ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 6 June 2024

Evaluating Image Fusion Techniques for Improved Low Light Surveillance

[1] Akshat Mishra, [2] Dipesh Kumar Yadav, [3] Nikhil Bathija

[1] [2] [3] SRM Institute of Science and Technology
Corresponding Author Email: [1] am6654@srmist.edu.in, [2] dr2034@srmist.edu.in, [3] nr2265@srmist.edu.in

Abstract— ***There has been a surge of interest towards deep learning based, image fusion methods in recent years. Through the process of image fusion, complimentary information is extracted from images that have been captured by multiple sensors. Irrelevant characteristics are screened out and the remaining relevant information is combined to enrich the detail and quality of these images. In the context of low light, image fusion techniques face difficulties while preserving the details and diminishing the noise produced in the resulting fused image. This occurs mainly due to the lack of visibility caused by insufficient lighting. Such conditions severely impact the fused images generated by the model. This research paper aims to conduct a comparative analysis of several state-of-the-art, deep learning based image fusion models for low light surveillance applications. Additionally, our paper will investigate the merits and challenges corresponding to each method in the context of low light image fusion. The results of our comparative analysis revealed that 'SwinFuse' exhibited superior performance when compared with other methods in preserving image details and reducing noise in the fused images.***

Index Terms— ***Image Fusion, Low Light Surveillance, Visible and Infrared Images.***

I. INTRODUCTION

In today's evolving world, the role played by surveillance systems cannot be downplayed. Surveillance systems play a vital role in the safety and security of the community. They act asthe primary toolsfor crimeprevention,threat detection and investigation. The clarity and resolution of the images captured is crucial to ensure the effectiveness of the surveillance systems, especially in conditions with limited visibility. Deep learning based image fusionoffers sophisticated solutions to deal with the limitations induced by low lighting [1]. This involvescombining multiple images captured by different sensors, such as visible and infrared image sensors and creating a resulting fused image that embodies enhanced visibility and betteroverall situational awareness for tasks such as object detection. However, image fusion under low light is still a challenging task. This is because factors such as noise, low contrast and limited visibility are induced in images that are captured under low light, all of which contribute to the degradation of the fused image. In recent years, new techniques have been Devised that are impervious to these aforementioned challenges. Majority of these techniques utilise visible (VIS) and infrared (IR) images for image fusion [2].

Infrared images have the capability of capturing thermal radiation that is emitted by objects. Hence, they have found usage in object detection in little to no light conditions. They are also impervious to conditions such as smoke, fog [3], haze and obstructions. However, IR images lack colour information, textural details, and have lower spatial resolution when compared to visible spectrum images. Whereas, Visible images offer detailed colour information and higher spatial resolution [2]. This in turn enhances the overall quality of the fused image. But in low light conditions, visible images are susceptible to reduced visibility and noise [4], limiting their usefulness in image fusion. Thus by utilising both visible and infrared images, we can compensate for each of their limitations [5] and create a more robust image fusion technique for low light surveillance applications.

The organisation of this paper is as follows. In the next section a comprehensive analysis of the related works is performed. Four state-of-the-art deep learning based image fusion methods are depicted in section III. Low Light Visible-Infrared Paired (LLVIP) Dataset [7] is used to perform experiments, and qualitative and quantitative comparisons are done in section IV. At last conclusions are drawn in section V.

II. LITERATURE SURVEY

Image fusion, which combines information from visible and infrared cameras, holds great promise in enhancing visibility, object detection, and overall situational awareness in surveillance applications [1]. These methods often utilise deep learning frameworks to extract features and optimise fusion strategies for different scales of source images [6]. Multiple state-of-the-art, deep learning-based image fusion methods have been designed to overcome these challenges. For example, DenseFuse, focuses on fusing infrared and visible images using a convolutional neural network (CNN)

[4]. Similarly, FusionDN introduces a unified densely connected network for image fusion, leveraging deep learning for feature extraction and fusion [8]. NestFuse, developed by, incorporates nest connections and spatial/ channel attention models to enhance fusion performance [2]. Moreover, advancements in image fusion technology have

led to the development of innovative approaches such as generative adversarial networks, total variational models, and adaptive algorithms that further enhance the quality and details of fused images [9]. These CNN or GAN-based fusion frameworks have shown exceptional fusion capabilities [2]. Furthermore, research efforts have explored decision-level fusion detection methods for visible and infrared images under low light conditions [10]. These methods aim to enhance object detection precision by effectively combining information from both image types

[10]. Additionally, studies have highlighted the advantages of RGB-NIR fusion for low-light imaging, demonstrating promising results in improving image quality in challenging lighting conditions [11]. Techniques like weighted sparse representation, gradient domain guided filter pyramid fusion, and spectral graph wavelet transforms have demonstrated superior performance in generating fused images with enhanced visual effects and detailed texture information [12].

