Multi-Resolution Image Fusion Using Various Techniques

Omkar Lonkar [1] & Dr Archana Ghotkar[2]
[1]Student, SCTR's Pune Institute of Computer Technology, Information Technology, Pune, Maharashtra, India
omkar.v.lonkar@gmail.com
[2]Associate Prof. and Head, Dept. of IT Engineering Pune Institute of Computer Technology, Pune, Maharashtra, India,
aaghotkar@pict.edu

Abstract

In recent years, the field of image fusion has seen substantial development, particularly with the rise of generative adversarial networks (GANs). Multi-Resolution Image Fusion is an approach used to combine images from numerous sources with different resolutions to create a single, more informative and detailed image. Researchers have focused on optimizing these networks to handle the complex task of fusing images in real-time, which is crucial for applications such as surveillance and disaster management. This research paper provides a comprehensive review of key image fusion approaches, including Principal Component Analysis (PCA), the Laplacian Pyramid approach, and the state-of-the-art Dual-Discriminator Conditional Generative Adversarial Network (DDcGAN). Comparative analysis of these techniques is done based on fusion performance, computational complexity, information preservation and amount of noise present in fused image.
Keywords: Multi-Resolution Image, Image fusion, Generative adversarial network (GAN), Remote sensing.

1. Introduction

Fusion of visible and infrared light has become popular in signal processing because of its widespread application in numerous fields like remote sensing computer vision, medical imaging, etc. diseases, military research, etc. Among these sensors, probably the most widely used types are infrared and visible light, along with wavelengths of 300-530 nm μm and 8-14 nm, respectively. The feature of this fusion of visible as well as infrared light is that visible light allows to represent rich details, while infrared imaging displays the captured infrared images in grayscale even in different lighting conditions. It also works best when it comes to hot targets or unacceptable situations. Due to their close compatibility, the fusion results can improve visual perception by revealing almost all the features of the target. As a result, their fusion has a dominant role in civilian and military applications. The key to image fusion for multi-source images is to extract the most dominant data from images in different views and combine them into a sole image fusion. So, the fused images are able to give a better detailed as well as comprehensive depiction whilst decreasing the amount of data. Therefore, various fusion methods have been suggested in recent times, along with neural network-based techniques and other fusion techniques. Recently, neural network-based image fusion techniques have become popular.

Multi-Resolution fusion of image is the combination of numerous source images of different resolutions creating synthetic images that provide more descriptive information. Infrared and visible light sensors have different characteristics due to different spectra. Infrared can identify electrical energy even in low light, while visible light can capture spatial details but may lose effectiveness in low light. So, by fusing these images we can generate an enhanced image with super-resolution. Here the image which is fused is enforced to maintain the thermal energy radiation in this infra-red taken image as well as these texture attributes in this visible image together.

On account of the constraints of hardware equipment as well as environments, these infrared images display blurred details and low resolution with contrast to its correlating visible images, because it is difficult to upgrade the clarity of infra-red taken images by improving the HW equipment. For the fusion of multi-resolution infra-red as well as visible images (e.g., images of various resolutions), an approach of down sampling visible images or up sampling infrared images before its fusion unavoidably resulting in thermal radiation data blurring or visible texture data deprivation. As a result, it leaves a difficult job to fuse multi-resolution visible as well as infrared images without overlooking the prime data. We are able to be able to get the most information from two images (infrared and visible) by fusing them instead of separate. So, the motivation to select this topic for the seminar was because of the knowledge gap in this problem. Though there are many solutions, the purpose of this is to find the best method for the image fusion giving its best result.

2. Literature Survey

Over the years, several methodologies in different fields have been proposed to tackle the challenge of fusing images with varying resolutions and characteristics, each aiming to amplify the quality of fused images while retaining critical details.

