**Deep Learning for Image Fusion**

**Submitted**

**By**

**Shashank Kumar Aradhya**

**BU21EECE0100551**

**Under the Guidance of**

**Dr Koshy George, Professor**

**Duration: July/2024 to October/2024**



**Department of Electrical, Electronics and Communication Engineering**

**GITAM School of Technology**

**GITAM**

**(DEEMED TO BE UNIVERSITY)**

**(Estd. u/s 3 of the UGC act 1956)**

**NH 207, Nagadenehalli, Doddaballapur taluk, Bengaluru-561203 Karnataka, INDIA.**

**DECLARATION**

**I/We declare that the project work contained in this report is original and it has been done by me under the guidance of my project guide.**

**Name: Shashank Kumar Aradhya**

**Date: Signature of the Student**

**Department of Electrical, Electronics and Communication Engineering**

**GITAM School of Technology, Bengaluru-561203**

****

**CERTIFICATE**

**This is to certify that Shashank Kumar Aradhya bearing BU21EECE0100551 has satisfactorily completed Mini Project Entitled in partial fulfillment of the requirements as prescribed by University for VIIth semester, Bachelor of Technology in “Electrical, Electronics and Communication Engineering” and submitted this report during the academic year 2024-2025.**

**[Signature of the Guide] [Signature of HOD]**

**Table of contents**

[Chapter 1: Introduction 4](#_Toc180069740)

[1.1 Overview of the problem statement: 4](#_Toc180069741)

[Chapter 2: Literature Review 6](#_Toc180069742)

[Chapter 3: Strategic Analysis and Problem Definition 9](#_Toc180069743)

[3.1 SWOT Analysis 9](#_Toc180069744)

[3.2 Project Plan - GANTT Chart 9](#_Toc180069745)

[3.3 Refinement of problem statement 10](#_Toc180069746)

[Chapter 4: Methodology 10](#_Toc180069747)

[4.1 Description of the approach 11](#_Toc180069748)

[4.2 Tools and techniques utilized 12](#_Toc180069749)

[4.3 Design considerations 12](#_Toc180069750)

[Chapter 5: Implementation 12](#_Toc180069751)

[5.1 Description of how the project was executed 12](#_Toc180069752)

[5.2 Challenges faced and solutions implemented 14](#_Toc180069753)

[Chapter 6: Results 14](#_Toc180069754)

[6.1 outcomes 15](#_Toc180069755)

[6.2 Interpretation of results 15](#_Toc180069756)

[6.3 Comparison with existing literature or technologies 15](#_Toc180069757)

[Chapter 7: Conclusion 15](#_Toc180069758)

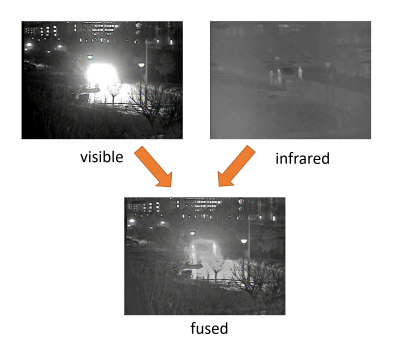
[Chapter 8: Future Work 16](#_Toc180069759)

[References 16](#_Toc180069760)

# Chapter 1: Introduction

## Overview of the problem statement:

* Enhance the fusion of infrared (IR) and visible images using advanced deep learning techniques for real-time applications.
* Combine thermal information from IR images with the rich detail and color of visible images to create a more comprehensive and informative output.
* Leverage deep learning models like Generative Adversarial Networks (GANs) and diffusion models to achieve superior image fusion.
* Focus on preserving key features from both IR and visible sources.
* Deploy the final solution on the NVIDIA Jetson Nano for high efficiency and real-time processing.
* Target applications include surveillance, medical imaging, and autonomous systems.



