引用格式：FENG Xin, YANG Jieming, ZHANG Hongde, et al. Infrared and Visible Image Fusion Based on Dual Channel Residual Dense Network[J]. Acta Photonica Sinica, 2023, 52(11):1110003
冯鑫,杨杰铭,张鸿德,等.基于双通道残差密集网络的红外与可见光图像融合[J].光子学报,2023,52(11):1110003

基于双通道残差密集网络的红外与可见光图像融合

冯鑫,杨杰铭,张鸿德,邱国航
(重庆工商大学 机械工程学院 制造装备机构设计与控制重庆市重点实验室,重庆 400067)

摘　要:为改善红外与可见光融合结果与源图像间的部分细节特征丢失问题,充分提取红外与可见光图像中的特征信息,提出了一种改进的双通道深度学习自编码网络进行红外与可见光图像融合。其中,双通道结构由密集连接和残差连接模块级联构成,并设置一种综合像素、结构相似度和梯度特征保留的损失函数,使该编码器结构可以充分提取红外与可见光图像的多层次特征,在融合层采用空间L1范数和注意力机制对级联双通道特征分别进行融合,最后设计对应的解码器对融合特征图像进行重构,获取最终的融合结果。通过与传统算法以及近年最新的深度学习算法进行实验对比,结果表明该方法在主观和客观上都具有优秀的综合性能。

关键词:红外与可见光图像融合;双通道网络;残差密集模块;注意力机制;自编码器
中图分类号:TP391　　**文献标识码**:A　　**doi**:10.3788/gzxb20235211.1110003

0　引言

图像融合是图像处理的中的一项重要任务,它需要运用融合策略将不同模态传感器获得的图像的重要特征信息进行融合,使得单幅图像具有源图像的显著特征[1]。红外与可见光图像融合是图像融合中一个具有挑战性的任务,可见光图像由传感器捕获反射光形成,符合人眼的成像规律,并且包含了丰富的纹理与细节信息,缺点是容易受到光照环境影响,穿透性不足;红外光图像由传感器捕获热辐射信息形成,相比于可见光图像,它具有较高的对比度,并且具有很强的穿透性,不易受到天气与光线环境影响[2]。将红外光图像与可见光图像的优势结合在一起,生成的具有高对比度且具有良好纹理信息的融合图像在夜间导航[3]、地下勘探[4]以及目标检测[5]等众多领域都具有良好的应用前景。

目前,图像融合算法分为传统和深度学习两个大类。传统方法主要通过在变换域或空间域手动进行活动水平测量以及融合规则的设计完成图像融合。例如在变换域中将多尺度变换与稀疏表示两者相结合的通用方法MST-SR[6],以及基于空间域的梯度转移融合方法(Gradient Transfer Fusion,GFT)[7]。尽管现有的传统算法已经能够取得较好的融合结果,但是传统方法在图像融合过程中需要手动进行融合规则的设计,该过程大大提高了整个算法的工作量。

深度学习方法凭借着强大的特征提取能力在计算机视觉任务中展现出了显著的性能优势,在图像融合任务中也不例外。基于深度学习的融合方法主要有特征提取、特征融合、图像重建三个步骤,各种算法完成这三个步骤的方式也不尽相同。LIU Yu等[8]通过训练卷积神经网络(Convolutional Neural Network,CNN)模型用于联合生成活动水平测量和融合规则,克服了传统图像融合方法所面临的困难,深度学习逐渐开始进入图像融合领域。MA Jiayi等[9]在损失函数中引入显著目标掩码,能够较好的提取红外与可见光图像中的显

基金项目:国家自然科学基金(No. 22178036),重庆市高校创新研究群体项目(No. CXQT21024),重庆市自然科学基金项目(No. CSTB2022NSCQ-MSX0271)
第一作者:冯鑫,149495263@qq.com
通讯作者:杨杰铭,2871600119@qq.com
收稿日期:2023-05-19;**录用日期**:2023-06-13

著区域以精确指导网络的优化，但其更加关注于显著特征而忽视了细节特征的发掘。2019年，MA J等[10]将图像的融合过程转化为生成器和判别器的博弈过程，将生成对抗网络（Generative Adversarial Network，GAN）模型应用到图像融合任务中。随后，MA Jiayi等又进一步的对该融合模型进行优化，分别提出了双鉴别器（Dual Discriminator Conditional Generative Adversarial Network，DDcGAN）[11]和一个多分类鉴别器（Multiclassification Constraints Generative Adversarial Network，MccGAN）[12]的生成对抗网络红外与可见光融合模型。然而基于生成对抗网络算法的模型结构复杂，训练的成本高，运用生成器获得包含更多源图像特征的融合图像比较困难。而在自编码网络方面，LI Hui等[13]提出将自动编码器框架（Auto-Encoder，AE）应用于图像融合，并使用密集连接层来获取更多的图像特征，其后，XU Han等[14]通过编码器将图像分解为场景信息和传感器的模态信息，通过不同的融合策略将其融合，获得了较好的视觉效果；LI Hui等[15]通过提出一种新的训练策略，分两阶段训练解码器和融合层，实现了自动编码器的端到端融合。上述方法中的神经网络结构大多都使用单一的骨干网络结构获得图像的特征，然而使用单一网络对图像进行特征提取可能会导致特征提取的固化，从而忽视掉部分特征。

基于上述分析，为了通过神经网络结构获得更多源图像的特征信息，本文结合双通道并联的神经网络思路设计了一种基于自动编码器网络框架的红外与可见光图像融合算法，并且在融合阶段，根据双通道特征的特点，对基于注意力模型的融合策略进行改进。

1 双通道级联残差密集网络框架

1.1 基于双通道级联的自动编码器框架

通用的基于自动编码器的红外与可见光图像融合框架如图1所示。该网络主要由编码器和解码器组成，原理是通过编码器对输入进行编码或者特征提取，再由解码器还原输入，属于自监督学习网络。在图像融合任务中，将编码器用于提取红外与可见光图像信息，中间加入融合层融合两者的特征，最后由解码器重建融合图像。在训练的过程中只需训练编码器的特征提取能力以及解码器的重建能力，不需要融合层，因此无需红外与可见光数据集进行训练，在网络训练完成后加入融合层即可获得融合图像。

