{fused}$ :

$$l_{fused} = \frac{\sum_{i=1}^n ((I_{fused} - I_{ir})_i^2 + \lambda \times (I_{fused} - I_{vis})_i^2)}{n} \quad (5)$$

其中,  $\lambda$  表示可见光平均绝对误差的权重, 通过对输入的两种源图像分别和融合图像计算平均绝对误差来保证源图像的细节同时进入融合图像。

对于网络重建图像, 对其计算损失也能够通过反向传播使得网络前半部的提取特征能力增强。因此, 重建图像的损失函数如下所示:

$$l_{de} = \frac{\sum_{i=1}^n ((De_{ir} - I_{ir})_i^2 + (De_{vis} - I_{vis})_i^2)}{n} \quad (6)$$

式 (6) 反映的是重建图像重构之后与源图像的差异, 差异越小, 说明网络前半部分的特征提取层的能力就越强。

最后, 将几种损失函数组合在一起, 得到最终的损失函数  $l$ :

$$l = \alpha \times l_{vsm1} + \beta \times l_{vsm2} + l_{fused} + l_{de} \quad (7)$$

### 1.3 网络模型训练

由于本文方法的训练模型和推理模型有所区别, 训练模型后部含有图像重构部分, 而推理模型则需要将图像重构部分去除以获得最大的推理速度。另外, 训练的过程中主要靠 loss 函数来加以指导, 本文的 loss 函数是通过输入数据提取一定的信息得到确定的优化方向, 因而达到了自监督训练, 现给出该网络的整体训练和部署流程如算法 1。

算法 1. 网络训练过程

输入. 红外图像  $I_{ir}$ , 可见光图像  $I_{vis}$ 。

输出. 融合图像  $I_{fused}$ 。

(1) 按照方法中的模型结构搭建融合网络 MSDNet\_train。

(2) 预处理两种输入图像:

1) 源图像转换成浮点格式;

2) 按照公式  $input = (image - mean) / std$  得到输入网络的图像, 其中  $image$  为源图像,  $mean = 127.5$ ,  $std = 127.5$ 。

3) 预处理后的两种源图像, 输入融合网络 MSDNet, 得到融合图像  $I_{fused}$ , 红外重建图像  $De_{ir}$ , 可见光重建图像  $De_{vis}$ 。

4) 将 Step 3 得到的图像, 按照式 (7) 中的损失函数公式, 计算损失 loss, 并设置网络进行反向传播以学习并优化网络参数  $params$ 。

5) 训练循环  $n$  个 epochs 后 (本文  $n = 20$ ), 得到最终的参数进行保存。

6) 将训练模型 MSDNet\_train 中  $I_{fused}$  之后的重构部分删除, 重新构建 MSDNet 的推理模型, 并读入相应的网络参数  $params$ , 得到最终的推理模型 MSDNet\_forward。

7) 将模型 MSDNet\_forward 部署到相应设备, 进行融合推理。

## 2 实验和分析

为了验证本文方法改进的有效性, 将其直接与 SDNet 方法<sup>[8]</sup>进行了比较, 此外将本文方法与 VGG<sup>[5]</sup>、ResNet50\_zca4<sup>[6]</sup>、ResNet50\_zca5<sup>[6]</sup>以及 RFN-Nest<sup>[7]</sup>方法进行详细对比, 说明本文方法的优越性。

### 2.1 实验数据集和训练参数选择

本文实验的训练集采用了 TNO 红外与可见光数据集<sup>[11]</sup>, 测试集选择了文献<sup>[12]</sup>中已使用过的 44 对灰度图像<sup>[12]</sup>。

实验环境上, 选择了 RTX3080 (显存 10 GB) 作为训练主要硬件, 以最大限度加快训练速度。在训练参数选择上, 根据 SDNet 的训练经验<sup>[8]</sup>, 本文选择了网络训练的循环次数 epochs 为 50, batch size 大小为 16。网络训练图像统一用双三次插值法缩放到  $288 \times 384$  的大小; 对于可见光平均绝对误差的权重  $\lambda$ , 为了得到最佳融合质量, 选择了与 SDNet 同样的参数大小<sup>[8]</sup>, 即  $\lambda = 0.5$  (式 (5)); 对于 VSM 损失函数的权重  $\alpha$  和  $\beta$ , 本文选择了

平均两个损失的权重,采用 $\alpha = \beta = 40$  (式 (7)).