Jiayi Ma *et al.* have put forth a new technique of Multi-Resolution Image Fusion in their study. This study proposes using a DDcGAN instead of a GAN which has a single discriminator. DDcGAN can effectively identify and retain the most valuable details from images from the source by using two discriminators while training to improve the accuracy [1]. Qi Jin et al. focuses on infrared as well as visible image fusion of flaming fires in forests. This study explores different image fusion techniques using GANs to find out the optimal technique for image fusion of flaming fires in forests. The study concludes with the optimal technique being multi-level as well as multi-classification GAN-based technique (MMGAN) for the purpose of union of infra-red and visible images of flaming fire in forests. It points out that MMGAN solves the problem that GANs usually disregard contrast ratio of visible light details as well as texture data of infrared light. The result has a more notable richer detailed texture and contrast ratio details, as MMGAN could maintain the exhaustive texture details of these infrared images as well as the contrast ratio of these visible images. [5]

Peter M. Atkinson *et al.* delve into the challenges faced in multi-resolution image fusion regarding the deprivation of valuable details whilst the process of image fusion. This Information Loss-Guided Multi-Resolution Image Fusion method directly addresses this challenge by guiding the fusion process to minimize information loss while preserving important details from each input image. This approach is built upon the multi-resolution decomposition framework, where input images are decomposed into different resolution levels (typically using wavelet transforms, Laplacian pyramids, or similar methods). The fusion process occurs at each resolution level, combining the most relevant features while minimizing the loss of information [7].

Yue Xi *et al.* aim to improve the quality and efficiency of infrared as well as visible fusion of images by developing a state-of-the-art approach called MFST. The approach introduces an accommodative fusion plan to better utilize multi-modal attributes, a central self-attention system to consider both global and local information during fusion, as well as the integration of saliency information to preserve prominent objects from infrared images. Testing on three different datasets demonstrates the method's strong generalization ability and effectiveness in fusing complex scenes. The authors plan to focus on these issues in future research to enhance the model further [6].

Rekha R. Nair *et al.* explore about multi modal fusion of the images in the medical sector. In medical imaging, different modalities like CT, MRI, PET scans as well as others supply with complementary details about the patient's anatomy and pathology. Multi-sensor medical image fusion combines these diverse modalities into a single image that retains critical features from each source, improving diagnostic accuracy. The fusion method in this study employs pyramid decomposition combined with Discrete Wavelet Transform (DWT), which allows for multi-resolution analysis of input images. This is crucial in medical imaging, where fine details like tissue boundaries and large-scale structures need to be preserved simultaneously [8].

