IMAGE FUSION EMERGING APPLICATIONS AND TECHNIQUES

Report submitted to GITAM (Deemed to be University) as a partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Electronics and Communication Engineering

Haripriya Kamsala
Regd. No: BU22EECE0100518
Sadiya Samrin K
Regd. No: BU22EECE0100317

GITAM
DEEMED TO BE UNIVERSITY

DEPARTMENT OF ELECTRICAL, ELECTRONICS AND COMMUNICATION ENGINEERING
GITAM SCHOOL OF CORE TECHNOLOGY
GITAM (DEEMED TO BE UNIVERSITY)
BENGALURU -561203
NOV 2025

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: 1. Haripriya Kamsala
2. Sadiya Samrin K

Date:26/11/2025 Signature of the Student

Department of Electrical, Electronics and Communication Engineering GITAM School of Technology, Bengaluru-561203

CERTIFICATE

This is to certify that Haripriya Kamsala bearing Regd. No.: BU22EECE0100518 and Sadiya Samrin bearing Regd. No.: BU22EECE0100317 has satisfactorily completed Mini Project Entitled in partial fulfillment of the requirements as prescribed by University for VII semester, Bachelor of Technology in “Electrical, Electronics and Communication Engineering” and submitted this report during the academic year 2025-2026.

[Signature of the Guide] [Signature of HOD]

Table of contents

Chapter 1: Introduction

1.1 Overview of the problem statement

In the modern digital era, the ability to capture and process high-quality images plays a vital role across diverse domains such as healthcare, surveillance, autonomous navigation, and remote sensing. However, a single imaging modality often fails to provide complete information about a scene due to limitations such as shallow depth of field, varying illumination, or restricted spectral sensitivity.

Image fusion has emerged as a solution to this challenge by combining information from multiple input images into a single enhanced image. Three important applications of image fusion are **multi-focus fusion**, **multi-exposure fusion**, and **thermal-RGB fusion**. Multi-focus fusion addresses the problem of shallow depth of field, producing an all-in-focus image. Multi-exposure fusion combines differently exposed images to preserve details in both shadows and highlights. Thermal-RGB fusion merges temperature information with visual details, which is particularly useful in low-light environments, medical imaging, and surveillance.

The lack of a unified framework for handling these fusion tasks and the need for systematic comparison of different algorithms form the core problem addressed in this project. The study aims to evaluate and optimize multiple fusion approaches, identifying their suitability for different application domains.

1.2 Objectives and Goals

The primary objectives of this project are as follows:

1. To study and analyze existing image fusion techniques across pixel, spatial, and transform domains.
2. To implement traditional approaches for multi-focus, multi-exposure, and thermal-RGB image fusion with suitable preprocessing and postprocessing methods.
3. To compare the performance of different fusion algorithms using qualitative and quantitative metrics.
4. To explore machine learning and deep learning-based methods for improved fusion performance.
5. To identify the most effective approaches for each application scenario and propose an optimized framework for image fusion.
6. To evaluate the potential real-world applications of the developed techniques in areas such as autonomous navigation computer vision, and security.

Through these goals, the project aims to contribute to the development of robust and adaptive image fusion techniques, offering insights into their practical deployment in emerging fields.

Chapter 2: Literature Review

Image fusion has been an active area of research for several decades, with applications expanding as imaging technologies advance. Early approaches relied on simple pixel-level techniques, while modern methods leverage advanced transform-based, machine learning, and deep learning models.

1. Image fusion — overview and foundational methods

Image fusion integrates complementary information from multiple imagery sources to form a composite image that is more informative for human interpretation or automated processing than any single input [1]. Foundational work framed fusion at multiple abstraction levels — pixel, feature, and symbol — each with different pre-processing and robustness requirements; pixel-level fusion preserves original information but is highly sensitive to misregistration, while feature- and symbol-level fusion afford more semantic integration at the cost of additional processing such as segmentation or classification [2]. Multiresolution transforms (pyramids, wavelets, contourlets, NSCT) became popular because they separate image content across scales, enasdsabling selective fusion of approximation and detail bands; statistical wavelet-domain modeling also enhanced robustness to noise and texture variations [3], [14]. Sparse representation (SR) approaches marked a shift to *data-driven* basis design: instead of fixed transform bases SR learns overcomplete dictionaries that better capture salient structures (edges, textures) of images and thereby often outperform fixed MSTs in subjective and objective metrics, albeit at higher computational cost and dictionary-dependence [5]. More recently, deep learning has produced end-to-end fusion frameworks capable of learning fusion policies and perceptual objectives directly, but this progress comes with new challenges: large annotated datasets are scarce for many fusion problems, learned models can be hard to interpret, and standardized evaluation remains incomplete [6], [9]. Across paradigms, two persistent practical needs are robust registration (especially for pixel-level fusion) and standardized metrics and benchmarks to allow fair comparison [2], [3], [13].

