

基于卷积神经网络的低照度可见光与近红外图像融合

唐超影^{1,2}, 浦世亮^{1*}, 叶鹏钊¹, 肖飞¹, 冯华君²

¹杭州海康威视数字技术股份有限公司海康威视研究院, 浙江 杭州 310051;

²浙江大学光电科学与工程学院, 浙江 杭州 310027

摘要 针对低照度应用场景,提出一种基于卷积神经网络的可见光与近红外融合算法,采用端到端网络实现了图像融合,所得融合图像能够兼顾近红外图像的信噪比与可见光图像的色彩。采集了真实场景下精确配准的近红外-可见光图像对作为训练集样本,提升了网络对实际数据的融合效果。通过对训练样本进行融合预处理,提升了网络对近红外图像中细节纹理信息的提取能力。各项测试表明,本文算法在主观感受和客观评价上均优于现有算法。

关键词 图像处理; 图像融合; 卷积神经网络; 近红外光

中图分类号 TN911.73

文献标志码 A

doi: 10.3788/AOS202040.1610001

Fusion of Low-Illuminance Visible and Near-Infrared Images Based on Convolutional Neural Networks

Tang Chaoying^{1,2}, Pu Shiliang^{1*}, Ye Pengzhao¹, Xiao Fei¹, Feng Huajun²

¹Hikvision Research Institute, Hangzhou Hikvision Digital Technology Co., Ltd., Hangzhou, Zhejiang 310051, China;

²College of Optical Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China

Abstract Herein, for low-illumination application scenes, an end-to-end convolutional neural network (CNN) is proposed for the fusion of near-infrared (NIR) and visible images. The fused image can combine the signal-to-noise ratio of an NIR image and the color of visible images. To verify the capability of CNN for practical fusion tasks, a real dataset with accurate registration was collected. Moreover, the training set was preprocessed via information fusion, thereby enabling the network to extract additional information from NIR images. Experimental results reveal that the proposed method is superior to existing fusion methods in terms of visual quality and quantitative measurements.

Key words image processing; image fusion; convolutional neural network; near-infrared light

OCIS codes 100.1758; 100.4996; 110.3080

1 引 言

环境照度是影响图像摄影效果的主要因素之一,过低的照度会导致图像噪声放大,信息被掩盖。针对这一问题,比较直接的方法是增加曝光时间或增大传感器像素感光面积。然而增加曝光时间会导致运动模糊,且难以用于视频拍摄;增大传感器像素感光面积则对硬件要求较高,在极低照度下图像摄影效果提升有限。另一种方式则是通过增加近红外补光得到高信噪比的近红外灰度图像,然后与彩色可见光图像进行融合,获得同时具有可见光色彩与

近红外图像信噪比的融合图像。由于人眼无法感知近红外波段,近红外补光并不会影响正常生活环境,因此,近红外-可见光融合技术在夜间监控等低照度场景中具有重要的应用价值。

不同于传统的长波红外与可见光亮度融合^[1],近红外与可见光融合的目的是获取融合后的彩色图像,如何保证融合图像色彩准确是融合的难点,若直接对亮通道融合再将其转换到色彩空间,融合图像的色彩会有严重偏差。学术界对于可见光与近红外融合最初是作为近红外上色问题来讨论的^[2-3],因此并没有考虑可见光噪声。近年来,人们逐渐将低照

收稿日期: 2020-04-14; 修回日期: 2020-04-30; 录用日期: 2020-05-11

基金项目: 浙江省博士后项目择优资助(zj2019047)

* E-mail: pushiliang, hri@hikvision.com

度可见光与近红外融合作为独立领域开展研究,传统融合算法一般将融合分为降噪与融合两步处理^[4-5]。例如:文献[5]使用三维块匹配算法(Block Matching 3D, BM3D)^[6]对可见光预降噪,再将降噪后的可见光与近红外图像进行融合。分步处理的问题在于降噪时没有利用近红外图像信息,导致降噪效果差,信息丢失多。文献[7]采用多尺度融合,在每一层上计算小波尺度图,然后基于尺度图对小波系数降噪以抑制高频噪声,最后融合降噪后的系数并作为下一层的可见高频信息向上传递,但在边缘区域容易产生光晕等瑕疵。文献[8]基于视网膜大脑皮层(Retinex)模型在照度图上进行融合,对于可见光噪声大的暗区域更多地融合近红外信息,从而降低融合图像噪声,然而通过融入近红外亮度来提升信噪比的方法也会导致色彩不准。Jee 等^[9]则是将融合操作转换为近红外辅助可见光降噪并叠加近红外纹理,其降噪算法的本质为导向滤波^[10],同样存在色彩不准的问题。