## 2.2 实验结果与评价

### 2.2.1 参数量、FLOPs 及平均测试时间对比

基于测试集中的 44 对图像,本文记录了 MSDNet、SDNet 与 RFN-Nest 融合方法的测试总时间,并计算出了单组图像平均测试时间,另外计算了参数量,及用统一的 224×224 图像尺寸计算 FLOPs 值 (见表 1).

表 1 参数量、FLOPs 值及平均测试时间

指标	RFN-Nest	SDNet	MSDNet
参数量 (M)	0.4185	0.0671	0.0067
FLOPs (G)	21.0466	3.3514	0.3372
平均时间 (s)	3.9890	0.2657	0.0798

相比于 SDNet,由于本文方法对其结构以及通道数进行了缩减,利用参数和反结构化技术,有效地降低了整体的参数量和 FLOPs 值 (约是其 10 倍),从而大大地减少了方法在单组图像上的平均测试时间 (提升了约 3.3 倍);同时相比于 RFN-Nest,本文方法在参数量上约是其 1/60,融合速度提升了约 50 倍,这说明了本文方法在提升效率上的有效性,可以保证图像的实时融合,为其在移动端的移植提供了可能.

### 2.2.2 主观效果和客观指标对比

本节分别从主观视觉分析和客观指标评价上对本文方法的有效性和优越性进行详细说明.

在有效性方面,利用消融实验的思想,与 SDNet 方法进行: (1) 不改动训练方法情况下的比较; (2) 添加参数结构化增强和 VSM 损失函数情况下的比较.

在优越性方面,与其余方法进行主观视觉效果和客观指标数据的综合比较.本文采用了单一指标信息熵 EN<sup>[13]</sup>、平均梯度 AG<sup>[14]</sup>、综合指标差异相关和 SCD<sup>[15]</sup>、互信息 MI<sup>[16]</sup>和交叉熵 CE<sup>[17]</sup>,前 5 项指标值越大,CE 指标值越小,说明融合效果越好<sup>[17]</sup>.

如图 3 所示,以 3 种不同场景下的图像对为例,给出了不同融合方法的融合结果,其中 ori 代表用 SDNet 的训练方式训练 MSDNet,rep 代表使用了参数结构化增强, vsm 代表利用提出的 VSM 损失函数训练 MSDNet.

(1) 结合 SDNet 和 MSDNet (ori),可以说明在相同的训练步骤下,两者的融合结果差距不大,这说明虽然对网络模型进行了精简,但是并没有过多影响图像融合的效果;与 MSDNet (ori+rep) 相比,融合图像的质量有所提高,尤其是在添加了 VSM 损失函数之后,融合

图像的质量有了明显的提升,其图像能很好地兼顾双光图像各自明显的明暗特征,如偏白的天空,照射的灯光等,而原始 SDNet 主观上更加偏重于红外特征,并且 MSDNet (ori+rep) 图像更加平滑且观感自然,图像的边缘特征也有了一定程度的增强,虽然相对于 SDNet 来说还是有些偏弱,这是因为采用的 VSM 损失函数更加注重融合结果的对比度.

(2) 本文方法相对于 VGG、ResNet50\_zca4、ResNet50\_zca5 以及 RFN-Nest 方法,在主观效果图上也具有明显的优越性,更加接近理想的融合图像.

