RB Journal of Lib & Information Science ISSN: 0972-2750
(UGC Care Group I Listed Journal) Vol-13 Issue-03 No.03: 2023

Infrared and Visible Image Fusion (IVF) using Latent Low-Rank Representation and Deep Feature Extraction Network

[1]T. Sandhya Kumari, [2]Gundala Sujatha, [3]Boddeda Sravya, [4]Gorle Vanaja, [5]Badireddi Sri Nanditha

[1]Associate professor, Department of Electronics and Communication Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

[2,3,4,5] Department of Electronics and communication Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

Abstract—- **The combination of infrared and visible images from different sensors can provide a more detailed and informative image. Visible images capture environmental detailsand texture, while infrared sensors can detect thermal radiation and create grayscale images that have high contrast. These images are useful for distinguishing between target and background in challenging conditions, such as at night or in inclement weather. When these two types of images are fused, they create high- contrast images with rich texture and target details. In thispaper, An effective image fusion technique has been developed, which utilizes Latent Low Rank Representation (LatLRR) method that decomposes the source images into latent low rank and salient parts to capture common and unique information respectively. The proposed network design incorporates the dense network and VGG-19 architectures for deep feature extraction of latent low- rank and salient parts, that minimize distortion while maintaining crucial texture and details in the output. Weighted average fusion strategies are used to combine these latent low-rank and salient parts, and the resulting fused features are used for feature reconstruction to generate a fused low-rank and salient part. These parts are integrated to yield a fused image output. The proposed approachout performs existing state-of-the-art methods on both visual characteristics and objective evaluation metrics.**

Index terms—**infrared images, visible images, Image Fusion, Latent Low Rank Representation, VGG-19 network, Dense network.**

I. INTRODUCTION

Image fusion is a technique for enhancing the data gained from images captured by several sensors. The aim is to yield a composite image which is richer in information and can be usedfor various applications. The most common type of image fusion is the fusion of infrared and visible images, which is regularly employed in various domains such as military and defence, and night vision. Visible images provide a highly detailed and realistic depiction of the object or scene being captured. On the other hand, infrared images have strong contrast and are capable to differentiate between targets and their surrounding backgrounds by detecting variations in radiation., even in Low-light or adverse climatic conditions. The combination of infrared and visible images yields a rich comprehensive image with enhanced images andextraction of meaningful information for subsequent analysis and applications. In recent years, Infrared and visible image fusion techniques have become increasingly popular in various fields such as target detection [1], target recognition, image enhancement [2], remote sensing [3], medical imaging [4], and industrial applications [5]. Early approaches for image fusion were based on mathematical transformations that analyzed activity level and formulated fusion rules either in the spatial domain or transformation domain. These techniques are known as traditional fusion methods, and they include various methods such as multi- scale transform-based [6], sparse representation-based, subspace-based [7], saliency- based and total variation- based approaches [8]. However, these traditional methods have several limitations such as requiring significant domain knowledge, being time-consuming to implement and not being able to capture complex relationships between input sources. Recent studies have demonstrated that combining infrared and visible images using feature extraction based fusion rules [9-11] can improve the overall quality of the fused image in relation to contrast, information content, and edge preservation. These fusion rules rely on statistical, saliency, and structural approaches to extract relevant features from both images. However, fusion techniques that utilize deep learning algorithms [12] has come into existence, that can offer more flexibility, automation and adaptability in the fusion process in recent years.

Deep learning techniques utilize various network branches to extract distinct features, enabling them to obtain more precise and specific features. With well-designed loss functions, these methods can learn an optimized feature fusion strategy that allows for adaptable feature fusion. This approach results in the development of more targeted features, leading to more effective and accurate results. However, there are some challenges associated with supervised learning-based methods [13], including their dependence on labelled data, limited generalization to new data, and lack of interpretability. In contrast, unsupervised learning-based methods [14] can overcome these challenges by not relying on labelled data and has the potential to be utilized in a broader spectrum of scenarios, making them more flexible and versatile in handling diverse input sources. Hence, this paper introduces an unsupervised deep fusion framework for fusing visible and infrared images. The proposed method utilizes Latent Low Rank Representation to decompose the source images into distinct low rank and salient parts, which are then combined using diverse fusion techniques The network architecture utilizes VGG-19 network and dense network structure to preserve important features and texture while reducing distortion. It outperforms traditional methods, making it a powerful approach to image fusion.

