

基于引导滤波的多分支注意力残差红外图像去噪网络

张 骏^{1,2,3}, 朱 标^{1,2,3}, 沈玉真^{1,2,3}, 张 鹏^{1,2,3}

- (1. 中航华东光电有限公司, 安徽 芜湖 241002;
2. 特种显示国家实验室, 安徽 芜湖 241002;
3. 国家特种显示工程技术研究中心, 安徽 芜湖 241002)

摘要: 目前红外图像广泛应用于各个领域,但受限于探测单元的非均匀性,使得红外图像具有低信噪比、视觉效果模糊的缺点,严重影响其在高端领域中的应用。常用的去噪算法无法兼顾降噪平滑和边缘细节的保持,针对这一问题,文中提出了一种基于引导滤波的多分支注意力残差去噪网络。根据引导滤波原理设计一种引导卷积模块,同时为了兼顾提取浅层和深层特征设计了多分支注意力残差模组。通过实验证明加入新模块后的网络不仅可以有效地实现红外图像降噪,而且能最大程度地保持图像中的边缘细节信息,提升视觉效果,同时在 PSRN 和 SSIM 指标上也有良好的表现。

关键词: 引导滤波器; 视觉注意力机制; U-NET

中图分类号: TP391 **文献标志码:** A **DOI:** 10.3788/IRLA20220060

Multi-drop attention residual infrared image denoising network based on guided filtering

Zhang Jun^{1,2,3}, Zhu Biao^{1,2,3}, Shen Yuzhen^{1,2,3}, Zhang Peng^{1,2,3}

- (1. Aviation Industry Corporation Huadong Photoelectric Company Limited, Wuhu 241002, China;
2. State Special Display Engineering Laboratory, Wuhu 241002, China;
3. National Special Display Engineering Research Center, Wuhu 241002, China)

Abstract: At present, infrared images are widely used in various fields, but limited by the non-uniformity of detector unit, the infrared image has the disadvantages of low signal-to-noise ratio and blurred visual effects, which seriously affect its application in advanced fields. Commonly used denoising algorithms cannot take into account the smoothing of denoising and the preservation of edge details. In response to the above problems, this paper proposes a multi-drop attention residual denoising network based on guided filtering. A guided convolution module is designed according to the principle of guided filtering and a multi-drop attention residual module is designed for both the extraction of shallow and deep features. Experiments have proved that the network after adding the new module can effectively reduce the noise of infrared images, and can maintain the edge detail information in the image to the greatest extent, improve the visual effect, and also have good performance on the PSRN and SSIM indicators.

Key words: guided filter; visual attention; U-NET

收稿日期:2022-01-18; 修订日期:2022-04-01

基金项目:安徽省重大专项(201903a05020057)

作者简介:张骏,男,中级工程师,硕士,主要从事特种光图像处理、信号处理及模式识别方面的研究。

0 引言

随着红外设备生产制造技术的成熟,红外成像技术应用的领域亦越发广泛。红外图像是红外热成像仪利用探测器将红外辐射能转换成电信号,经放大处理,转换成标准视频信号在显示器上显示的红外热图形^[1]。由于红外图像是通过“测量”物体向外辐射的热量获得的,与可见光图像相比具有分辨率差、对比度低、信噪比低、视觉效果模糊和信息量少等缺点,需要经过一系列图像处理后,才能获得舒适的视觉成像效果^[2]。其中尤以图像降噪、增强最为重要。红外图像在成像的各个环节都会受到外界的影响而产生噪音,如热传导效应及空气的散射,导致图像边缘模糊,对比度低;温度分布不均匀会引起散斑噪声等。这些噪音严重影响视觉效果以及后续高级应用的实现,因此基于红外图像的去噪方法具有重要的研究价值^[1,3-4]。目前提升红外图像质量的方法主要分为硬件和软件两种方法^[5]。硬件方法多集中在提升探测器材料或者制作工艺,通过抑制各探测器单元的非均匀性大幅度提升红外图像的质量,但经济成本较高,研发周期长,难以广泛应用^[6]。因此基于软件的提升方法便成为最有效的改善方法。

目前常用的图像去噪算法有传统去噪算法和基于深度学习的去噪算法^[7-10]。传统去噪算法是利用噪声自身特性设计特定的滤波器或者特定模型达到去噪的效果^[11-13],代表性的算法有 BM3D(block matching and 3D filtering)^[14-15]算法以及 K-SVD 算法^[16]。基于深度学习的去噪算法其本质属于外部先验法^[8,10],与传统的外部先验方法相比,深度学习类算法去噪的效果要更为突出。2017年,Zhang等^[17]提出的深层卷积神经网络去噪算法(denoising convolutional neural networks, DnCNN)^[17-20]。该方法引入残差技术,通过预测残差图像,即预测噪声和潜在干净图像之间的差异进行去噪。2018年,JingWen Chen等在其文章^[21]中提出了利用 GAN 对噪声分布进行建模,并通过建立的模型生成噪音样本,与干净图像集合构成训练数据集,训练去噪网络进行盲去噪^[21]。2018年,Zhang Kai 等人在论文^[22]中提出了多个降级的超分辨率网络(Single Convolutional Super-Resolution Network for multiple degradations, SRMD),通过该算法可以处理多

个退化空间不均匀的退化类型,同时提出的维度拉伸策略可以满足不同维度的输入。

文中提出一种基于引导滤波的多分支注意力残差红外去噪网络(guided-filter denoising network, GFD-Net),该网络结合视觉注意力机制(visual attention)^[23]选择核心单元(selective kernel unit, SKUnit)^[24],提出了多分支注意力残差模块(multi-drop attention resnet, MAR),同时根据引导滤波原理设计^[25]了引导卷积模块组件(guided-filter convolution, GFC)。通过实验证明,相比现有算法,文中提出的算法在实现红外图像去噪的同时最大程度保留红外图像的边缘细节,有效提升图像的视觉效果,并且在峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性(structural similarity, SSIM)指标上取得了较好的表现。

1 网络原理及结构

红外图像因其特殊的成像方式,往往具有分辨率小、对比度差、视觉效果模糊等特点。红外图像去噪过程中不仅需要完成去噪,同时还要最大程度保留红外图像的边缘细节,有效提升图像视觉效果。因此在设计 CNN 网络时,需要兼顾空间信息和语义信息,通过空间信息提升图像的整体视觉效果,通过语义信息保留图像上下文信息以及边缘细节信息。文中根据上述的要求设计了 GFDNet(guided-filter denoising network)网络。网络的宗旨在于通过引入多分支注意力残差模块(multi-drop attention resnet, MAR)和引导卷积模块(guided-filter convolution, GFC),完成多尺度特征提取和融合,实现端到端的去噪网络。其中 MAR 模块通过空间和通道注意力机制来捕获全局信息(包括全局空间信息和全局的上下文信息),通过 SKNet 模块自适应感受野获取局部信息,以兼顾空间信息和语义信息;GFC 模块以 MAR 提取的特征信息和低尺度的重建特性信息作为输入信息,相较于普通的标准拼接运算,输入信息更加丰富、多样,同时利用引导滤波原理在完成去噪功能的基础上最大程度的保留了图像的边缘细节。