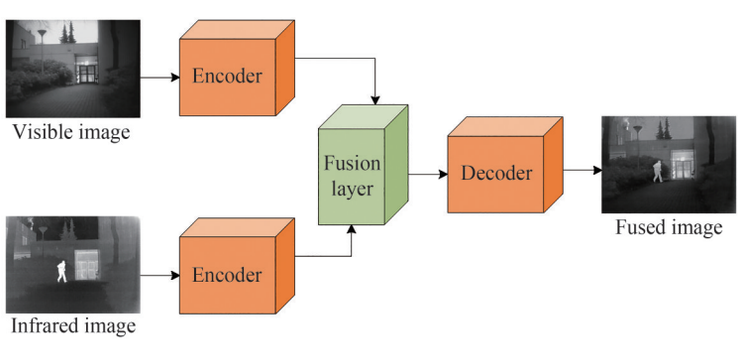


图1 自动编码器结构框架

Fig.1 Auto-Encoder structure framework

红外与可见光图像的多尺度特征是决定融合图像质量的重要因素。因此，本文引入双通道级联网络（Dual Path Networks，DPN）[16]的设计思想，结合残差网络（Residual Network，ResNet）[17]与密集连接网络（Densely Convolutional Network，DenseNet）[18]的优势来设计编码器的主干网络结构，该网络结构如图2所示。

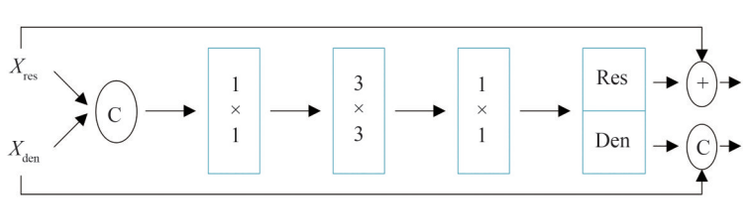


图2 编码器网络结构

Fig.2 Encoder network structure

图中“+”表示线性相加，“C”表示维度拼接。Res为残差特征，残差连接模块使用跳层连接，将同一残差块的输入特征连接到输出中，以保证在保留输入特征上具有较好的效果。Den为密集连接块特征，密集连接模块将输入特征与之前每一层的输出特征进行连接，使得每个密集连接块能够获得之前所有模块的信息，该模块擅长在之前模块特征的基础上发掘新的特征。双通道网络模块将输入分成两条路径，同时进行密集连接与残差连接，能够有效的实现特征的重用和发掘，相比于残差连接和密集连接拥有更低的计算成本和更高的参数效率。

基于残差连接与密集连接网络强大的特征提取能力，本文提出一种双通道并联模块设计自动编码器的编码器与解码器结构（Dual Paths Cascade Auto-Encoder，DPCAE），其详细的网络架构如图3所示。编码器由一层卷积层和双通道级联模块构成，通过残差以及密集连接通道级联的方式提取并传递源图像的特征信息，实现通道间的信息共享。融合层采用空间L1范数和注意力机制分别融合残差与密集通道特征。解码器根据编码器的特点，对密集以及残差特征进行不同的处理，高维的密集特征层数更深，则使用更多的卷积采样层进行特征的还原，低维的残差特征层数较浅，则减少卷积层的数量，通过这种方式将不同通道和层次的特征相结合，从而获得源图像中更加详细丰富的细节信息。

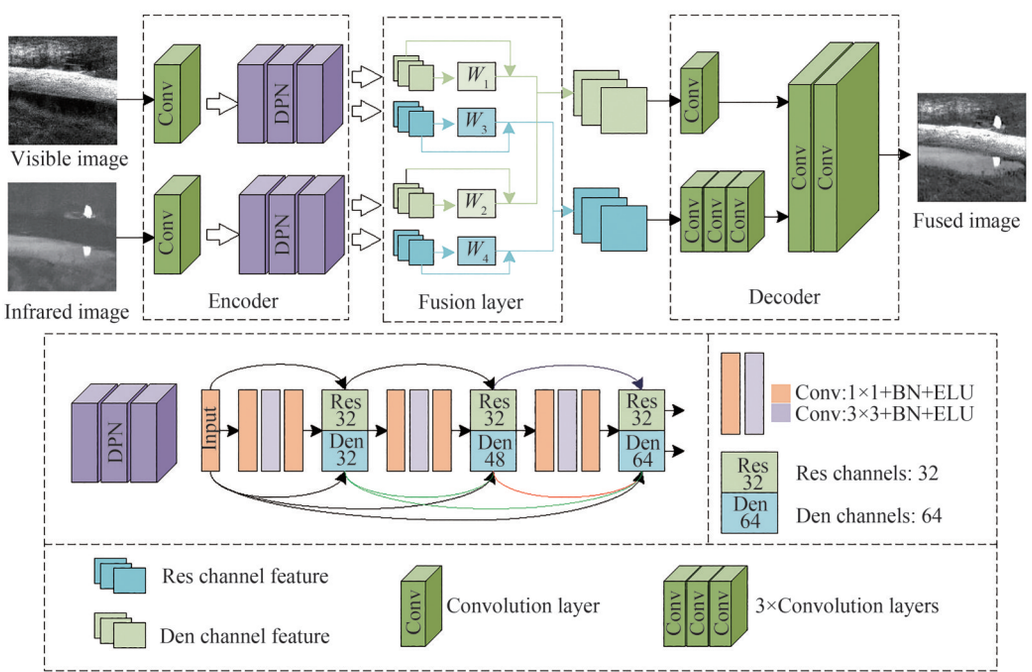


图3　基于双通道级联的残差密集网络融合框架

Fig.3　A Residual dense network fusion framework based on dual-channel cascade

双通道级联模块由三个双通道层构成，每个双通道层的输入与输出层采用步长为1，卷积核为1×1的卷积层调节特征维度；中间层采用卷积核大小为3×3，步长为1的卷积层进行特征提取。与此同时，为了减少参数量，提高网络的效率同时防止过拟合，在卷积层之后加入池化（Batch Normalization，BN）[19]操作，然后通过ElU激活函数提高网络的效果。根据双通道的编码器结构，将解码器的结构设计为双通道分别处理后结合，图像信息通过融合之后，输出的残差特征通道数为32，密集特征通道数为64，分别通过1层和3层卷积降低特征维度，随后对残差通道与密集通道进行维度拼接，经过两层卷积之后在最后一层使用tanh激活函数。为了避免图像信息的丢失，在整个网络结构尽量避免了使用数据归一化运算，同时较多使用激活函数提高网络的拟合效果，这也是ELU激活函数的优点，能够在进行少量归一化运算的条件下达到Relu激活函数的效果[20]。