目前,低照度可见光与近红外融合研究还主要集中在传统算法,基于深度学习的研究并不多,其主要原因之一在于获得高质量训练集较为困难。文献[11]通过施加高斯噪声仿真噪声图像,但实际场景中低照度可见光噪声较为复杂,器件噪声等固有噪声占比较大,因此对于真实数据融合效果不佳。另一方面,可见光和近红外图像对配准精度要求很高,大多现有的数据集难以满足要求,例如文献[12]提供了使用尺度不变特征变换算法(SIFT)进行配准的可见光-近红外图像数据集,但配准精度无法达到训练要求。

目前学术界对基于神经卷积网络(CNN)的低照度可见光与近红外融合的研究较为理想,多采用仿真数据集训练。然而实际低照度场景的噪声形态与仿真噪声有较大差异。近年来,不少研究学者将仿真噪声向实际情况逼近^[13-15],但选用的噪声模型还是以期望为零的指数族分布为主,这些噪声模型可以很好地

模拟中低增益下的噪声,却难以模拟增益更大时的噪声形态。除了存在仿真噪声与实际噪声差异问题,现有算法多将降噪与融合分开处理,在降噪时并没有利用近红外信息,导致可见光信息丢失。

本文提出了一种端到端的低照度可见光与近红外融合网络,将降噪与融合同步处理,避免了降噪时可见光信息丢失。针对实际双光融合需求,采集了精确配准的实际低照度可见光与近红外样本集,提升了网络对实际数据的融合效果。近红外样本经加权预处理后,在近红外样本中融入部分可见光信息,从而使网络能够学会从近红外图像中提取到更多的高频信息,同时保证颜色准确。网络采用平均绝对误差(MAE)与感知损失函数^[16]对融合图像进行约束,前者用于保证色彩准确性,后者则用于保证融合结果的视觉感受。训练时先使用室内照片进行初步训练,照片相较于室外场景色彩更加丰富,有助于网络对色彩的处理。此外,室内照片反射特性相对简单,利于网络快速收敛。之后再使用室外场景样本对初步训练的网络精调,以提升对实际数据的处理效果。实验结果表明,本文提出的融合网络能够同时兼顾近红外信噪比与可见光色彩。

2 算法介绍

2.1 数据集

近年来研究表明基于仿真噪声图像训练的网络,在处理实际问题时往往效果欠佳,相较于仿真噪声(图 1(b),噪声模型^[15]),实际低照度场景由于截断效应和器件噪声、热噪声占比较高因素,噪声形态与仿真模型的差异较大,器件噪声等固有噪声会导致图像出现对比度降低,如图 1(a)所示。仿真噪声模型的主要问题在于未考虑固有噪声,噪声模型可表示为

$$y \sim N(\mu = x, \sigma^2 = M(x)), \quad (1)$$

其中 y 表示噪声像素值, x 为无噪声像素值, $N(\mu, \sigma^2)$ 为高斯分布, $M(x)$ 为 x 的函数,即仿真噪声像

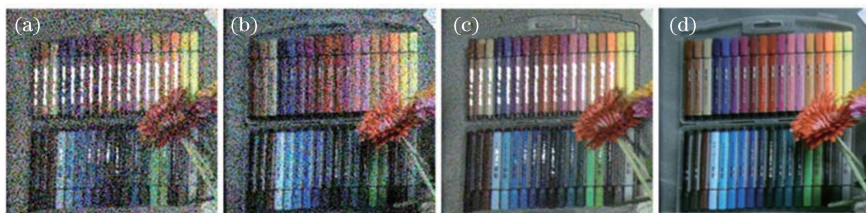


图 1 实际噪声与仿真噪声对比。(a)实际噪声图像;(b)仿真噪声图像;

(c)连续 100 frame 实际噪声图像的平均结果;(d)连续 100 frame 仿真噪声图像的平均结果

Fig. 1 Comparison between real noisy images and simulation noisy images. (a) Real noisy image; (b) simulation noisy image; (c) average of 100-frame real noisy images; (d) average of 100-frame simulation noisy images