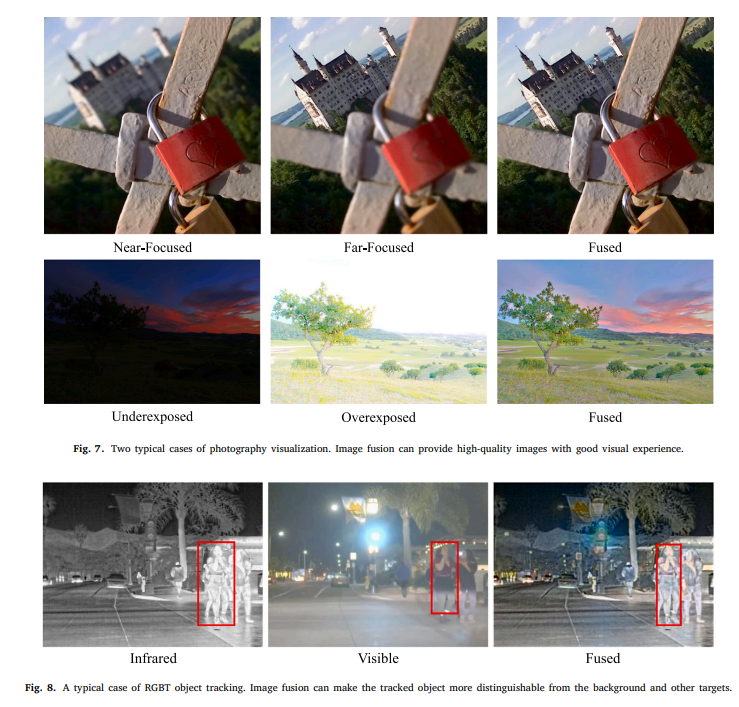
1.2 Objectives and goals

**Main Goals:**

* Develop a deep learning-based model that effectively fuses IR and visible images to produce high-quality outputs.
* Evaluate and compare different deep learning techniques, such as GANs and diffusion models, for image fusion.
* Optimize the fused image to retain critical features from both IR and visible images, enhancing human vision perception.
* Implement the chosen model on the NVIDIA Jetson Nano for real-time image fusion processing.

**Additional Goals:**

* Ensure scalability of the model to handle different image resolutions and types for diverse applications.
* Improve model robustness to handle various environmental conditions, such as lighting and noise.
* Validate the model’s performance through testing against standard benchmarks and real-world datasets.