3. Proposed Methods

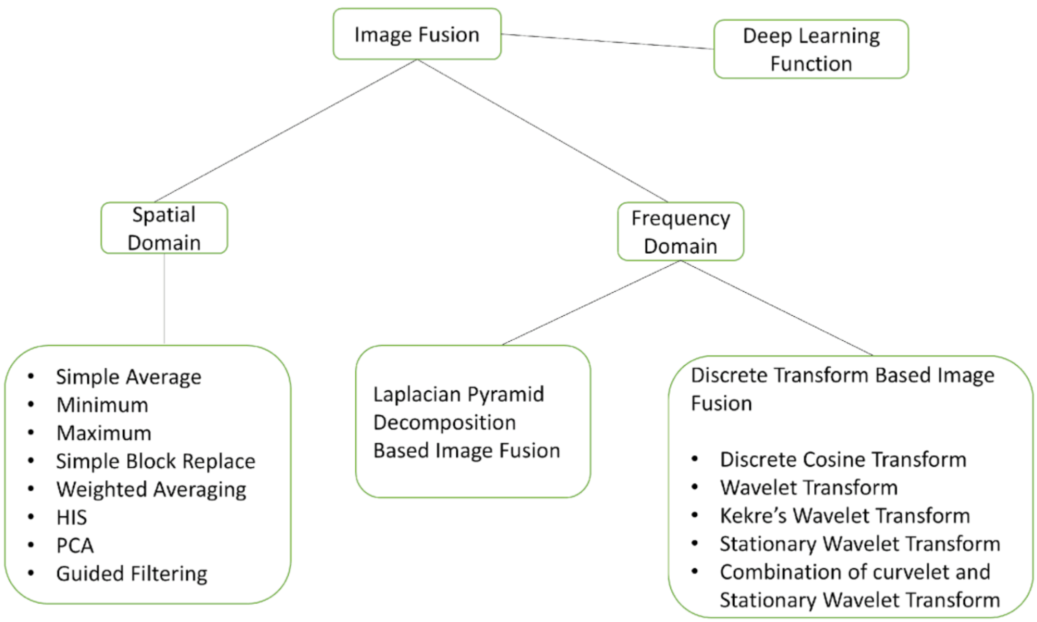

Fig. 1. Image Fusion Techniques

Fig 1 showcases numerous image fusion techniques. The techniques are divided into three main domains which are spatial domain, frequency domain and the deep learning established function domain. For this paper we will be exploring Principal Component Analysis (PCA) technique from the Spatial Domain, Laplacian Pyramid technique from the Frequency Domain along with image fusion using DDCGAN technique from the Deep Learning Based Domain. In this research paper PCA and Laplacian Pyramid techniques of image fusion have been implemented and tested with further comparisons been done on numerous parameters.

Dataset Used: The dataset consists of image pairs, each comprising a visible and an infrared version of the same scene, which were utilized to evaluate the performance of the image fusion techniques.

PCA (Principal-Component-Analysis) Technique

PCA is a widely used statistical method which modifies correlated data to a section of uncorrelated variables referred to as principal components. In image fusion, PCA plays a valuable part in combining information from multiple images while reducing redundancy and preserving essential details.

The process begins by representing the source images in a high-dimensional space. These images are then decomposed into principal components based on their variance, with the first main component capturing the highest variance and the subsequent components representing decreasing levels of variance. The fusion is typically performed by retaining the most significant principal component and combining it with complementary information from the other components of the images.

In the context of fusion of images, PCA helps by reducing the dimensionality of the dataset, thus compressing the dominant attributes of both images to a single image fusion. This major upper hand of PCA is its ability to upgrade image clarity by maintaining the most significant information from the input images while discarding less relevant details. This technique is particularly effective in applications like remote sensing, where the fusion of multispectral as well as panchromatic images is required for improvement of spatial resolution and spectral fidelity.

Advantages:

- Computationally efficient and relatively easy to implement.
- Reduces dimensionality and preserves the most significant features.

Disadvantages:

- May lead to information loss, especially in detailed regions.
- Does not explicitly consider spatial information.

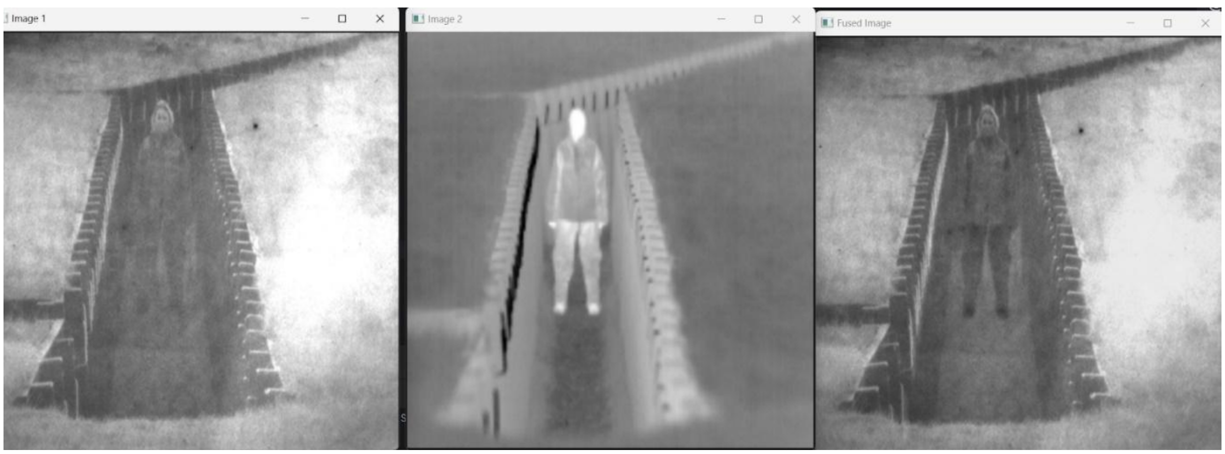

Fig. 2. Image Fusion using PCA

Fig 2 demonstrates the implementation of image fusion of infra-red as well as visible images using this PCA technique. Image1 (from the left) is the higher resolution visible image while Image2 is the relatively lower resolution infra-red image. Image3 is the fusion of Image1 and Image2. Even though the texture detail is preserved in the fused image, it suffers from the lack of vital thermal radiation data from the infra-red image.