2. Multi-focus image fusion

Multi-focus fusion specifically addresses limited depth-of-field by merging images focused at different depth planes into an all-in-focus image. Piella (2003) offered a general multiresolution (MR) fusion framework from pixels to regions, formalizing how MR decomposition plus coefficient-level fusion yields coherent results and how region-based strategies can reduce pixel-wise artifacts [7]. Traditional transform-domain approaches (DWT, NSCT, curvelet) dominated because they preserve multiscale structure, though they can be sensitive to shift variance and misregistration; redundant transforms such as NSCT or SWT mitigate some of those sensitivities [8], [14]. To better capture local structure, sparse representation (SR) methods emerged: Zhang et al. (2018) reviewed SR-based multi-sensor fusion, detailing representation models, dictionary learning strategies (global/adaptive/coupled) and activity measures; SR often yields better edge and texture preservation than fixed MSTs and is more robust to small misalignments due to overlapping patches, but its effectiveness strongly depends on dictionary quality and it increases computational expense [5]. Feature-based innovations also advanced the field. Liu et al. (2015) demonstrated that **dense SIFT** descriptors provide a powerful local activity measure: using sliding windows, DSIFT produces reliable initial decision maps which are then refined via feature matching and local focus comparison, improving the handling of mis-registered pixels and object edges relative to many earlier methods [9]. Hybrid methods combine strengths: Aymaz & Köse (2019) applied **super-resolution** preprocessing to increase source image detail, then used Stationary Wavelet Transform (SWT) + PCA to fuse subbands; this hybrid reduces edge loss and spatial distortion and yields visually clearer results, at the cost of extra computation and sensitivity to interpolation artifacts if super-resolution is mismatched [10].

3. Multi-exposure image fusion (MEF)

Multi-exposure fusion aims to merge images taken at different exposures to produce HDR-like results while preserving both informative content and visual realism. Classical MEF pipelines follow decomposition → activity measurement (exposure, contrast, saturation) → fusion, but struggle to guarantee natural color and consistent appearance across scenes [11]. With deep learning, unsupervised and perceptual strategies emerged to handle the lack of ground truth: early unsupervised networks (e.g., DeepFuse variants) optimized MEF-specific non-reference losses such as MEF-SSIM to preserve structure, but often produced desaturated or pale color outputs because many pipelines fuse only luminance channels [12]. Han et al. (2022) addressed this by introducing DPE-MEF, a deep perceptual enhancement network composed of a **detail enhancement** module (to mine structural and informative content) and a **color enhancement** module (to learn color–brightness relationships), thus balancing informativeness and visual realism. DPE-MEF reported superior perceptual quality and competitive speed on modern GPUs, while highlighting the broader issues: limited real ground truth for HDR fusion, and reliance on synthetic or pseudo-ground truth for supervised training [13]. Other MEF works explored patch-wise structural decomposition, gradient-domain fusion to avoid ghosting, and GAN-based approaches (MEF-GAN) to improve realism; yet color fidelity and artifact suppression remain active challenges [11], [12], [13]

4. Thermal (IR) + RGB fusion

Fusing thermal and RGB imagery harnesses complementary strengths: thermal is robust under poor lighting and can reveal heat-based features (useful in inspection and surveillance), while RGB provides high spatial resolution and color detail. Robust cross-modal matching and registration are prerequisites: Jiang et al. (2021) surveyed multimodal image matching and emphasized the importance of descriptors and deep correspondence models for reliable alignment across modalities [1]. Practical sensor studies — e.g., Alexander & Lunderman's

FLIR One Pro reliability study (2021) and application reports on image-based monitoring of infrastructure — underscore that sensor characteristics (resolution, dynamic range, calibration) and environmental factors strongly affect fusion performance [14], [15]. In algorithmic development, RGBT tracking surveys (Feng & Su, 2024) and deep multimodal biomedical surveys (Duan et al., 2024) show that adaptive fusion strategies (attention/gating, three-stream networks, relation

propagation modules) often outperform naive concatenation: gating or attention mechanisms learn to weight modalities by reliability (e.g., thermal at night, RGB by day) and reduce cross-modal contamination [4], [5], [16]. Still, thermal-RGB fusion faces persistent challenges: modality gap in appearance/statistics, sensor misalignment (parallax, differing FOV), resolution mismatch, limited labeled datasets across real-world conditions, and computational constraints for real-time deployment. Domain-specific works (applications in cultural heritage inspection, structural nondestructive testing, and robotics) demonstrate gains but also reveal that calibration and pre-processing (denoising, radiometric correction, geometric alignment) are essential to practical success [18], [19].

Chapter 3: Strategic Analysis and Problem Definition

3.1 SWOT Analysis

A SWOT analysis was conducted to evaluate the strengths, weaknesses, opportunities, and threats associated with this project on image fusion.