素均值为无噪声像素。图 1(c)(d)所示分别为连续 100 frame 实际低照度图像和仿真噪声图像(原始数据)的平均结果,可以明显看出,实际噪声图像依旧存在较大的噪声。使用 200 frame 实际低照度图像和仿真噪声图像的平均结果几乎无差别,即残余的噪声为时域不变噪声,其中仿真噪声由于期望为零,平均后基本被消除。

考虑到仿真噪声与实际噪声间的差异问题,越来越多的图像降噪^[13]、去模糊以及增强等复原网络使用真实数据集。相比而言,近红外与低照度可见光融合的真实数据集更难获取,困难主要体现在两方面:一是图像配准精度。实验中发现,为了得到高质量的融合结果,近红外与可见光图像的空间坐标误差需要达到亚像素级。若使用存在空间偏差的样本进行训练,网络在融合图像前还需要对异源的近红外与可见光图像进行配准,这无疑增加了训练难

度。实际上,单独的多光谱图像配准都难以达到足够的精度,例如我们尝试过利用 PWC-Net^[17]来训练可见光与近红外图像的配准,由于两者光谱特性不同,配准结果并不理想。二是目标图像的定义。所谓目标图像定义就是以什么样的图像作为融合图像,目前大多使用的是无噪声、清晰的可见光图像,即采用类似降噪等常规复原网络。然而将清晰的可见光图像作为目标融合图像会导致网络功能最终变成辅助降噪近红外图像,融合结果缺乏近红外信息。图 2(a)和(b)给出标准 24 色卡的可见光图像和近红外图像。为此,FusionNet^[11]在损失函数中增加了网络输出与输入近红外图像间的差异约束项,使网络学会从近红外图像中提取融合信息,但此方式容易导致输出图像的色彩产生偏差,同时近红外与可见光信息不匹配区域的融合效果欠佳,如图 2(c)所示。

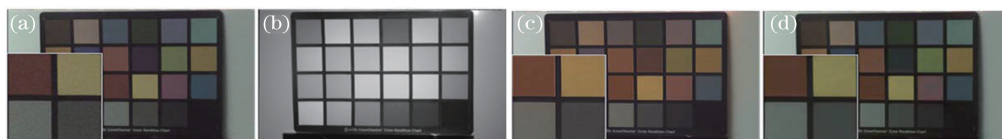


图 2 标准 24 色卡。(a)输入可见光图像;(b)输入近红外图像;(c)FusionNet^[11]结果;(d)本文结果

Fig. 2 Standard 24 color card. (a) Visible input image; (b) NIR input image; (c) result of FusionNet^[11]; (d) our result

针对数据集问题,本研究采用相机获取真实低照度可见光与近红外图像样本对,相机型号为 HIK iDS-2CD9736-AMS,通过分光棱镜将可见光与近红外光分别导入对应传感器,其中可见光与近红外图像间仅存在微小的刚性空间变换,利用棋盘格校准获取配准矩阵并应用于所有图像样本对。近红外补光波长为 850 nm,图像大小为 2000 pixel × 1500 pixel。目标融合图像仍旧采用清晰的无噪声

可见光图像,对近红外图像进行预处理,使网络能够同时兼顾近红外信息与可见光色彩准确性。本文方法对标准 24 色卡的测试结果如图 2(d)所示,可以看到融合图像色彩保持程度有大幅提升。考虑到室外场景色彩较为单一,本研究采用的数据集分为两部分:一部分为室内拍摄的打印图片,图像为色彩丰富的 DIV2K_HR 数据集,即在室内对 DIV2K_HR 数据集的打印图片进行拍摄,如图 3(a)~(c)所示,



图 3 实际数据集示例。(a)室内图片噪声可见光图像;(b)室内图片近红外图像;(c)室内图片目标图像;
(d)室外噪声可见光图像;(e)室外近红外图像;(f)室外目标图像

Fig. 3 Examples of real dataset. (a) Noisy visible image of photographs taken indoors; (b) NIR image of photographs taken indoors; (c) target image of photographs taken indoors; (d) outdoor noisy visible image; (e) outdoor NIR image; (f) outdoor target image

共 200 对;另一部分为室外场景,如图 3(d)~(f)所示,共 100 对。根据曝光时间比例,将获得的低照度可见光图像的亮度调整至与对应高照度图像接近的亮度。需要注意的是,为方便观察,本文出现的图像均为经过图像信号处理(ISP)后的结果,实际输入网络的数据为 Raw 格式数据。

