

## Article

# Multiple Degradation Skilled Network for Infrared and Visible Image Fusion Based on Multi-Resolution SVD Updation

Gunnam Suryanarayana <sup>1,\*</sup>, Vijayakumar Varadarajan <sup>2,\*</sup>, Siva Ramakrishna Pillutla <sup>3</sup>, Grande Nagajyothi <sup>4</sup> and Ghanya Kotapati <sup>5</sup>

<sup>1</sup> Department of Electronics and Communications, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada 520007, India

<sup>2</sup> Department of Computer Science Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

<sup>3</sup> School of Electronics Engineering, VIT-AP University, Amaravati 522237, India

<sup>4</sup> Department of Electronics and Communications, Golden Valley Integrated Campus, Madanapalli 517325, India

<sup>5</sup> Department of Computer Science Engineering, Sree Vidyanikethan Engineering College, Tirupati 517102, India

\* Correspondence: surya\_gunnam@yahoo.co.in (G.S.); v.varadarajan@unsw.edu.au (V.V.)

**Abstract:** Existing infrared (IR)-visible (VIS) image fusion algorithms demand source images with the same resolution levels. However, IR images are always available with poor resolution due to hardware limitations and environmental conditions. In this correspondence, we develop a novel image fusion model that brings resolution consistency between IR-VIS source images and generates an accurate high-resolution fused image. We train a single deep convolutional neural network model by considering true degradations in real time and reconstruct IR images. The trained multiple degradation skilled network (MDSNet) increases the prominence of objects in fused images from the IR source image. In addition, we adopt multi-resolution singular value decomposition (MRSVD) to capture maximum information from source images and update IR image coefficients with that of VIS images at the finest level. This ensures uniform contrast along with clear textural information in our results. Experiments demonstrate the efficiency of the proposed method over nine state-of-the-art methods using five image quality assessment metrics.

**Keywords:** multiple degradation model; deep networks; image fusion; multi-resolution singular value decomposition

**MSC:** 34A08; 26A33; 34A08; 34A34; 34B10



**Citation:** Suryanarayana, G.; Vijaykumar, V.; Pillutla, S.R.; Nagajyothi, G.; Kotapati, G. Multiple Degradation Skilled Network for Infrared and Visible Image Fusion Based on Multi-Resolution SVD Updation. *Mathematics* **2022**, *10*, 3389. <https://doi.org/10.3390/math10183389>

Academic Editors: Yazhou Yao and Xiaoshui Huang

Received: 22 July 2022

Accepted: 9 September 2022

Published: 19 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Multi-modal images have huge demand in many areas, such as medical imaging, target recognition, and remote sensing [1]. This is due to the advancement in sensor technology. In particular, the infrared (IR) and visible (VIS) sensor combination has distinct advantages. IR sensors capture thermal radiation and represent as IR images. On the contrary, VIS images hold textural details captured from the reflected light on VIS sensors. By considering the mapped IR and VIS sensor data, the fused IR-VIS images have inherent features with good visual understating of a scene [2].

Image fusion aims to extract the vital features from multiple source images and merge them into a single image. In this regard, various algorithms have been developed with certain fusion rules. These are based on saliency detection [3–5], sparse representation theory [6–8], multi-scale transforms [9,10], convolutional neural networks [11,12], subspace methods [13,14], machine and deep learning [15–17], hybrid methods [18,19], and other fusion strategies [20,21].

In spite of the appreciable performance achieved by these methods, the IR-VIS image fusion at different resolutions still remains an unsolved research problem. IR images usually suffer from poor resolution when compared with the corresponding VIS data. However, the existing methods [2–4,6,7,9–16,18–22] do not consider resolution differences between IR and VIS source images, leading to inaccurate fusion results. Under this condition, there are two possible cases while fusing the IR and VIS images with different resolution levels. For intuitiveness, these cases are illustrated in Figure 1. Let the dimensions of IR and VIS images be  $M \times N$  and  $sM \times sN$ , respectively. In case 1, the VIS image is downsampled with a factor  $s$  before applying the fusion process [23]. The fused output in this procedure still suffers from low resolution. For consistency and also to compare with other methods, the fused image is forcibly converted to an HR image. However, this cannot recover the loss of VIS textures caused by the downsampling operation prior to image fusion.

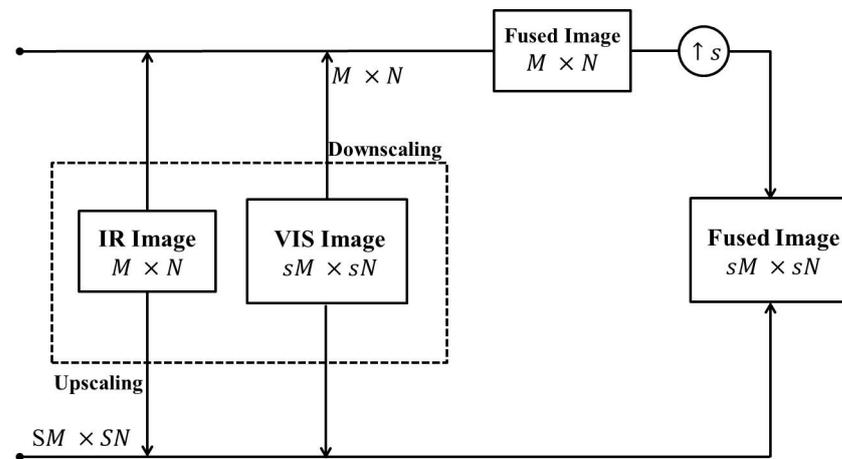


Figure 1. Different resolution IR-VIS image fusion.

To address this, we confine to case 2 in our work and eliminate inaccuracy and blur in the fused image. Here, the IR image is first upsampled to  $sM \times sN$  before using it as the source image [23]. Hence, both IR and VIS images have the same HR dimensions before applying the fusion process. With the proper upscaling of IR images, both IR and VIS source images contribute to the fused image with improved resolution.

In this paper, we address the problem of image fusion with different resolution VIS-IR source images. We learn a deep CNN by assuming multiple true degradations in real time to improve practicability. The learned multiple degradations skilled network (MDSNet) provides good scalability to recover the IR images with low resolution (LR). Moreover, it creates resolution compatibility between IR-VIS images for the fusion process. Correspondingly, the multi-resolution singular value decomposition captures information from IR-VIS source images in horizontal, vertical and diagonal directions at different levels. The contrast correction on IR image coefficients with that of VIS image coefficients at the finest level ensures better fused images. The rest of this paper is organized as follows. In Section 2, we discuss the degradation model, network structure and loss function of MDSNet. In Section 3, we propose our IR-VIS fusion method by detailing MDSNet upscaling and MRSVD updating. In Section 4, the proposed method is validated over the existing methods based on quantitative metrics and qualitative analysis. Finally, conclusions are presented in Section 5.

## 2. MDSNet

In recent times, CNNs have shown unprecedented success in extensive applications. However, we limit our discussion on CNNs to super-resolution (SR) imaging in this paper. Existing SR CNN methods apply bicubic downsampling on an HR image, which do not consider true degradation, thus leading to poor performance. To address this

issue, we employ the MDSNet for SR that can handle multiple degradations and improve the practicability [24,25]. In general, the degradation model of an LR image  $p$  from the corresponding HR image  $q$  can be viewed as

$$p = (q \otimes b) \downarrow_s + n, \quad (1)$$

where  $\otimes$  represents convolution operation,  $b$  is a blur kernel,  $\downarrow_s$  indicates a downsampling operation with a factor  $s$ , and  $n$  is additive Gaussian noise. To obtain better insights on CNNs in the context of SR, we analyze the approximate solution to Equation (1) using the maximum a posteriori (MAP) framework as

$$\hat{q} = \arg \min_q \frac{1}{2\sigma^2} \|(q \otimes b) \downarrow_s - p\|^2 + \lambda\phi(x). \quad (2)$$