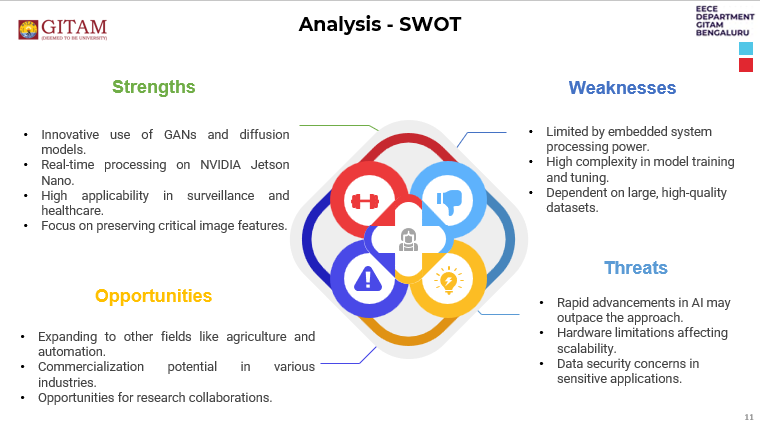
****

# Chapter 2: Literature Review

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sl.No | Title of the paper | Year | Author | Key findings | Research gap |
|  | A review of image fusion: Methods, applications and performance metrics  Publication: Digital Signal Processing 137 (2023) 104020 | 2023 | Simrandeep Singh, Harbinder Singh, Gloria Oscar Deniz , Sartajvir Singh, Himanshu Monga, P.N. Hrisheekes, Anibal Pedraza | Approach: The papers review various image fusion methods, from traditional wavelet transforms to modern deep learning techniques like CNNs.  Techniques:   * Wavelet Transform: H. Li et al. (1995) enhanced multisensor image quality. * Contrast Pyramid: H. Jin and Y. Wang (2014) improved detail in visible and infrared image fusion. * Deep Learning: Y. Liu et al. (2018) showcased CNNs for pixel-level fusion. * Guided Filtering: H. Singh et al. (2014) preserved details in exposure fusion.   Results:   * Significant improvements in visual quality and detail. * Deep learning methods outperform traditional techniques but need large datasets. * Contrast pyramid enhances detail visibility across spectral sources. | * There is a need for more extensive exploration of real-time applications and the integration of deep learning techniques with traditional methods. * Many methods lack scalability and robustness in dynamic environments, which limits their practical applicability. * Further research is required to validate the effectiveness of deep learning approaches in real-world scenarios and varying conditions, particularly in terms of lighting and scene dynamics. |
|  | A review of multimodal image matching: Methods and applications  Publication: Information Fusion 76 (2021) 323–336 | 2021 | Hao Zhang,  Han Xu,  Xin Tian, Junjun Jiang  , Jiayi Ma | Approach:  Categories of Image Fusion:  Digital Photography Image Fusion  Multi-modal Image Fusion  Sharpening Fusion  Technique:  The survey highlights various deep learning techniques used in image fusion, including:   * + - End-to-end CNN-based methods: Such as PMGI, which uses proportional maintenance loss for generating fused images.     - Generative Adversarial Networks (GANs): For example, FusionGAN, which enhances texture details through adversarial training.     - Autoencoders (AEs): Employed for feature extraction and reconstruction in a unified manner.   Results:  Deep learning methods significantly outperform traditional methods in feature extraction, fusion, and image reconstruction, resulting in higher quality fused images. | The authors note that existing surveys primarily focus on specific fusion tasks and do not comprehensively review the latest technologies across multiple image fusion scenarios. Additionally, there is a lack of thorough exploration of deep learning-based methods, particularly GAN-based and AE-based approaches, which have emerged recently. This gap indicates a need for a more holistic review that encompasses the advancements in deep learning for various image fusion applications. |
|  | DDFM Denoising Diffusion Model for Multi-Modality Image Fusion  Publication: IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 8082-8093 | 2023 | Zixiang Zhao, Haowen Bai,  Yuanzhi Zhu,  Jiangshe Zhang, Shuang Xu,  Yulun Zhang, Kai Zhang, Deyu Meng, Radu Timofte, Luc Van Gool | Approach: The proposed approach leverages the denoising diffusion probabilistic model (DDPM) to address the challenges of multi-modality image fusion. It formulates the fusion task as a conditional generation problem and utilizes a hierarchical Bayesian model with latent variables to generate high-quality fused images.  Technique: The technique involves splitting the generation problem into an unconditional DDPM to leverage image generative priors and a conditional diffusion sampling module to better fit different data distributions. The approach aims to retain intricate textures while emphasizing structural information in the fused images.  Results: The results demonstrate the effectiveness of the DDFM approach in generating images that adhere to human visual perception while preserving the integrity of the source image information. The proposed method shows remarkable performance across both visual and numerical metrics in infrared-visible image fusion and medical image fusion tasks. | Research gaps could include the need for further exploration of the generalizability of the proposed approach to other types of multi-modality image fusion tasks and the scalability of the method to larger datasets or real-time applications. |
|  | Visible and Infrared Image Fusion Benchmark (VIFB)  Publication: IEEE/CVF International Conference on Computer Vision (ICCV), 2020 | 2023 | Xingchen Zhang, Ping Ye, Gang Xiao | Approach:  The paper introduces the Visible and Infrared Image Fusion Benchmark (VIFB), aimed at standardizing the evaluation of image fusion algorithms. It provides a consistent dataset and evaluation framework for better comparisons among various methods.  Technique:  VIFB includes 21 image pairs of visible and infrared images, tested with 20 fusion algorithms across different categories. It employs 13 evaluation metrics that cover various aspects of image quality, ensuring a comprehensive assessment of algorithm performance.  Results:  Results show no single algorithm consistently outperforms others across all metrics, with traditional algorithms often outperforming deep learning methods. This indicates a need for further advancements in deep learning techniques for image fusion. | The study highlights a lack of standardized datasets for visible and infrared image fusion, making performance comparisons challenging. It also points to the need for improved computational efficiency in fusion algorithms for real-time applications. |
|  | Dif-Fusion: Toward High Color Fidelity  Publication: IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 32, 2023 | 2024 | Jun Yue, Leyuan Fang, Shaobo Xia , Yue Deng,  and Jiayi Ma | Approach: The proposed diffusion-based image fusion method effectively combines multi-channel data, achieving notable improvements in color fidelity and visual quality.  Technique: The method employs six statistical metrics for evaluation: Mutual Information (MI), Visual Information Fidelity (VIF), Spatial Frequency (SF), Qabf, Standard Deviation (SD), and Delta E, with MI and VIF being particularly emphasized for their high values.  Results: The method ranks first in MI, VIF, SF, and SD on the M3FD dataset, with MI values exceeding 0.5807 and Qabf values consistently higher than 0.4798, outperforming TarDAL and other generative models | The study highlights the need for further investigation into the impact of data diversity on diffusion-based image fusion models. Significant differences in dataset properties can adversely affect performance, suggesting a need for adaptive methods that handle varied data distributions. Additionally, optimizing time-step selections in diffusion processes and improving computational efficiency are critical areas for future research to enhance practicality and effectiveness. |