Strengths:

- Ability to integrate complementary information for improved visualization.
- Wide applicability in healthcare, surveillance, autonomous vehicles, and remote sensing.
- Systematic study of three distinct applications (multi-focus, multi-exposure, thermal-RGB).
- Combination of traditional and modern (ML/DL) approaches allows for comprehensive evaluation.

Weaknesses:

- Image registration errors may degrade fusion quality.
- Traditional methods often fail to adapt to non-linear variations.
- Deep learning approaches require large training datasets and significant computational resources.
- Implementation complexity may increase with multi-modal datasets.

Opportunities:

- Growing demand for real-time, reliable fusion in AI-driven systems.
- Potential to extend framework into hardware for real-world deployment.
- Contribution to ongoing research in medical imaging, security, and autonomous navigation.
- Possibility of optimizing algorithms for cross-application adaptability.

Threats:

- Availability of suitable datasets for all three applications may be limited.
- Risk of overfitting when using deep learning models on small datasets.
- Competing technologies such as hyperspectral imaging may reduce dependency on fusion.
- High computational requirements may pose challenges for real-time applications.

3.2 Project Plan

The project is planned in phases, each focusing on progressive development. A GANTT chart is used to visualize task scheduling and dependencies.

Phase	Activity	Duration	Timeline
Phase1	Problem understanding & Literature survey	10 weeks	Week 1-10
Phase 2	Dataset collection & preprocessing methods	3 weeks	Week 2-5
Phase 3	Implementation of traditional fusion techniques (multi-focus, multi-exposure, thermal-RGB)	3 weeks	Week 6-8
Phase 4	Post-processing & performance evaluation (metrics, comparison with benchmarks)	5 weeks	Week 8-13
Phase 5	Implementation of ML/DL approaches	5 weeks	Week 9-13
Phase 6	Comparative analysis & unified framework design	2 weeks	Week 12-13
Phase 7	Documentation, final report & presentation	2 weeks	Week 14–15

3.3 Refinement of Problem Statement

Initial understanding of the problem identified image fusion as a critical tool for combining multiple input images to generate an output with richer information content. However, different application domains (multi-focus, multi-exposure, thermal-RGB) pose distinct challenges such as focus-level detection, illumination balancing, and multi-modal data alignment.

The refined problem statement is as follows:

"To design, implement, and evaluate a unified framework for image fusion that can adapt to multiple application domains — specifically multi-focus, multi-exposure, and thermal-RGB fusion — by systematically comparing traditional, machine learning, and deep learning approaches, while employing preprocessing and postprocessing techniques to enhance performance and robustness.

Chapter 4: Methodology

4.1 Description of the approach

The methodology for this project is structured to systematically implement, evaluate, and optimize image fusion techniques across multiple applications. The approach is divided into three main stages:

1. **Preprocessing:**
   - Image registration to align input images for multi-focus, multi-exposure, and thermal-RGB datasets.
   - Noise reduction using filters (e.g., Gaussian, median) to improve fusion quality.
   - Intensity normalization and contrast adjustment to handle illumination variations.
2. **Fusion Techniques:**

   Traditional Methods:
   - Multi-scale approaches such as Laplacian pyramid and wavelet transform.
   - Pixel-level and feature-level fusion using gradient, edge, and morphological operations.
   - Weighted averaging and principal component analysis (PCA)-based fusion.

   Machine Learning and Deep Learning Methods (future phases):
   - Supervised learning techniques (SVM, Random Forest) for feature selection and fusion decisions.
   - Convolutional Neural Networks (CNN) and advanced deep learning architectures for end-to-end fusion.
3. **Postprocessing and Evaluation:**
   - Contrast enhancement, sharpening, and color correction to improve visual quality.
   - Quantitative evaluation using metrics such as Feature Mutual Information (FMI), Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and entropy.
   - Comparative analysis to determine the best-performing method for each application domain.

4.1.1 Preprocessing Pipeline

- Resizing images to the same dimensions
- Grayscale conversion (if needed)
- Noise removal (Gaussian smoothing, median filtering)
- Histogram equalization or normalization

4.1.2 Fusion Algorithm

Multi-Focus Image Fusion

Description of the Algorithm

The algorithm fuses two or more partially focused images of the same scene into a single all-in-focus image using morphological focus detection, trimap generation, alpha matting, and weighted fusion. This approach ensures that only the sharp, in-focus regions from each source image are retained.

1. Input Handling

- Reads a set of multi-focus grayscale (or converted-to-grayscale) images of the same scene.
- Ensures all input images are of the same size (resized if necessary).

2. Focus Measure Calculation (Morphological Filtering)

- A focus detection measure is applied using morphological filters (e.g., gradient, Laplacian, or variance operators).
- Each pixel is assigned a sharpness score, indicating whether it belongs to a focused or defocused region.

3. Trimap Generation

- Based on the focus measure, a trimap is created with three regions:
- Foreground → definitely in-focus.
- Background → definitely out-of-focus.
- Unknown region → uncertain focus.