Here, the first term  $\frac{1}{2\sigma^2} \|(q \otimes b) \downarrow_s - p\|^2$  represents data fidelity and the second term  $\lambda\phi(x)$  indicates regularization with a trade-off  $\lambda$ . The above MAP problem in Equation (2) conveys that the SR reconstruction process should obey the degradation model while estimating HR images. In addition, it is clear that the estimated SR image  $\hat{q}$  is a function of LR input  $p$ , blur operator  $b$ , and standard deviation  $\sigma$  representing noise level and  $\lambda$ .

$$\hat{q} = f(p, b, \sigma; \theta), \quad (3)$$

where  $\theta$  represents the MAP inference parameter. However, the existing SR networks [5,8,17] limit the single image SR problem to interpolation, neglecting the blurring operation and noise level as  $\hat{q} = f(p; \theta)$ . However, this does not cover true degradations, resulting in poor performance. So, we consider the three inputs  $p, b$ , and  $\sigma$  together as in Equation (3) while modeling MDSNet. The sizes of  $p, b$ , and  $\sigma$  are different, which causes dimensionality mismatch while feeding the CNN. Hence, we apply dimensionality stretching and convert all the input parameters to have a similar size. Let us assume an LR image of size  $M \times N$  with  $C$  number of channels as  $M \times N \times C$ , a blur operator of size  $r \times r$ , and a noise level  $\sigma$ . First, the  $r \times r$  blur matrix is vectorized to  $r^2 \times 1$  and projected onto  $k$ -dimensional space using the principal component analysis (PCA) approach. The obtained vector is then concatenated with the noise level and denoted by a vector  $t$ , which is finally stretched into  $k + 1$  degradation maps of size  $M \times N$  represented as  $M \times N \times (k + 1)$ . Now, these maps have similar size compared to LR images, thus enabling MDSNet capable of handling three inputs simultaneously.

The MDSNet is designed for multiple degradations for improving the real-time SR performance. We describe the architecture of MDSNet in the following.

- The network takes  $M \times N \times (C + k + 1)$  images as input by considering  $M \times N \times C$  LR images concatenated with  $M \times N \times (k + 1)$  degradation maps. The MDSNet achieves non-linear mapping using multiple convolutional layers in cascade form [17]. Each layer includes convolution operation (CONV), batch normalization (BN) [26] and rectifier linear unit (ReLU) [27].
- The first “CONV+BN+ReLU” layer operates on  $M \times N \times (C + k + 1)$  input images.
- All the middle layers are similar to each other, performing “CONV+BN+ReLU” operations.
- The last layer consists of “CONV” operation alone, which produces output HR image patches. The last layer is followed by a sub-pixel convolution layer [28] to convert HR image patches to a single HR image  $sM \times sN \times C$ .
- We use Adam optimizer [29] to minimize the loss function,

$$L(\theta) = \frac{1}{2K} \sum_{i=1}^N \|f(p_i, D_i; \theta) - q_i\|^2, \quad (4)$$

where  $K$  is the total number of images including the degradation maps  $D_i$  and  $p_i, q_i$  are the  $i$ -th LR and HR images, respectively. For all scale factors  $s$ , we set a total of

12 convolution layers with 128 feature maps in each layer. The training is performed separately for different scaling factors. We use the trained model to enhance the resolution of the input LR-IR image in our proposed method.

### 3. Proposed Method

In this paper, we aim to solve the different resolution IF problem of IR and VIS images. Let us denote the ratio of IR image resolution and VIS image resolution by a factor  $s$ . That is, the IR image has dimensions  $M \times N$  and that of the VIS image is  $sM \times sN$ . Given an LR-IR image  $IR_p$  and an HR-VIS  $VIS_q$  image, the fusion process of our proposed MDSNet-MRSVD algorithm is outlined in Figure 2.

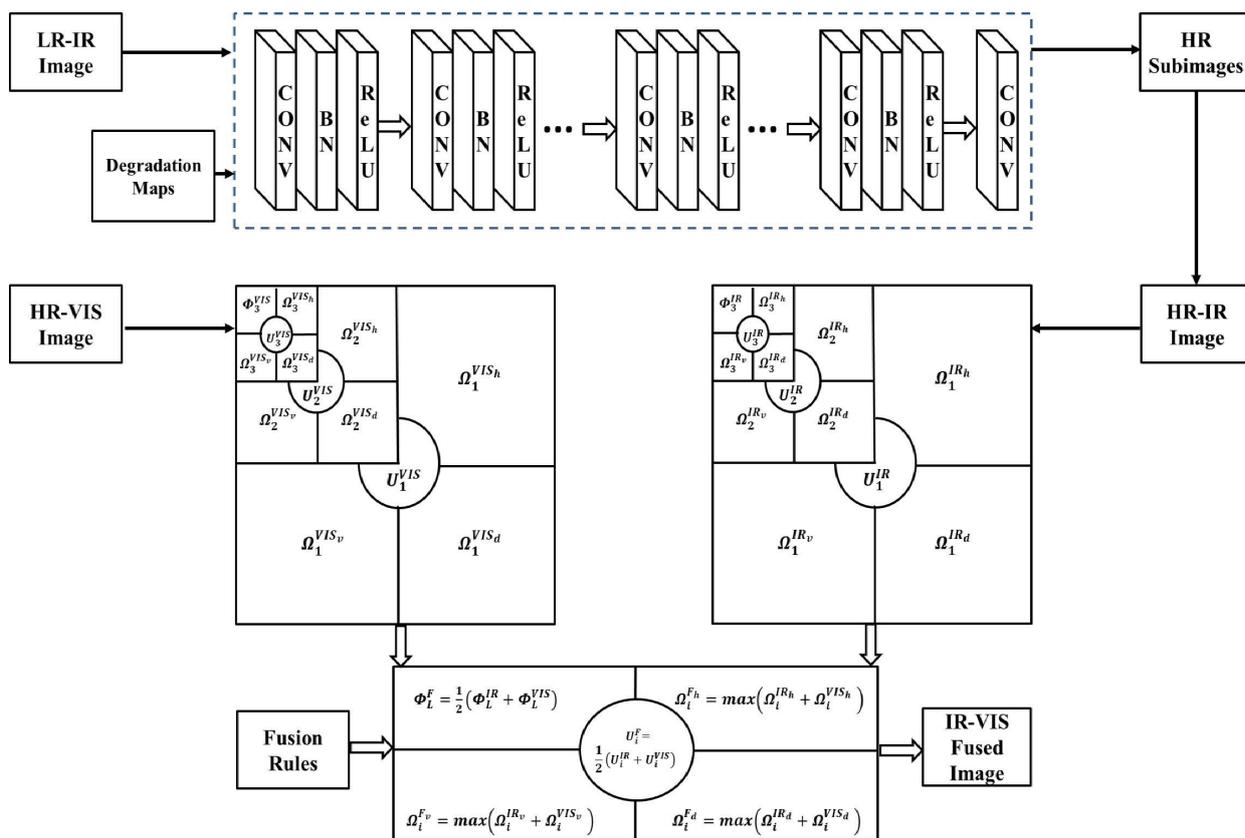


Figure 2. Proposed IR-VIS image fusion methodology.

#### 3.1. LR-IR Image Upscaling Using MDSNet

Since the IR image  $IR_p$  has a lower resolution, it is first subjected to MDSNet to generate the corresponding HR-IR image  $IR_q$ . The MDSNet is trained with numerous HR-LR example patches where the LR images are synthesized using Equation (1). However, the sizes of LR images, blur kernel, and noise are different, which causes dimensionality mismatch while feeding the CNN. Hence, we apply dimensionality stretching and convert all the input parameters to have the same size.

To synthesize HR images using MDSNet, it is required to specify the blur operator and noise level for defining the degradation maps. For blur operation, isotropic Gaussian kernel with a fixed width and anisotropic Gaussian kernel with probability density function having zero mean and variable covariance are used. The blur matrix is vectorized and projected onto  $k$ -dimensional space using the PCA approach. The obtained blur vector is concatenated with the additive white Gaussian noise and stretched into  $M \times N \times (k + 1)$  degradation maps. By considering these maps, MDSNet can handle multiple degradations in real time and improve the practicability.