2.2 样本预处理

图像融合的目的是最大化可见光与近红外图像信息,但融合过多的近红外信息会导致色彩偏差。为了使网络融合近红外图像信息而不偏色,本研究对训练样本进行加权预处理,类似于对样本打标签。首先将目标可见光图像进行频域分解,分别得到高频部分 f_{VIS_high} 和低频部分 f_{VIS_low} ,表达式为

$$f_{VIS} = f_{VIS_low} + f_{VIS_high} \quad (2)$$

然后根据文献[18]中的 Retinex 模型,将可见光图像分解为反射分量 R 和照度分量 T ,即

$$f_{VIS} = R \circ T, \quad (3)$$

式中:“ \circ ”表示“元素乘法”。最后,根据照度分量 T 计算可见低频分量 f_{VIS_low} 的权重,将可见光图像与近红外图像 f_{NIR} 进行加权,即

$$w_{NIR} = \{f(T) \cdot f_{VIS_low} + [1 - f(T)] \cdot f_{NIR}\} + \alpha \cdot f_{VIS_high}, \quad (4)$$

式中: $f(T) \in [0, 1]$ 为照度分量的减函数,即照度分量越大权重越低,从而仅在可见光亮度较低的区域

引入近红外亮度,本文采用的是线性函数; α 为高频叠加系数,每张加权图像随机选取 $[0, 0.3]$ 中的值。近红外加权图像数量占全部训练集的一半。使用亮度减函数的目的是为了减少近红外光谱对于融合结果的影响,不同谱段的近红外反射特性不同,对同一场景的成像结果也不同,由于光谱响应主要反映在低频分量,因此仅对可见光较暗的区域引入近红外亮度,以降低因近红外光谱变化带来的影响。

通过加权预处理,近红外图像加入了部分目标可见光信息,起到了约束近红外信息的作用,能够使得网络学会从近红外图像中提取更多近红外高频信息,同时又不会破坏色彩准确度。

2.3 网络结构

实验中发现,由于低照度可见光噪声大,非多尺度网络的输出结果往往存在低频色彩噪声。因此,采用感受野更大的 U-net 结构^[19]。本文中 U-net 包含 4 层 16 倍上下采样,每一层由一个卷积层 (Convolutional Layer)、批标准化 (Batch Normalization, BN)层和下(上)采样层组成,上下采样分别由反卷积 (Deconvolution) 和最大池化 (Max-pooling)实现。其中最上层卷积特征层数为 128,之后每下采样一次,特征层数就翻倍一次。将可见光与近红外图像输入网络后直接由连接层 (Concat Layer)连接,网络结构如图 4 所示。

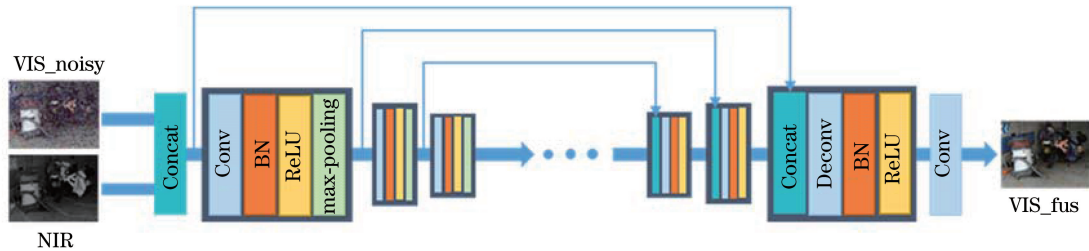


图 4 本文网络结构示意图

Fig. 4 Structure of our networks

2.4 损失函数

目前降噪网络多采用 MAE 或感知损失函数 (Perceptual Losses)^[16]。前者能够得到更平坦的降噪效果,有较好的峰值信噪比(PSNR),但细节纹理损失较多,后者则能带来更好的视觉效果。低照度可见光图像融合同样包含降噪任务,不同之处在于细节纹理更多来自于近红外图像,网络需要根据近红外图像对可见光图像进行降噪并融入近红外信息。因此,我们的损失函数同时考虑图像降噪与高频信息提取,由 MAE 与感知损失函数两部分组成。