网络结构如表1所示，其中，size表示卷积核的大小，stride表示步长，input与output的值表示输入与输出的通道数，activation表示该层使用的激活函数，×3表示三个相同的层。

表1 网络结构表

Table 1 The table of network structure

		Layer	Size	Stride	Input	Output	Activation
Encoder		Conv	3	1	1	16	ELU
		Dual path block×3	—	—	—	—	—
Decoder	Res	Conv	3	1	32	16	ELU
	Den	Conv	3	1	64	64	—
		Conv	3	1	64	32	—
		Conv	3	1	32	16	ELU
	General	Conv	3	1	32	16	—
		Conv	3	1	16	1	Tanh
Dual path block		Conv	1	1	Res:16 Den:16	32 32	ELU
		Conv	3	1	Res:32 Den:32	32 32	ELU
		Conv	1	1	Res:32 Den:32	32 32	ELU

1.2 损失函数

损失函数的设置对于图像融合的结果非常重要，自编码网络通过损失函数计算重建图像与源图像的损失来监督学习过程，为了更好的完成图像的重建，本文引入像素损失L_{pix}、结构相似性损失L_{ssim}以及梯度损失L_{gra}来综合设置损失函数L_{total}。其中像素损失主要用于约束源图像和融合结果之间的总体像素水平，结构相似度用于约束两者之间的亮度对比度和结构信息，梯度损失用于约束梯度信息和纹理特征。损失函数如式(1)所示，它们的关系为

$$L_{\mathrm{total}}=\alpha L_{\mathrm{pix}}+\beta L_{\mathrm{ssim}}+\gamma L_{\mathrm{gra}} \tag{1}$$

式中，α、β、γ分别用于调节像素损失、结构相似性损失和梯度损失之间的平衡，使训练的解码器和编码器在应对红外与可见光图像时能够平衡的获得红外与可见光信息，其取值由消融性实验获得。

像素损失L_{pix}通过融合图像与源图像的像素误差计算，表示为

$$L_{\mathrm{pix}}=\frac{1}{HW}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}\left(Y(i,j)-X(i,j)\right)^2 \tag{2}$$

式中，H和W分别表示图像的高和宽，Y表示输出图像，X表示输入图像。

图像的梯度特征中包含了大量的纹理特征信息，在训练自动编码器时可以通过减小梯度损失来达到保留源图像纹理细节信息的目的，梯度损失L_{gra}表示为

$$L_{\mathrm{gra}}=\frac{1}{HW}\left\|\nabla Y-\nabla X\right\| \tag{3}$$

式中，∇表示对图像进行梯度计算，参考文献[21]的处理，运用Sobel算子对图像进行梯度计算。

结构相似性损失L_{ssim}主要通过亮度、对比度、结构来对比输出图像与源图像之间的差异，结构相似性损失可以通过式(4)计算。

$$L_{\mathrm{ssim}}=1-\mathrm{SSIM}(X,Y) \tag{4}$$

$$\mathrm{SSIM}(X,Y)=\frac{2\mu_X\mu_Y+C_1}{\mu_X^2+\mu_Y^2+C_1}\cdot\frac{2\sigma_X\sigma_Y+C_2}{\sigma_X^2+\sigma_Y^2+C_2}\cdot\frac{\sigma_{XY}+C_3}{\sigma_X\sigma_Y+C_3} \tag{5}$$

式中，SSIM表示输入与输出图像的结构相似度[22]，μ_X、μ_Y为图像X、Y的均值，σ_X、σ_Y为X、Y的标准差，σ_{XY}是X与Y的协方差，C_1、C_2、C_3是用于保证公式稳定性的足够小的常数，结构相似性从更符合人类视觉感知的图像内容和结构角度出发，与像素损失和梯度损失相互补足。

1.3 融合策略

自动编码图像融合框架属于非端到端的图像融合框架，编码器的输出通过融合层后的特征通道数不

变，通过设计合适的融合策略将红外与可见特征融合之后将作为解码器的输入。目前，融合策略常见的主要如加法、最大值和平均值等方法。但是，加法、最大值和均值的融合策略是比较直接的图像融合策略，这类融合策略相对来说粗糙，对于一些目标和背景复杂的图像融合质量不佳，并且，由于编码器提取出的双通道特征具有不同的特点，因此本文在双通道网络的基础上改进了融合策略，通过空间L1范数和注意力机制对残差与密集通道的特征分别进行融合。

1）残差通道特征的融合策略：残差通道输出特征的特点是突出显著信息，目标特征明显，并且特征的层次低，通道数少。因此，在融合残差通道特征时更强调特征图的空间信息，将L1范数与softmax运算引入该融合策略，通过L1范数计算初始活动水平度量的值，随后运用块平均算法计算得到最终活动水平测量值C_i^m，如式(6)所示。

$$C_i^m(x,y)=\frac{\sum_{a=-r}^{r}\sum_{b=-r}^{r}\left\|\phi_i^m(x+a,y+b)\right\|_1}{(2r+1)^2} \tag{6}$$

式中，ϕ_i^m表示第i个特征图（包括红外特征图与可见光特征图，即$k=2$）的m（$m=1,2,\cdots,M=32$）特征维度，(x,y)表示图中对应坐标。r为区块的大小，根据参考文献[9]设置$r=1$。最后通过softmax回归运算得到特征图的权重信息w_i^m，如式(7)，通过权重计算得到网络的融合特征图f_{res}^m，如式(8)所示。

$$w_i^m(x,y)=\frac{C_i^m(x,y)}{\sum_{i=1}^{k}C_i^m(x,y)} \tag{7}$$

$$f_{\text{res}}^m(x,y)=\sum_{i=1}^{k}w_i^m(x,y)\times\phi_i^m(x,y) \tag{8}$$

2）密集通道特征的融合策略：由于编码器提取的红外与可见光信息往往是多通道的，在融合目标图像时需要考虑到空间信息和通道信息，而基于空间L1范数的融合策略更加侧重于计算空间信息的融合。因此，本文参考文献[23]的处理方式，引入通道注意力机制进行通道信息的融合。通过通道均值运算和softmax回归运算获得红外与可见光特征通道的权重信息w_i^n，如式(9)所示。

$$w_i^n(x,y)=\frac{\sum_{n-r}^{n+r}\phi_i^n(x,y)}{\sum_{i=1}^{k}\sum_{n-r}^{n+r}\phi_i^n(x,y)} \tag{9}$$