MDSNet is optimized by minimizing the loss function in Equation (4) using Adam optimizer [29]. The learned knowledge is used to produce the HR-IR image with factor  $s$ . In other words, the size of HR-IR is  $sM \times sN$ , which is similar to that of the HR-VIS image.

### 3.2. MRSVD Coefficients Generation and Update

The implementation of MRSVD resembles the discrete wavelet transform (DWT) except the finite impulse response filters are replaced with SVDs [30,31]. However, MRSVD performs similar to DWT but with reduced computations. Moreover, its basis is data dependent unlike a fixed basis in DWT, discrete cosine transform and discrete Fourier transform. Hence, it can capture maximum information from the source IR and VIS images, thereby proving the finest details in the fused image.

The MRSVD of the HR-VIS image is given by

$$\left\{ \Phi_L^{VIS}, \left\{ \Psi_i^{VIS_v}, \Psi_i^{VIS_h}, \Psi_i^{VIS_d} \right\}_{i=1}^L, \left\{ U_i^{VIS} \right\}_{i=1}^L \right\} = \text{MRSVD}(VIS_q) \tag{5}$$

where  $\Phi_L^{VIS}$  denotes the  $L$ -th level approximation coefficient of the HR-VIS image  $VIS_q$ ,  $\left\{ \Psi_i^{VIS_v}, \Psi_i^{VIS_h}, \Psi_i^{VIS_d} \right\}_{i=1}^L$  are the detailed coefficients of  $VIS_q$  at different levels  $i$  in the vertical, horizontal and diagonal directions.  $\left\{ U_i^{VIS} \right\}_{i=1}^L$  represents the eigenvectors of  $VIS_q$  corresponding to  $i = 1, 2, \dots, L$ .

Similarly, the MRSVD of HR-IR can be written as

$$\left\{ \Phi_L^{IR}, \left\{ \Psi_i^{IR_v}, \Psi_i^{IR_h}, \Psi_i^{IR_d} \right\}_{i=1}^L, \left\{ U_i^{IR} \right\}_{i=1}^L \right\} = \text{MRSVD}(IR_q) \tag{6}$$

where  $\Phi_L^{IR}$  denotes the  $L$ -th level approximation coefficient of the HR-IR image  $IR_q$ ,  $\left\{ \Psi_i^{IR_v}, \Psi_i^{IR_h}, \Psi_i^{IR_d} \right\}_{i=1}^L$  are the detailed coefficients of  $IR_q$  at different levels  $i$  in the vertical, horizontal and diagonal directions.  $\left\{ U_i^{IR} \right\}_{i=1}^L$  represents the eigenvectors of  $IR_q$  corresponding to  $i = 1, 2, \dots, L$ .

The HR-IR image  $IR_q$  is obtained from its LR-IR  $IR_p$  using MDSNet. In the process of approximation, a non-uniform illumination may be induced in the resulting image due to the external training database considered. To address this issue, we correct the HR-IR image MRSVD approximation coefficient in Equation (5) using the corresponding VIS image coefficient in Equation (6).

$$\zeta = \frac{\max(U_L^{VIS})}{\max(U_L^{IR})} \tag{7}$$

where  $\zeta$  is the correction factor for updating the  $\Phi_L^{IR}$  coefficient using  $U_L^{VIS}$  and  $U_L^{IR}$  at the  $L$ -th level. Now, the updated IR approximation coefficient can be obtained as  $\Phi_L^{IR} = \zeta \Phi_L^{IR}$  and thus representing the new set of IR coefficients as  $\left\{ \Phi_L^{IR}, \left\{ \Psi_i^{IR_v}, \Psi_i^{IR_h}, \Psi_i^{IR_d} \right\}_{i=1}^L, \left\{ U_i^{IR} \right\}_{i=1}^L \right\}$ .

### 3.3. Fusion Process

In this section, the fusion process of HR-IR and HR-VIS images using approximation coefficients, detailed coefficients and eigenvector matrices is presented. From Equation (5), the  $VIS_q$  is decomposed into  $\left\{ \Phi_L^{VIS}, \left\{ \Psi_i^{VIS_v}, \Psi_i^{VIS_h}, \Psi_i^{VIS_d} \right\}_{i=1}^L, \left\{ U_i^{VIS} \right\}_{i=1}^L \right\}$ , and from Equation (6),  $IR_q$  is decomposed and updated as  $\left\{ \Phi_L^{IR}, \left\{ \Psi_i^{IR_v}, \Psi_i^{IR_h}, \Psi_i^{IR_d} \right\}_{i=1}^L, \left\{ U_i^{IR} \right\}_{i=1}^L \right\}$ . The approximation coefficients and the eigen matrices are fused by following the average

strategy, whereas the detailed coefficients of the fused image are obtained by taking the maximum value of IR and VIS image detailed coefficients,

$$\Phi_L^F = \frac{1}{2}(\Phi_L^{IR} + \Phi_L^{VIS}) \tag{8}$$

$$\Omega_i^{Fh} = \max(\Omega_i^{IRh} + \Omega_i^{VISh}) \tag{9}$$

$$\Omega_i^{Fv} = \max(\Omega_i^{IRv} + \Omega_i^{VISv}) \tag{10}$$

$$\Omega_i^F = \max(\Omega_i^{IRh} + \Omega_i^{VISh}) \tag{11}$$

$$U_i^F = \frac{1}{2}(U_i^{IR} + U_i^{VIS}). \tag{12}$$

Here,  $\Phi_L^F$  denotes the  $L$ -th level approximation coefficient of the fused image. The set  $\{\Psi_i^{IRv}, \Psi_i^{IRh}, \Psi_i^{IRd}\}$  represents the detailed coefficients of the fused image in the vertical, horizontal and diagonal directions. In addition,  $U_i^F$  is the fused eigen matrix corresponding to the IR and VIS images. Finally, we apply the inverse MRSVD process and reconstruct the HR fused image  $F_q$  using these fused coefficients  $\{\Phi_L^F, \{\Psi_i^{Fv}, \Psi_i^{Fh}, \Psi_i^{Fd}\}_{i=1}^L, \{U_i^F\}_{i=1}^L\}$ .

#### 4. Results and Discussion

In this section, we present the experimental settings of the proposed fusion algorithm, methods and metrics used for comparison. Finally, we present a qualitative and quantitative analysis to validate our MDSNet-MRSVD approach over the existing state-of-the-art methods.

##### 4.1. Datasets and Settings

The resolution ratio between IR and VIS input images is set to  $s = 4$ . However, the dimensions  $M \times N$  for IR image and  $sM \times sN$  for VIS image vary with the input test images. The problem of super-resolving IR images can be solved using  $s = 4, s = 3$ , or  $s = 2$ . In general, small scaling factors result in a better enhancement of IR images, and thus, considering large upscaling factors is a challenging task in real time. For this reason, we employ MDSNet with  $s = 4$  to perform super-resolution on IR input images.

The training data should be as close as possible to the data on which predictions are to be made. In our use case, MDSNet training involves blurry and LR data from a security camera. However, we first select a clean large-scale HR image dataset with DIV2K [5] 800 images and BSDS [32] 500 images. Out of these 1300 images, 1200 images are randomly selected for training, and the remaining 100 images are used as validation data. Next, the LR images are synthesized using Equation (1) to generate LR-HR image patch pairs. For this, the width of the blur kernel is set to  $[0.2, 4]$  for isotropic Gaussian, the angle range of eigenvectors is set to  $[0, \pi]$  to determine the space of anisotropic Gaussian kernel, and the noise range is set to  $[0, 75]$ . The PCA approach is implemented by setting  $k = 15$  dimensional space. The loss function in Equation (4) is minimized using the Adams approach [29], with an initial learning rate and mini-batch size set to  $10^{-3}$  and 8, respectively. The learning rate reduces by a factor of 10 when the loss is becoming saturated.