MAE 为网络输出与目标真值间的 L1 范数,即

$$L_{L1} = \frac{1}{W \times H} \| \mathbf{Y} - \mathbf{Y}^* \|_1, \quad (5)$$

其中 W 和 H 为图像的宽和高。

感知损失函数同样为网络输出与目标真值的函数,这是因为实验中发现,若将网络输出与输入的近红外图像进行约束,融合结果容易产生偏色现象,且不容易收敛。此外,实际场景中由于反射特性不同,近红外图像与可见光图像信息并不匹配,包含许多可见光有但近红外没有的信息。如果过度依赖近红外信息,会导致融合结果出现信息丢失或类似多帧融合问题中的“鬼影”。感知损失函数选用预训练好

的 VGG-19^[20] 网络, 选取 ‘conv1_2’, ‘conv2_2’, ‘conv3_2’, ‘conv4_2’ 和 ‘conv5_2’ 特征层的特征用于计算损失值, 令 $\Phi_k(\mathbf{Y})$ 表示 VGG-19 第 k 层的特征提取函数, 取各层特征值的 L1 范数, 感知损失函数可以表示为

$$L_P = \sum_k \lambda_k \|\Phi_k(\mathbf{Y}) - \Phi_k(\mathbf{Y}^*)\|_1, \quad (6)$$

其中 λ_k 为不同层提取特征值的权重。

完整的损失函数即为

$$L = L_{L1} + \beta \cdot L_P, \quad (7)$$

其中 β 为感知函数权重比例。

3 实验结果分析

3.1 主观效果

本文网络训练使用 Adam 优化算法, Batch size 为 32, 训练图像块大小为 128 pixel × 128 pixel。训

练时先使用室内照片数据集进行初步训练, 然后再由室外数据集进行调优。分别与传统低照度可见光与近红外融合算法 IFNR^[5] 以及基于 CNN 的 FusionNet^[11] 进行比较。对于 FusionNet, 训练集对文献[4]提供的数据集添加仿真噪声, 噪声模型满足高斯-泊松分布^[15]。具体地来讲, 对图像进行反白平衡、反 Gamma 等反向 ISP 处理后转换为 Raw 数据, 反向 ISP 的各参数与实际拍摄图像的参数相同, 然后在 Raw 数据上施加仿真噪声, 最后正向 ISP 处理得到噪声图像。网络收敛后同样再由本文数据集仿真噪声数据进行调优。图 5 为本文复现的 FusionNet 对于仿真噪声测试集融合结果, 图 5(a)~(c) 分别为输入仿真噪声图像、无噪声可见光真值以及 FusionNet 结果。可以看到, 对于仿真噪声图像, FusionNet 基本能够达到文中要求的效果。



图 5 复现 FusionNet 对仿真测试集的融合结果。(a)输入仿真噪声图像;(b)无噪声可见光图像;(c)融合结果

Fig. 5 Results of FusionNet for simulation dataset. (a) Simulation noisy input image; (b) noise-free visible image; (c) result of fusion

然而仿真图像训练的网络对实际噪声图像的处理效果欠佳, 本研究分别对比了低、中、高三种增益下的图像融合效果。首先是低增益场景, 图 6(a)~(c) 分别为噪声可见光图像、近红外图像与高照度无

噪可见光图像, 图 6(d)~(e) 分别为 IFNR、FusionNet 和本文算法的图像融合结果。噪声可见光图像中蓝色框内的黄色区域完全被噪声湮没, 信息基本丢失, 近红外图像的对应区域则包含这部分



图 6 实际噪声图像 1 融合结果比较。(a)实际噪声图像;(b)输入近红外图像;(c)无噪声可见光图像;

(d) IFNR 融合结果;(e) FusionNet 融合结果;(f) 本文结果

Fig. 6 Visual comparison of image fusion for real noisy image 1. (a) Real noisy visible image; (b) NIR input image; (c) noise-free visible image; (d) result of IFNR; (e) result of FusionNet; (f) our result

信息,三种算法均能够根据近红外图像恢复信息。但在噪声方面,使用仿真噪声训练的 FusionNet [图 6(e)]并不能很好地抑制实际噪声。传统算法虽然能较好抑制噪声,但图像清晰程度上略低于网络融合结果。另一方面,从图中蓝色框中可以看出,本文算法融合结果缺少近红外阴影,这是由于为了保证色彩准确性而限制了近红外低频信息的引入。同样,IFNR 算法融合结果的近红外阴影也较少。而 FusionNet 结果中保留了近红外阴影,从信息最大化的角度来看 FusionNet 的结果最优,但过多引入近红外信息会导致色彩不准,从图 6~9 中可以发现,FusionNet 融合图像色彩与实际色彩偏差较大。