III. METHODOLOGY

In this section, we will describe the parameters of our study. We selected 4 state-of-the-art deep learning image fusion models. A uniform dataset was used for the training and evaluation of these models.

A. Dataset Description

For our research, we utilised the Low Light Visible-Infrared Paired (LLVIP) dataset [7]. This dataset is specifically designed for low light surveillance applications and comprises visible and infrared image pairs as illustrated in Fig.1. It has captured a total of 26 different scenes through 16,836 VIS-IR image pairs. This allows for a more comprehensive analysis and comparison of fusion methods under challenging lighting conditions [13]. All captured scenes depict pedestrians, automobiles and various others object on the road between 1800 hours and 2200 hours.

The LLVIP dataset offers a diverse range of low light scenarios, capturing the complexities of real-world surveillance environments. The utilisation of paired visible and infrared images in image fusion is vital in preserving image details, reducing noise and enhancing overall visibility in low light conditions.

By leveraging the LLVIP dataset, our study aims to provide a comprehensive comparative analysis of various state-of-the-art, deep learning based image fusion methods.

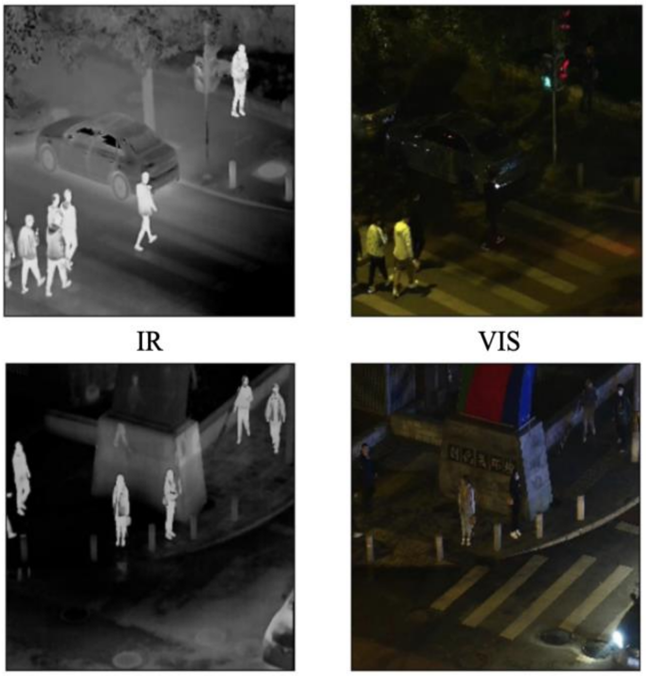

Fig. 1. Sample IR-VIS image pairs from LLVIP Dataset.

B. Image Fusion Methods

Models were trained on identical sets of grayscale, resized (256x256) images from the LLVIP dataset. The training parameters for each model were kept identical to maintain consistency.

C. Deep Image Decomposition Fusion (DIDFuse)

DIDFuse makes use of an auto-encoder architecture to translate visible and infrared images into distinct feature maps. These feature maps are designed to capture the low-frequency background information and high-frequency detail information [4] present in the images. The fundamental principle of DIDfuse revolves around the utilisation of a loss function that ensures the similarity of background features and dissimilarity of detail features between the original images. During the [4] fusion process, the background and detail feature maps are amalgamated independently through a specialised module. Subsequently, the decoder is employed to reconstruct the fused image with the objective of preserving the highlighted targets from the infrared image (maintained by the similarity in background features) and the intricate texture details from the visible image (maintained by the dissimilarity in detail features) as shown in Fig.2.