本文计算了几种 MSDNet 和其余几种方法的指标值,为了降低极端数据对结果比较的影响,对去最值后的均值进行对比,如表 2 所示,本文所使用的 SDNet 为自行训练,训练数据和超参数均和 MSDNet 相同,其余对比方法中的参数设置均和本文方法保持一致.根据表 2 可得以下结论.

(1) 随着参数结构化增强、VSM 损失函数的添加,MSDNet 方法在 5 项指标均值上都在不断提升,且当添加 VSM 损失函数时,提升的幅度总体较大.

(2) 本文方法在大部分指标均值上虽然没有超过 RFN-Nest 方法,但从主观视觉角度考虑,该方法对于源图像信息的综合,过多地偏向红外图像,这可能是导致指标值高的原因,而本文方法相对于 SDNet 一类的网络来说,优势较为明显,最终优化得到的 MSDNet 取得了其余方法中的最优值,相对于 SDNet,MSDNet 在 EN、SCD、MI 和 CE 指标上分别提升了 1.24%、9.79%、1.24% 和 6.76%,虽然在 AG 指标上还有所不足,但也取得了比 RFN-Nest 更高的指标值,说明了方法具有一定的边缘保持能力.

## 3 结论与展望

一种良好的网络结构和合适的融合策略对于提取和组合图像信息具有非常重要的作用.本文在 SDNet 的研究基础上,为了进一步提高融合网络模型的计算效率,改善方法融合的效果,优化并设计了一种 VSM 指导下的 MSDNet 轻量型融合网络.实验验证了本文方法的有效性和优越性,所得结论如下.

(1) 本文对 SDNet 的结构以及通道数进行了缩减,并使用了参数结构化及反结构化操作,使得网络能够更加快速地提取到充分的源图像特征.

(2) 选择了基于显著值图 VSM 指导的损失函数,来指导模型的训练,从而改善了融合图像的视觉效果.