1.1 多分支注意力残差模块

近年来,深度学习与视觉注意力机制结合的应用越发广泛。目前常用的视觉注意力机制主要有空间

域注意力机制 (spatial attention)^[26] 和通道域注意力机制 (channel attention)^[27]。

Spatial attention 表现在图像中就是对特征映射图中不同位置的关注程度不相同。其数学本质为: 针对大小为 $H \times W \times C$ 的特征图, 一个有效的空间注意力对应大小为 $H \times W$ 的矩阵, 矩阵中每个元素是特征映射图相应位置的权重, 通过逐像素相乘 (pixel-wise multiply) 实现关键位置的标识, 如公式 (1) 所示:

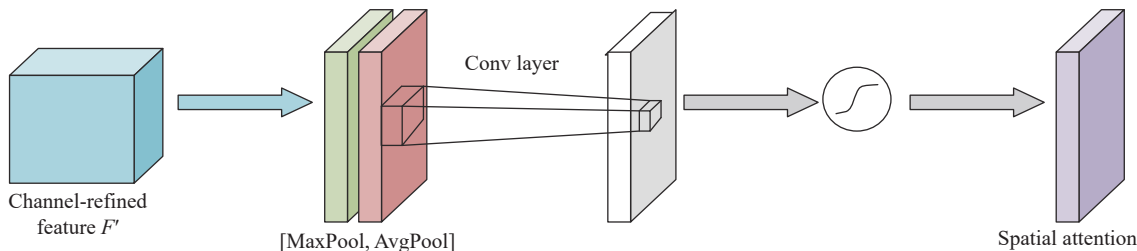


图 1 Spatial attention 结构图

Fig.1 Diagram of spatial attention structure

Channel attention 注意力机制主要分布在通道中, 表现在图像上就是对不同的图像通道关注程度不同, 其数学本质为: 针对大小为 $H \times W \times C$ 的特征映射图, 一个有效的通道注意力对应一个大小 $1 \times 1 \times C$ 的矩阵, 矩阵中每个元素是特征映射图对应通道的全部像素的权重, 通过逐通道相乘 (channel-wise multiply) 实现对关键通道的标识, 如公式 (2) 所示:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j)$$

$$s_c = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z))$$

$$W_1 \in R^{\frac{C}{r} \times C}, W_2 \in R^{C \times \frac{C}{r}}$$

$$X_C = F_{scale}(u_c, s_c) = s_c \times u_c \quad (2)$$

式中: F_{sq} 操作, 本质是全局均值池化, 使得 $H \times W \times C$ 的特征映射图转变为 $1 \times 1 \times C$ 的序列; F_{ex} 操作完成通道依赖性操作, 这里采用了两个核为 1×1 的 conv 的操作代替全连接操作, 其中 W_1 为降维层参数, 降维比例

$$M_s(F) = \sigma(\text{conv}^{3 \times 3}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(\text{conv}^{3 \times 3}([F_{avg}^s; F_{max}^s]))$$

$$\sigma(\cdot) = F \otimes \text{sigmoid}(\text{conv}^{3 \times 3}(\cdot)) \quad (1)$$

式中: F 表示特征映射图; $\text{AvgPool}(\cdot)$ 表示平均池化操作; $\text{MaxPool}(\cdot)$ 表示最大池化操作, 结果分别为 F_{avg}^s , F_{max}^s ; $\text{conv}^{3 \times 3}$ 表示核为 3×3 的卷积操作; \otimes 表示逐像素相乘。

其模块结构如图 1 所示。

为 r , W_2 为升维层参数。替代后, 整个操作的非线性增强, 可以拟合更复杂的通道间的依赖性, 同时减少了参数数量和运算量; F_{scale} 操作表示逐通道相乘操作。

其结构如图 2 所示。

2019 年, Li Xiang 提出了选择核心网络 (selective kernel network, SKNet)^[24], 其设计思想来自于皮质神经元根据不同的刺激可以动态调节自身的感受野 (receptive field)。整个网络是结合了 SENet (squeeze-and-excitation networks)、合并和运行映射 (merge-and-run mappings) 以及 attention on inception block 思想的产物。

SKNet 网络是一种非线性方法, 聚合来自多个卷积核的信息, 以实现神经元的自适应感受野大小。整个网络由三个操作组成: Split, Fuse 和 Select。Split 操作生成具有各种卷积核大小的多个路径。Fuse 操作的基本思想是使用门运算来控制来自多个分支的信

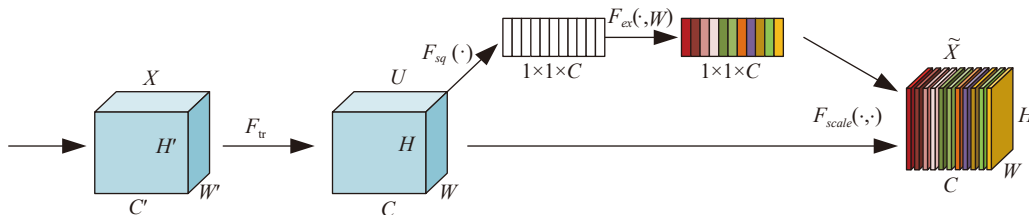


图 2 Channel Attention 结构图

Fig.2 Diagram of channel attention structure

息流。不同分支携带不同尺度的信息流向下层的神经元,通过门运算实现不同分支信息的整合。这里使用 Attention 运算实现门运算。Select 操作使用多个权重矩阵对不同分支信息流进行加权操作,然后求和得到最终的输出。以两路分支为例,如公式 (3) 所示:

$$\begin{aligned}
 U_c &= \tilde{U}_c + \hat{U}_c \\
 S_c &= F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \\
 z &= F_{fc}(s) = \text{Relu}(\text{Norm}(s)) \\
 a_c &= \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}}, b_c = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}} \\
 V_c &= a_c \times \tilde{U}_c + b_c \times \hat{U}_c, a_c + b_c = 1 \quad (3)
 \end{aligned}$$

式中: $U_c = \tilde{U}_c + \hat{U}_c$ 表示 Split 操作,生成不同卷积核的两个路径 \tilde{U}_c 和 \hat{U}_c ; S_c, z 两个公式表示 Fuse 操作; $F_{gp}(\cdot)$ 表示全局平均池化操作; H, W 分别表示特征映射的宽和高; z 表示一个紧凑的特征,其目的是为精确和自适应选择提供指导; $F_{fc}(\cdot)$ 表示全连接层。最后两个公式表示 Select 操作, A, B 分别表示不同的权重矩阵; a_c, b_c 分别表示 \tilde{U}_c, \hat{U}_c 的 Attention 结果,使用 softmax 运算求解 a_c, b_c ,通过紧凑特征 z 引导出不同空间尺度; V_c 为最终的特征映射。