Laplacian Pyramid Technique

The Laplacian Pyramid is a multi-resolution, multi-scale technique used for image fusion, particularly known for its ability to preserve both high-frequency as well as low-frequency data from the images from the source. This technique is taken from the hierarchical decomposition of the image to different layers or levels, each representing a specific frequency band. This approach ensures that finer details and broader features are preserved during the fusion process.

The fusion process using the Laplacian Pyramid typically involves the following process:

1. **Pyramid Decomposition**: The insert images are first decomposed into a series of layers using Gaussian and Laplacian pyramids. The Gaussian pyramid smooths the image by reducing its resolution progressively,

creating lower-frequency representations at each level. From the Gaussian pyramid, the Laplacian pyramid is constructed by subtracting adjacent Gaussian levels, capturing the high-frequency details (edges and textures) that represent finer image details at each scale.

2. **Fusion at Each Level**: Once the pyramids are created for each source image, the its corresponding levels of the Laplacian pyramids are fused. Different fusion strategies can be applied at each level, but the most common one is to choose the highest value of each pixel from the correlating layers of the source images. This method ensures that the sharpest details from each source image are retained in the fused result.

3. **Pyramid Reconstruction**: After the fusion process, the fused Laplacian pyramid is reconstructed back into a single image by reversing the decomposition process. This involves adding the fused Laplacian layers back together with the smoothed Gaussian images, producing the final fused image that contains both fine details and smooth transitions.

Advantages:

- High contrast details and overall image structure is preserved. Especially useful in medical imaging and remote sensing
- Maintains spatial information well, especially edges and textures.
- Effective in capturing and preserving both global and local details.

Disadvantages:

- Computationally expensive compared to simpler techniques like PCA.
- May suffer from artifacts or loss of sharpness during reconstruction

Fig. 3. Image Fusion using Laplacian Pyramid

Fig 3 demonstrates the implementation of image fusion of infra-red as well as visible images applying this Laplacian Pyramid technique. Image1 (from the left) is the higher resolution visible image while Image2 is the relatively lower resolution infra-red image. Image3 is the fusion of Image1 and Image2. The fused image preserves the thermal data but lacks texture details.

Image Fusion Using DDcGAN (Dual- Discriminator- Conditional -Generative- Adversarial- Network)

DDcGAN is proposed for the fusion of infrared together with visible images of various resolutions. This network forms a hostile game with two discriminators along with a generator. This generator focuses to produce an authentic fused image derived from a particularly designed content deprivation to fool these two discriminators, whilst they strive to differentiate the structure distinctions between these two source images as well as the fused image, consecutively, on top of the content deprivation.