近红外与可见光图像由于反射特性的不同,存在许多信息不匹配的区域,尤其是可见光图像有而

近红外图像没有的区域。对于这些区域,希望融合图像能够最大程度地保持可见光信息。图 7 为信息不匹配示例,框图为放大图。由于输入近红外图像图 7(b)中缺乏输入可见光图像图 7(a)中的行人,因此 IFNR 算法无法从近红外图像中获取行人的高频信息,致使因降噪丢失的信息无法恢复,融合结果较为模糊[图 7(d)]。FusionNet 由于损失函数中包含近红外图像的损失项,网络会引入较多近红外低频信息,最终导致融合结果出现虚影[图 7(e)]。相比上述算法,本文损失函数中仅对输出结果与目标真值进行约束,近红外高频约束则通过样本预处理实现,因此,融合结果能够最大程度保证可见光信息且不受近红外亮度的干扰,如图 7(f)所示。



图 7 图像内容不匹配时融合结果的比较。(a)实际噪声图像;(b)输入近红外图像;
(c)无噪声可见光;(d) IFNR 融合结果;(e) FusionNet 融合结果;(f)本文结果

Fig. 7 Visual comparison of content mismatched image fusion. (a) Real noisy visible image; (b) NIR input image;
(c) noise-free visible image; (d) result of IFNR; (e) result of FusionNet; (f) our result

图 8 和图 9 为实际高、中噪声数据融合结果的比较。实际应用中,随着增益的增加,噪声分布不再是简单的高斯-泊松分布,器件噪声等固有噪声占比变大,如图 8(a)中的条纹噪声。同时,由于截断效应,图像对比度也会降低,如图 8(a)低照度图像和图 8(c)高照度图像。对于大噪声数据,IFNR 需要增大可见光图像的降噪强度,并提高近红外图像的融合比例,导致图像对比度产生变化,如图 8(d)所示。FusionNet 一方面难以处理实际大噪声,融合结果中条纹噪声较大,图像的对比较低,另一方面由于将降噪和融合分步处理,可见光信息丢失较多,如图 8(e)所示。本文算法则能够较好地处理实际噪声,如图 8(f)所示,融合结果在对比度和色彩方面均表现良好。此外,实际低照度图像存在蓝色偏紫等的问题,如图 8(a)中电瓶车挡风布,主要是由于噪声截断导致图像红色通道

均值上移。仿真数据由于无法模拟这些实际问题,最终的融合结果难以恢复这些颜色。

实际低照度场景应用中,经常会出现可见光图像被噪声湮没,而近红外图像无信息的情况,如图 8 中红色条纹以及图 9 中的 Hello Kitty 图案。对比图 8(d)~(f)红色条纹可发现,IFNR 与 FusionNet 均会导致条纹变淡和边缘模糊,本文算法能够在不影响可见光原有信息的前提下去除噪声。但当可见光信噪比更低,以至于难以区分是真实信息还是噪声时,本文算法会优先去除噪声,在一定程度上丢失信息。如图 9 所示,长曝光图像图 9(c)中的 Hello Kitty 图案在噪声图像图 9(a)中完全被噪声湮没,近红外图像图 9(b)中缺少对应信息。相较于 IFNR [图 9(d)]和 FusionNet [图 9(e)],本文算法结果噪声更低,但 Hello Kitty 图案信息相对较少。