通过公式 (3) 可以发现通过 softmax 运算,将不同感受野的信息进行融合,分支中不同注意力产生不同的有效感受,最终通过加权的形式实现自适应的感受野大小。在参考文献 [24] 中作者给出模块结构图,文中不在赘述。

文中提出的 (multi-drop attention resnet, MAR), 采用并行分支残差结构,整个单元的结构如图 3 所示,输入特征首先进行核为 3×3 的卷积运算,再分别执行 channel attention、spatial attention 和 SKNet 三种操作,其次将

channel attention 和 spatial attention 的运算结果进行 concat 操作,再执行核为 1×1 的卷积运算,其结果再与 SKNet 的运算结果进行 concat 操作,并对合并后的结果执行核为 3×3 的卷积运算,最后将 3×3 卷积的运算结果与输入特征进行残差连接,如公式 (4) 所示:

$$\begin{aligned}
 F' &= \text{conv}^{3 \times 3}(F) \\
 F_{ca} &= CA(F') \\
 F_{sa} &= SA(F') \\
 F_{sk} &= SK(F') \\
 O &= F + \text{conv}^{3 \times 3}([\text{conv}^{1 \times 1}([F_{ca}; F_{sa}]); F_{sk}]) \quad (4)
 \end{aligned}$$

式中: F 表示输入特征图; O 表示输出特征图; $\text{conv}^{3 \times 3}(\cdot)$ 表示核为 3×3 的卷积操作; $\text{conv}^{1 \times 1}(\cdot)$ 表示核为 1×1 的卷积操作; $CA(\cdot)$ 表示 channel attention 运算; $SA(\cdot)$ 表示 spatial attention 运算; $SK(\cdot)$ 表示 SKNet 运算。

对比标准卷积和空洞卷积运算无法获取全局特征信息, MAR 模块中 spatial attention 集中关注于位置信息,通过最大池化和平均池化得到两个不同的特征描述,实现特征图在全局空间维度上的压缩; channel attention 可看做是一个特征提取器,主要关注于输入特征图中有意义的信息特征,更多地可理解为全局上下文语义特征的提取^[28]; 上述两种操作更多聚焦特征图的全局性,而 SKNet 具备自适应感受野的能力^[24],因此,在不同尺寸的输入特征图中,可较优地获取局部特征信息,对比于标准卷积和空洞卷积运算 SKNet 更加灵活,适应性更强;同时笔者也注意到上述的三种运算都不约而同地采用了降维运算,虽然降低了运算量,但也使得特征的梯度信息明显下降,伴随着多尺度网络深度的增加,尤其是下采样操作,特征梯度消失情况也愈发明显,同时全局特征信息也会逐步减少,因此在模块中使用了残差结构,通过增加上层特征弥补特征信息,解决特征梯度消失的问题^[27]。

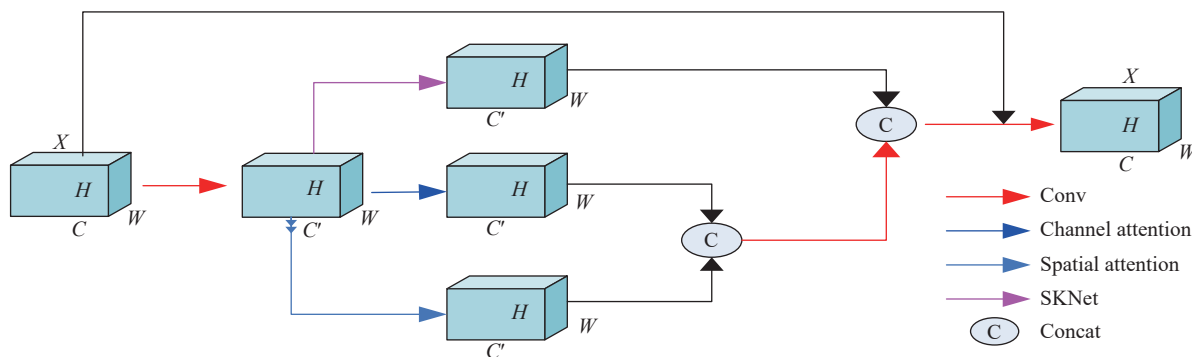


图 3 多分子注意力残差模块结构图

Fig.3 Diagram of multi-drop attention resnet structure

1.2 引导卷积模块

2009 年, 学者何凯明提出了引导滤波 (guided filter) 并将其应用在暗通道图像去雾算法中^[29, 25]。引导滤波原理类似于双边滤波, 属于可以保持边缘的滤波算法, 如公式 (5) 所示:

$$q_i = a_k I_i + b_k \quad \forall i \in w_k \quad (5)$$

式中: q 为输出图像; I 为引导图像; a_k 和 b_k 是以像素 k 为中心的窗口的线性函数的不变系数; w_k 为以像素 k 为中心的窗口。对公式 (5) 取梯度得到公式 (6):

$$\nabla q = a \nabla I \quad (6)$$

在一定范围内, 当引导图 I 较为缓和时, a_k 趋近于 0, 输出图像 q 接近于 b_k , 而当引导图 I 较为陡峭时, a_k 值变大, 输出图像 q 接近于 $a_k \times I$ 。

设真值图像为 p , 则系数 a 和 b 可以通过线性回归得方式求解, 如公式 (7):

$$f(a_k, b_k) = \arg \min \sum_{i \in w_k} (a_k I_i + b_k - p_i)^2 \quad (7)$$

对公式 (7) 进行最小二乘法求解, 得到公式 (8):

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_i + b_k - p_i)^2 + \gamma a_k^2) \quad (8)$$

式中: γa_k^2 是最小二乘法引入的乘法项。

根据上述公式, 文中设计了一种引导卷积模块 (guided filter convolution, GFC)。一般图像被认为具有低频区域 (low-frequency regions), 即平滑区域或平坦区域 (smooth or flat areas) 和高频区域 (high-frequency regions), 即边缘和噪声 (lines edges and noise)。在低