The MRSVD of HR-VIS and HR-VIS images in Equations (5) and (6) are implemented with  $L = 2$  levels, and the corresponding  $\zeta$  is computed for the 2nd level as shown in Equation (7). The  $VIS_q$  coefficients and the updated  $IR_q$  coefficients are fused using Equations (8)–(12), which are followed by the inverse MRSVD with  $L = 2$  levels. For testing the MDSNet-MRSVD approach, we select 20 IR-VIS registered HR image pairs from the TNO Human Factors [22] dataset. The IR and VIS images corresponding to each image pair in the dataset have the same resolution. In our experiments, we downscale the IR images to LR by the factor  $s = 4$ .

#### 4.2. Methods and Evaluation Metrics

We compare the proposed method with nine state-of-the-art IF methods, including a deep networks based super-resolution model in the stationary wavelet domain [17], the adversarial game model [22], total variation minimization (TVM) [33], gradient transfer–TVM problem [20], multi-sensor IF model with one level SVD [34], wavelets approach [35], image guided filtering [36], curvelet transform [37], and Laplacian pyramid model [38]. All the methods are implemented using Matlab software on a system with Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz and 16 GB RAM.

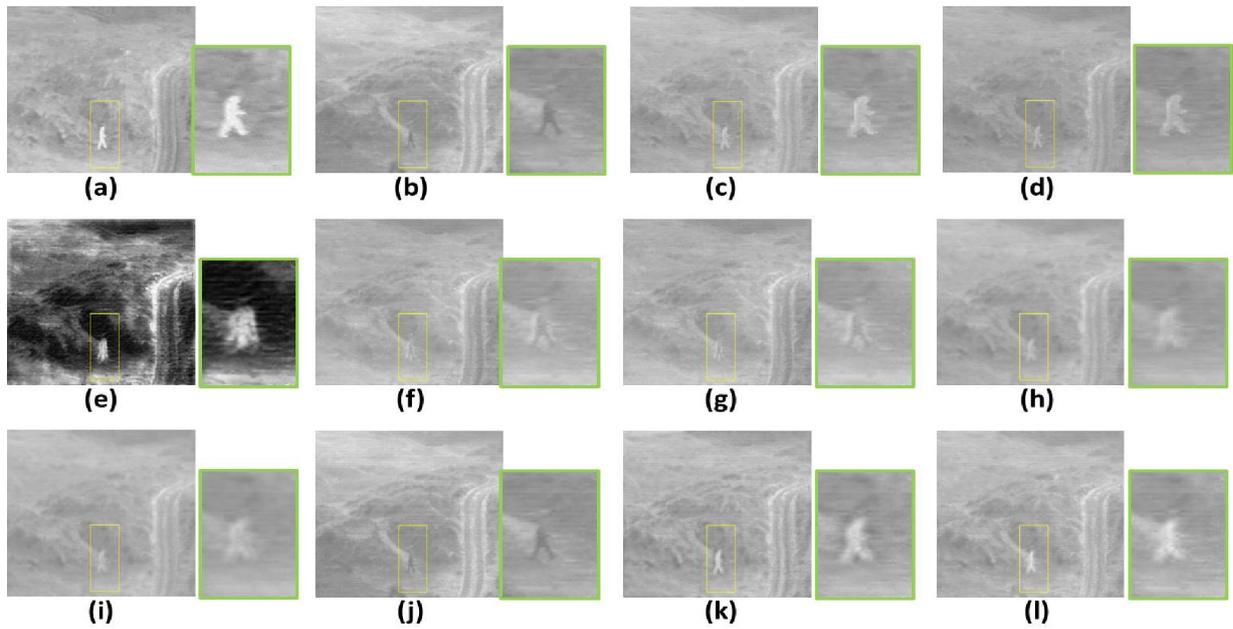
For the comprehensive evaluation of different IF methods, we adopt the peak signal-to-noise ratio (PSNR) [39] and structural similarity index measure (SSIM) [40]. In addition, the total IF performance  $Q^{AB/F}$ , IF loss  $L^{AB/F}$ , and IF artifacts  $N^{AB/F}$  are also considered for in-depth IF performance analysis [41]. It is noted that the sum of  $Q^{AB/F}$ ,  $L^{AB/F}$ , and  $N^{AB/F}$  should be unity [42]. The metrics SD, EN, and SF depend on the fused image only, whereas PSNR, SSIM,  $Q^{AB/F}$ ,  $L^{AB/F}$ , and  $N^{AB/F}$  depend on both fused and IR and VIS source images. Theoretically, PSNR, SSIM, SD, SF, EN, and  $Q^{AB/F}$  should be high, and  $L^{AB/F}$  and  $N^{AB/F}$  should be low for better IF performance.

#### 4.3. Qualitative Results Analysis

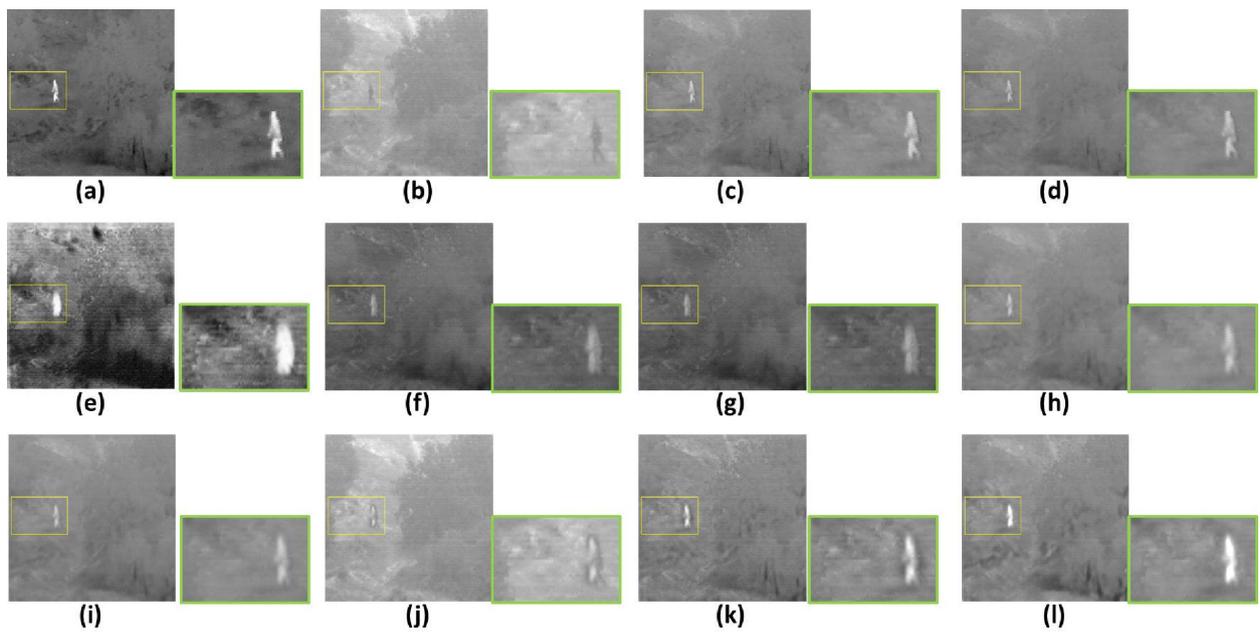
For qualitative comparison of different state-of-the-art IF methods, the visual IF results on seven typical IR-VIS image pairs are presented in Figures 3–9. In Figures 3 and 4, the fused images of the proposed method, existing methods [17,20,22,33–38] on test image pair1-*duine sequence* and test image pair4-*tree sequence* are given, respectively. Figure 3a,b represent the IR-VIS source images of the *duine sequence* and Figure 4a,b represent the IR-VIS source images of the *tree sequence*. Figures 3c and 4c show the results of the proposed MDSNet-MRSVD method, whereas Figures 3d–l and 4d–l are the fused images from the methods in [17,20,22,33–38]. Here, we have selected the captains present in the *duine sequence* and *tree sequence* as our region of interest. It can be noticed that the thermal information is well preserved in our result. In Figure 3, the captains contrast is higher in Figure 3c than in Figure 3f–j. The fused images in Figure 3d,e,k,l also enhance the contrast, but it comes with an additional blurring.