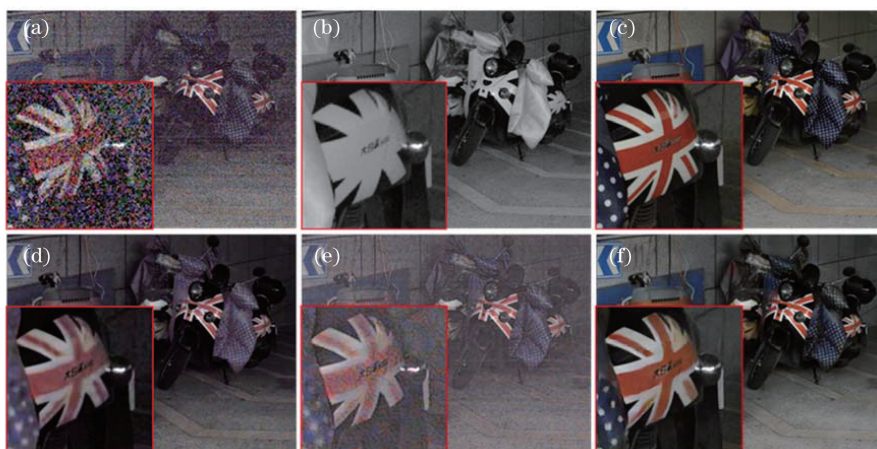


图 8 实际噪声图像 2 测试比较;(a)实际噪声图像;(b)输入近红外图像;

(c)无噪声可见光图像;(d) IFNR 融合结果;(e) FusionNet 融合结果;(f)本文结果

Fig. 8 Visual comparison of image fusion for real noisy image 2. (a) Real noisy visible image; (b) NIR input image; (c) noise-free visible image; (d) result of IFNR; (e) result of FusionNet; (f) our result

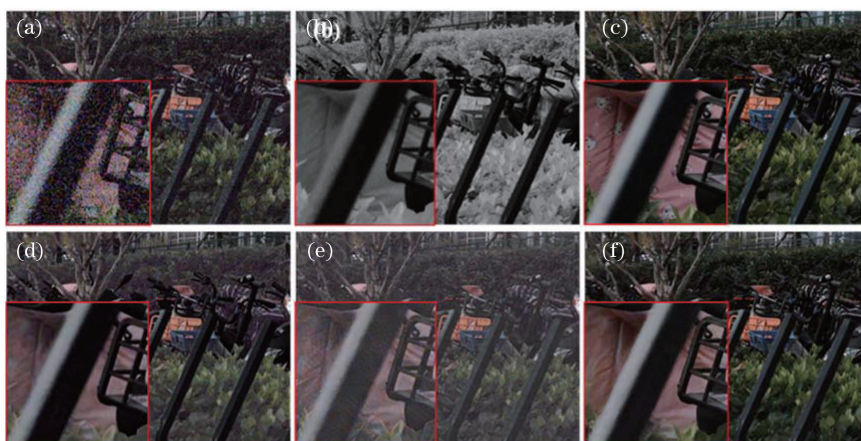


图 9 实际噪声图像 3 测试比较;(a)实际噪声图像;(b)输入近红外图像;

(c)无噪声可见光图像;(d) IFNR 融合结果;(e) FusionNet 融合结果;(f)本文结果

Fig. 9 Visual comparison of image fusion for real noisy image 3. (a) Real noisy visible image; (b) NIR input image; (c) noise-free visible image; (d) result of IFNR; (e) result of FusionNet; (f) our result

从上面的对比结果上来看,本文算法对于中、高增益场景的融合结果更优,主要体现在:融合图像噪声、色彩准确度、可见光信息保留程度、图像对比度以及近红外信息量等方面。

3.2 客观指标

对于融合问题理论上并没有合适的客观评价指

标,还是以主观评价为主。由于降噪问题也属于融合的一部分,并且训练以无噪声可见光图像作为真值,因此,本研究还比较了融合图像的 PSNR 和结构相似度(SSIM),见表 1,用于评价融合图像噪声程度,PSNR 和 SSIM 指标均是值越大表明图像越接近无噪声可见光图像。

表 1 各算法 PSNR 与 SSIM 指标比较

Table 1 Comparison of PSNR and SSIM of different algorithms

Image	IFNR ^[5]		FusionNet ^[11]		Our result	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Fig. 6	32.068	0.949	27.950	0.890	30.600	0.919
Fig. 8	25.618	0.729	23.591	0.630	33.120	0.885
Fig. 9	27.622	0.770	24.405	0.652	34.858	0.890

从结果上来看, FusionNet 对于实际噪声数据处理效果有限, 特别是相较于其在仿真噪声测试集上的表现(平均 PSNR 为 32.603, SSIM 为 0.905)。当噪声水平较高时, 传统算法逐渐难以应对, 本文算法对于大噪声真实场景能够较好地抑制噪声、保持细节。

表 2 各算法无参考评价结果比较

Table 2 Comparison of blind image quality assessment of different algorithms

Image	Input	IFNR ^[5]	FusionNet ^[11]	Our result
Fig. 6	6.176	5.629	6.084	5.932
Fig. 8	8.025	5.439	6.512	4.985
Fig. 9	7.770	5.856	6.271	5.410