II. LITERATURE SURVEY

Image processing has witnessed significant research and practical applications of conventional techniques for fusing infrared and visible images. However, recent developments in deep learning have revolutionized field of image processing, enabling significant advancements in various applications. By employing deep neural networks, the image fusion process can be converted into a training process, which has led to remarkable improvements in the standard of fused images. The literature contains numerous techniques aimed at optimizing the standard of infrared and visible image fusion. This literature review will examine various techniques proposed in the survey to strengthen the standard of fused images, discussing their benefits and drawbacks.

Encoder and Decoder network is a popular method for image fusion that involves taking two input images and passingthem through two encoding functions to obtain their respective encoded functions. The encoded features are then fused using a weighted average strategy and passed through a decoder to reconstruct the fused image. The results demonstrate the efficiency of the fusion process in producing images that meet the desired criteria in many cases. However, the main disadvantage of image fusion using encoder and decoder network is its computational complexity, especially during the training phase.

The approach of fusing infrared and visible images using Latent Low Rank Representation [15] involves merging several images taken at varying focal lengths to generate a final image output with a greater level of sharpness and clarity. The technique involves decomposing the input images into latent low rank parts and salient parts, where the low-rank representation captures the common information from each image, while the salient parts contain the unique information that distinguishes each image from the others. Once the low rank and salient parts are obtained, theyare fused using different strategies. Using a latent low rank representation for image fusion has several advantages, including improved fusion quality with less noise and artifacts, robustness to misalignments datasets, but it may require appropriate decomposition and fusion strategies for quality results.

An effective approach for integrating the infrared and visible images has been introduced using deep learning methodology [16]. This technique involves performing a decomposition process on each of the images to separate out the low-frequency components, which are the base parts, and the high-frequency components, which are the detail content. The base parts and detail content are then fused using different fusion strategies respectively and the final fused image is obtained by combining the fused base part and the fused detail content. The performance of this type of fusion is subject to various factors, including: the quality of the source images, the complexityof the detail content, and the effectiveness of the fusion algorithm. However, in general, image fusion techniques that incorporate deep learning methods have shown to perform well in preserving important details while reducing noise and artifacts resources.

For information fusion, Yu Liu and Xun chen [17] proposed “Convolutional neural networks (CNNs) based visible and infrared image fusion”. This paper introduces a productive and effective technique that combines the information from both infrared and visible images to yield a single image that preserves the best features from both images. A pair of infrared and visible images serve as the input to the neural network, that are first pre-processed to remove noise and enhance their contrast. The CNN network receives the pre-processed images as input, which consists of five layers. The fusion process aims to retain the salient features from both images while suppressing the noise and redundant information. The fused image obtained from this process has improved contrast, enhanced edges, and better texture details than the individual input images. The fusion performance of CNN-based methods is generally better than traditional fusion methods as they can learn the optimal weights for feature combination and handle complex nonlinear relationships between the input images.

The VIF-Net framework [18] consists of two main modules: a feature extraction module and a fusion module. The feature extraction module uses a dense block to capture the distinctive characteristics of the source images. The fusion module includes a fusion layer and a feature reconstruction layer. The feature extraction module extract significant properties from the input images by passing them through a dense block, which is composed of several convolutional layers. The fusion module consists of a fusion layer, which integrates the extracted features using a fusion strategy, and feature reconstruction layer, generates the final fused image by reconstructing the fused features. The advantage of the framework includes its ability to fuse input images without requiring ground-truth information. However, this method is less efficient in terms of computational resources compared to other methods.