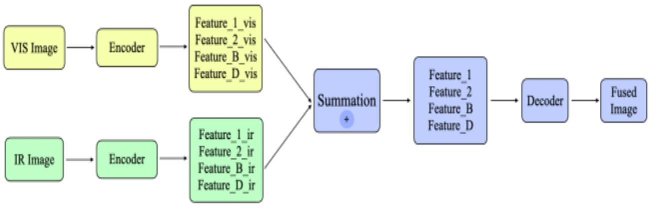

Fig. 2. The architecture of the DIDF use Network

D. SwinFusion

SwinFusion introduces an innovative method for image fusion by utilising the Swin Transformer architecture. Unlike traditional approaches, SwinFusion utilises Swin Transformer blocks to extract shallow and deep features from each source image, enabling a detailed analysis at various scales to capture essential information [14]. Instead of employing a simple process for combining these features, SwinFusion incorporates an "attention" mechanism that prioritises critical details from both images as shown in Fig.3. This allows features from any region of the images to influence the final fused image and capture long-range relationships [15]. This approach enhances accuracy and information preservation compared to conventional fusion techniques [16].

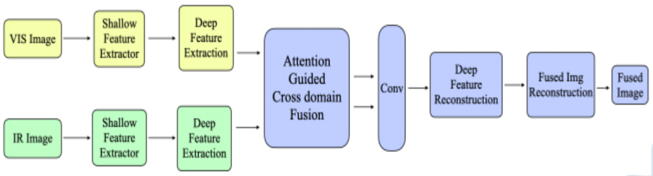
VIS Image
Shallow Feature Extractor
Deep Feature Extraction
IR Image
Shallow Feature Extractor
Deep Feature Extraction
Attention Guided Cross domain Fusion
Conv
Deep Feature Reconstruction
Fused Img Reconstruction
Fused Image

Fig. 3. The architecture of the Swinfuse Network

E. Dual-branch Network for Infrared and Visible Image Fusion

In our study, we compared and evaluated four deep learning based image fusion algorithms based on their effectiveness in low light surveillance applications. All the Dual-branch Network for Infrared and Visible Image Fusion utilises Dense Convolutional Blocks (DCBs) in each branch to ensure smooth encoding. This method is very efficient in extracting features crucial for low light fusion The visible branch captures high-level semantic information and spatial details through shallow features, while the infrared branch focuses on extracting deeper feature maps containing thermal information [4]. Through a feature fusion strategy like channel concatenation, the network merges these distinct features as shown in Fig.4, ensuring the preservation of both spatial and thermal details in the final fused image.

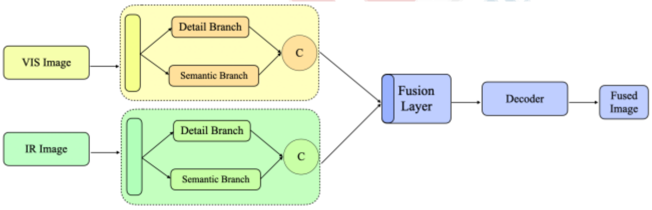

Fig. 4. The architecture of the Dual Branch fusion Network

F. DenseFuse

DenseFuse introduces a unique approach to convolutional neural networks by integrating dense blocks within its encoder architecture. These dense blocks enable feature reuse, where feature maps from each convolutional layer are concatenated and passed as input to all subsequent layers within the block. This dense connectivity enhances feature propagation, leading to a more comprehensive extraction of features from input images. Subsequently, a fusion strategy is employed to integrate complementary feature maps, followed by a decoder network that reconstructs the final fused image [4] [17] [18] as shown in Fig.5.

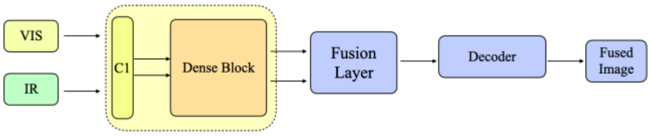

Fig. 5. The architecture of the Densefuse Network

Table I: Quantitative results od different methods. The largest value is shown in bold, and the second largest value is underlined

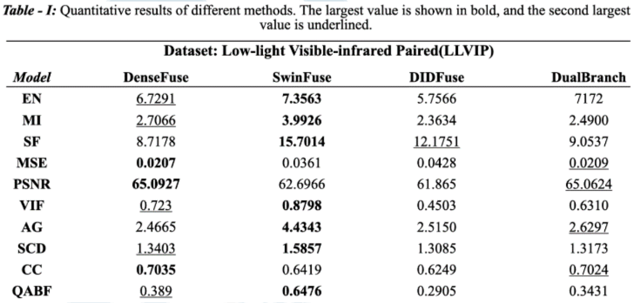

Table - I: Quantitative results of different methods. The largest value is shown in bold, and the second largest value is underlined.