频区域梯度往往较小, 趋近于 0, 而高频区域梯度变换较大, 通常是边缘或者是噪声。设 $f_m \in R^{w \times h}$ 为第 m 层的特征映射图, 尺度为 $w \times h$, $f_{m+1} \in R^{w_1 \times h_1}$ 为第 $m+1$ 层的特征映射图, 尺度为 $w_1 \times h_1$, $\nabla(\cdot)$ 表示求梯度函数, 因为 f_m, f_{m+1} 是同一图像不同尺度的特征映射图, 具有相同的梯度趋势。因此当 f_{m+1} 为低频区域时 $\nabla(f_{m+1})$ 趋近于 0, $\nabla(f_m)$ 也趋近于 0, 而当 f_{m+1} 为高频区域时 $\nabla(f_{m+1})$ 会被平滑, 达到去噪的效果, 与引导滤波原理一致。但 f_m, f_{m+1} 的尺度不同, 因此对 f_{m+1} 进行上采样操作统一尺度, 同时使用 $\text{Relu}(\text{BN}(\text{conv}(\cdot)))$ 代替以像素 k 为中心的窗口的线性函数, 以增强非线性, 更好地拟合每个特征信息之间的梯度依赖性。其公式如下:

$$\begin{aligned} q &= f_1(\cdot) \cdot f_2(\cdot) + f_3(\cdot) \\ f_1(\cdot) &= \text{Relu}(\text{BN}(\text{conv}(In^a))) f_1 \in R^{c \times w \times h} \\ f_2(\cdot) &= \text{upsample}(\text{Relu}(\text{BN}(\text{conv}(In^b)))) f_2 \in R^{c \times w \times h} \\ f_3(\cdot) &= \text{Relu}(\text{BN}(\text{conv}(In^a))) f_3 \in R^{c \times w \times h} \end{aligned} \quad (9)$$

式中: $In^a \in \mathfrak{R}^{w \times h}$; $In^b \in \mathfrak{R}^{w_1 \times h_1}$ 对应两路输入特征图; In^a 的尺寸大小为 $w \times h$; In^b 的尺寸大小为 $w_1 \times h_1$; R 为空间区域; c 为卷积的输出通道数。为了减少模块的计算量, 对公式 (9) 中采用公式 $f_4(\cdot) = \text{Relu}(\text{BN}(\text{conv}(In^a)))$ $f_4 \in R^{(c+c) \times w \times h}$, 即将 $f_4(\cdot)$ 的 conv 的输出通道数扩大 1 倍, 这样就可以用 $f_4(\cdot)$ 代替 $f_1(\cdot), f_3(\cdot)$ 。公式 (9) 可以变换成公式 (10):

$$\begin{aligned} f_1(\cdot) &= \text{Relu}(\text{BN}(\text{conv}(In^a))) f_1 \in R^{(c+c) \times w \times h} \\ f_2(\cdot) &= \text{upsample}(\text{Relu}(\text{BN}(\text{conv}(In^b)))) f_2 \in R^{c \times w \times h} \\ q &= f_1(\cdot)[1 : c] \cdot f_2(\cdot) + f_1(\cdot)[c + 1 : c + c] \end{aligned} \quad (10)$$

GFC 模块的结构图如图 4 所示。

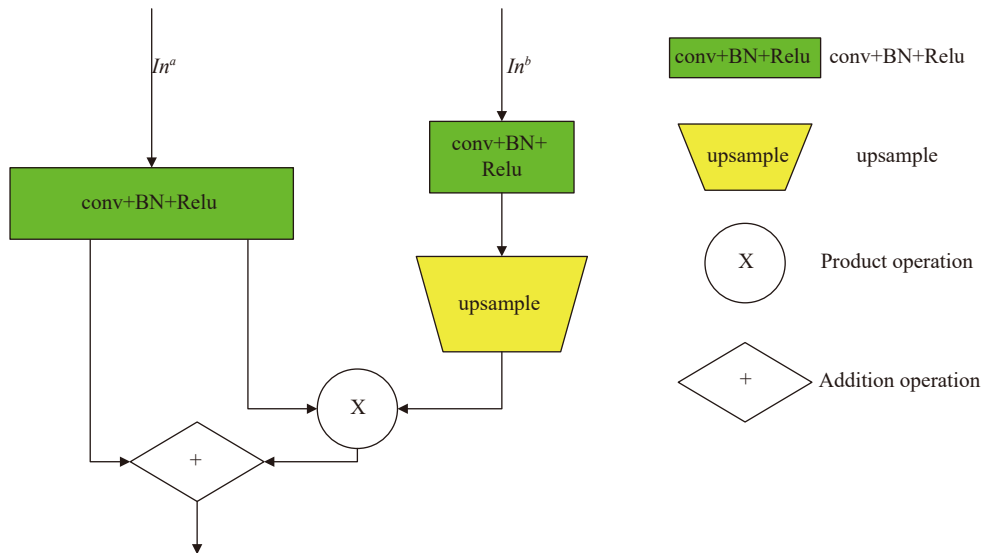


图 4 引导卷积模块结构图

Fig.4 Diagram of GFC block structure

1.3 网络结构

文中设计的网络结构如图 5 所示, 整个网络采用 U 型编解码形式。

编码网络 (encoder Net) 采用双路径的形式, 其中一路为 bottleneck 的残差模块形式^[30]。通过一系列的池化操作实现多尺度化, 用来快速提取多尺度特征; 另一路采用 MAR+ ASPP 模块 (最低层采用 MAR+ conv 形式)^[31], 常规 UNet 网络中的跳跃连接 (skip connection), 仅仅是单尺度特征图的转移, 无法获取更多的上下文信息。文中网络充分利用 skip connection, 采用 MAR+ ASPP 的形式, 以获取更多单尺度特征图的上下文的空间特征, 其中 ASPP 模块采用三个空洞卷积 (dilated convolution, DConv) 并联, 后接一个卷积核为 1×1 的 conv 操作, 三个 DConv 的空洞卷积系数分别是 2, 4, 8, 卷积核是 3×3, 通过 ASPP 模块捕获不同空洞比例下特征图的上下文特征。

在图像识别和超分辨率中, 残差模块得到了广泛使用^[30]。文中算法在编码网络中引入残差模块。设编码网络中第 i 个 bottleneck 模块的输出为 f_i , bottleneck 模块可以表示为 $bottleneck(f) = M(f) + f$, 其中 $M(\cdot)$ 表示对输入的特征映射图进行卷积, 归一化等系列操

作; 则 f_i 可以用公式 (11) 表示:

$$\begin{aligned}
 f_i &= bottleneck_k(f_{i-1}) = \\
 &M(f_{i-1}) + f_{i-1} = \\
 &M(f_{i-1}) + M(f_{i-2}) + f_{i-2} = \\
 &\dots = \\
 &\sum_{j=1}^{i-1} M(f_j) + f_0
 \end{aligned}
 \tag{11}$$

随着网络深度的增加, 残差模块也在同步增加, 对应的卷积操作也在增加, 因此可以提取出更多的不同尺度的特征信息, 学习更多的残差噪声^[17,21]。

解码网络 (decoder Net) 的设计上基本遵循了 U 型网络的解码网络的设计思, 但在不同尺度特征融合处进行了修改, 常规的解码网络只是将低分辨率特征进行上采样, 再与跳跃连接 (skip connection) 的特征进行拼接。这种编码方式聚焦局部特征的重建, 无法兼顾全局的空间信息, 虽然可以实现去噪功能, 但图像模糊, 边缘细节丢失严重, 整体的视觉效果差。针对上述不足, 文中网络在设计解码网络进行了如下的改变: (1) 在 skip connection 中加入了 MAR, ASPP 模块, 通过 MAR 模块中的注意力机制获取更多的全局空间信息, 利用 MAR 模块的选择核心 (selective kernel) 思想自适应感受野的特性, 提取局部语义信息, 同时使