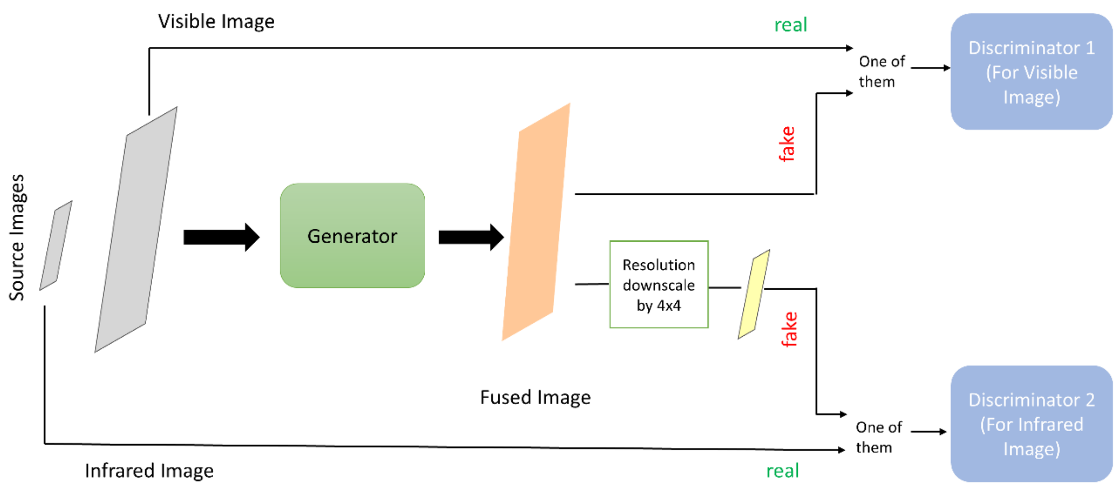

Fig. 4. Framework of DDcGAN [10]

The DDcGAN architecture builds upon the framework of a Conditional Generative Adversarial Network (CGAN), where the generator network is trained for the creation of a image fusion, and two discriminator networks are used to assess the quality of the generated image. These components of the DDcGAN technique are as listed below:

1. **Generator Network**: This generator produces the fused image by learning a mapping from the source images (e.g., visible and infrared) to a fused output. The network is conditioned on both source images, which means it takes them as inputs and produces a image fusion that incorporates the most relevant data from both. The generator's objective is to create an image that not only captures high-frequency attributes (such as edges and textures) but also maintains the semantic content of the scene.

2. **Dual Discriminators**: The dual-discriminator setup includes two separate discriminator networks that evaluate the image fusion quality from various perspectives:

- **Local Discriminator**: This local discriminator focuses on evaluating fine attributes, such as edges and textures, ensuring that the fused image preserves sharpness and clarity.

- **Global Discriminator**: The global discriminator assesses the overall structure and coherence of the fused image, making sure that large-scale features, such as shapes and context, are preserved.

The two discriminators operate in parallel, each providing feedback to the generator during the training process. By balancing the input from both discriminators, the generator learns to produce image fusion that excel in local detail as well as in global consistency.

3. **Loss Functions**: The training of DDcGAN is driven by adversarial loss, which motivates the generator to create realistic fused images which are difficult to distinguish from actual images. On top of adversarial loss, the network may incorporate pixel-wise loss (such as mean squared error) to ensure that the image fusion remains faithful to the original insert images in terms of intensity and texture distribution.

4. **Multi-Resolution Fusion**: One of the main advantages of DDcGAN is in its capability to handle images of varying resolutions. By using a multi-scale architecture, the network can capture and preserve attributes at various scales, ensuring that the image fusion contains fine-grained data as well as broad contextual data. This makes it is especially well-suited for applications where input images have different resolutions or contain diverse types of information (e.g., combining low-resolution thermal images with high-resolution visible images).

Fig 5 Image Fusion Using DDcGAN

Fig 5 demonstrates the implementation of this image fusion of infra-red as well as visible images using this PCA technique. Image1 (from the left) is the higher resolution visible image while Image2 is the relatively lower resolution infra-red image. Image3 is the fusion of Image1 and Image2. Both visible textures and thermal information are preserved in this image fusion.

Fig 6 Image Fusion using DDcGAN

Fig 6 shows another example of image fusion using DDcGAN.

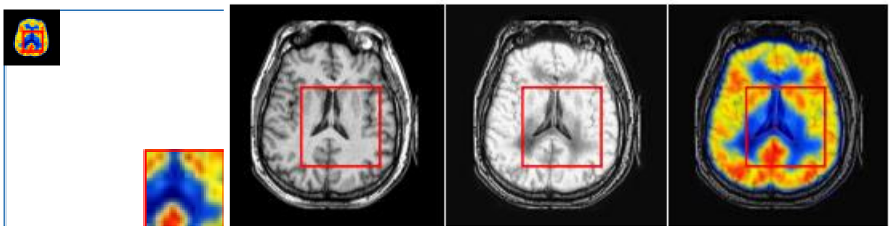

Fig 7 Fusion of PET scans and MRI images

Fig 7 shows the image fusion of PET scan and image (medical imaging). Image 1 (from the left) is of a PET scan. These images are generally of lower resolution. Image 2 is of the source MRI scan. Image 3 shows the intensity channel of this image fusion while Image 4 is the actual fused image. The basic structure of this MRI image is fairly preserved in this image fusion.