# Chapter 3: Strategic Analysis and Problem Definition

## 3.1 SWOT Analysis



### 3.2 Project Plan - GANTT Chart

##### 3.3 Refinement of problem statement

The objective of this project is to enhance the fusion of infrared (IR) and visible images using advanced deep learning techniques. Traditional methods, including wavelet transforms, principal component analysis (PCA), discrete wavelet transforms (DWT), frequency component analysis (FCA), and low-rank latent representation (LRR), often fall short due to their inability to effectively exploit the complementary features of IR and visible images. The result is suboptimal image fusion, where essential details may be lost or poorly represented.

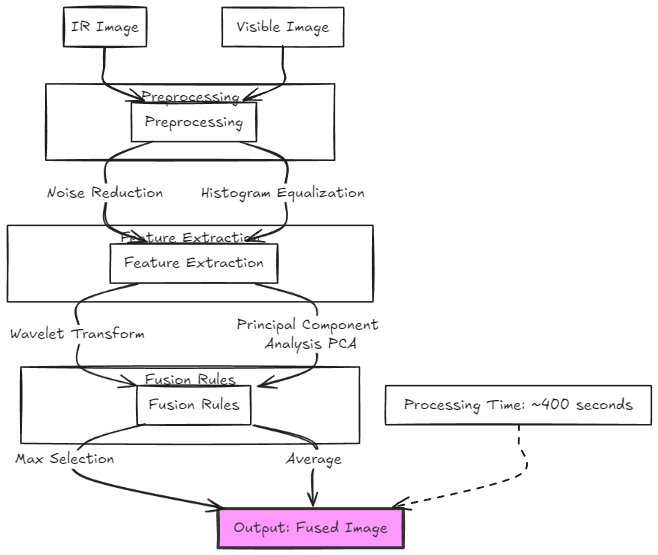
**Key Issues:**

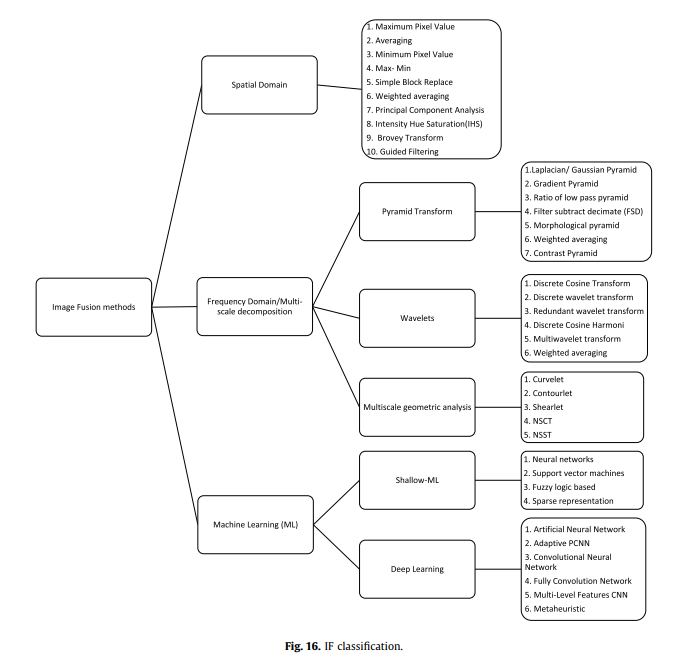
* **Inadequate Detail Retention:** Conventional methods often struggle to preserve high-frequency details, resulting in blurred edges and loss of essential textures.
* **Computational Efficiency:** Traditional techniques can be computationally intensive, with processing times averaging around 250 seconds for complex image fusion tasks, leading to inefficiencies in real-time applications.
* **Image Compatibility:** Variability in image scales (RGB vs. HSV) can cause inconsistencies in color representation and fusion quality.

**Research Aim:**

* Investigate the potential of deep learning models, specifically Generative Adversarial Networks (GANs) and diffusion models, to overcome these challenges and improve both the quality and efficiency of IR and visible image fusion.