In Figures 5–7, the results of the proposed method and the existing methods [17,20,22,33–38] on test image pair7-*airplane in trees*, test image pair10-*two men in front of house*, and test image pair13-*lake* are evaluated. Figure 5a,b represent the IR-VIS source images of *airplane in trees*, Figure 6a,b represent the IR-VIS source images of *two men in front of house*, and Figure 7a,b represent the IR-VIS source images of *lake*. Figures 5c, 6c and 7c show the results of the proposed MDSNet-MRSVD method, whereas the fused images of the methods in [17,20,22,33–38] are shown in Figures 5d–l, 6d–l and 7d–l. For *airplane in trees*, our region of interest is airplane, and for *two men in front of house* and *lake*, we highlight the tree branches. Here, the fine textures are greatly recovered in our results. For example, the tree branches corresponding to the test image pair10 and test image pair13 are clearly visible in Figures 6c and 7c, whereas the results in Figures 6d–l and 7d–l suffer from non-uniform illumination along the textures.

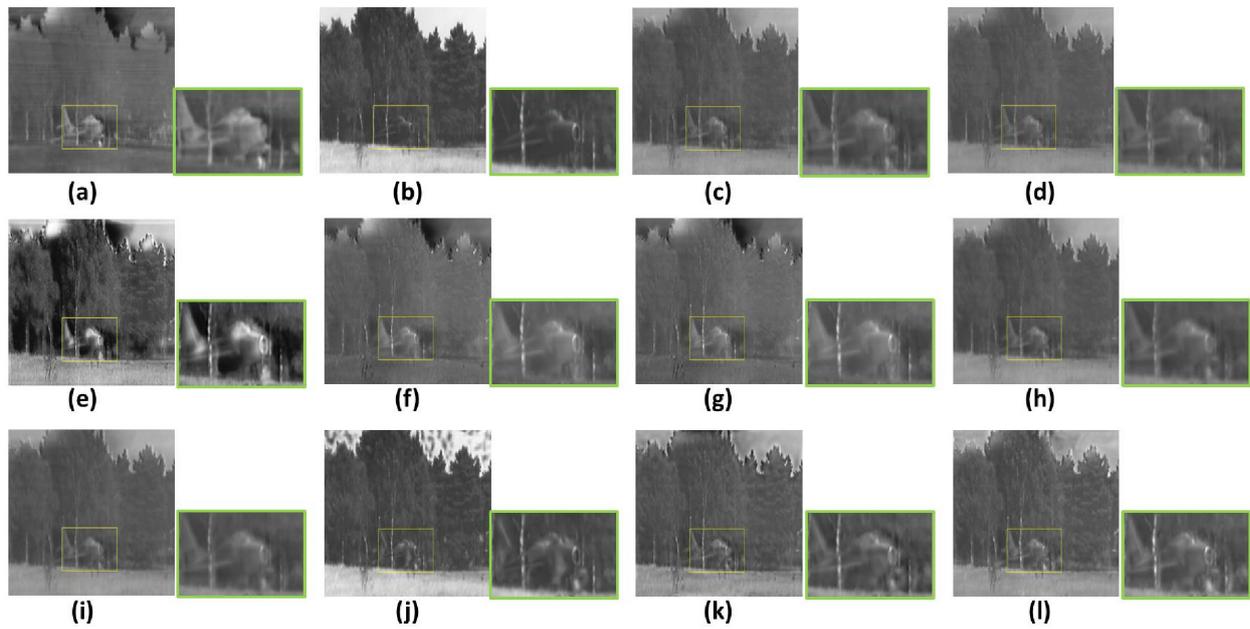
Finally, Figures 8 and 9 compare the proposed method with the existing methods [17,20,22,33–38] using the results on test image pair16-*soldier in trench* and test image pair19-*marne*. Figure 8a,b represent the IR-VIS source images of *soldier in trench* and Figure 9a,b are the IR-VIS source images of *marne*. Figures 8c and 9c show the results of the proposed MDSNet-MRSVD method. The fused images of the methods in [17,20,22,33–38] are shown in Figures 8d–l and 9d–l. Here, we have selected the stairs present in *soldier in trench* and *marne* as our region of interest. We can notice that the stairs are clearer for our method in Figures 8c and 9c compared to the existing methods. Moreover, our results provide good thermal information with minimized blurring effect.



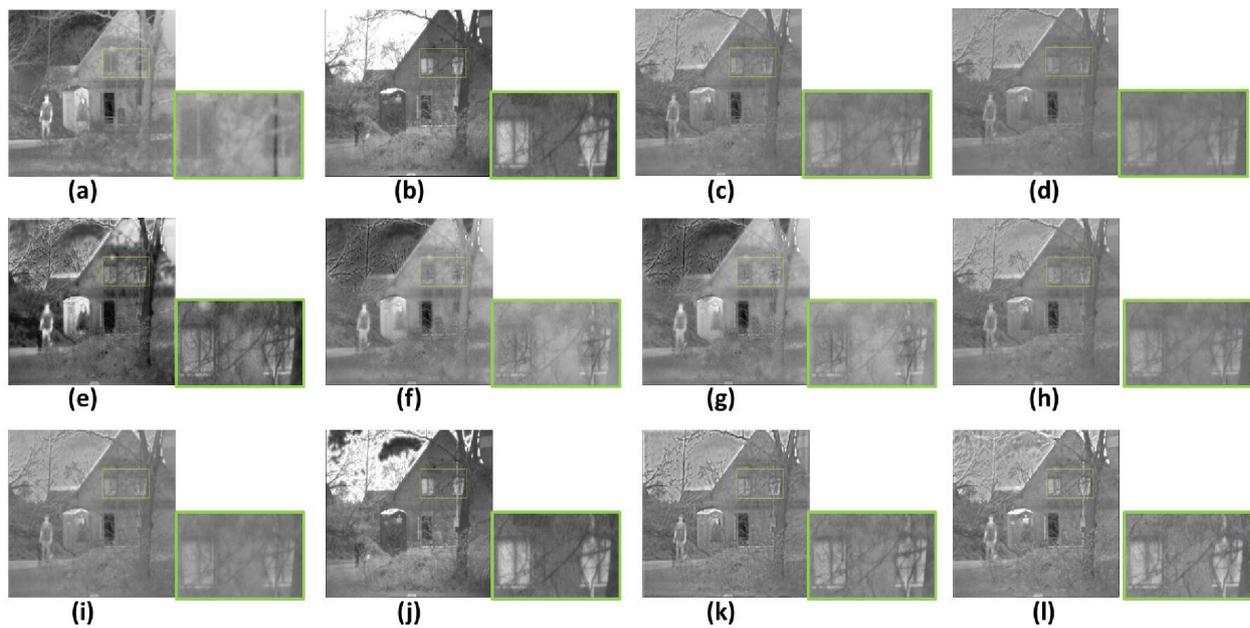
**Figure 3.** Visual IF results of different methods on *duine sequence*: (a) IR image, (b) VIS image, (c) Proposed method, (d) [17], (e) [22], (f) [33], (g) [20], (h) [34], (i) [35], (j) [36], (k) [37], and (l) [38].



**Figure 4.** Visual IF results of different methods on *tree sequence*: (a) IR image, (b) VIS image, (c) Proposed method, (d) [17], (e) [22], (f) [33], (g) [20], (h) [34], (i) [35], (j) [36], (k) [37], and (l) [38].



**Figure 5.** Visual IF results of different methods on *airplane in trees*: (a) IR image, (b) VIS image, (c) Proposed method, (d) [17], (e) [22], (f) [33], (g) [20], (h) [34], (i) [35], (j) [36], (k) [37], and (l) [38].



**Figure 6.** Visual IF results of different methods on *two men in front of house*: (a) IR image, (b) VIS image, (c) Proposed method, (d) [17], (e) [22], (f) [33], (g) [20], (h) [34], (i) [35], (j) [36], (k) [37], and (l) [38].