式中，ϕ_i^n表示第i个特征图的n（$n=1,2,\cdots,N=64$）特征维度，设置$r=1$（边界条件：$n-r=0$时取1，$n+r=65$时取64）。通过权重信息w_i^n进行权重计算即可得到融合特征图f_{den}^n，如式(10)所示。

$$f_{\text{den}}^n(x,y)=\sum_{i=1}^{k}w_i^n(x,y)\times\phi_i^n(x,y) \tag{10}$$

最后，分别将残差融合特征图$f_{\text{res}}^{1:M}$以及密集融合特征图$f_{\text{den}}^{1:N}$进行维度拼接得到双通道融合结果特征F，如式(11)所示，式中$\oplus$表示维度拼接。

$$F=f_{\text{res}}^{1:M}(x,y)\oplus f_{\text{den}}^{1:N}(x,y) \tag{11}$$

2 实验结果与分析

2.1 实验平台及参数设置

在训练自编码器时，根据自编码网络本身不具备图像融合的能力并且不会对最终的融合性能造成影响的这一特点，由于MS-COCO数据集相对于已配准的红外与可见光数据集的场景的覆盖面更广，因此选用MS-COCO数据集[24]的5 000图像作为训练集，并且将图像的大小缩放为224×224。选用TNO数据集[25]的42组已配准的红外与可见光数据集作为测试数据。训练参数设置为：训练批量为32，训练迭代次数为10，初始学习率为10^{-4}并随着迭代次数下降，损失函数中α设为10，β设为30，γ设为15。在训练过程中去掉融合层，输入为5 000张彩色图像；在测试过程中加入融合层，输入为成对的红外与可见光图像，并获得融合结果，同时支持彩色图像的融合，方法为将RBG图像转化为YCrCb空间实现融合后还原色彩通道。

本次实验的硬件平台为Win10操作系统；AMD Ryzen 7 3700X CPU，主频为3.6 GHz；GPU为NVIDIA GeForce RTX 3080。软件平台为VS Code，MATLAB 2019a。代码环境为Python 3.7版本；Pytorch 1.11.0版本。

2.2 消融性实验分析

为了证明本文重点提出的双通道模块的有效性，在实验过程中测试了单独使用残差连接或者单独使用密集通道（参考文献[13]）结合融合层算法的融合性能，由于单个通道不能使用双通道的融合策略，为了控制融合策略的影响，统一使用L1范数作为融合策略进行实验。

从TNO数据集中选取4组红外与可见光图像，从上到下分别为“ground”、“people”、“house-1”、“house-2”。对四组图像进行融合的结果如图4所示，图4(a)为可见光图像，图4(b)为对应的红外图像，图4(c)、(d)、(e)分别为单独使用残差通道、密集通道以及使用双通道级联的融合图。以“house-2”为例进行主观评价可以发现，通过单个通道进行融合后的图像可以提取源红外与可见光图像的大部分信息，但是单个通道提取的特征信息不如双通道的详尽，比如图中红框标出的云朵信息，残差通道提取出的亮度信息有所丢失，密集通道提取的边缘轮廓信息不足，而双通道结构则保留了更多的信息。同时，为了验证本文融合策略的有效性，从TNO数据集中选取“car”组数据进行对比分析，结果如图5所示。不难看出，在这组图像中，除去最大值策略的融合结果较少的保留了红外图像特征之外，其余几种融合策略都较好的融合了源图像的信息，其中双通道融合策略相对优秀，在保留细节特征的同时图像的对比度更好，更加清晰。