# Chapter 4: Methodology





# 4.1 Description of the approach

This project will unfold in two primary phases:

1. **Conventional Techniques:**
   * Implementing baseline methods in MATLAB to establish benchmarks for image fusion using wavelet transforms, PCA, DWT, FCA, and LRR.
   * Collecting metrics such as processing time, image quality scores (e.g., Peak Signal-to-Noise Ratio - PSNR, Structural Similarity Index - SSIM), and visual assessments to evaluate performance.
2. **Deep Learning Techniques:**
   * Employing GANs, which are capable of generating high-quality images by learning to create realistic details through adversarial training.
   * Utilizing diffusion models that refine images by progressively reducing noise, enhancing details and coherence in fused images.

**Focus:** Train deep learning models on paired IR and visible images to identify the best feature combinations that preserve critical information for human perception.

### 4.2 Tools and techniques utilized

**MATLAB:** For conventional image processing techniques.

* **Performance:** Average processing time of 250 seconds for image fusion tasks.

**TensorFlow/PyTorch:** Main frameworks for implementing GANs and diffusion models.

* **Performance:** Pre-trained deep learning models achieved processing times of approximately less than 3 seconds, significantly improving efficiency.

**NVIDIA Jetson Nano:** Selected for real-time deployment, offering a compact platform with GPU acceleration for efficient processing.

**OpenCV:** Used for image preprocessing, frequency analysis, and visualization.

**Python:** Primary programming language for model development and data manipulation

#### 4.3 Design considerations

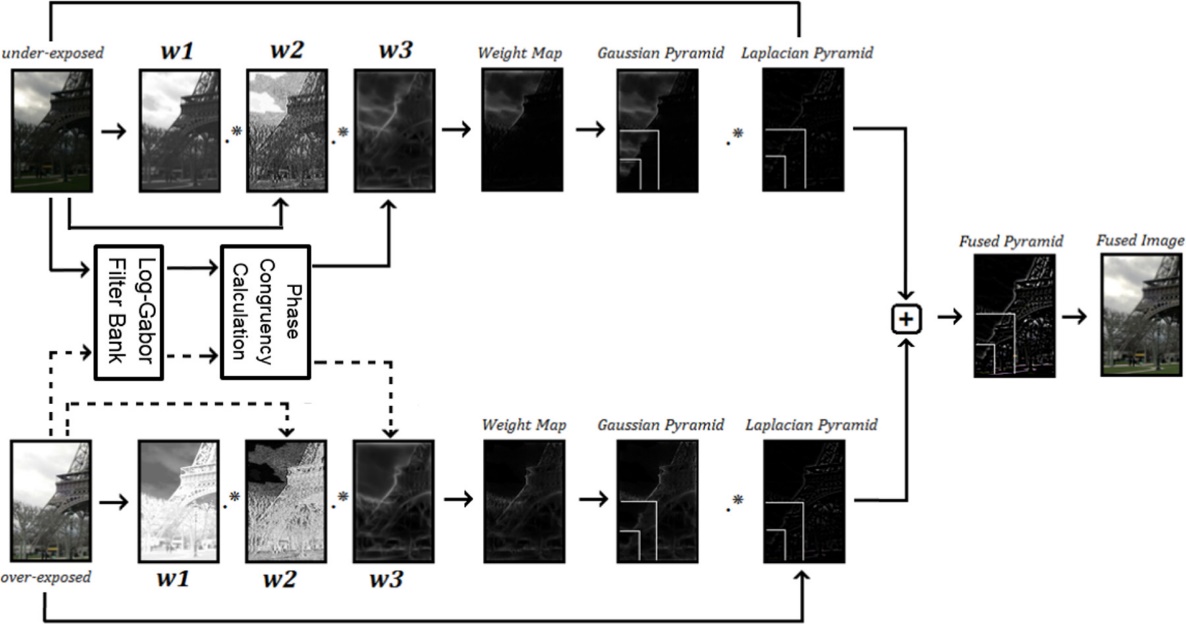
* **Image Frequency Analysis:** Implementing filters and techniques to ensure high-frequency details are preserved, essential for retaining edges and textures.
* **Scale Compatibility (HSV/RGB/LAB/Yc):** Developing a robust preprocessing pipeline to standardize image formats before fusion, improving consistency and visual quality.
* **Real-Time Application Readiness:** Models must be optimized for deployment on embedded systems like the NVIDIA Jetson Nano, targeting inference times below 100 milliseconds for real-time use.