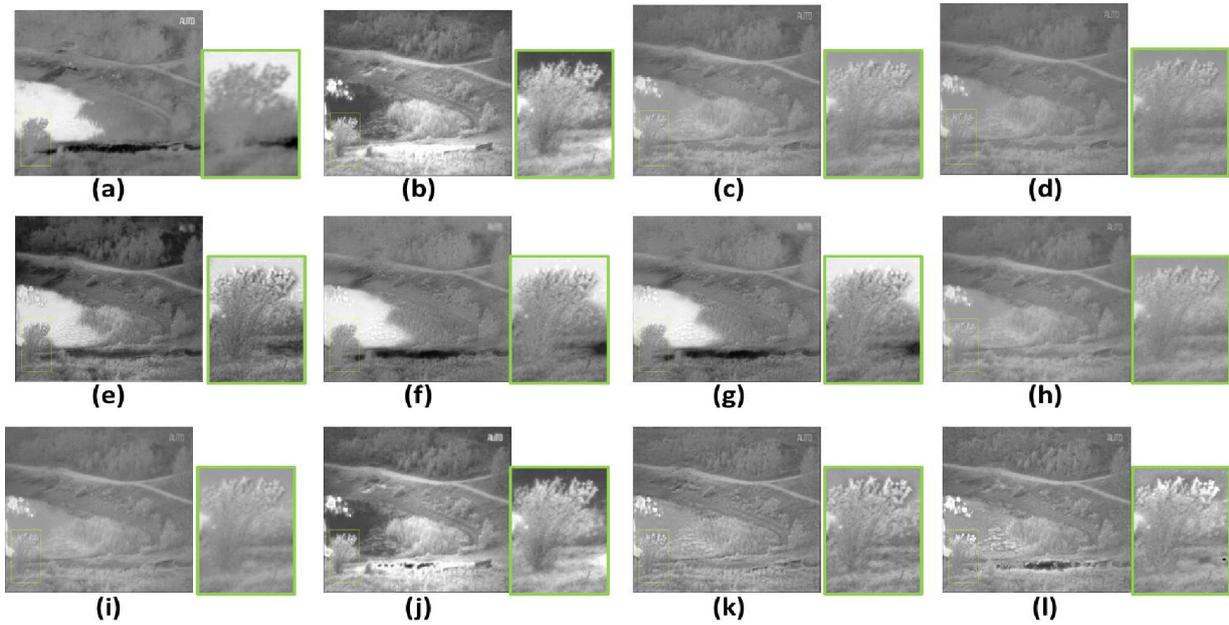


Figure 7. Visual IF results of different methods on lake: (a) IR image, (b) VIS image, (c) Proposed method, (d) [17], (e) [22], (f) [33], (g) [20], (h) [34], (i) [35], (j) [36], (k) [37], and (l) [38].

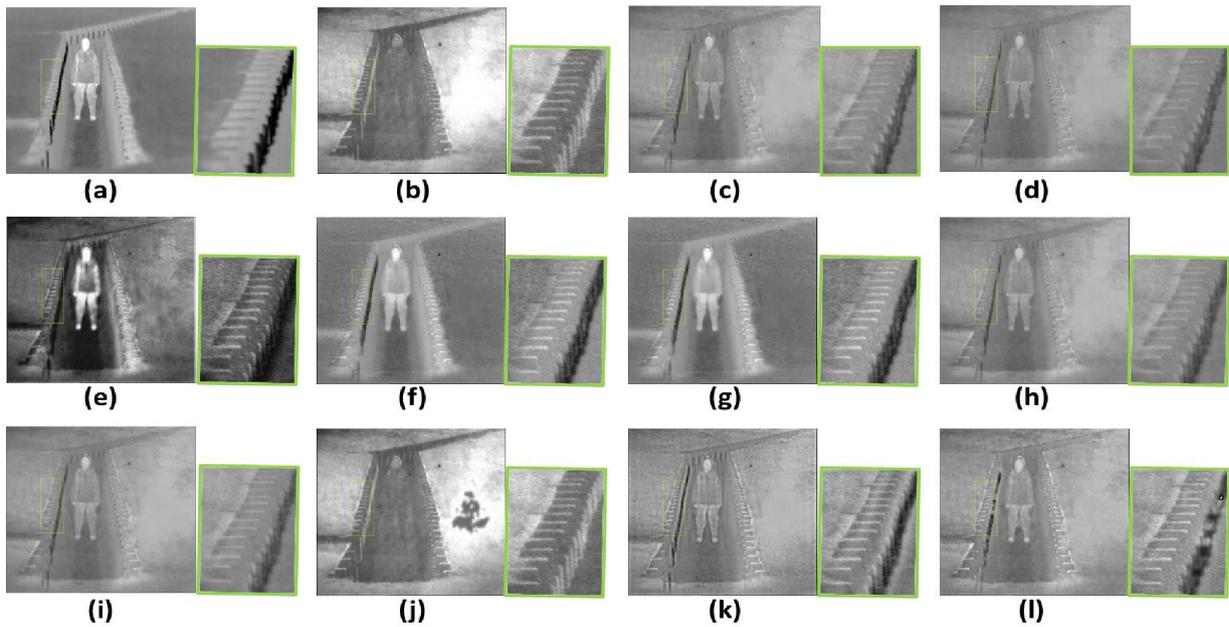
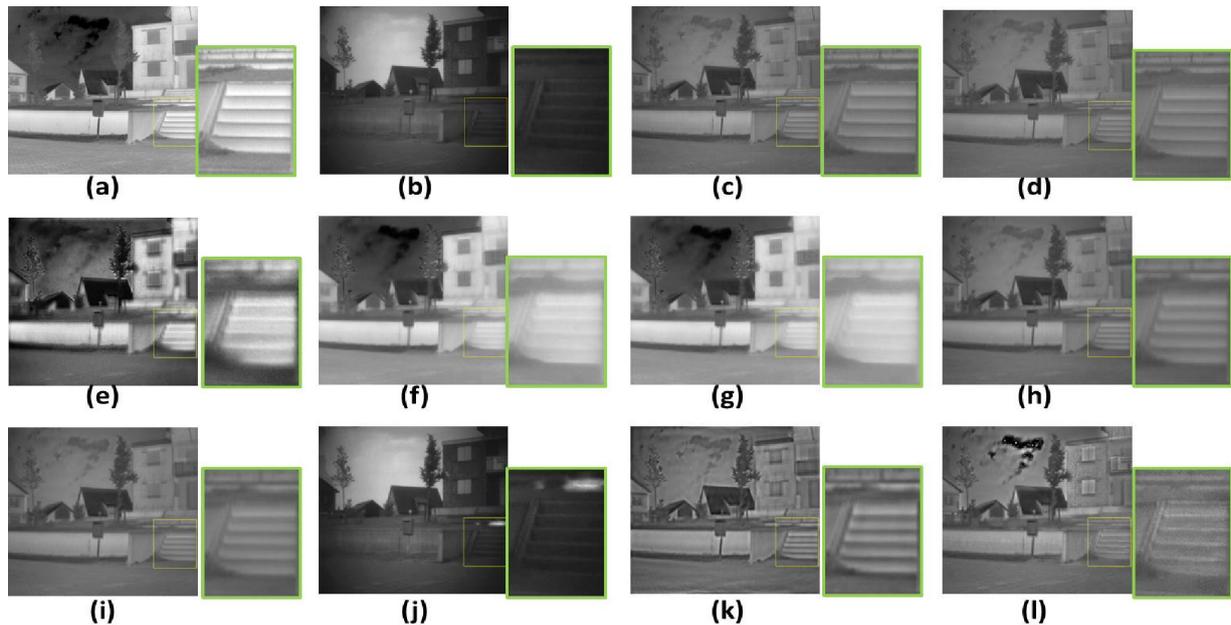


Figure 8. Visual IF results of different methods on soldier in trench: (a) IR image, (b) VIS image, (c) Proposed method, (d) [17], (e) [22], (f) [33], (g) [20], (h) [34], (i) [35], (j) [36], (k) [37], and (l) [38].



**Figure 9.** Visual IF results of different methods on *marne*: (a) IR image, (b) VIS image, (c) Proposed method, (d) [17], (e) [22], (f) [33], (g) [20], (h) [34], (i) [35], (j) [36], (k) [37], and (l) [38].