Dataset: Low-light Visible-infrared Paired(LLVIP)

Model	DenseFuse	SwinFuse	DIDFuse	DualBranch
EN	6.7291	**7.3563**	5.7566	7172
MI	2.7066	**3.9926**	2.3634	2.4900
SF	8.7178	**15.7014**	12.1751	9.0537
MSE	**0.0207**	0.0361	0.0428	0.0209
PSNR	**65.0927**	62.6966	61.865	65.0624
VIF	0.723	**0.8798**	0.4503	0.6310
AG	2.4665	**4.4343**	2.5150	2.6297
SCD	1.3403	**1.5857**	1.3085	1.3173
CC	**0.7035**	0.6419	0.6249	0.7024
QABF	0.389	**0.6476**	0.2905	0.3431

G. Performance Metrices

To evaluate the effectiveness of the various image fusion techniques under low light conditions, we employed a variety of qualitative metrics. Each of these metrics quantifies different features of the fused image. Metrics like Entropy (EN) and Mutual Information (MI) compute the valuable information that is preserved in the fused image from the input images. The fine details and structural coherence in the image is quantified using Spatial Frequency (SF). Error-based metrics like Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR), quantify the fidelity of the fused image. Additionally, Visual Information Fidelity (VIF) metric considers visual perception factors (visual appeal) of the fused image. Average Gradient (AG) and Edge Information (QABF) Focus on sharpness, and structural integrity in the fused image which is crucial for surveillance tasks such as object recognition. Finally, similarity metrics like Coefficient Correlation (C C), and Sum of Correlation of Differences (SCD) measure the consistency between the input and the fused image.

IV. EXPERIMENTAL RESULTS

In this section, we will present the results of our comparative analysis on the LLVIP dataset. The corresponding fused images of each model are depicted in Figure 6. This is followed by a quantitative analysis of the results of each model depicted in Table-I.

The attention guided cross domain approach in the fusion layer of SwinFuse is able to fuse the corresponding VIS and

IR features of low light images more efficiently when compared to the other methods.

The results in Table-1 validate this.SwinFuse outperformed all the other models by having the highest scores in seven of the ten metrics that were considered for evaluation. It had the best scores for EN, MI, SF, VIF, AG, SCD and QABF.

SwinFuse was followed by DenseFuse which had the highest scores in the remaining three metrics.

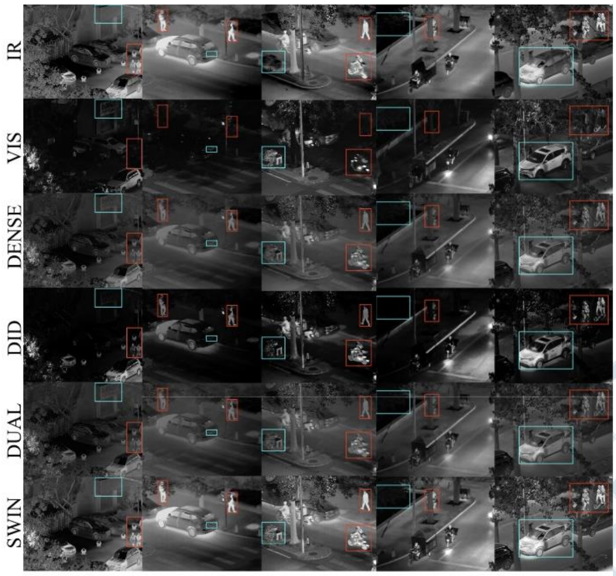

Fig. 6. Quantitative results for different methods. Areas marked by red and blue boxes are features retained from IR and VIS image respectively.

DenseFuse had the best scores in MSE, PSNR, CC and the second best scores in EN, MI, VIF, SCD and QABF.

DenseFuse performed extremely well in error-based metrics like MSE and PSNR. It also has competing scores for similarity metrics like CC and SCD.

By utilising dense blocks in its encoder, DenseFuse was able to capture features from the input images effectively when compared with DIDFuse and Dual Branch Fusion.