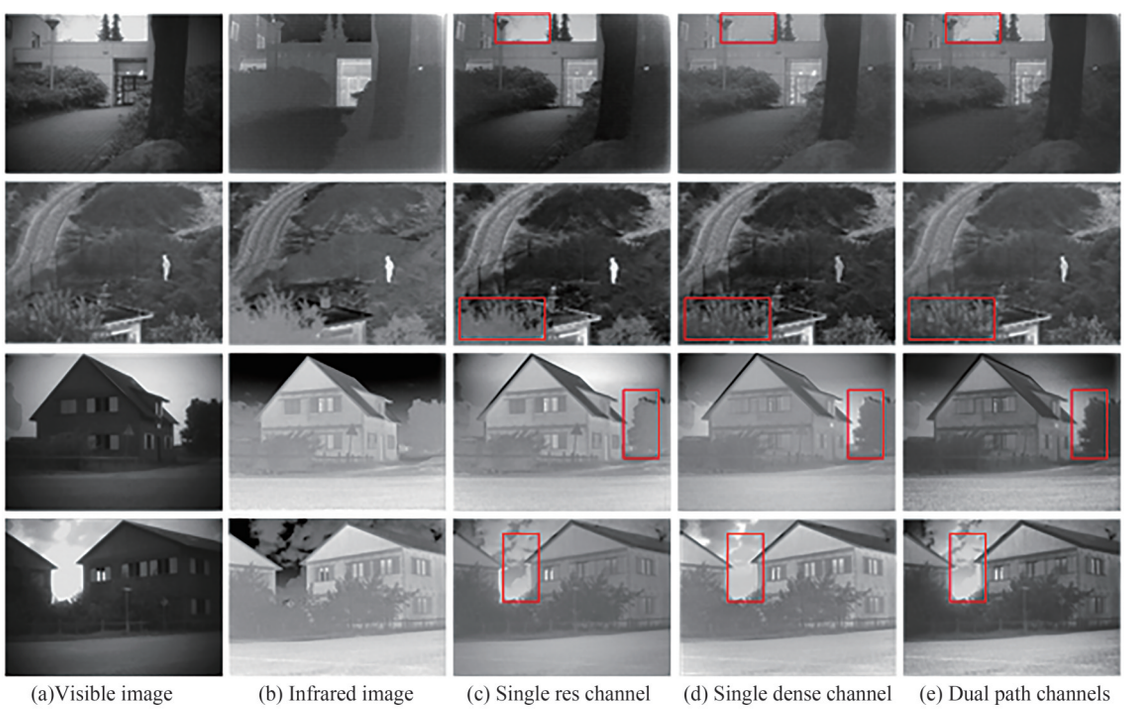


图4 不同通道的红外与可见光融合结果

Fig.4 Infrared and visible light fusion results of different channels

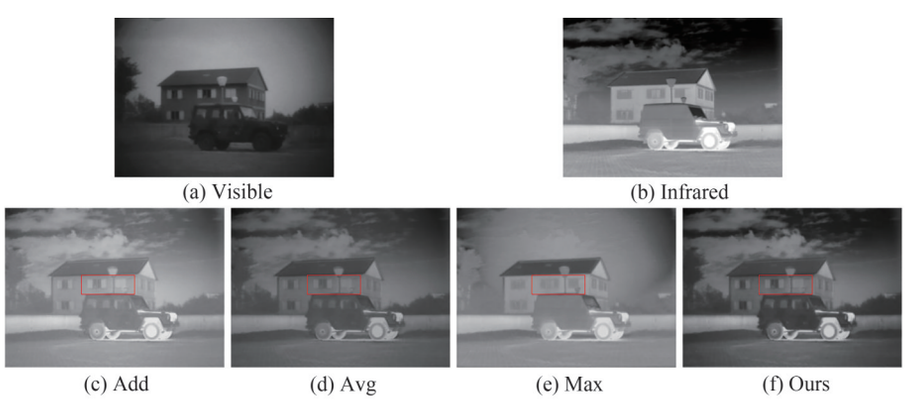


图5 不同融合策略的红外与可见光融合结果

Fig.5 Infrared and visible light fusion results of different fusion strategies

为了能够更加直观地体现消融性实验的结果，从TNO数据集中选取8组数据分别对双通道网络模块和融合策略进行客观评价分析，为了控制实验变量，对于不同通道都使用相同的L1范数融合策略，对于不同融合策略则都使用相同的双通道网络结构。实验结果如表2所示，本文提出的双通道自编码器结构和融合策略在SSIM[26]、SCD[27]和PSNR[28]指标上都带来了融合性能的提升，与主观描述相符合。

表2　消融性实验评价指标平均定量值

Table 2　The table of mean quantitative values of experimental evaluation indexes for ablation

Metrics	Different channels			Different fusion strategies			
	Res	Dense	Dual Path	Add	Avg	Max	Dual path attention
SSIM	0.906 0	0.912 3	**0.923 6**	0.912 3	0.905 9	0.891 8	**0.928 9**
SCD	1.712 9	1.702 8	**1.718 9**	1.696 5	1.555 4	1.535 0	**1.860 5**
PSNR	63.601 1	65.391 8	**65.682 8**	59.621 6	64.248 6	62.230 5	**66.072 7**

2.3　融合图像的主观评价

为了验证本文提出方法的有效性，将本文的方法与7种红外与可见光融合方法进行比较，包括2种传统方法：MST-SR[6]、GFT[7]和5种深度学习方法，分别是基于卷积神经网络框架的IFCNN[29]、PMGI[30]、U2Fusion[31]；基于自动编码器框架的DenseFuse[13]，Res2Net[21]，基于生成对抗网络框架的FusionGAN，从TNO数据集中选取“pedestrian”组数据进行对比分析。各种算法的对比图如图6所示，从对比图中不难看出，图中的算法都能够完成红外与可见光融合任务，反映出图中行人的信息，但是一些算法在树枝的纹理特征方面的信息不足，比如GFT、DenseFuse、Res2Net、FusionGAN方法，而本文提出的算法，在双通道注意力

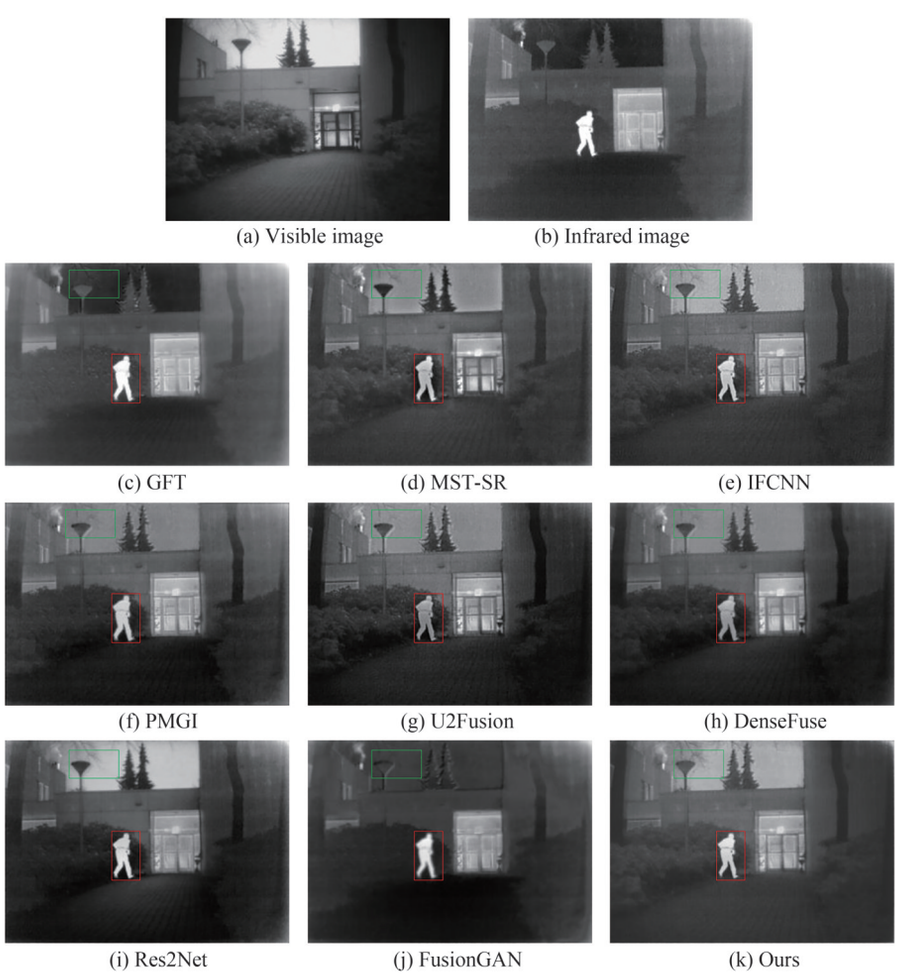
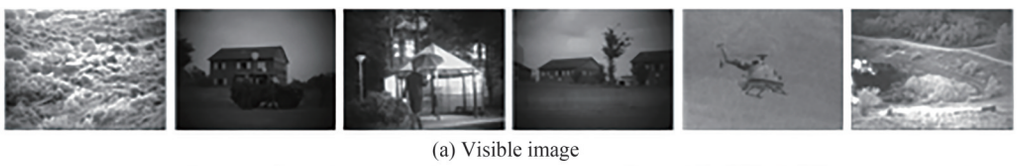