4. Results & Discussion

.4.1 Equations

4.1.1 PSNR

$$\mathbf{PSNR = 10^2. \log_{10}}(\frac{MAX^2}{MSE})$$

Where:

- **MAX** is the highest possible pixel value of the selected image.
- **MSE** is Mean Squared Error connecting the fused image with one of their input/reference images

$$\mathbf{MSE} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (I(i,j) - K(i,j))^2$$

- **I(i,j)** refers to the Pixel value of their input/reference image.
- **K(i,j)** refers to the Pixel value of their fused image.
- **m,n**: Dimensions of the image (height and width).

The above formulas outline the equations needed to calculate the Peak- Signal-to-Noise- Ratio (PSNR), a key parameter evaluating the effectiveness for these techniques.

4.1.2 Entropy

Entropy quantifies the amount of information or detail present in an image. When fusing images from different modalities (infrared and visible), the goal is to preserve as much useful detail as possible from both sources.

$$\boldsymbol{Entropy = -\sum_i pi.\log_2 pi}$$

Where:

- $\boldsymbol{pi}$ is the frequency of the i^{th} unique pixel in an image divided by the total pixel frequency count of the image

4.2 Tables

Table 1. Analysis of the Techniques

Serial No.	Technique	Visual Quality	PSNR (Peak Signal-to-Noise Ratio)	Information Preservation	Computational Efficiency
1	PCA (Principal Component Analysis)	Moderate: Image artifacts may appear due to variance loss	Moderate to High, typically around 30-35 dB	Limited: Information from one source might be lost due to dimensionality reduction	High: Fast with low computational cost
2	Laplacian Pyramid	High: Produces better edge preservation and clarity.	High, typically 35-40 dB	Good: Retains more details from the original images, especially edges	Moderate: Requires multi-scale decomposition, which can be computationally expensive.
3	DDcGAN (Dual-Discriminator Conditional GAN) [10]	Excellent: Provides highly realistic and detailed images.	Very High: Can achieve over 40 dB	Excellent: Multi-level constraints allow for the preservation of high-level semantic and low-level detail information.	Low: Computationally intensive, requires significant training time and resources (GPU).

Table 1 provides a clear comparison of these fusion techniques, highlighting their strengths and limitations. While PCA offers high computational efficiency, it sacrifices some information and image quality.

The Laplacian Pyramid offers a balance, preserving more details at the cost of higher computational demands. DDcGAN, on the other hand, excels in visual quality and information preservation, but its computational cost is high, making it suitable for tasks where quality is of importance, and computational resources are available.