V. CONCLUSION

In this paper, we successfully conducted a comparative study of four state-of-the-art, deep learning based image fusion models on the LLVIP dataset to test their applications in low light surveillance. We analysed the architectures of all four models and explained them briefly. Ten evaluation metrics were utilised and were used to compare the effectiveness of each model. Our results were definitive and concluded that out of the four models, SwinFuse had the best performance in fusing images in low visibility. We

conclude that models that utilise a specialised fusion technique like the attention guided cross domain fusion of SwinFuse are cable of fusing low light images more efficiently than the models that utilise standardised fusion techniques like summation, L1 norm and weighted average.

In the future, we can extend this work by focusing on specific low-light surveillance scenarios and factor in adverse weather conditions. We can also explore real-time processing of the fusion models, which is crucial for practical applications.

REFERENCES

[1] D. Zhu, W. Zhan, Y. Jiang, X. Xu, and R. Guo, "MIFFuse: A Multi-Level Feature Fusion Network for Infrared and Visible Images," IEEE Access, vol. 9, pp. 130778-130792, 2021.

[2] H. Li, X.-J. Wu, and T. Durrani, "NestFuse: An Infrared and Visible Image Fusion Architecture Based on Nest Connection and Spatial/Channel Attention Models," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 12, pp. 9645-9656, Dec. 2020.

[3] K. Beier, R. Boehl, J. Fries, W. Hahn, D. Hausamann, V. Tank, G. Wagner, and H. Weisser, "Measurement and modeling of infrared imaging systems at conditions of reduced visibility (fog) for traffic applications," Proceedings of SPIE - The International Society for Optical Engineering, vol. 2223, pp. 175-186, 1994.

[4] H. Li and X.-J. Wu, "DenseFuse: A Fusion Approach to Infrared and Visible Images," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2614-2623, May 2019.

[5] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," in Information Fusion, vol. 8, no. 2, pp. 119-130, 2007.

[6] A. Toet, M. A. Hogervorst, S. G. Nikolov, J. J. Lewis, T. D. Dixon, D. R. Bull, and C. N. Canagarajah, "Towards cognitive image fusion," Information Fusion, vol. 11, no. 2, pp. 95-113, 2010.

[7] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou, "LLVIP: A Visible-infrared Paired Dataset for Low-light Vision," 2021 IEEE/ CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, pp. 3489-3497, 2021.

[8] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: A Unified Densely Connected Network for Image Fusion," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 12484-12491, Apr. 2020.

[9] H. Li, X.-J. Wu, and J. Kittler, "Infrared and Visible Image Fusion using a Deep Learning Framework," in 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, pp. 2705-2710, 2018.

[10] Z. Hu, Y. Jing, and G. Wu, "Decision-level fusion detection method of visible and infrared images under low light conditions," EURASIP Journal on Advances in Signal Processing, vol. 2023, no. 38, 2023.

[11] S. Jin, B. Yu, M. Jing, Y. Zhou, J. Liang, and R. Ji, "DarkVisionNet: Low-Light Imaging via RGB-NIR Fusion with Deep Inconsistency Prior," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 1, pp. 1104-1112, 2022.

[12] Y. Liu, B. Yan, R. Zhang, K. Liu, G. Jeon, and X. Yang, "Multi-Scale Mixed Attention Network for CT and MRI Image Fusion," Entropy, vol. 24, p. 843, 2022.

[13] R. Liu, J. Liu, Z. Jiang, X. Fan, and Z. Luo, "A bilevel

integrated model with data-driven layer ensemble for multi-modality image fusion," IEEE Transactions on Image Processing, vol. 30, pp. 1261-1274, 2020.

[14] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, and B. Guo, "Swin transformer: hierarchical vision transformer using shifted windows," arXiv:2103.14030, 2021.

[15] Y. Xu, S. Zhou, and Y. Huang, "Transformer-Based Model with Dynamic Attention Pyramid Head for Semantic Segmentation of VHR Remote Sensing Imagery," Entropy, vol. 24, p. 1619, 2022.

[16] S. Liang, Z. Hua, and J. Li, "Transformer-based multi-scale feature fusion network for remote sensing change detection," Journal of Applied Remote Sensing, vol. 16, no. 4, p. 046509, Nov. 2022.

[17] Z. Shen, J. Wang, Z. Pan, Y. Li, and J. Wang, "Cross attention-guided dense network for images fusion," arXiv:2109.11393v2, 2021.

[18] Y. Zang, "Ufa-fuse: a novel deep supervised and hybrid model for multi-focus image fusion," arXiv:2101.04506v4, 2021.