图6　“pedestrian”的融合结果

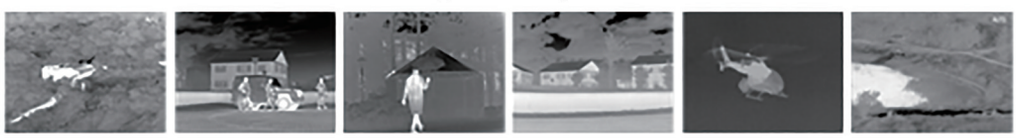
Fig.6　Fusion results of “pedestrian”

融合策略下相对完整的提取并融合了可见光中的树枝特征，效果略优于其他算法。

单一场景的融合图像对比的结果是比较片面的，本文从TNO数据集中选用六种不同场景下的图像融合结果扩大对比实验。实验结果如图7所示。从图中不难发现，GFT、FusionGAN方法的融合图像中物体的轮廓不清晰，缺乏可见光图像的信息；PMGI和U2Fusion方法能够较好的反映目标的特征，但是引入了过多的噪声，图像的质量不佳；MST-SR、IFCNN、DenseFuse方法的图像含有较为清晰的轮廓信息及目标特征，但仍会出现部分信息不明显的情况。而本文方法在大多数融合场景下都能够维持红外与可见光图像间的信息平衡，获得较好的融合结果。



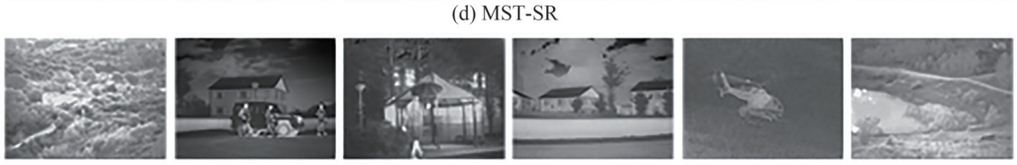
(a) Visible image



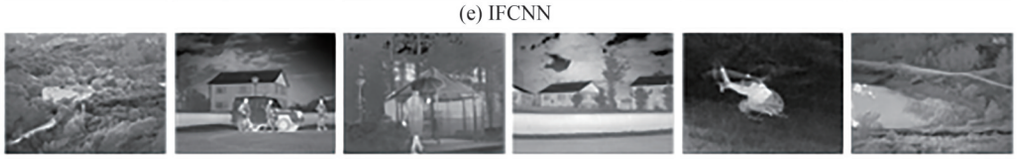
(b) Infrared image

(c) GFT

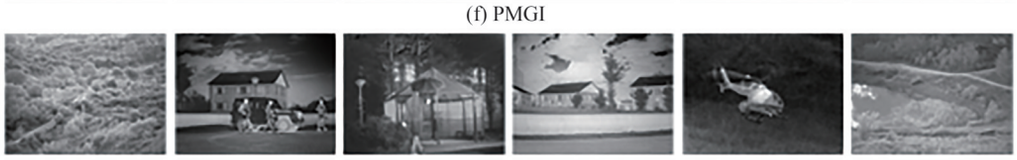
(d) MST-SR



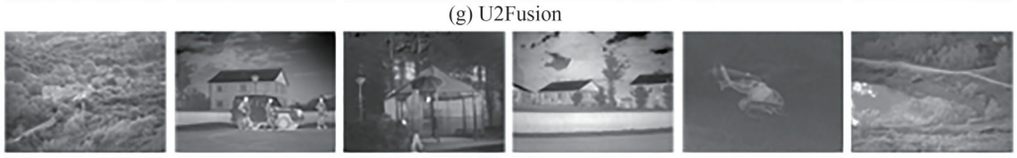
(e) IFCNN



(f) PMGI



(g) U2Fusion



(h) DenseFuse

(i) Res2Net

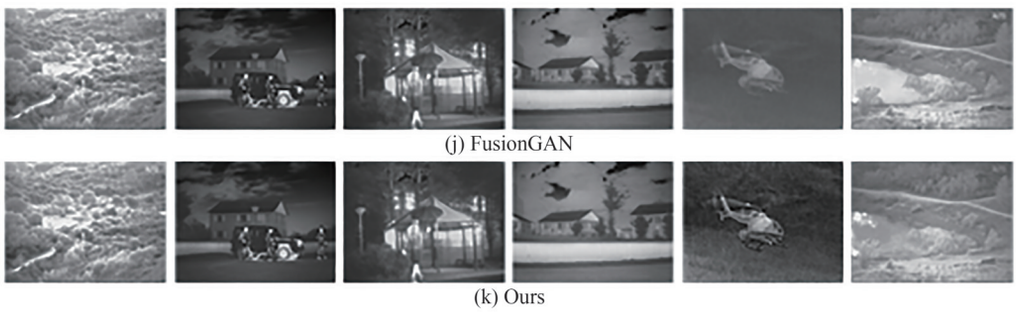


图7　各算法红外与可见光融合结果

Fig.7　Infrared and visible light fusion results of each algorithm

2.3　融合图像的客观评价

为了更进一步的验证本文提出方法的有效性，对主观评价中的算法采用6种典型的客观评价指标进行融合结果质量评价，分别为标准差（Standard Deviation，SD）[32]、相关系数（Correlation Coefficient，CC）、峰值信噪比（Peak signal-to-noise ratio，PSNR）、多尺度结构相似度（Multi-Scale Structural Similarity，MS-SSIM）、差异相关性总和SCD和互信息（Mutual Information，MI）[33]。本文从TNO数据集中选取10组融合图像进行对比实验，通过折线图直观地表示出来，如图8所示。融合结果的平均定量值如表3所示，从图表中可以看出，本