In summary, we examine several image fusion techniques, each with its advantages and disadvantages, to strengthen the standard of fused images from multiple source images.

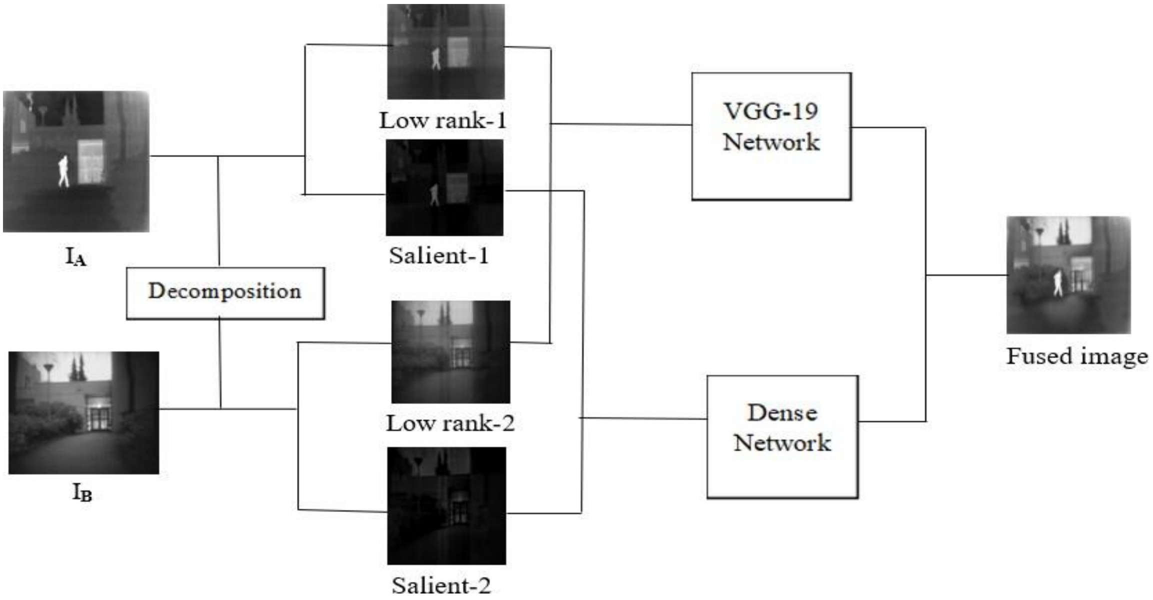

Fig. 1. Block diagram of proposed network architecture for IVF.

III. PROPOSED METHOD

The network architecture for IVF framework is represented in Fig.1, and comprises of four primary components: Image decomposition, feature extraction, fusion and feature reconstruction.

Decomposition of Source Images

The infrared image and visible image are represented as I_A and I_B respectively as shown in Fig.2. Latent low rank representation is a mathematical technique used in image processing to extract and separate the common and unique content present in the source images. This technique is built upon the notion that image data can be expressed as a composite of a low-rank matrix and a sparse matrix . The low rank matrix representsthe common information shared by all the images in the dataset,while the sparse matrix represents the unique information that distinguishes each image from the others.

The Latent Low Rank Representation is solved by using the optimization method as follows:

$$\min_{Z,L,E} ||Z||_* + ||L||_* + \lambda||E||_1 \qquad (1)$$

$$s.t., X = XZ + LX + E$$

The optimization method decompose a given data matrix X into three parts: a low rank part represented by matrix Z, a salient part represented by matrix L, and a sparse noisy part represented by matrix E. The decomposition is obtained by minimizing the sum of the nuclear norm of Z, the nuclear norm of L, and the L1-norm of E, subject to the constraint that X can be represented as the sum of XZ, LX, and E. The balance coefficient λ determines the significance assigned to each of the three components in the decomposition. The inexact Augmented Lagrangian Multiplier (ALM) is used to solve Eq.(1), which results in obtaining the low-rank part XZ and the salient part LX as per Eq.(1).