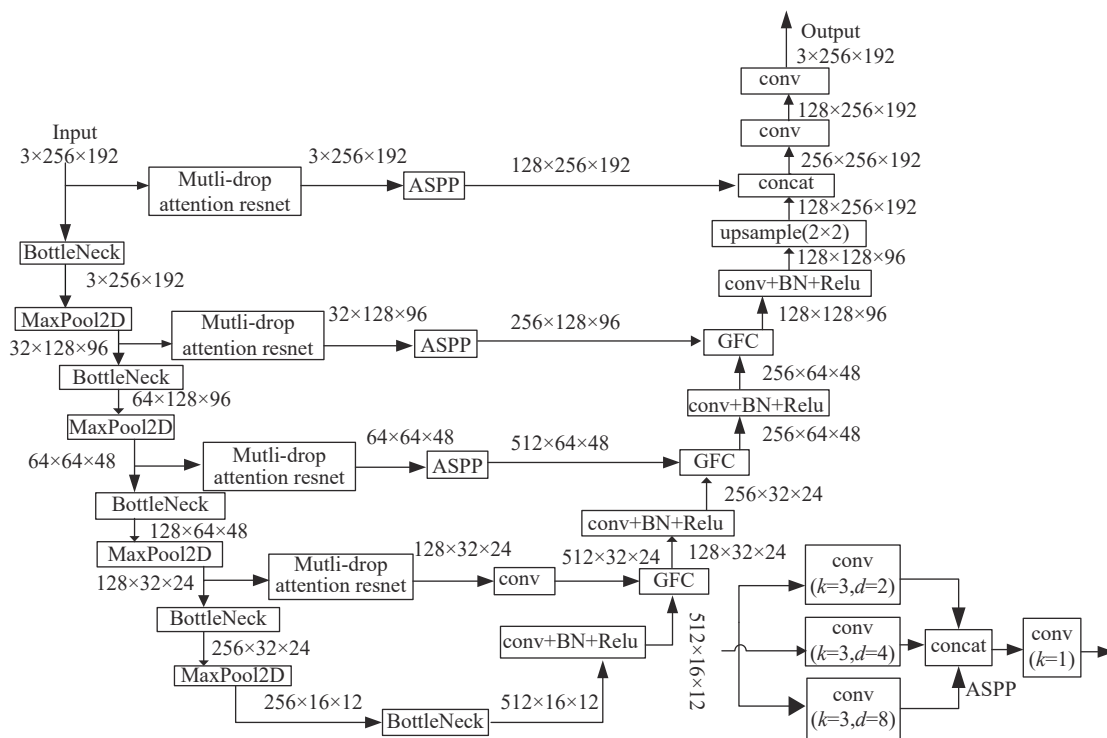


图 5 GFDNet 结构图

Fig.5 Diagram of GFDNet structure

用 ASPP 模块获取特征间的关联信息^[31]。(2) 使用 GFC 模块取代了拼接操作, 将改进后的 skip connection 的输出和低尺度的重建特征图作为 GFC 模块的输入, 兼顾了低尺度的语义信息和高尺度的空间信息和关联信息, 保证了特征信息的丰富性和多样性, 同时根据引导滤波特性完成去噪, 并保留更多的边缘细节。

1.4 损失函数

文中网络的损失函数采用混合损失函数的加权线性形式, 如公式 (12) 所示:

$$Loss = \beta L_{guided} + (1 - \beta)L_{content} \quad (12)$$

式中: L_{guided} 表示边缘细节损失函数; $L_{content}$ 表示语义损失函数; β 是一个权重因子, 用于平衡两种不同损失函数, 在文中网络中, $\beta = 0.14$ 。

图像降噪后的边缘细节的恢复是整个算法重要的组成部分, 文中网络通过设计的 GFC 模块保留图像的边缘细节, 因此边缘细节损失函数 L_{guided} 更多是对 GFC 模块重要性的体现, 在 1.2 节, 公式 (7) 已经对 GFC 模块的损失函数进行了描述, 不难发现公式 (7) 对应为 L_2 范数损失函数, 如公式 (13) 所示:

$$L_{guided} = L^2 = \frac{1}{N} \sum_{p \in P} (x(p) - y(p))^2 \quad (13)$$

式中: $x(\cdot)$ 为预测图像; $y(\cdot)$ 为目标图像; N 表示样本的数量; P 为图像的尺寸; p 为像素位置。

$L_{content}$ 语义损失函数对应整个图像的整体信息, 图像降噪算法最基础的要求是保证降噪后的图像与

原图像在整体上保持一致。因此语义损失函数是整个算法最基础, 也是最重要的损失函数之一。文中网络采用平均结构相似性 (MSSIM) 作为语义损失函数, 如公式 (14) 所示:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

$$MSSIM(x, y) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N SSIM(x_i, y_j)$$

$$L_{content} = 1 - MSSIM \quad (14)$$

式中: μ_x 为 x 的均值; μ_y 为 y 的均值; σ_x^2 为 x 的方差; σ_y^2 为 y 的方差, σ_{xy} 为 x 和 y 的协方差; M, N 表示样本的数量; $c_1 = (k_1L)^2$ 和 $c_2 = (k_2L)^2$ 为两个常数; $L = 255$, $k_1 = 0.01$, $k_2 = 0.03$ ^[32-33]。

将公式 (13) 和公式 (14) 代入公式 (12) 即可得到最终的损失函数。

2 实验及分析

2.1 数据集

当前红外图像噪声数据集较少, 常用的方法是在干净的红外图像中添加的噪声, 通过改变噪声水平来丰富数据集, 但现实中的红外图像噪声类型复杂, 强度也不尽形同。因此, 此次实验使用了艾睿光电公司提供的红外图像降噪数据集, 该数据集通过各种红外成像仪器进行数据采集, 拍摄了不同时间、不同场景、不同环境下 14 bit 的红外数据, 数据集包含 2000 对噪声与干净图像对, 图像的分辨率为 256×192 。如图 6 所示。