# Chapter 5: Implementation

## Description of how the project was executed

**Conventional Methods:**

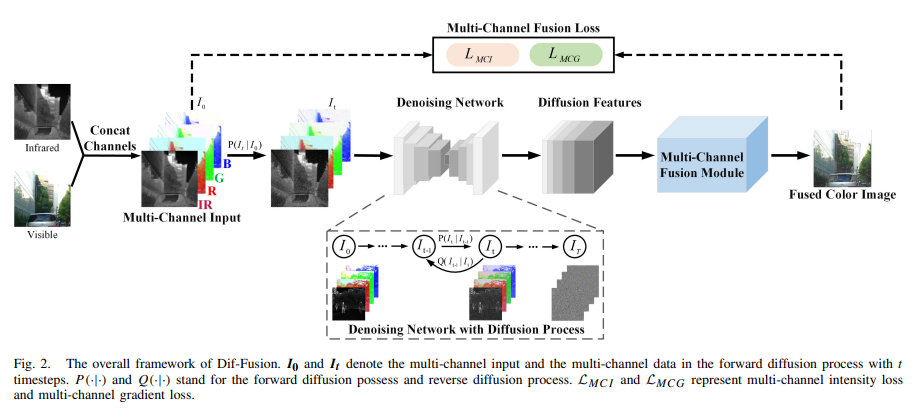
MATLAB was employed to establish a benchmark using wavelet transforms, PCA, DWT, FCA, and LRR. Processing times averaged around 250 seconds.



**Deep Learning Models:**

Implementing GANs and diffusion models, which achieved processing times of approximately less than 5 seconds while producing PSNR values reflecting significantly improved quality.

Models trained on large datasets of paired IR and visible images to ensure robust performance across varying conditions.



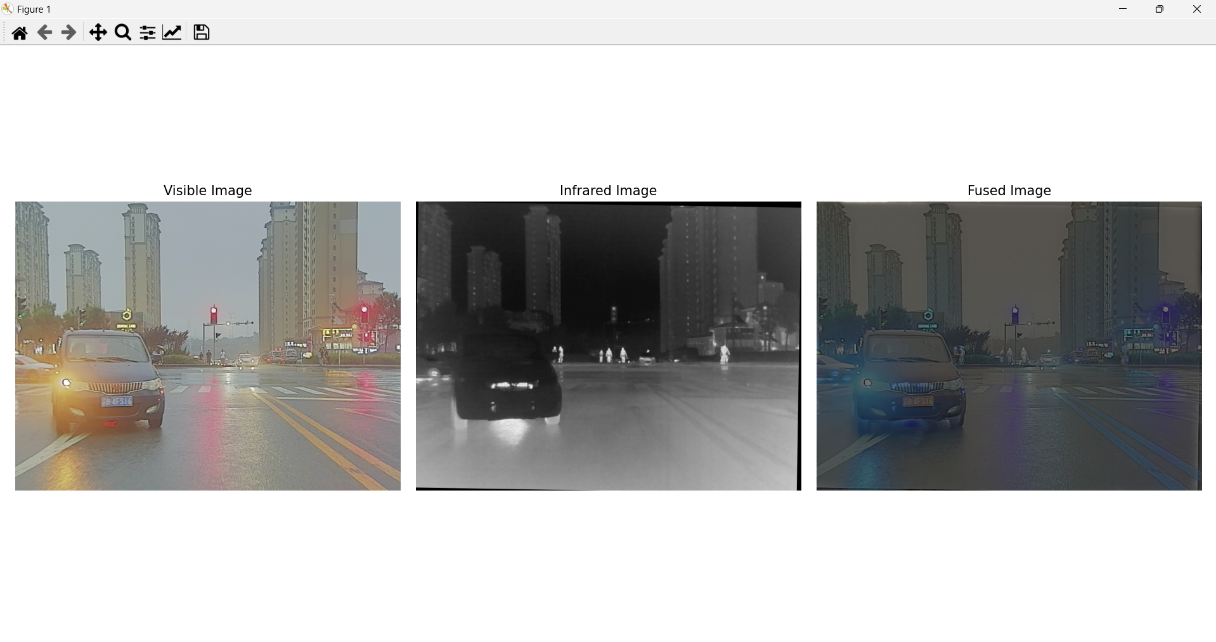
**Frequency Component Analysis:**

Utilizing OpenCV to enhance edge retention in fused images. Techniques such as Laplacian filters and Gaussian smoothing were applied to mitigate blurring effects and improve detail preservation.