In image fusion, the low-rank and sparse components of two or more images obtained from various modalities, such as infrared and visible images, are fused by using several techniques to generate a composite image with improved visualquality. The low-rank component captures the structural content of the images, while the sparse component contains the unique features of each image that are not present in the other images. The fusion of the low-rank and sparse features of images can result in a composite image that carries greater information content than either of the original images.

Feature Extraction of Low Rank Parts using VGG-19 Network

The low-rank part of an image represents the background or the underlying structure of the image, which is typically smoother and more uniform than the salient part. This component contains information about the overall structure and composition of the scene.

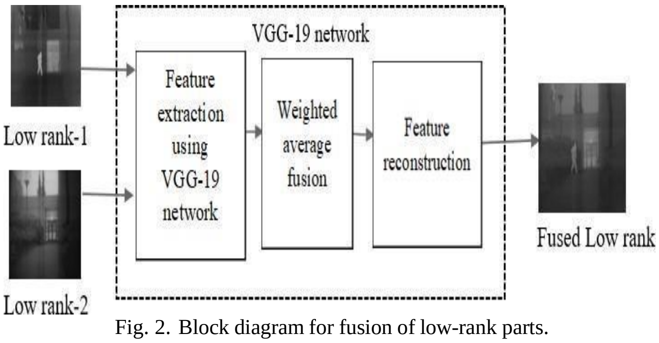

Fig. 2. Block diagram for fusion of low-rank parts.

VGG-19 is a complex neural network architecture that learns to capture important characteristics from each input image and combine them to generate a fused image of superior quality. Hence for the fusion of low rank parts, an effective fusion methodology is proposed which incorporates VGG-19 network architecture to extract deep features. This process is depicted in Fig.2. once the features are extracted, they can be combined by using a weighted average fusion approach to obtain a more compact and informative feature representation. Finally, the fused features can be used for feature reconstruction to generate a fused low-rank part. The calculation of the fused low-rank part ($Flrr$) is accomplished as follows:

$$F_{lr}(i,j) = w_1 I_{1\_lrr}(i,j) + w_2 I_{2\_lrr}(i,j) \qquad (2)$$

Where I_{1_lrr} and I_{2_lrr} represents the low-rank parts. The coordinates (i,) indicate the position of the coefficients for $I1\_lrr$, $I2\_lrr$, and $Flrr$, respectively. The weight values assigned to the coefficients of $I1\_lrr$ and $I2\_lrr$ are denoted by $w1$ and $w2$, respectively.

Feature Extraction of Salient Parts using Dense Network

The salient part of an image represents the foreground and distinctive features of the image, such as objects, textures, and patterns. This salient component can be considered as the "highlights" of the image, and it includes information regarding the specific details and characteristics that make the image unique.

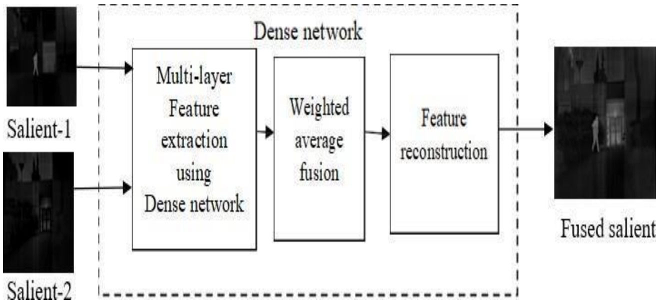

Fig 3: Block diagram for fusion of salient parts

The dense block allows the network to learn more complex and diverse features from the input images. Specifically, the dense block enables the network to reuse and combine features from previous layers, which helps to capture and preserve more information about the input images. Hence for the fusion of salient parts, an effectivefusion approach is proposed which uses Dense network that enables the network to capture both basic and advanced characteristics, which are then used to produce a fused image. The procedure depicted in Fig:3 involves feeding salient parts, which are obtained from the source images into adense block, that has five convolutional layers. This block is used to capture complex features from the input. The resulting features are then combined using weighted average fusion technique. The fused features can be utilized for the feature reconstruction to generate a fused salient part. The fused salient part (Fs) is calculated as follows:

$$F(i,j) = w_1 I_{1\_s}(i,j) + w_2 I_{2\_s}(i,j) \qquad (3)$$

Where I_{1_s} and I_{2_s} represents the salient parts. The coordinates (i,) indicate the position of the coefficients for I_{1_s}, I_{2_s} and Fs, respectively. The weight values assigned to thesecoefficients of I_{1_s} and I_{2_s} are denoted by $w1$ and $w2$, respectively.

Reconstruction of the fused image

By integrating the fused low rank component and fused salient component, a single fused image is produced that exhibits improved visual quality. The fused image (F) is calculated as follows:

$$(i,j) = F_{lrr}(i,j) + F_s(i,j) \qquad (4)$$

where F_{lrr} is fused latent part and F_s is fused salient part.

IV. RESULTS AND DISCUSSION

This section presents the outcome of a qualitative and quantitative analysis carried out on the fused images generated using five different fusion techniques. The dataset used for this analysis comprises 15 pairs of infrared and visible images. In qualitative analysis, the images are fused using MATLAB software and the resulting output images are analyzed based on their visual quality, fusion accuracy, and preservation of important features.

A quantitative analysis is conducted to further assess the fused images obtained from the five methods. This analysis is performed using image quality metrics, including mutual information (MI), edge retentiveness (QAB/F), phase congruency (PC), non-linear correlation information entropy (QNCIE), and universal image quality index (UIQI), to evaluate the standard of the output fused images

Qualitative Analysis

A qualitative analysis is conducted on the output fused images produced from five different fusion techniques on 15 pairs of visible and infrared images, each capturing various scenes from"https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029". The fusion methods compared with the proposed method include encoder and decoder network, latent low rank representation, VGG-19 network, dense network, and CNN method. It is observed that the fusion method that uses encoder and decoder shows unsatisfactory results as it introduces significant artificial noise in the fused image. Similarly, VGG-19 and latent low rank representation methods produce artifacts that are visually similar. While image fusion methods relying on CNN and densenetwork result in a fused image with prominent targets, but the background appears distorted.

However, the proposed method emphasizes, thermal targets and retains textural details, resulting in the most optimal fusion performance compared to the aforementioned techniques.