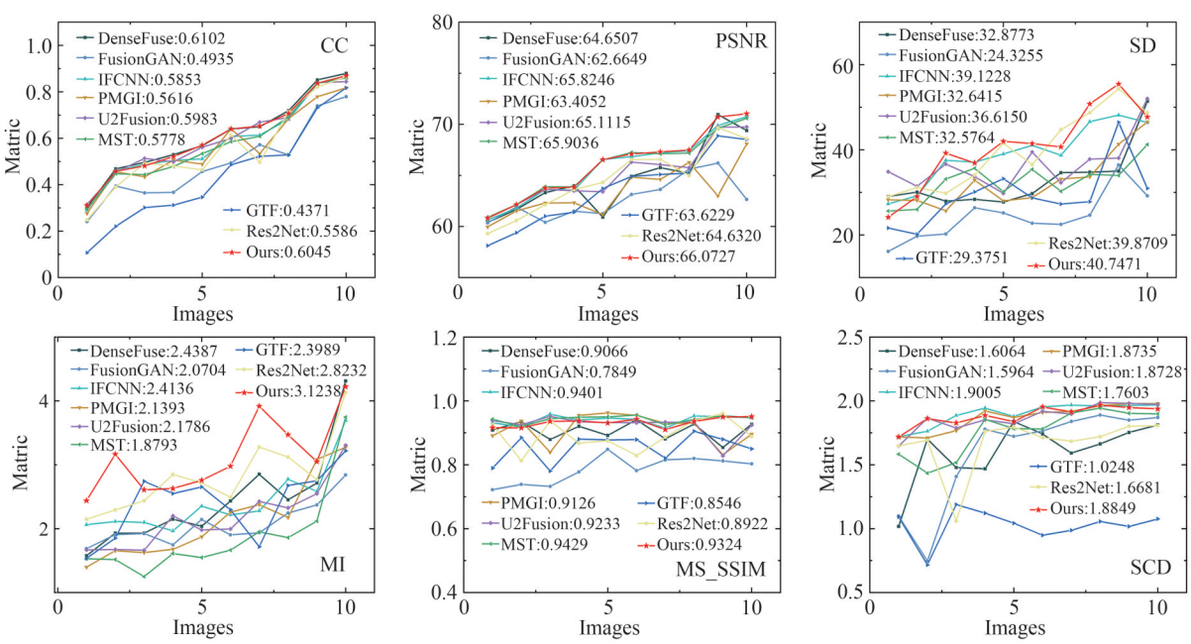


图8　客观实验折线图

Fig.8　Objective experimental line chart

表3　10组评价指标平均定量值

Table 3　The average quantitative value of each evaluation index

Metrics	Methods								
	MST-SR	GTF	IFCNN	PMGI	U2Fusion	DenseFuse	Res2Net	FusionGAN	Ours
SD	32.576 4	29.375 1	39.122 8	32.641 5	36.615 0	32.877 3	39.870 9	24.325 5	**40.747 1**
CC	0.577 8	0.437 1	0.585 3	0.561 6	0.598 3	**0.610 2**	0.558 6	0.493 5	0.604 5
PSNR	65.903 6	63.622 9	65.824 6	63.405 2	65.111 5	64.650 7	64.632 0	62.664 9	**66.072 7**
MS-SSIM	**0.942 9**	0.854 6	0.940 1	0.912 6	0.923 3	0.906 6	0.892 2	0.784 9	0.932 4
SCD	1.760 3	1.024 8	**1.900 5**	1.873 5	1.872 8	1.606 4	1.668 1	1.596 4	1.884 9
MI	1.879 3	2.398 9	2.413 6	2.139 3	2.178 6	2.438 7	2.823 2	2.070 4	**3.123 8**

文所提出的算法相比较于典型传统方法以及近年的深度学习方法在SD、PSNR、MI三个指标中处于领先，在其余指标中也处于中上水平，具有最好的综合质量。也说明本算法在保留源图像信息、减少噪声的引入同时提高融合图像的质量上有较好的效果，与主观评价相符合。

3 结论

本文提出了一种基于双通道残差密集网络红外与可见光图像融合方法，用于改善红外与可见光融合图像与源图像间的部分细节特征丢失的问题，充分提取红外与可见光图像中的特征信息，同时在网络结构的基础上，改进了融合策略。通过将该方法与单通道的网络结构对比，证明双通道并联的结构的确提升了图像特征提取的效率。最后将本文的方法与传统以及深度学习融合方法进行对比与分析，结果表明该方法在保留源图像更多细节特征的同时减少噪声和其他干扰的引入，有助于后续的高级视觉应用。后续工作中将继续优化网络参数、优化融合策略、简化网络结构并获取更好的融合效果。

参考文献

[1] MA Jiayi, MA Yong, LI Chang. Infrared and visible image fusion methods and applications: a survey[J]. Information Fusion, 2019, 45: 153-178.

[2] ZHANG HAO, XU Han, TIAN Xin, et al. Image fusion meets deep learning: a survey and perspective[J]. Information Fusion, 2021, 76: 323-336.

[3] DAS S, ZHANG Yunlong. Color night vision for navigation and surveillance[J]. Transportation Research Record, 2000, 1708(1): 40-46.

[4] SUN Jiping, FAN Weiqiang. Mine dual-band image fusion in MS-ADoG domain combined with ReNLU and VGG-16[J]. Acta Photonica Sinica, 2022, 51(3): 0310002.
孙继平, 范伟强. MS-ADoG域结合ReNLU与VGG-16的矿井双波段图像融合算法[J]. 光子学报, 2022, 51(3): 0310002.

[5] CAO Yanpeng, GUAN Dayan, HUANG Weilin, et al. Pedestrian detection with unsupervised multispectral feature learning using deep neural networks[J]. Information Fusion, 2019, 46: 206-217.

[6] LIU Yu, LIU Shuping, WANG Zengfu. A general framework for image fusion based on multi-scale transform and sparse representation[J]. Information Fusion, 2015, 24: 147-164.

[7] MA Jiayi, CHEN Chen, LI Chang, et al. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. Information Fusion, 2016, 31: 100-109.

[8] LIU Yu, CHEN Xun, HU Peng, et al. Multi-focus image fusion with a deep convolutional neural network[J]. Information Fusion, 2017, 36: 191-207.