4. Alpha Estimation (Matting)

- Alpha values (weights between 0 and 1) are estimated for pixels in the unknown region using an image matting algorithm.
- This produces a smooth transition mask, reducing boundary artifacts between focused and unfocused regions.

5. Weighted Fusion

- Using the alpha matte, the in-focus regions from different images are blended together.
- Sharp areas are preserved from each input image, while blurred regions are suppressed.

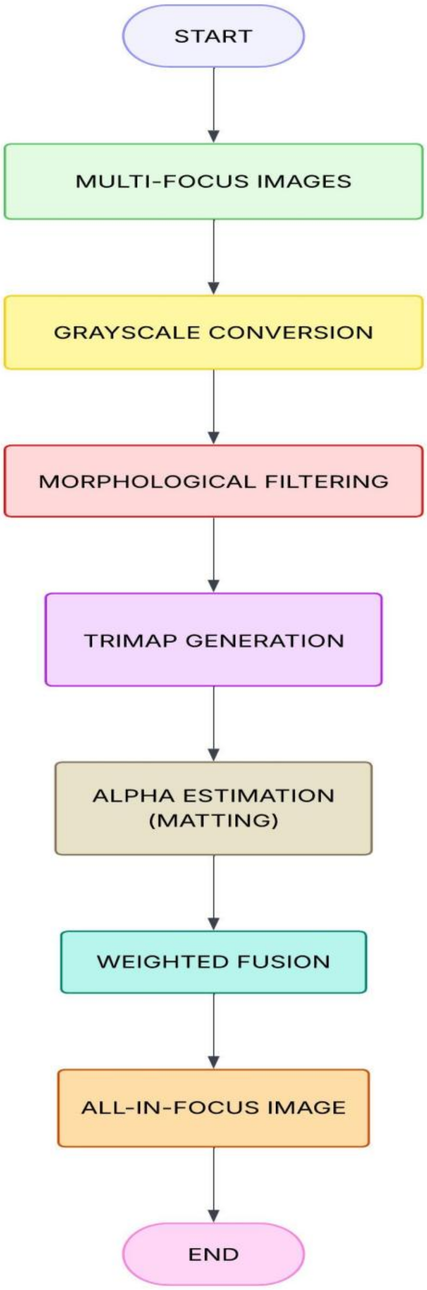
6. Final All-in-Focus Image Generation

- The fused image is normalized to 8-bit (uint8).
- The final all-in-focus image is saved with a suffix such as _fused.png.

Advantages

- Preserves sharp details from all source images.
- Eliminates blur and depth-of-field limitations of optical lenses.
- Reduces blocking artifacts by using trimap + alpha matting instead of hard decision maps.
- Smooth blending at focus boundaries improves visual quality.
- Effective for digital photography, optical microscopy, and surveillance applications.

Flowchart : -



Multi-Exposure Image Fusion:

Description of the Algorithm

The algorithm performs multi-exposure image fusion using guided filtering to combine multiple input images with different exposurSe levels into a single, visually balanced fused image.

Step-by-Step Workflow

1. Input Handling

Reads all the preprocessed input images (under-exposed, normally exposed, and over-exposed).
Ensures that all input images are the same size (resized if required).

Converts images into grayscale versions for computing weight maps, while still keeping the RGB channels for the final fusion.

2. Weight Map Generation

Three weight maps are computed for each input image:
Well-exposedness map → gives higher weight to pixels closer to mid-intensity, ensuring proper brightness.
Contrast map → highlights regions with stronger edges and texture.
Saturation map → emphasizes colorful and visually rich regions.
These maps are normalized and combined to generate the final weight map for each input image.

3. Multi-scale Decomposition

Each input image is decomposed into base layers (low-frequency structure) and detail layers (high-frequency texture).
Similarly, the weight maps are refined using guided filtering to avoid artifacts and halo effects.