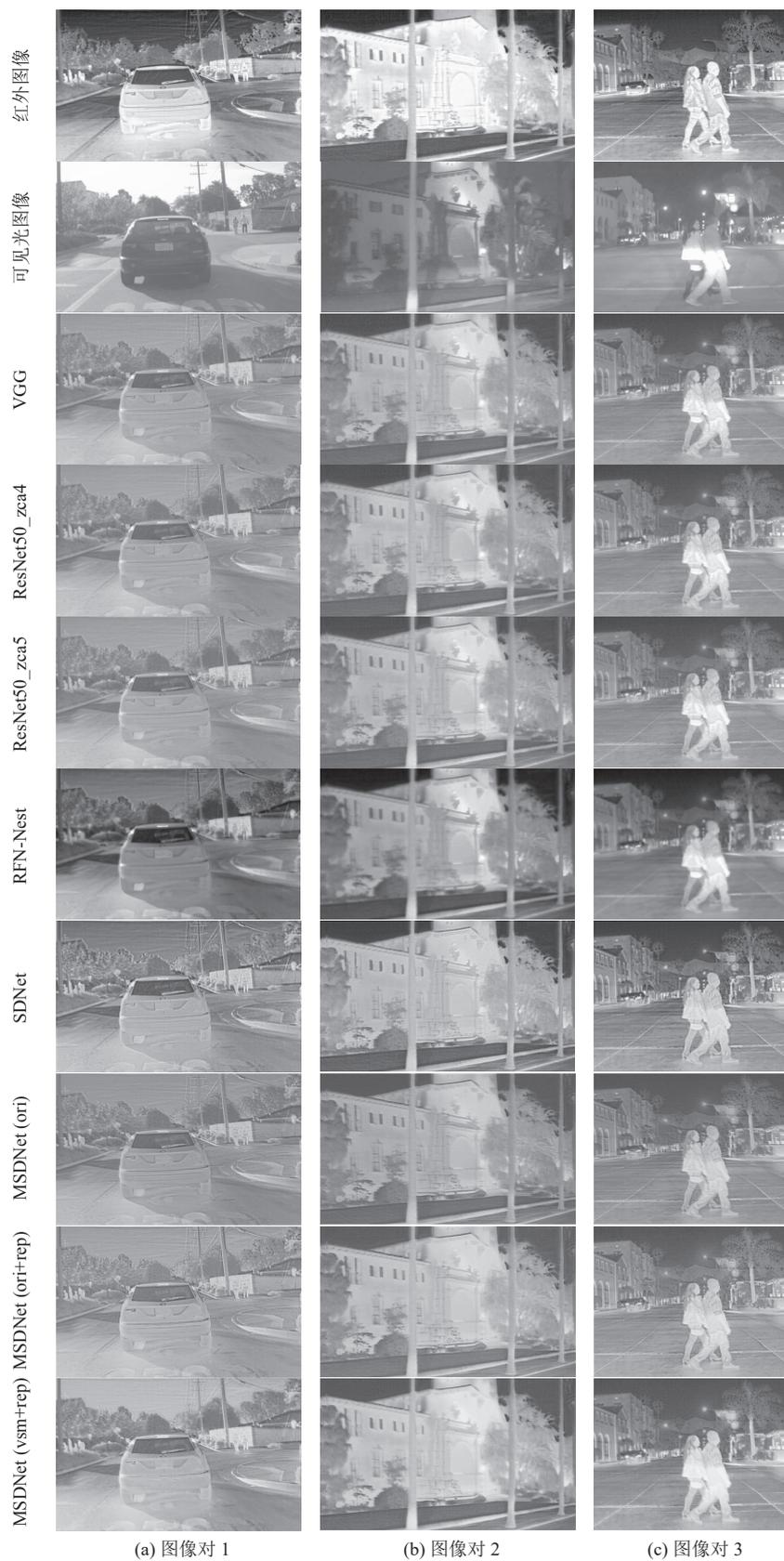


图3 不同融合方法的结果图

表2 不同融合方法的各项均值比较

实验类型	方法	EN	AG	SCD	MI	CE
对比实验	VGG	6.8394	3.8099	1.3566	13.6787	1.6082
	ResNet50_zca4	6.8422	3.6625	1.3492	13.6844	1.6228
	ResNet50_zca5	6.8330	3.6473	1.3503	13.6661	1.6365
	RFN-Nest	<b>7.3224</b>	3.2325	<b>1.7204</b>	<b>14.6449</b>	<b>1.3777</b>
消融实验	SDNet	6.8646	<b>4.9388</b>	1.3175	13.7293	1.6752
	MSDNet (ori)	6.6632	3.8367	1.1454	13.3264	2.1404
	MSDNet (ori+rep)	6.7781	4.5216	1.2048	13.5562	1.8993
	MSDNet (vsm+rep)	<u>6.9499</u>	4.5340	<u>1.4465</u>	<u>13.8998</u>	<u>1.5619</u>

注: 粗体表示最优值, 下划线表示SDNet一类方法中的最优值

结合具体指标和融合图像来做综合判断, 本文改进后的 MSDNet 无论在性能上还是实际效果上都有较好的优势, 这对于移动端等一些低算力平台来说, 是一种不错的融合方法。