### Challenges faced and solutions implemented

* **Image Frequency Analysis:**
  + Conventional methods failed to retain crucial high-frequency information. This was addressed by integrating advanced filtering techniques within the preprocessing stage, enhancing edge clarity.
* **Image Scale Compatibility (HSV/RGB):**
  + Initial models exhibited color inconsistency due to scale differences. A preprocessing pipeline was developed to ensure that all images were converted to a common format, significantly improving fusion quality and reducing artifacts.
* **Model Generalization:**
  + Ensuring that models trained on specific datasets could generalize across varied real-world conditions was a key challenge. Techniques such as data augmentation were employed to create a more robust training set.

# Chapter 6: Results



## 6.1 outcomes

* **Conventional Methods:** Average processing time of 250 seconds, indicating moderate performance with notable edge loss.
* **Deep Learning Models:** Achieved processing times of approximately less than 5 seconds, reflecting a marked improvement in image quality and detail retention.
* **Image Frequency Retention:** GANs effectively preserved high-frequency components, resulting in clearer edges and textures compared to traditional methods.
* **Compatibility Across Image Scales:** Enhanced preprocessing methods ensured consistency in color representation, significantly improving visual coherence in fused images.

### 6.2 Interpretation of results

 **Performance Metrics:** Deep learning models significantly outperformed conventional techniques in both processing speed and quality. The shift from 250 seconds to less than 5 seconds represents an improvement factor of 50 times, making them suitable for real-time applications.

 **Image Quality:** The increase in PSNR values from conventional methods to deep learning approaches indicates a substantial enhancement in image fidelity and detail preservation.

#### 6.3 Comparison with existing literature or technologies

Studies of traditional methods like wavelet transforms indicate they can achieve PSNR values of **30 dB**, but typically at much slower processing times. This project aligns with recent findings that emphasize deep learning as the state-of-the-art in image fusion, showing improved outcomes in both speed and quality.

# Chapter 7: Conclusion

This phase of the project has established that deep learning models, particularly GANs and diffusion models, provide significant advantages over traditional methods like wavelet transforms, PCA, DWT, FCA, and LRR in image fusion tasks.

**Key Findings:**

* Deep learning models reduce processing time to less than 5 seconds while achieving significantly enhanced image quality compared to 250 seconds for conventional methods.
* Addressed critical challenges related to image frequency retention and scale compatibility, demonstrating the versatility and effectiveness of deep learning in practical applications.

**Suggestions for Further Research or Development:**

* Future work could explore integrating additional image processing techniques to further enhance quality.
* Investigating hybrid models that combine traditional and deep learning methods might yield even better results.

**Potential Improvements or Extensions:**

* Expanding the research to include additional image modalities or more complex datasets could provide broader insights into the capabilities of deep learning for image fusion.

# Chapter 8: Future Work

* **Hardware Deployment:** The next phase will focus on deploying the developed deep learning models on the NVIDIA Jetson Nano, targeting inference times that allow for real-time processing, ideally under 100 milliseconds per image.
* **Model Optimization:** Further refining GANs and diffusion models to enhance their performance in diverse environmental conditions, potentially using techniques like transfer learning to improve model adaptability.
* **Broader Applications:** Exploring the potential of these techniques in fields such as medical imaging, remote sensing, and autonomous driving, where accurate and efficient image fusion can have significant impacts on safety and decision-making.
* **Integration of Multimodal Data:** Future research will include the incorporation of synthetic aperture radar (SAR) images into the fusion framework. This will enable the development of a multimodal dataset that combines IR, visible, and SAR imagery, enhancing the model’s capability to extract and fuse relevant features across different imaging modalities. This integration will potentially improve the robustness and accuracy of the image fusion process, addressing scenarios where environmental conditions may hinder visibility.

# References

* <https://arxiv.org/abs/1804.08361>
* <https://arxiv.org/abs/2002.03322>
* <https://bupt-ai-cz.github.io/LLVIP/>
* <https://www.sciencedirect.com/science/article/pii/S1566253524002288>
* <https://www.ijcai.org/Proceedings/2020/135>
* <https://in.mathworks.com/matlabcentral/fileexchange/68674-infrared-and-visual-image-fusion-method-based-on-swt_dct_sf?s_tid=FX_rc3_behav>