4. Fusion Strategy

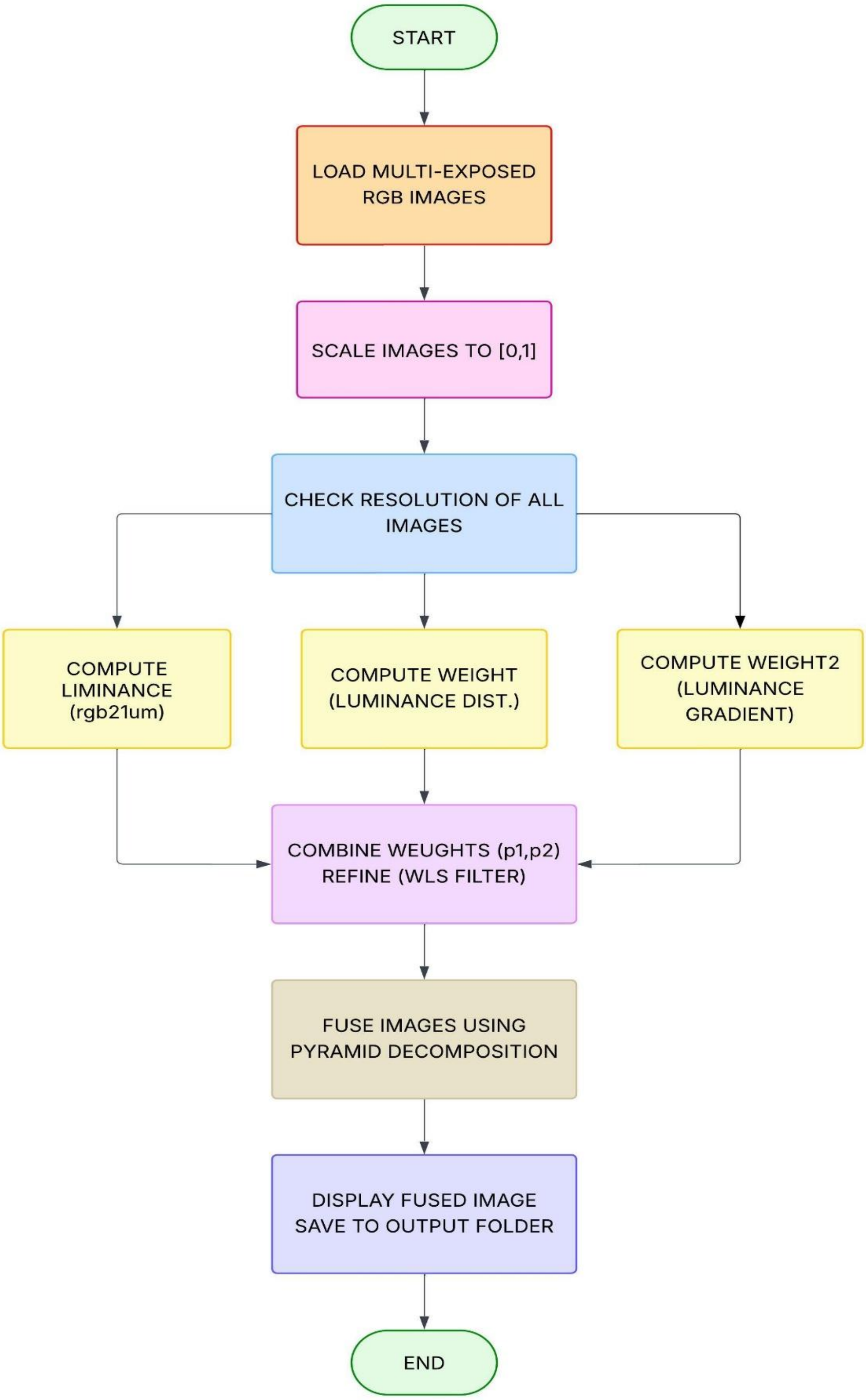
Base layers are fused by weighted averaging, ensuring a smooth and balanced illumination.
Detail layers are fused by taking maximum contributions from high-frequency details, preserving sharpness and texture.

5. Reconstruction

The fused base and fused detail layers are recombined to form the final fused image.
The output image has well-balanced brightness, rich colors, and preserved details from all input exposures.

6. Output Handling
The fused image is normalized and converted into a standard 8-bit format.
Finally, it is saved in the specified folder with a unique file name.

Flowchart;-


START
LOAD MULTI-EXPOSED RGB IMAGES
SCALE IMAGES TO [0,1]
CHECK RESOLUTION OF ALL IMAGES
COMPUTE LIMINANCE (rgb21um)
COMPUTE WEIGHT (LUMINANCE DIST.)
COMPUTE WEIGHT2 (LUMINANCE GRADIENT)
COMBINE WEUGHTS (p1,p2) REFINE (WLS FILTER)
FUSE IMAGES USING PYRAMID DECOMPOSITION
DISPLAY FUSED IMAGE SAVE TO OUTPUT FOLDER
END

Thermal and RGB Fusion

Description of the Algorithm

The algorithm fuses visible RGB images (for detailed textures and colors) and thermal infrared images (for heat signature and night vision) using **2D Discrete Wavelet Transform (DWT)**.