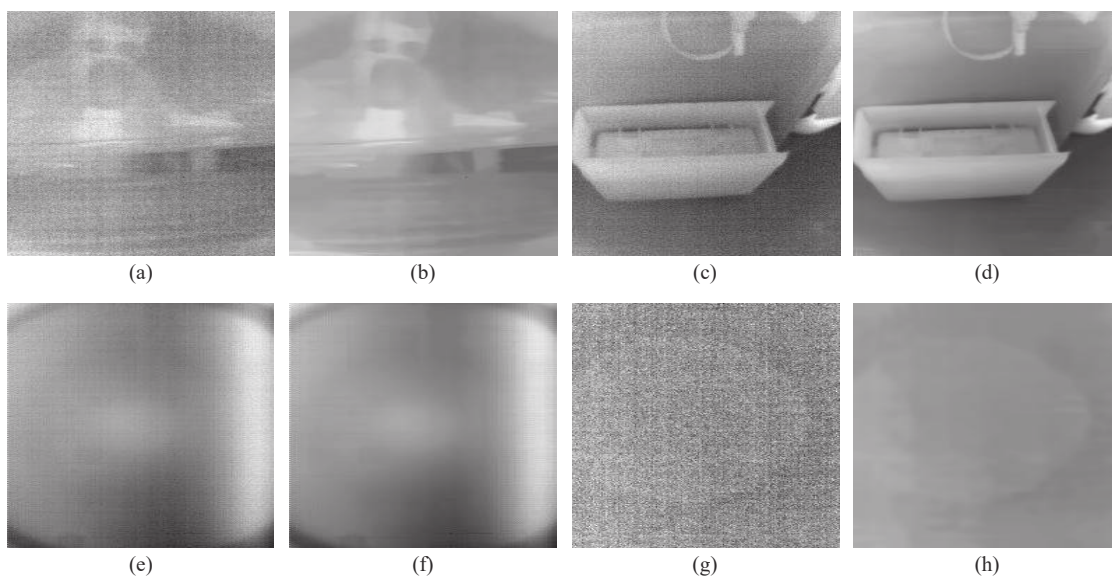


图 6 艾睿红外噪声数据 ((a)、(c)、(e)、(g) 为红外噪声图像, (b)、(d)、(f)、(h) 为红外干净图像)

Fig.6 Iraytek infrared noise data ((a), (c), (e), (g) are infrared noise images, (b), (d), (f), (h) are infrared clean images)

为了增加训练样本的多样性,同时还选择了艾睿公司的 T3 热红外模组、FLIR One Pro 拍摄了 500 张干净的红外图片,分别对图像增加了 $\sigma = 15$ 的高斯噪声以及均值为 0, 方差为 0.04 的均匀分布的随机乘性噪声,添加后的数据集数量到达 3000 对图像对。

2.2 评估指标

目前图像降噪算法的主流定量评估是使用峰值信噪比 (peak signal-to-noise ratio, PSNR), 结构相似性 (structural similarity, SSIM) 的值来测试去噪效果, 定性评估是使用视觉图形来显示恢复的干净图像^[32-33]。

PSNR 是一种像素级别的差异, 通常用对数分贝单位表示。如公式 15 所示:

$$PSNR = 20 \times \lg \left(\frac{N_{\max}}{\sqrt{MSE}} \right)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (15)$$

式中: I 和 K 分别为估计量和被估计量; i 和 j 均为像素点; N_{\max} 为图形点颜色的最大值, 在实验中设为 255。

SSIM 指标可以分别从亮度、对比度、结构三个方面评估退化图像和参考图像之间的相似性, 在公式 (14) 中有所定义。

2.3 实验参数

实验参数设置如下: 将数据集按 7 : 2 : 1 比例分为训练样本、验证样本及测试样本。单个图像样本的分辨率为 256×192 , 在数据增强上, 使用垂直/水平翻转。采用 Xavier 函数初始化卷积权重, 优化器 (Optimizer) 选择 Adam, 学习率 (Learning Rate) 设置为 0.000 1, 批次设置 (Batch-Size) 设置为 8, 迭代次数为 700 次, 整个训练用时约为 24 小时。网络框架采用 Pytorch 1.8, 训练平台为 Inter Core i7-7700HQ CPU@2.80 GHz, 内存为 16 GB, 显卡为 NVIDIA GeForce GTX 1050Ti (4 G)。

2.4 实验结果分析

整个实验分为损失权重选择实验、有效去噪实验和消融实验。

(1) 损失权重选择实验

公式 12 中, 损失权重 β 起到了平衡边缘损失函数和语义损失函数的作用。通过不同权重因子在测试数据中的 PSNR 和 SSIM 的指标值, 选定合适的权重因子。实验以艾睿数据集中的 100 张数据作为测试

数据, 在迭代 200, 400 次时, 分别以 0.1, 0.12, 0.14, 0.16, 0.2, 0.25, 0.5 作为权重因子进行测试。

从表 1 中的数值可以发现, 权重因子在 0.16 和 0.14 时 PSNR 指标近乎一致, 差距较小。而在 SSIM 指标上 0.14 时最高, 0.16 处的 SSIM 要略低。因此, 文中选择 0.14 作为损失权重因子。

表 1 损失权重对比表

Tab.1 Comparison table of loss weight

Loss weight	200 iteration		400 iteration	
	PSNR/dB	SSIM	PSNR/dB	SSIM
0.1	15.29	0.3729	21.69	0.6729
0.12	15.34	0.3841	21.83	0.6841
0.14	15.39	<u>0.4005</u>	22.24	<u>0.7013</u>
0.16	<u>15.41</u>	0.3991	<u>22.29</u>	0.7004
0.2	15.48	0.3739	22.01	0.6954
0.25	15.36	0.3714	22.11	0.6932
0.5	15.17	0.3692	21.78	0.6689

(2) 有效去噪实验

为了验证网络有效性, 选择 BM3D 算法^[11]、CBDNet^[34]、MIRNet^[35]、RIDNet^[36] 及 DPIR^[37] 进行对比, 训练集采用数据集中的 2100 对作为训练集, 600 对作为验证集, 测试数据采用两种不同的数据, 艾睿数据集中的 300 张; OTCBVS 数据集中的 100 张图片。评价指标采用峰值信噪比 (PSNR) 和结构相似性 (SSIM)。实验结果如图 7、8 所示, 评价指标如表 2 所示。

艾睿数据集对比分析: 从表 2 中的指标可以看出, MIRNet 的峰值信噪比 (PSNR) 最高, RIDNet、DPIR 和文中提出的算法的 PSNR 基本一致, 略低于 MIRNet 算法, CBDNet 和 BM3D 算法表现的指标数较差; 结构相似性 (SSIM) 指标上, 所提算法最高, DPIR 和 MIRNet 算法的指标值略低于文中算法, 而 BM3D、CBDNet 及 RIDNet 算法的指标值差距较大。通过图 7 观察, CBDNet、RIDNet、MIRNet 图中存在着少量噪点, 细节保持较好, 而 BM3D 明显过度平滑, 边缘细节丢失严重, 文中提取的算法很好的平衡了噪点平滑和细节保留, 视觉效果上更加贴近 Groundtruth 图像。