4.3 Results and discussion

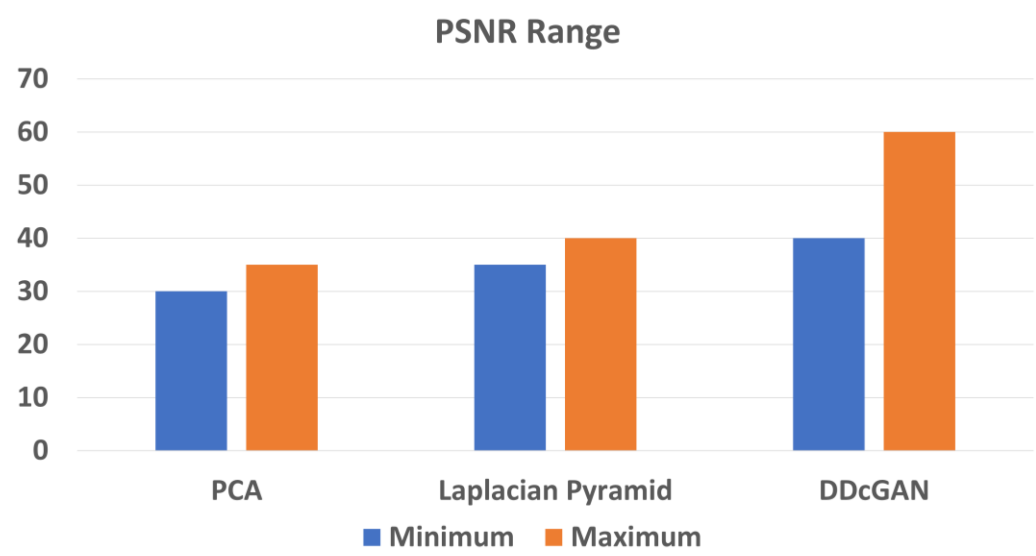

Fig. 8. PNSR Comparison

Figure 8. illustrates the PSNR range (minimum and maximum) for three image fusion methods: PCA, Laplacian Pyramid, and DDcGAN using a general dataset. DDcGAN exhibits the highest PSNR , demonstrating superior image fusion superiority & eminence in comparison with other approaches.

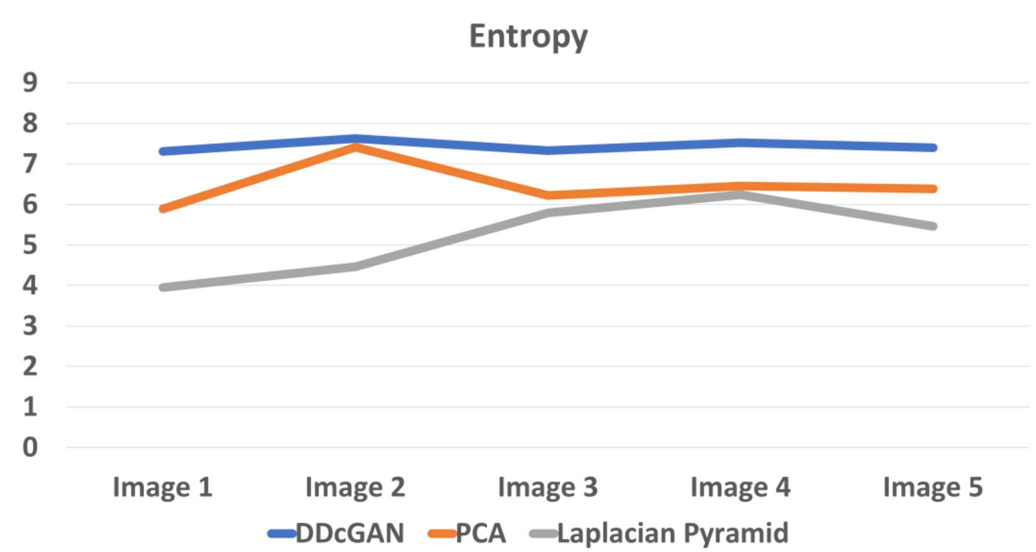

Fig. 9. Entropy Comparison

Figure 9. presents the entropy measurements of fused images generated by the three distinct image fusion techniques. A sample of five image pairs were selected from the dataset and processed using each of the fusion methods. The results consistently indicate that the DDcGAN approach yields fused images with higher entropy (Avg: 7.4) compared to the other techniques. This increased entropy suggests that DDcGAN provides superior feature representation and enhanced detail retention in the fused images. It's important to note that an increase in entropy might sometimes be influenced by noise. However, the high PSNR ratio observed with the DDcGAN method mitigates this concern.

5. Conclusion

The primary purpose of this study is to inspect different fused image techniques and find out optimal techniques. From this paper it has become conclusive that Deep Learning based techniques have gone way ahead of any traditional methods. Of the three techniques reviewed in this paper, DDcGAN is overwhelmingly better. DDcGAN outperforms other methods by generating high quality fused images with realistic textures and fine details. The use of a generative adversarial network ensures that the fused image not only looks natural but also incorporates intricate image features that may be lost in traditional approaches. The DDcGAN-based fusion method consistently outperformed both the Laplacian Pyramid and PCA techniques. Its superior PSNR values suggest that DDcGAN is highly effective in preserving the original signal quality, resulting in fused images with minimal noise and distortion. In addition, the higher entropy values with low noise achieved by DDcGAN imply enhanced feature representation and a greater retention of intricate details, confirming its capability to integrate diverse information from the source images more effectively than the traditional methods.