Step-by-Step Workflow

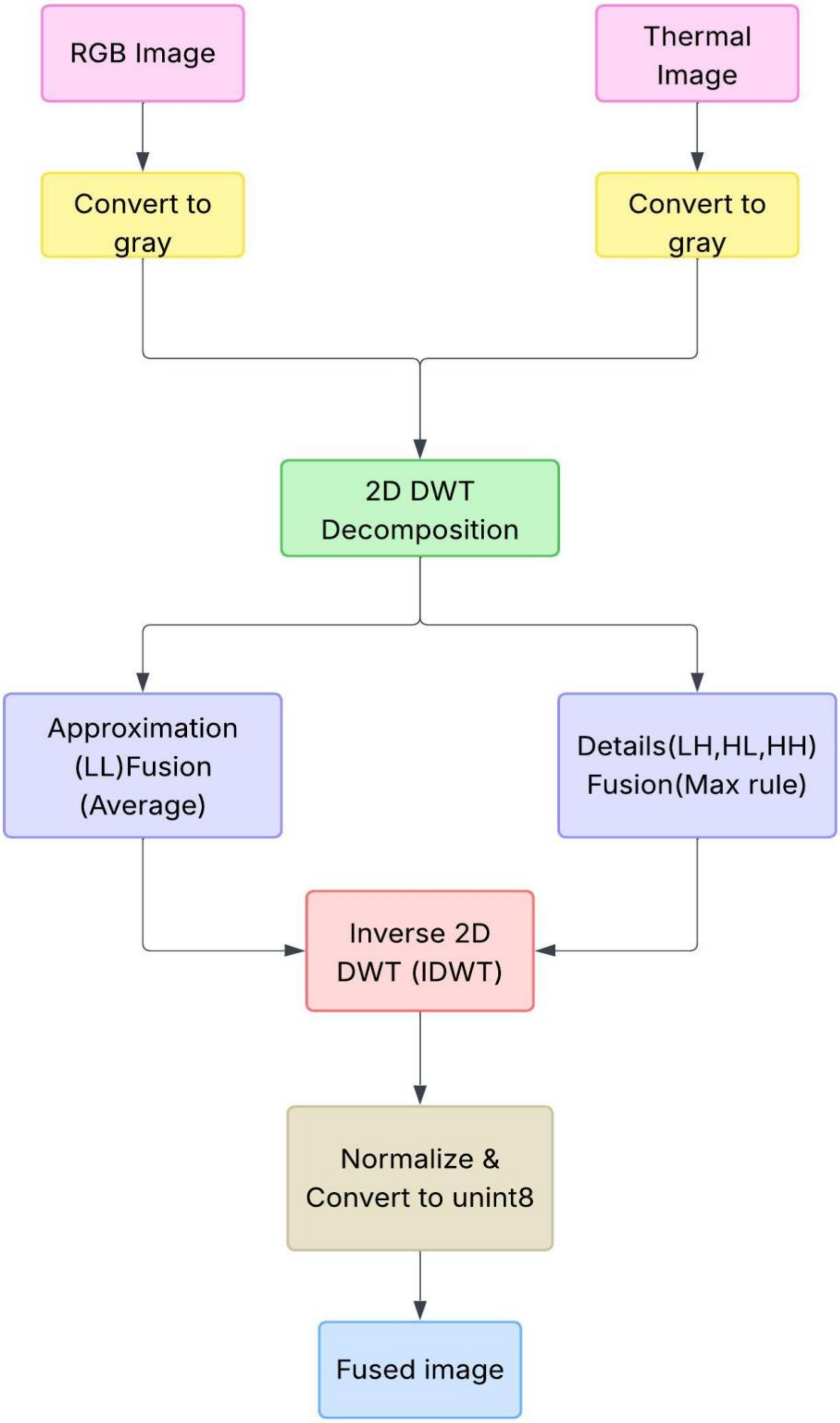
1. **Input Handling**
   - Reads all preprocessed RGB and their corresponding thermal images.
   - Ensures both are the same size (resized if needed).
   - Converts thermal images to grayscale if they are RGB.
2. **Wavelet Decomposition**
   - Both RGB (converted to grayscale) and thermal images are decomposed using **DWT with Daubechies-2 (db2)**.
   - This produces **four subbands** for each image:
     - Approximation coefficients (**LL**) → low-frequency (overall image structure).
     - Horizontal detail (**LH**).
     - Vertical detail (**HL**).
     - Diagonal detail (**HH**).
3. **Fusion Strategy**
   - **Approximation (LL)**: averaged between RGB & thermal ((a1+a2)/2) → balances structure.
   - **Detail coefficients (LH, HL, HH)**: maximum selection → keeps the strongest edge/feature.
4. **Inverse DWT**
   - The fused coefficients are recombined using **inverse DWT** to form the final fused image.
5. **Normalization & Saving**
   - The fused image is normalized to 8-bit (uint8).
   - The final fused image is saved with the suffix _fused.png.

Advantages

- **Preserves structural details** (edges) from both modalities.
- **Retains thermal information** (good for night/low-light).

Simple, computationally efficient, and effective for autonomous vehicle perception, surveillance, and medical imaging.

Flow Chart:


RGB Image
Thermal Image
Convert to gray
Convert to gray
2D DWT Decomposition
Approximation (LL)Fusion (Average)
Details(LH,HL,HH) Fusion(Max rule)
Inverse 2D DWT (IDWT)
Normalize & Convert to unint8
Fused image

:

Multi-Exposure Image Fusion:

Description of the Algorithm

The algorithm performs color multi-exposure image fusion using a deep learning-based End-to-End Multi-Exposure Fusion (E2EMEF) network.It combines multiple color images captured at different exposure levels (under, normal, and overexposed) into a single, high-quality fused image with balanced brightness, enhanced contrast, and preserved texture details.The model processes each color channel (R, G, and B) individually and then merges them to produce the final fused color image.