OTCBVS 数据集对比分析: OTCBVS 数据集目前主要应用与目标检测中, 图像噪点较多, 视觉效果较

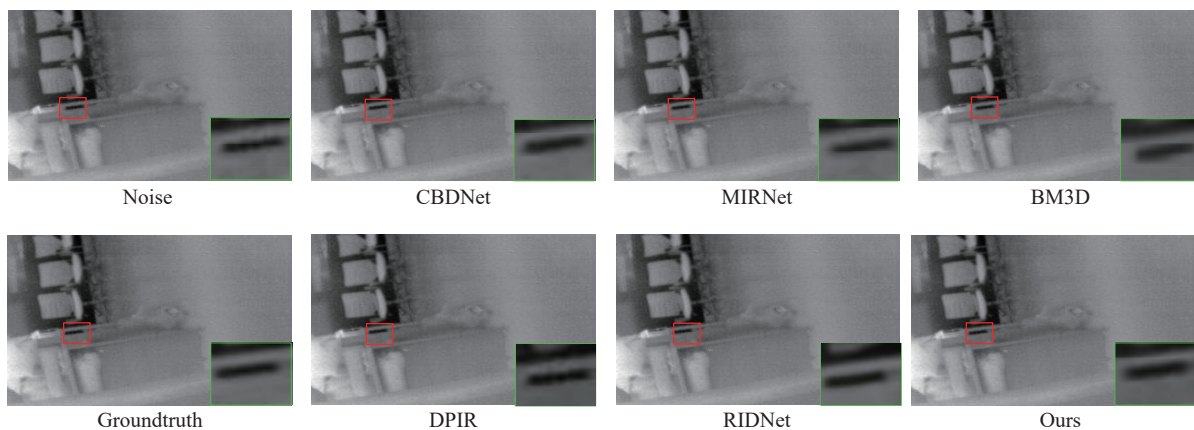


图 7 艾睿数据集实验结果图, 右下角绿色区域为红色区域的放大情况

Fig.7 Experimental results of the Iraytek dataset, the green area in the lower right corner is the magnification of the red area

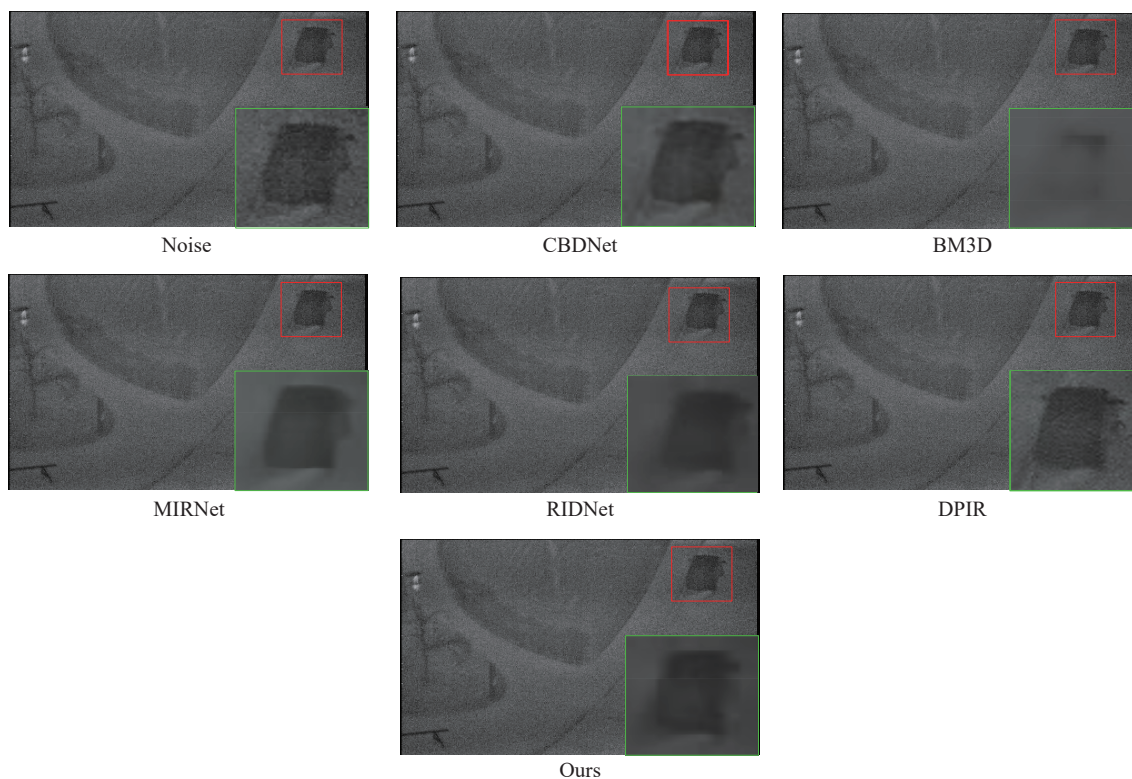


图 8 OTCBVS 数据集实验结果图, 右下角绿色区域为红色区域的放大情况

Fig.8 Experimental results of OTCBVS dataset, the green area in the lower right corner is the magnification of the red area

表 2 有效去噪实验数据对比表

Tab.2 Comparison table of effective denoising experimental data

Algorithm	Year	Iraytek dataset(300)		OTCBVS(100)	
		PSRN/dB	SSIM	PSRN/dB	SSIM
BM3D	2007	27.43	0.7814	—	—
CBDNet	2019	29.86	0.8517	27.87	0.7765
RIDNet	2019	30.67	0.8993	32.73	0.7821
DPIR	2020	30.78	0.9009	<u>32.86</u>	0.7839
MIRNet	2020	<u>31.52</u>	0.9032	32.77	<u>0.7842</u>
Ours		30.71	0.9057	32.71	0.7827

差,同时 OTCBVS 没有 Groundtruth 图像,在实验中以 BM3D 算法的结果作为 Groundtruth 图像。从指标中可以看出, RIDNet、DPIR、MIRNet 及文中算法在 PSRN 和 SSIM 指标上基本相近,差异较小。CBDNet 算法表现较差;视觉感受上,从图 8 可以发现, CBDNet 算法中的存在明显的噪点, RIDNet、DPIR、MIRNet 及文中算法在去噪效果和细节保持均有不错的表现,但前三者在整体对比度有所增强。

(3) 消融实验

文中提出的模型以 UNet 结构为主体结构,编解码过程中分别引入 MAR 和 GFC 两个模块,通过消融实验,验证新增加的两个模块可以有效的提升 PSRN 和 SSIM 指标。

从表 3 中的结果可以看出,采用模型 1 的方案在实验中结果较差, PSRN 指标仅为 27.96 dB, SSIM 指标为 0.8253;对比模型 1 和模型 5 可以清晰的发现,使用 MAR 和 GFC 模块后, PSRN 和 SSIM 指标有了大幅提升。这也验证了两个模块在红外图像的去噪中起到了积极的作用。

表 3 艾睿测试集下消融实验指标对比表

Tab.3 Comparison table of ablation experiment indexes under Iraytek testset

Model structure	Iraytek testset	
	PSRN/dB	SSIM
Model 1.(UNet + conv + concat)	27.96	0.8253
Model 2.(UNet + conv + GFC)	28.14	0.8432
Model 3.(UNet+bottleneck+concat)	28.86	0.8628
Model 4.(UNet+ bottleneck + GFC)	29.69	0.8861
Model 5.(UNet +MAR + GFC)	30.71	0.9057