4 结 论

提出了一种基于卷积神经网络的近红外与低照度可见光融合算法, 实际采集了近红外-低照度可见光图像样本对, 样本更加符合低照度下可见光噪声的分布形态。对近红外图像的预处理保证了网络在提取近红外高频信息的同时不会破坏可见光信息。网络采用端到端结构, 将降噪与融合同时处理, 避免了可见光图像降噪时信息丢失。实际测试集的测试结果表明, 本文算法在主观效果与各项客观指标上均有优异的表现, 推进了深度学习在实际近红外与低照度可见光融合问题上的应用。

参 考 文 献

- [1] Zhou X L, Jiang Z T. Infrared and visible image fusion combining pulse-coupled neural network and guided filtering [J]. *Acta Optica Sinica*, 2019, 39(11): 1110003.
周晓玲, 江泽涛. 结合脉冲耦合神经网络与引导滤波的红外与可见光图像融合 [J]. *光学学报*, 2019, 39(11): 1110003.
- [2] Lim C B, Salud M M, Uttoh B, et al. Coloring high-resolution near-infrared images using chromatic cues from low-resolution references [C] // 2015 International Conference on Humanoid, December 9-12, 2015, Cebu City, Philippines. New York: IEEE Press, 2015: 1-5.
- [3] Son C H, Zhang X P. Near-infrared coloring via a contrast-preserving mapping model [J]. *IEEE Transactions on Image Processing*, 2017, 26(11): 5381-5394.
- [4] Son C H, Zhang X P. Layer-based approach for image pair fusion [J]. *IEEE Transactions on Image Processing*, 2016, 25(6): 2866-2881.
- [5] Son C H. Near-infrared fusion via a series of transfers for noise removal [J]. *Signal Processing*, 2018, 143: 20-27.
- [6] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering [J]. *IEEE Transactions on Image Processing*, 2007, 16(8): 2080-2095.
- [7] Su H N, Jung C. Multi-spectral fusion and denoising of RGB and NIR images using multi-scale wavelet analysis [C] // 2018 24th International Conference on Pattern Recognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE Press, 2018: 1779-1784.
- [8] Jung T Y, Son D M, Lee S H. Visible and NIR image blending for night vision [C] // International Conference on Image Processing, Computer Vision and Pattern Recognition (ICCV). Las Vegas: CSREA Press, 2019: 110-113.
- [9] Jee S, Kang M G. Sensitivity improvement of extremely low light scenes with RGB-NIR multispectral filter array sensor [J]. *Sensors*, 2019, 19(5): 1256.
- [10] He K M, Sun J, Tang X O. Guided image filtering [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(6): 1397-1409.
- [11] Jung C, Zhou K L, Feng J W. FusionNet: multispectral fusion of RGB and NIR images using two stage convolutional neural networks [J]. *IEEE Access*, 2020, 8: 23912-23919.
- [12] Matthew B, Sabine S. Multi-spectral SIFT for scene category recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2011, Providence, RI, USA. New York: IEEE, 2011: 12218780.
- [13] Claus M, Gemert J V. ViDeNN: deep blind video denoising [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 16-17, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1843-1852.

- [14] Brooks T, Mildenhall B, Xue T F, et al. Unprocessing images for learned raw denoising[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 11028-11037.
- [15] Mildenhall B, Barron J T, Chen J W, et al. Burst denoising with kernel prediction networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2502-2510.
- [16] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[M]//Computer Vision - ECCV 2016. Cham: Springer International Publishing, 2016: 694-711.
- [17] Sun D Q, Yang X D, Liu M Y, et al. PWC-net: CNNs for optical flow using pyramid, warping, and cost volume [C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8934-8943.
- [18] Fu X Y, Zeng D L, Huang Y, et al. A weighted variational model for simultaneous reflectance and illumination estimation [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2782-2790.
- [19] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [M] // Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2020-03-24]. <https://arxiv.org/abs/1409.1556>.
- [21] Wang Z, 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.
- [22] Zeng H, Zhang L, Bovik A C. Blind image quality assessment with a probabilistic quality representation [C] //2018 25th IEEE International Conference on Image Processing (ICIP), October 7-10, 2018, Athens, Greece. New York: IEEE Press, 2018: 609-613.