Step-by-Step Workflow

1. Input Handling

- The algorithm begins by **reading all input images** from the specified directory.
- Each image is loaded in **color (BGR)** format and then converted to **RGB** for consistent channel representation.
- The pixel values are **normalized to a [0,1] range** to prepare them for processing in the neural network.
- All images are verified to be of the same spatial dimensions before fusion.

2. Model Preparation

- The pre-trained **E2EMEF deep learning model** is loaded using **PyTorch**.
- The computation device is automatically selected based on system capability — either **GPU (CUDA)** or **CPU**.
- The model is set to **evaluation mode** (model.eval()) to ensure stable inference without training updates.

3. Channel-wise Fusion

- Since the fusion is performed per channel, the algorithm splits each image into its **R, G, and B components**.
  For each channel:
  - The stack of exposure images for that channel is created as a 3D array of shape **(N, H, W)**, where *N* is the number of exposures.
  - Each channel tensor is resized to a **low-resolution (128×128)** version using bilinear interpolation for coarse-level fusion.
  - Both **low-resolution** and **high-resolution** tensors are passed through the E2EMEF network to produce a fused channel output.

4. Fusion Strategy

- The E2EMEF network learns to merge input exposures in a data-driven manner, optimizing both illumination and detail retention.
- Unlike traditional weighted or guided-filter methods, this approach directly predicts the **fused output** that combines:
    - **Well-exposed regions** from each input,
    - **Sharp edges** and **fine details** from high-frequency content,
    - **Balanced global brightness** and color consistency across all channels.