通过上述的两个实验证明,文中提出的 GFDNet 网络不仅可以实现降噪功能,而且可以最大程度的保留红外图像中边缘细节,同时可以有效提升视觉效果。

3 结 论

针对现有的热红外图像去噪算法,无法同时兼顾降噪平滑和细节保留的现象,文中提出一种基于引导滤波的多分支注意力残差深度网络,通过该网络不仅可以有效的完成红外图像的降噪功能,而且可以最大程度保留边缘细节信息。降噪后的红外图像更加贴

近人眼的视觉效果。实验结果表明,文中提出的方法在 PSRN 和 SSIM 指标上取得了较好的结果。但是该网络也存在不足,如降噪后的红外图像在整体对比度上有所下降,图像增强效果不明显等,后续将针对上述的不足继续深入研究。

参考文献:

- [1] Yuan Jijun. Review of infrared detector development [J]. *Laser & Infrared*, 2006, 36(12): 1099-1102. (in Chinese)
- [2] Liu Xinle. Research on fusion method of thermal infrared image and visible light image[D]. Chengdu: University of Electronic Science and Technology of China, 2013. (in Chinese)
- [3] Zhang Shengwei, Xiang Wei, Zhao Yaohong. Stripe noise removal method for infrared images based on guided filtering [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2017, 29(8): 1434-1443. (in Chinese)
- [4] Cao Y P, Tisse C L. Single-image-based solution for optics temperature-dependent nonuniformity correction in an uncooled long wave infrared camera [J]. *Optics Letters*, 2014, 39(3): 646-648.
- [5] Chen Shihong, Chen Rongjun. Infrared image enhancement algorithm based on iteration deep convolution network [J]. *Laser & Infrared*, 2021, 51(1): 114-121. (in Chinese)
- [6] Zhou Yi, Chai Xuliang, Tian Yuan, et al. Studies on InAs/GaAsSb mid-wavelength interband cascade infrared focal plane arrays [J]. *Journal of Infrared and Millimeter Waves*, 2019, 38(6): 745-750. (in Chinese)
- [7] Liu Liping, Qiao Lele, Jiang Liucheng. Overview of image denoising methods [J]. *Journal of Frontiers of Computer Science and Technology*, 2021, 15(8): 1418-1431. (in Chinese)
- [8] Tian Chunwei, Fei Lunke, Zheng Wenxian, et al. Deep learning on image denoising: An overview [J]. *Neural Networks*, 2020, 131(4): 251-275.
- [9] Liu B, Liu J. Overview of image denoising based on deep learning [J]. *Journal of Physics: Conference Series*, 2019, 1176(2): 022010.
- [10] Shizuo Kaji, Satoshi Kida. Overview of image to image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in medical imaging [J]. *Radiological Physics and Technology*, 2019, 12: 235-248.
- [11] Peng Zhenming, Chen Yingpin, Pu Tian. et al. Image denoising based on sparse representation and regularization constraint: A review [J]. *Journal of Data Acquisition & Processing*, 2018,

- 33(1): 1-11. (in Chinese)
- [12] Liu Chengshi, Zhao Zhigang, Li Qiang, et al. Enhanced low-rank representation image denoising algorithm [J]. *Computer Engineering and Applications*, 2020, 56(2): 216-225. (in Chinese)
- [13] Wang Hongyu, Chen Dongmei, Wang Hui. Image denoising model based on low rank and nonlocal sparse representation [J]. *Journal of Yanshan University*, 2017(3): 272-277. (in Chinese)
- [14] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, et al. Image denoising by sparse 3D transform-domain collaborative filtering [J]. *IEEE Transactions on Image Processing*, 2007, 16(8): 2080-2095.
- [15] Dabov K, Foi A, Katkovnik V, et al. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space[C]//Proceedings of the 2007 International Conference on Image Processing, 2007, 9: 313-316.
- [16] Yaniv Romano, Michael Elad. Improving K-SVD denoising by post-processing its method-noise[C]//2013 IEEE International Conference on Image Processing, 2013.
- [17] Zhang Kai, Zuo Wangmeng, Chen Yunjin, et al. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising [J]. *IEEE Transactions on Image Processing*, 2017, 26(7): 3142-3155.
- [18] Yu S, Park B, Jeong J. Deep iterative down-up CNN for image denoising[C]//Processings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019, 7: 2095-2103.
- [19] Liu Yichang, Ma Wei, Xu Shibiao. Edge-fidelity image denoising based on convolutional neural network [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2020(11): 1822-1831. (in Chinese)
- [20] Sun Haoze, Chang Tianqing, Zhang Lei, et al. Fast armored target detection based on lightweight network [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2019(7): 1110-1121. (in Chinese)
- [21] Chen Jingwen, Chen Jiawei, Chao Hongyang. et al. Image blind denoising with generative adversarial network based noise modeling[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, 7: 3155-3164.
- [22] Zhang Kai, Zuo Wangmeng, Zhang Lei. et al. Learning a single convolutional super-resolution network for multiple degradations[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018,
- [23] Itti Laurent, Christof Koch. Computational modelling of visual attention [J]. *Nature Reviews Neuroscience*, 2001, 2: 194-203.
- [24] Li Xiang, Wang Wenhai, Hu Xiaolin. et al. Selective kernel networks[C]//CVPR2019, 2019.
- [25] He Kaiming, Sun Jian, Tang Xiaoou. Guided image filtering [J]. *IEEE Transactions on Software Engineering*, 2013, 35(6): 1397-1409.
- [26] Zhu Xizhou, Cheng Dazhi, Zhang Zheng, et al. An empirical study of spatial attention mechanisms in deep networks[C]//ICCV2019, 2019.
- [27] Zhang Yulun, Li Kunpeng, Li Kai. et al. Image super-resolution using very deep residual channel attention networks[C]//ECCV2018, 2018: 294-310.
- [28] Chen Long, Zhang Hanwang, Xiao Jun, et al. SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning[C]//CVPR, 2017.
- [29] He Kaiming, Sun Jian, Tang Xiaoou. Single image haze removal using dark channel prior [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(12): 2341-2353.
- [30] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//CVPR, 2016.
- [31] Chen Liang Chieh, George Papandreou, Iasonas Kokkinos, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [32] 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.
- [33] Wang Zhou, Li Qiang. Information content weighting for perceptual image quality assessment [J]. *IEEE Transactions on Image Processing*, 2011, 20(5): 1185-1198.
- [34] Guo Shi, Yan Zifei, Zhang Kai, et al. Toward convolutional blind denoising of real photographs[C]//CVPR, 2019.
- [35] Zamir S W, Arora A, Khan S H, et al. Learning enriched feature for real image restoration and enhancement[C]//ECCV, 2020: 492-511.
- [36] Saeed Anwar, Nick Barnes. Real image denoising with feature attention[C]//IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [37] Zhang Kai, Li Yawei, Zuo Wangmeng, et al. Plug-and-play image restoration with deep denoiser prior [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(10): 6360-6376.