[9] MA Jiayi, TANG Linfeng, XU Meilong, et al. STDFusionNet: an infrared and visible image fusion network based on salient target detection[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-13.

[10] MA J, YU W, LIANG P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion[J]. Information Fusion, 2019, 48: 11-26.

[11] MA Jiayi, XU Han, JIANG Junjun, et al. DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion[J]. IEEE Transactions on Image Processing, 2020, 29: 4980-4995.

[12] MA Jiayi, ZHANG Hao, SHAO Zhenfeng, 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: 1-14.

[13] LI Hui, WU Xiaojun. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623.

[14] XU Han, WANG Xinya, MA Jiayi. DRF: Disentangled representation for visible and infrared image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-13.

[15] LI Hui, WU Xiaojun, KITTLER J. RFN-Nest: an end-to-end residual fusion network for infrared and visible images[J]. Information Fusion, 2021, 73: 72-86.

[16] CHEN Y, LI J, XIAO H, et al. Dual path networks[J]. arXiv, 2017. http://arxiv.org/abs/1707.01629.

[17] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA: IEEE, 2016: 770-778.

[18] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE, 2017: 2261-2269.

[19] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. arXiv, 2015. http://arxiv.org/abs/1502.03167.

[20] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by Exponential Linear Units (ELUs)[J]. arXiv, 2016. http://arxiv.org/abs/1511.07289.
[21] SONG Xu, WU Xiaojun, LI Hui, et al. Res2NetFuse: a fusion method for infrared and visible images[J]. arXiv, 2022. http://arxiv.org/abs/2112.14540.
[22] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
[23] LI Hui, WU Xiao Jun, 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.
[24] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]. European Conference on Computer Vision, Springer, Cham, 2014: 740-755.
[25] TNO image fusion dataset [EB/OL]. https://figshare.com/articles/TNO Image Fusion Dataset/1008029.
[26] MA Kede, KAI Zeng, ZHOU Wang. Perceptual quality assessment for multi-exposure image fusion [J]. IEEE Transactions on Image Processing, 2015, 24(11): 3345-3356.
[27] ASLANTAS V, ENDES 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: 1890-1896.
[28] SHEIKH H R, SABIR M F, BOVIK A C. A statistical evaluation of recent full reference image quality assessment algorithms[J]. IEEE Transactions on Image Processing, 2006, 15(11): 3440-3451.
[29] ZHANG Yu, LIU Yu, SUN Peng, et al. IFCNN: a general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020, 54: 99-118.
[30] ZHANG Hao, XU Han, YANG Xiao, et al. Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12797-12804.
[31] XU Han, MA Jiayi, JIANG Junjun, et al. U2Fusion: a unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 502-518.
[32] YUN Jiangrao. In-fibre Bragg grating sensors[J]. Measurement Science and Technology, 1997, 8(4): 355.
[33] QIU Huihong, ZHANG Dali, YAN Pingfan. Information measure for performance of image fusion [J]. Electronics Letters, 2002, 38: 313-315.

Infrared and Visible Image Fusion Based on Dual Channel Residual Dense Network

FENG Xin, YANG Jieming, ZHANG Hongde, QIU Guohang

(*School of Mechanical Engineering, Key Laboratory of Manufacturing Equipment Mechanism Design and Controlof Chongqing, Chongqing Technology and Business University, Chongqing* 400067, *China*)

Abstract: In the infrared and visible image fusion task, the visible image contains a large amount of texture and background information, while the infrared image contains obvious target information. The two complement each other and can effectively and comprehensively represent the visual information of a scene. In order to improve the problem of partial feature loss between infrared and visible fusion image and source image, and fully extract the feature information in infrared and visible image, this paper proposes an improved dual channel deep learning auto-encoder network for infrared and visible image fusion. The encoder is composed of three cascaded dual channel layers, and they are composed of the cascaded residual and dense connection modules. The source image is divided into two paths and input the residual connection network and the dense connection network at the same time. The residual connection network has a good effect in highlighting the target features. And the dense connection is good at preserving the texture details of the source image, so the encoder structure can fully extract the multi-level features of infrared and visible images. In the design of fusion layer, the spatial L1 norm and the channel attention mechanism are respectively used to fuse the cascades of residuals and dense channel features. The spatial L1 norm fusion strategy uses the L1 norm to calculate the value of activity level measurement and lays more emphasis on the fusion of spatial information. The channel attention mechanism obtains the weight graph of each channel through the global pooling operation. The information contained in each channel can be measured by weight

so that the channel information can be fused effectively. Finally, the corresponding decoder is designed to reconstruct the fusion feature image, and the decoder processes the dense and residual features differently according to the characteristics of the encoder. The dense feature layers in high dimension are deeper, so more convolutional sampling layers are used to restore the features; the residual feature layers in low dimension are sharer, so the number of convolutional layers is reduced. In this way, the features of different channels and levels are combined to obtain the final fusion result. In the network training stage, the fusion layer is removed, and 5 000 images are randomly selected from the ImageNet data set as the training set for the auto-encoder network. Meanwhile, the sum of pixel loss, gradient loss and structural similarity loss is used as the loss function to guide the optimization of network parameters. In the experimental phase, the network structure and fusion strategy of the ablation experiment. In terms of network structure, the comparison with single residual or dense channel network proves that the two-channel network structure indeed improves the feature extraction ability. In terms of fusion strategy, the comparison with classical fusion strategies such as addition, mean and maximum proves that the dual path fusion strategy can give play to the advantages of the dual channel structure. It can effectively integrate the salient features and detail features of the source image. Finally, the proposed method is compared with the traditional and the latest deep learning algorithms in recent years. The results show that the proposed method can better reflect the target features and background contour information subjectively, and can maintain the information balance between infrared and visible images in most fusion scenes, so as to obtain high-quality fusion images. In the objective indicators CC, PSNR and MI is in the lead, the rest of the indicators are also in the middle level, with excellent comprehensive performance.

Key words: Infrared and visible image fusion; Dual channel parallel network; Residual dense module; Attention model; Auto-encoder network

OCIS Codes: 100.1455; 100.2000; 100.2980; 100.3020

Foundation item: National Natural Science Foundation of China (No. 22178036), Chongqing University Innovative Group Research Project (No. CXQT21024), Chongqing Natural Science Foundation (No. CSTB2022NSCQ-MSX0271)