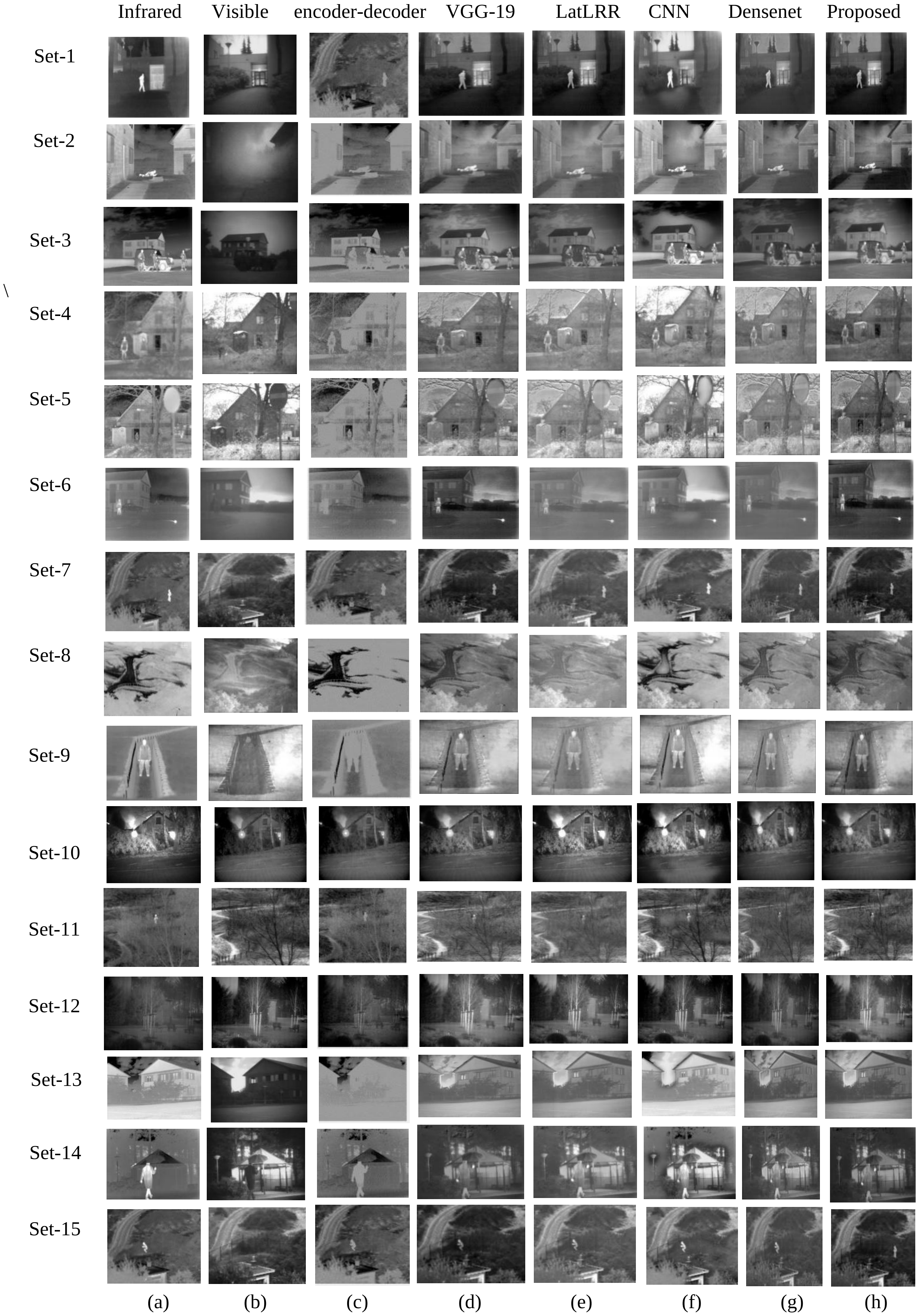

Fig 4: (a) infrared image, (b) visible image. Simulation results using (c) encoder-decoder network, (d) VGG-19 network, (e) Latent Low-Rank representation, (f) CNN, (g) dense network, (h) proposed method

Quantitative Analysis

To compare the effectiveness for the proposed fusion methodology with other approaches, we used five different metrics to quantitatively assess the standard of the resulting images.These metrics include mutual information (MI), edge retentiveness (QAB/F), phase congruency (PC), non-linear correlation information entropy (QNCIE), and universal image quality index (UIQI).

The mutual information (MI) score quantifies the level of information shared between the source images and the fused image. A higher MI score denotes that our method is able to capture more information from the source images.

$$\mathrm{MI} = \sum_{i_A \in I_A}\sum_{i_F \in I_F}(i_A, i_F)\log_2\frac{(i_A,i_F)}{p(i_A)p(i_F)} + \sum_{i_B \in I_B}\sum_{i_F \in I_F} p(i_B, i_F)\log_2\frac{p(i_B,i_F)}{(i_B)p(i_F)} \qquad (5)$$

Where (i_A, i_F) denotes the joint probability distribution of I_A and I_B . (i) is the marginal probability distribution.