In contrast, while the Laplacian Pyramid and PCA methods demonstrated acceptable performance, their lower PSNR and entropy values indicate that they may not be as proficient in maintaining both image fidelity and detail richness. These results suggest that, although conventional fusion techniques remain viable, advanced deep learning approaches like DDcGAN offer significant improvements in generating high-quality fused images.

References

[1] Jiayi Ma, Han Xu, Junjun Jiang, Xiaoguang Mei, and Xiao-Ping Zhang." DDcGAN: A Dual-Discriminator Conditional Generative Adversarial Network for Multi-Resolution Image Fusion" *IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 29, 2020.*

[2] Han Xu, Pengwei Liang, Wei Yu, Junjun Jiang and Jiayi Ma. "Learning a Generative Model for Fusing Infrared and Visible Images via Conditional Generative Adversarial Network with Dual Discriminators." *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*

[3] X. Jin et al., "A survey of infrared and visual image fusion methods," Infr. Phys. Technol., vol. 85,pp.478–501,Sep.2017

[4] Huafeng Li , Yueliang Cen, Yu Liu , Xun Chen and Zhengtao Yu." Different Input Resolutions and Arbitrary Output Resolution: A Meta Learning-Based Deep Framework for Infrared and Visible Image Fusion." *IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, 2021*.

[5] Qi Jin, Sanqing Tan, Gui Zhang, Zhigao Yang, Yijun Wen, Huashun Xiao and Xin Wu. "Visible and Infrared Image Fusion of Forest Fire Scenes Based on Generative Adversarial Networks with Multi-Classification and Multi-Level Constraints."

[6] Xiangzeng Liu, Haojie Gao, Qiguang Miao 1, Yue Xi, Yunfeng Ai and Dingguo Gao." MFST: Multi-Modal Feature Self-Adaptive Transformer for Infrared and Visible Image Fusion."

[7] Qunming Wang, Wenzhong Shi, and Peter M. Atkinson. "Information Loss-Guided Multi-Resolution Image Fusion". *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 58, NO. 1, JANUARY 2020.*

[8] Rekha R. Nair, Tripty Singh. "Multi-sensor medical image fusion using pyramid-based DWT: a multi-resolution approach".

[9] R. Nanmaran, S. Srimathi, G. Yamuna, S. Thanigaivel, A. S. Vickram, A. K. Priya, Alagar Karthick, J. Karpagam, V. Mohanavel,and M. Muhibbullah. Investigating the Role of Image Fusion in Brain Tumor Classification Models Based on Machine Learning Algorithm for Personalized Medicine

[10] W. Liao, X. Huang, F. Coillie, S. Gautama, A. Pizurica, W. Philips, H. Liu, T. Zhu, M. Shimoni, G. Moser, and D. Tuia, "Processing of Multiresolution Thermal Hyperspectral and Digital Color Data: Outcome of the 2014 IEEE GRSS Data Fusion Contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 2984-2996, June 2015*.

[11] Harpreet Kaur, Deepika Koundal, Virender Kadyan. "Image Fusion Techniques: A Survey". *Archives of Computational Methods in Engineering (2021) 28:4425–4447*

[12] Hao Zhang, Han Xu, Xin Tian, Junjun Jiang, Jiayi Ma. "Image fusion meets deep learning: A survey and perspective". *Information Fusion 76 (2021) 323-336*

[13] Hui Li, Xiao-Jun Wu. "DenseFuse: A Fusion Approach to Infrared and Visible Images", *IEEE TRANSACTIONS ON IMAGE PROCESSING, 2018*

[14] Jiao Du, Weisheng Li, Ke Lu and Bin Xiao, An Overview of Multi-Modal Medical Image Fusion, Neurocomputing, *http://dx.doi.org/10.1016/j.neucom.2015.07.160*

[15] Hui Li, Xiao-Jun Wu, and Josef Kittler. "MDLatLRR: A novel decomposition method for infrared and visible image fusion", *IEEE TRANSACTIONS ON IMAGE PROCESSING, 2020*