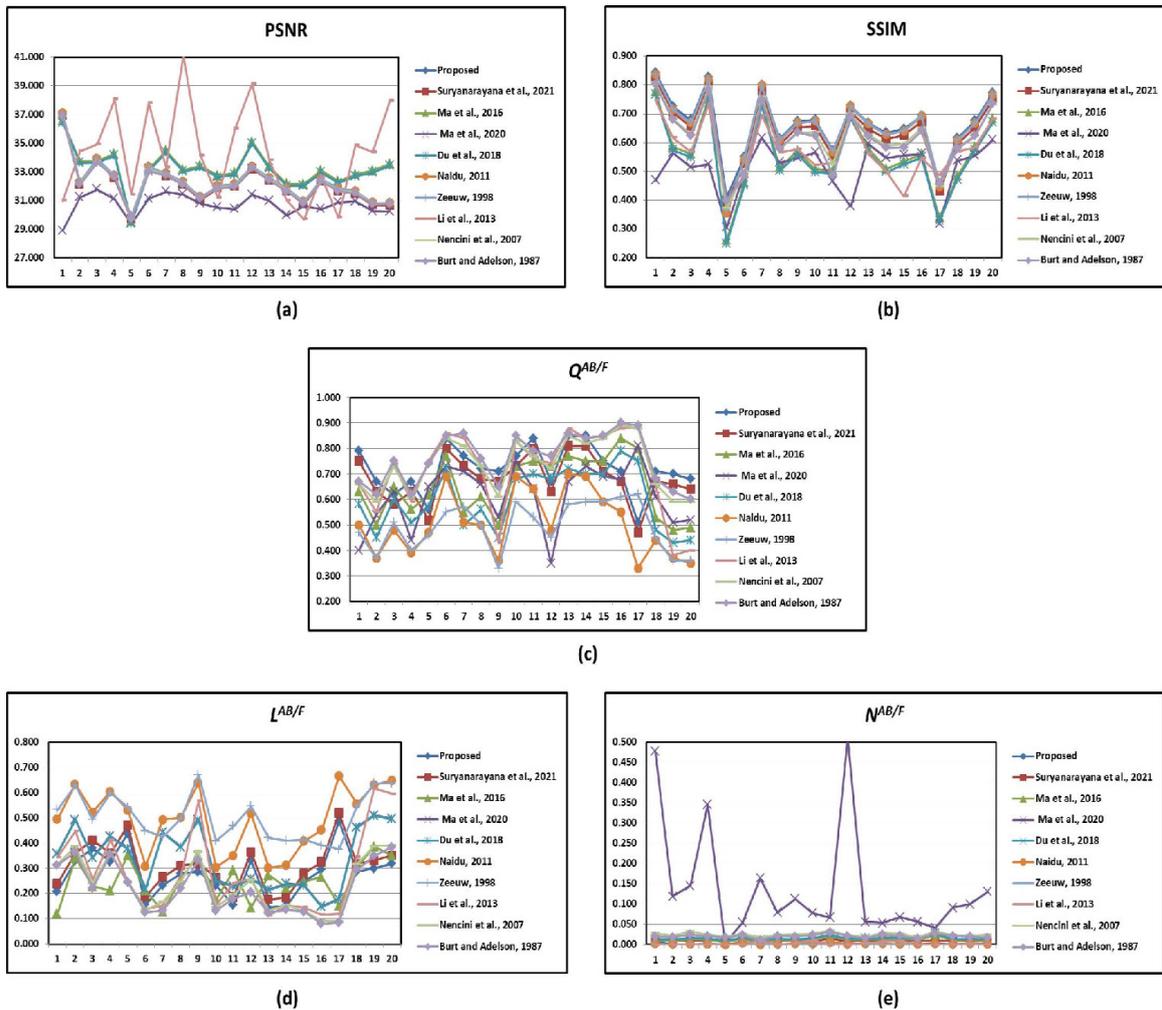
4.4. Quantitative Results Analysis

We first performed qualitative analysis in Figures 3–9. It is based on human visual perception and usually focuses on the contrast, edge details, and sharpness of an image. However, the visual results alone are insufficient to prove the significance of the proposed method. Therefore, we also present quantitative analysis which is consistent with visual results and is based on PSNR, SSIM,  $Q^{AB/F}$ ,  $L^{AB/F}$  and  $N^{AB/F}$ . In Table 1, we show different metrics of the proposed method averaged over 20 IR-VIS test image pairs. The individual metric results of different fusion methods are summarized in Figure 10.

As seen, the proposed method produces outstanding results in terms of PSNR, SSIM,  $Q^{AB/F}$ , and  $N^{AB/F}$ . The method in [36] and proposed method are capable of generating PSNR values above 34. In addition, the proposed method has the highest SSIM index of 0.663 compared to other methods. The methods in [37,38] and the proposed method have  $Q^{AB/F}$  values of 0.759, 0.738, and 0.720, respectively. In addition to the proposed method, the methods in [35,36] are advantageous with smaller  $N^{AB/F}$  indices. In terms of  $L^{AB/F}$ , the proposed method is comparable to the methods in [22,37,38].

**Table 1.** Average PSNR, SSIM,  $Q^{AB/F}$ ,  $L^{AB/F}$  and  $N^{AB/F}$  indices of different methods.

Method	PSNR	SSIM	$Q^{AB/F}$	$L^{AB/F}$	$N^{AB/F}$	$Q^{AB/F} + L^{AB/F} + N^{AB/F}$
Proposed	34.03	0.663	0.720	0.278	0.002	1
[17]	33.87	0.642	0.680	0.310	0.010	1
[22]	30.71	0.513	0.613	0.250	0.138	1
[33]	33.16	0.549	0.646	0.338	0.016	1
[20]	33.28	0.560	0.651	0.336	0.013	1
[34]	32.31	0.653	0.505	0.493	0.002	1
[35]	32.32	0.656	0.495	0.504	0.001	1
[36]	34.34	0.574	0.712	0.284	0.003	1
[37]	32.22	0.623	0.738	0.237	0.025	1
[38]	32.17	0.617	0.759	0.220	0.021	1



**Figure 10.** Metric results of different methods [17,20,22,33–38] on 20 IR-VIS test image pairs: (a) PSNR (b) SSIM (c)  $Q^{AB/F}$  (d)  $L^{AB/F}$  (e)  $N^{AB/F}$ .

The average runtime values corresponding to the proposed method and the existing methods [17,20,22,33–38] are given in Table 2. It can be noticed that the methods in [35,36,38] are the fastest methods that can execute in 1 s. The proposed method and the methods in [20,34,37] consume less than 10 s. In contrast, the methods in [17,22,33] are the slowest methods with execution times of 14.52, 32.56, and 12.79 s, respectively. Even though the methods in [35,36,38] are the fastest methods, the method in [35] yields very poor fusion performance  $Q^{AB/F}$ , resulting in high fusion loss  $L^{AB/F}$ . The methods in [36,38] suffer from poor SSIM indices with more noise compared to the proposed method. In particular, when better visual quality is desired with high PSNR, SSIM indices and low fusion loss, a slight increase in computational time is acceptable in view of improved computational facilities.

**Table 2.** Average runtime (in seconds) comparison of different methods on 20 IR-VIS test image pairs.

Proposed	[17]	[22]	[33]	[20]
9.26	14.52	32.56	12.79	3.29
[34]	[35]	[36]	[37]	[38]
7.41	0.24	0.70	9.16	0.07

Thus, comparing all metrics together, the proposed method can be seen as the best method in terms of SSIM, second best performed in terms of PSNR and  $N^{AB/F}$ , and third best performed with  $Q^{AB/F}$  metrics. Additionally, it gives comparable results in view of the  $L^{AB/F}$  metrics as well.

## 5. Conclusions

In this correspondence, we propose a IR-VIS image fusion algorithm based on MDSNet and MRSVD updation. The proposed method does not demand source images with the same spatial resolution but still produces a high spatial resolution. The multi-degradation handling capabilities of MDSNet and IR image coefficient correction using MRSVD enrich the fused image with clear visible textures and thermal radiation information. Extensive visual comparisons on various IR-VIS test images demonstrate the qualitative superiority of our method. The proposed method yields good results on PSNR, SSIM, and  $Q^{AB/F}$  with less fusion loss and artifacts as well.