The edge retentiveness (QAB/F) metric quantifies the extent to which edges from the source images are present in the fused image. This metric is important as edges carry important information in images and losing them can result in a loss of detail.

$$Q^{AB/F} = \frac{\sum_{=1}^{N}\sum_{j=1}^{M}(Q^{AF}(i,j)w^{A}(i,j) + Q^{BF}(i,j)w^{B}(i,j)}{\sum_{i}^{N}\sum_{j}^{M}(w^{A}(i,j) + w^{B}(i,j))} \qquad (6)$$

where $Q^{AF}(i,j) = {Q_g}^{AF}(i,j){Q_o}^{AF}(i,j)$, ${Q_g}^{AF}(i,j)$ and ${Q_o}^{AF}(i,j)$ are the edge strength and orientation preservation value at the location (i,j), respectively. N and M are size of the image, and $Q^{B}(i,j)$ is similar to $Q^{AF}(i,j)$, $w^{A}(i,j)$, $w(i,j)$ represent the weights of $Q^{AF}(i,j)$, and $Q^{BF}(i,j)$ respectively.

Phase congruency (PC) measures the structure of the fused image, which is crucial for preserving the integrity of the source images.

$$PC = (P_p)^{\alpha}(P_M)^{\beta}(P_m) \qquad (7)$$

Where p represents the phase, while M and m denote the maximum and minimum moments, respectively, and $\alpha = \beta = \gamma = 1$.

Non-linear correlation information entropy (QNCIE) quantifies the extent of non-linear correlation between the input images and the fused image. The significance of this metric lies in its ability to evaluate the degree of resemblance between the original input images and the fused output

$$Q^{NCIE} = 1 + \sum_{i=1}^{3}\frac{\lambda_i}{3}\log_{256}\left(\frac{\lambda_i}{3}\right) \qquad (8)$$

Where λ_i is the eigen value of the nonlinear correlation matrix.

Universal image quality index (UIQI) assesses the quality of the fused image based on three criteria: correlation loss,luminance, and contrast. A high UIQI score indicates that our method has the ability to maintain these three aspects of the input images in the fused image.

$$\mathrm{UIQI} = \left[\frac{4\sigma_{I_A I_F}\mu_{I_A}\mu_{I_F}}{(\sigma_{I_A}^2+\sigma_{I_F}^2)(\mu_{I_A}^2+\mu_{I_F}^2)} + \frac{4\sigma_{I_B I_F}\mu_{I_B}\mu_{I_F}}{(\sigma_{I_B}^2+\sigma_{I_B}^2)(\mu_{I_B}^2+\mu_{I_B}^2)}\right]/2 \qquad (9)$$

Where μ and σ denote the mean and standard deviation respectively, $\sigma_{I_A I_F}$ is the cross-correlation between I$_A$ and I$_F$.

TABLE 1: Quantitative assessments comparison of different fusion methods

	Encoder and Decoder	Latent Low Rank	DL Vgg19	CNN	Dense Network	Proposed Method
MI	1.995	1.652	1.8551	**3.2641**	2.2888	1.738
QABF	0.504	0.495	0.493	0.5442	**0.546**	0.455
UIQI	0.2089	0.481	0.1668	0.1450	0.198	**0.501**
QNCIE	0.1110	**0.112**	0.11	0.1106	0.1097	0.1089
PC	0.6697	0.6702	0.6694	0.6707	0.6687	**0.6714**

The proposed architecture outperforms state-of-the-art methods in objective evaluation, as shown by extensive experimental results. PC (0.6714) and UIQI (0.501) metrics of the proposed method have higher values compared to other fusion method, which implies that the proposed method produces fused images that have higher quality in relation to structural content, luminance, and contrast than other methods.

It is important to consider all metrics when evaluating image fusion methods, and the proposed method excels other methods across multiple metrics. Although the other methods may perform well on some metrics, their overall performance is inconsistent when considering all metrics. Therefore, we can confidently state that the proposed method demonstrates a higher level of performance than the existing approaches in terms of objective evaluation.