由于本文仅在红外与可见光图像数据集进行了测试比较, 未考虑进行其他类型图像的融合, 因此下一步将进行本文方法对于其他类型图像的适用性研究。

### 参考文献

- 李萌萌. 图像融合技术在图形融合中的应用研究. 硅谷, 2014, 7(6): 36–37, 51.
- 齐海生, 荣传振, 肖力铭, 等. 基于双树复小波变换与引导滤波的红外与可见光图像融合算法. 通信技术, 2019, 52(2): 330–336. [doi: [10.3969/j.issn.1002-0802.2019.02.012](https://doi.org/10.3969/j.issn.1002-0802.2019.02.012)]
- 王宇庆, 王索建. 红外与可见光融合图像的质量评价. 中国光学, 2014, 7(3): 396–401. [doi: [10.3788/CO.20140703.0396](https://doi.org/10.3788/CO.20140703.0396)]
- 王曙. 深度学习算法研究及其在图像分类上的应用 [硕士学位论文]. 南京: 南京邮电大学, 2016.
- Li H, Wu XJ, Kittler J. Infrared and visible image fusion using a deep learning framework. Proceedings of the 24th International Conference on Pattern Recognition. Beijing: IEEE, 2018. 2705–2710.
- Li H, Wu XJ, Durrani TS. Infrared and visible image fusion with ResNet and zero-phase component analysis. Infrared Physics & Technology, 2019, 102: 103039. [doi: [10.1016/j.infrared.2019.103039](https://doi.org/10.1016/j.infrared.2019.103039)]
- Li H, Wu XJ, Kittler J. RFN-Nest: An end-to-end residual fusion network for infrared and visible images. Information Fusion, 2021, 73: 72–86. [doi: [10.1016/j.inffus.2021.02.023](https://doi.org/10.1016/j.inffus.2021.02.023)]
- Zhang H, Ma JY. SDNet: A versatile squeeze-and-decomposition network for real-time image fusion. International Journal of Computer Vision, 2021, 129(10): 2761–2785. [doi: [10.1007/S11263-021-01501-8](https://doi.org/10.1007/S11263-021-01501-8)]
- Ding XH, Zhang XY, Ma NN, *et al.* RepVGG: Making VGG-style ConvNets great again. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13728–13737.
- Achanta R, Hemami S, Estrada F, *et al.* Frequency-tuned salient region detection. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 1597–1604. [doi: [10.1109/CVPR.2009.5206596](https://doi.org/10.1109/CVPR.2009.5206596)]
- Toet A, Hogervorst MA. Progress in color night vision. Optical Engineering, 2012, 51(1): 010901. [doi: [10.1117/1.OE.51.1.010901](https://doi.org/10.1117/1.OE.51.1.010901)]
- Davis JW, Keck MA. A two-stage template approach to person detection in thermal imagery. Proceedings of the 7th IEEE Workshops on Applications of Computer Vision (WACV/MOTION 2005) —Vol. 1. Breckenridge: IEEE, 2005. 364–369. [doi: [10.1109/ACVMOT.2005.14](https://doi.org/10.1109/ACVMOT.2005.14)]
- Roberts JW, van Aardt JA, Ahmed FB. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. Journal of Applied Remote Sensing, 2008, 2(1): 023522. [doi: [10.1117/1.2945910](https://doi.org/10.1117/1.2945910)]
- Cui GM, Feng HJ, Xu ZH, *et al.* Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. Optics Communications, 2015, 341: 199–209. [doi: [10.1016/j.optcom.2014.12.032](https://doi.org/10.1016/j.optcom.2014.12.032)]
- Aslantas V, Bendes E. A new image quality metric for image fusion: The sum of the correlations of differences. AEU-International Journal of Electronics and Communications, 2015, 69(12): 1890–1896. [doi: [10.1016/j.aeu.2015.09.004](https://doi.org/10.1016/j.aeu.2015.09.004)]
- Qu GH, Zhang DL, Yan PF. Information measure for performance of image fusion. Electronics Letters, 2002, 38(7): 313–315. [doi: [10.1049/el:20020212](https://doi.org/10.1049/el:20020212)]
- Haghighat MBA, Aghagolzadeh A, Seyedarabi H. A non-reference image fusion metric based on mutual information of image features. Computers & Electrical Engineering, 2011, 37(5): 744–756. [doi: [10.1016/j.compeleceng.2011.07.012](https://doi.org/10.1016/j.compeleceng.2011.07.012)]

(校对责编: 孙君艳)