**Author Contributions:** Conceptualization, S.R.P.; Methodology, G.S.; Supervision, V.V.; Writing—original draft, G.S.; Writing—review & editing, G.N. and G.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Hervella, Á.S.; Rouco, J.; Novo, J.; Ortega, M. Retinal microaneurysms detection using adversarial pre-training with unlabeled multimodal images. *Inf. Fusion* **2022**, *79*, 146–161. [\[CrossRef\]](#)
- Jin, X.; Jiang, Q.; Yao, S.; Zhou, D.; Nie, R.; Hai, J.; He, K. A survey of infrared and visual image fusion methods. *Infrared Phys. Technol.* **2017**, *85*, 478–501. [\[CrossRef\]](#)
- Zhang, X.; Ma, Y.; Fan, F.; Zhang, Y.; Huang, J. Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition. *JOSA A* **2017**, *34*, 1400–1410. [\[CrossRef\]](#) [\[PubMed\]](#)
- Shibata, T.; Tanaka, M.; Okutomi, M. Visible and near-infrared image fusion based on visually salient area selection. In *Digital Photography XI, Proceedings of the International Society for Optics and Photonics, San Francisco, CA, USA, 9–10 February 2015*; SPIE: Bellingham, WA, USA, 2015; Volume 9404, p. 94040G.
- Suryanarayana, G.; Tu, E.; Yang, J. Infrared super-resolution imaging using multi-scale saliency and deep wavelet residuals. *Infrared Phys. Technol.* **2019**, *97*, 177–186. [\[CrossRef\]](#)
- Li, S.; Yin, H.; Fang, L. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3450–3459. [\[CrossRef\]](#)
- Zhang, Q.; Liu, Y.; Blum, R.S.; Han, J.; Tao, D. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Inf. Fusion* **2018**, *40*, 57–75. [\[CrossRef\]](#)
- Suryanarayana, G.; Dhuli, R. Image resolution enhancement using wavelet domain transformation and sparse signal representation. *Procedia Comput. Sci.* **2016**, *92*, 311–316. [\[CrossRef\]](#)
- Pajares, G.; De La Cruz, J.M. A wavelet-based image fusion tutorial. *Pattern Recognit.* **2004**, *37*, 1855–1872. [\[CrossRef\]](#)
- Dogra, A.; Goyal, B.; Agrawal, S. From multi-scale decomposition to non-multi-scale decomposition methods: A comprehensive survey of image fusion techniques and its applications. *IEEE Access* **2017**, *5*, 16040–16067. [\[CrossRef\]](#)
- Liu, Y.; Chen, X.; Cheng, J.; Peng, H.; Wang, Z. Infrared and visible image fusion with convolutional neural networks. *Int. J. Wavelets Multiresolut. Inf. Process.* **2018**, *16*, 1850018. [\[CrossRef\]](#)
- Kong, W.; Zhang, L.; Lei, Y. Novel fusion method for visible light and infrared images based on NSST-SF-PCNN. *Infrared Phys. Technol.* **2014**, *65*, 103–112. [\[CrossRef\]](#)
- Bavirisetti, D.P.; Xiao, G.; Liu, G. Multi-sensor image fusion based on fourth order partial differential equations. In *Proceedings of the 2017 20th International Conference on Information Fusion (Fusion), Xi'an, China, 10–13 July 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 1–9.
- Patil, U.; Mudengudi, U. Image fusion using hierarchical PCA. In *Proceedings of the 2011 International Conference on Image Information Processing, Shimla, India, 3–5 November 2011*; IEEE: Piscataway, NJ, USA, 2011; pp. 1–6.

15. Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [[CrossRef](#)]
16. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **2018**, *42*, 158–173. [[CrossRef](#)]
17. Suryanarayana, G.; Chandran, K.; Khalaf, O.I.; Alotaibi, Y.; Alsufyani, A.; Alghamdi, S.A. Accurate magnetic resonance image super-resolution using deep networks and Gaussian filtering in the stationary wavelet domain. *IEEE Access* **2021**, *9*, 71406–71417. [[CrossRef](#)]
18. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [[CrossRef](#)]
19. Ma, J.; Zhou, Z.; Wang, B.; Zong, H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys. Technol.* **2017**, *82*, 8–17. [[CrossRef](#)]
20. Ma, Y.; Chen, J.; Chen, C.; Fan, F.; Ma, J. Infrared and visible image fusion using total variation model. *Neurocomputing* **2016**, *202*, 12–19. [[CrossRef](#)]
21. Ma, J.; Chen, C.; Li, C.; Huang, J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **2016**, *31*, 100–109. [[CrossRef](#)]
22. Ma, J.; Xu, H.; Jiang, J.; Mei, X.; Zhang, X.P. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. *IEEE Trans. Image Process.* **2020**, *29*, 4980–4995. [[CrossRef](#)]
23. Zhou, Z.; Wang, B.; Li, S.; Dong, M. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters. *Inf. Fusion* **2016**, *30*, 15–26. [[CrossRef](#)]
24. Chen, R.; Zhang, H.; Liu, J. Multi-attention augmented network for single image super-resolution. *Pattern Recognit.* **2022**, *122*, 108349. [[CrossRef](#)]
25. Zhang, K.; Zuo, W.; Zhang, L. Learning a single convolutional super-resolution network for multiple degradations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3262–3271.
26. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
27. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. [[CrossRef](#)]
28. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
29. Na, G.S. Efficient learning rate adaptation based on hierarchical optimization approach. *Neural Netw.* **2022**, *150*, 326–335. [[CrossRef](#)]
30. Zhang, X.; Zhang, C.; Wei, Z. Carbon price forecasting based on multi-resolution singular value decomposition and extreme learning machine optimized by the moth–flame optimization algorithm considering energy and economic factors. *Energies* **2019**, *12*, 4283. [[CrossRef](#)]
31. Suryanarayana, G.; Dhuli, R. Super-resolution image reconstruction using dual-mode complex diffusion-based shock filter and singular value decomposition. *Circuits Syst. Signal Process.* **2017**, *36*, 3409–3425. [[CrossRef](#)]
32. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the Eighth IEEE International Conference on Computer Vision, ICCV 2001, Vancouver, BC, Canada, 7–14 July 2001; IEEE: Piscataway, NJ, USA, 2001; Volume 2, pp. 416–423.
33. Du, Q.; Xu, H.; Ma, Y.; Huang, J.; Fan, F. Fusing infrared and visible images of different resolutions via total variation model. *Sensors* **2018**, *18*, 3827. [[CrossRef](#)]
34. Naidu, V. Image fusion technique using multi-resolution singular value decomposition. *Def. Sci. J.* **2011**, *61*, 479. [[CrossRef](#)]
35. Zeeuw, P. *Wavelets and Image Fusion*; CWI: Amsterdam, The Netherlands, 1998; Volume 444.
36. Li, S.; Kang, X.; Hu, J. Image fusion with guided filtering. *IEEE Trans. Image Process.* **2013**, *22*, 2864–2875.
37. Nencini, F.; Garzelli, A.; Baronti, S.; Alparone, L. Remote sensing image fusion using the curvelet transform. *Inf. Fusion* **2007**, *8*, 143–156. [[CrossRef](#)]
38. Burt, P.J.; Adelson, E.H. The Laplacian pyramid as a compact image code. In *Readings in Computer Vision*; Elsevier: Amsterdam, The Netherlands, 1987; pp. 671–679.
39. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 2366–2369.
40. Zhao, C.; Shao, G.; Ma, L.; Zhang, X. Image fusion algorithm based on redundant-lifting NSWMDA and adaptive PCNN. *Optik* **2014**, *125*, 6247–6255. [[CrossRef](#)]
41. Shreyamsha Kumar, B. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform. *Signal Image Video Process.* **2013**, *7*, 1125–1143. [[CrossRef](#)]
42. Petrovic, V.; Xydeas, C. Objective image fusion performance characterisation. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Beijing, China, 17–21 October 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 2, pp. 1866–1871.