V. CONCLUSION

This project is entitled “Infrared andVisible Image Fusion using Latent Low Rank Representation and Deep Feature Extraction Network” adopts a comprehensive integrated deep fusion approach called the Visible and Infrared image fusionnetwork, extract deep features in an adaptive manner, fuse them together, and then reconstruct them. The fusion outputs not only maintain the sharp contrast between the thermal objects and the surroundings but also include abundant texture details.

The unsupervised framework demonstrates that the network architecture possesses a high degree of proficiency in retaining prominent features and textural details, without any apparent distortions or artifacts.

REFERENCES

1. Han, J.; Bhanu, B. Fusion of color and infrared video for moving human detection. Pattern Recognit. 2007,40, 1771–1784. [CrossRef]
2. Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. IEEE Comput 2001, 21, 34–41. [CrossRef]
3. Simone, G.; Farina, A.; Morabito, F.C.; Serpico, S.B.; Bruzzone, L. Image fusion techniques for remote sensing applications. Inf. Fusion 2002, 3, 3–15. [CrossRef]
4. Hanna, B.V.; Gorbach, A.M.; Gage, F.A.; Pinto, P.A.; Silva, J.S.; Gilfillan, L.G.; Elster, E.A. Intraoperative assessment of critical biliary structures with visible range/infrared image fusion. J. Am. Coll. Surg. 2008,206, 1227–1231. [CrossRef]
5. Sanchez, V.; Prince, G.; Clarkson, J.P.; Rajpoot, N.M. Registration of thermal and visible light images of diseased plants using silhouette extraction in the wavelet domain. Pattern Recognit. 2015, 48, 2119–2128.
6. Li, S.; Kang, X.; Hu, J. Image fusion with guided filtering. IEEE Trans. Image Process. 2013, 22,2864–2875. [PubMed]
7. 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.
8. J. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, Information Fusion 31 (2016) 100–109.
9. Teku Sandhya Kumari, Koteswara Rao Sanagapallea, and Santi Prabha Inty. "A two-stage processing approach for contrast intensified image fusion." World Journal of Engineering 17.1 (2020): 68-77.
10. Teku Sandhya Kumari, S. Koteswara Rao, and I. Santi Prabha. "A compendious analysis of feature-extraction algorithms to frame fusion rules." International Journal of Computing and Digital System (2021).
11. Teku Sandhya Kumari, S. Koteswara Rao, and I. Santi Prabha. "Adaptive window-based fractal dimension estimation for weight maps in contrast improved multi-sensor fusion." Journal of Engineering Science and Technology 15.2 (2020): 1319-1337.
12. H. Xu, J. Ma, J. Jiang, X. Guo, H. Ling, U2fusion: A unified unsupervised image fusion network, IEEE Transactions on Pattern Analysis and Machine Intelligence (2020).
13. Prabhakar K R, Srikar V S, Babu R V. DeepFuse: A Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs[C]//2017 IEEE International Conference on Computer Vision(ICCV). IEEE, 2017: 4724-4732.
14. Hui Li and Xiao-Jun Wu. DenseFuse: A Fusion Approach to Infrared and Visible Images. IEEE Transactions on Image Processing, 28(5):2614– 2623, 2018.
15. Yong Ma, Haojie Li, and Baocai Yin. "Infrared and visible image fusion using Latent Low-Rank Representation." Information Fusion, vol. 36, (2017),pp. 191-207.
16. H. Li, X. Wu, and J. Kittler. “Infrared and visible image fusion using a deep learning framework,” in Proc. 2018 24th Int. Conf. Pattern Recognit., Beijing, 2018, pp. 2705–2710.
17. Y. Liu, X. Chen, H. Peng, and Z. Wang, “Multi-focus image fusion with a deep convolutional neural network,” Inf. Fusion, vol. 36, pp. 191–207, Jul. 2017.
18. R. Hou, D. Zhou, R. Nie, D. Liu, L. Xiong, Y. Guo, C. Yu, VIF-net: an unsupervised framework for infrared and visible image fusion, IEEE Transactions on Computational Imaging 6 (2020) 640–651.