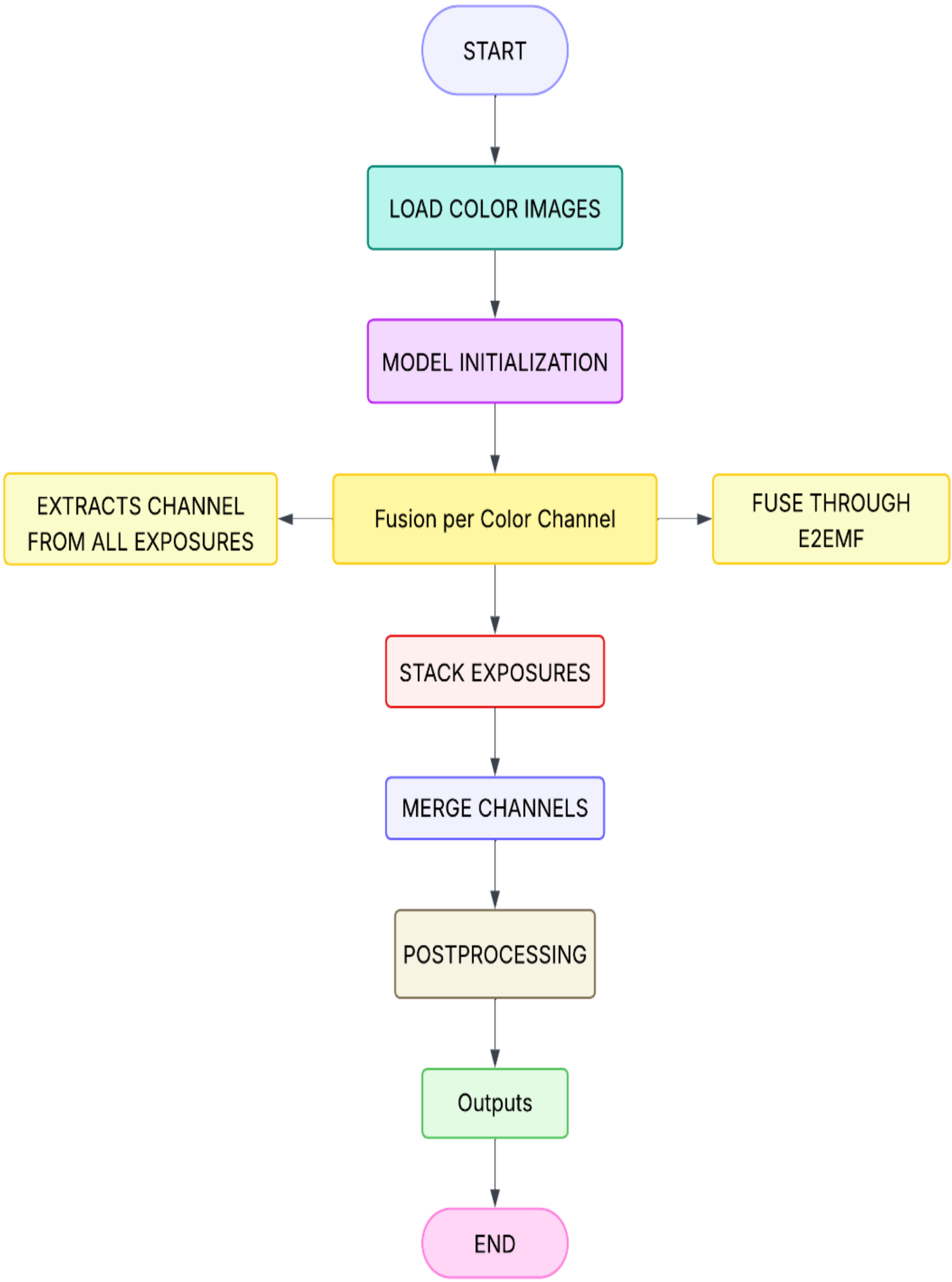
5. Reconstruction

- The fused outputs of the three channels (R, G, and B) are recombined using **channel stacking** to form the final RGB fused image.
- The fused image values are clipped to the valid [0,1] range, then scaled and converted back to **8-bit integer format (0–255)**.
- The final image is converted from **RGB** to **BGR** format for saving using OpenCV functions.

6. Output Handling

- The final fused color image is saved automatically in the **output directory** under the name fused_result_color.png.
- A message confirming successful image fusion and storage location is displayed in the console.
- The resulting fused image can also be **displayed in a window** for visualization using OpenCV (cv2.imshow).

Flowchart :-


START
LOAD COLOR IMAGES
MODEL INITIALIZATION
EXTRACTS CHANNEL FROM ALL EXPOSURES
Fusion per Color Channel
FUSE THROUGH E2EMF
STACK EXPOSURES
MERGE CHANNELS
POSTPROCESSING
Outputs
END

Multi-Exposure Image Fusion:

Description of the Algorithm

The algorithm performs multi-exposure image fusion using CNN-extracted feature maps to combine multiple images with different exposure levels into a single fused image. It utilizes deep feature-based weights to enhance important regions, balance illumination, and preserve color and texture details effectively.

Step-by-Step Workflow

1. **Input Handling**
   - The user selects a folder containing multiple input images with different exposure levels.
   - All images are loaded into a 4D array and displayed for visual verification.
   - Each input image is resized or adjusted to ensure uniform dimensions before fusion.

2. **CNN Feature Extraction**
   - The algorithm extracts deep features from each input image using a pretrained CNN model.
   - These CNN features highlight texture, contrast, and well-exposed regions across all input images.
   - The extracted feature maps are normalized so their values lie within a consistent range for accurate weighting.

3. **Weight Map Generation**
   - The CNN-derived feature maps are treated as weight maps, indicating the importance of each pixel.
   - These weights ensure that brighter, more detailed, and better-exposed regions contribute more to the final fused image.

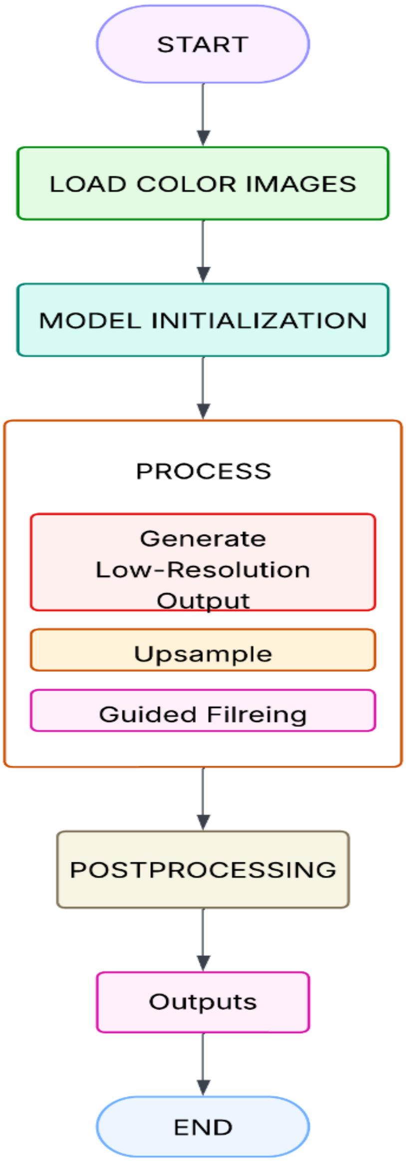
4. **Fusion Strategy**
   - Each RGB channel of the input images is multiplied by its corresponding CNN weight map.
   - The weighted channels are added together to form the proposed fused image.
   - This step combines structural information, color details, and brightness from all exposures.

5. **Postprocessing and Normalization**
   - The fused output is refined using postprocessing techniques to enhance smoothness and color balance.
   - The image is normalized with mat2gray() to maintain pixel values between 0 and 1.
   - This ensures proper visualization and consistent brightness across the entire image.

6. **Output Handling and Visualization**

   - The fused image is displayed with the title “Fused Image (CNN-based)”.
   - Histograms of all input images are plotted to show their intensity distributions.
   - The histogram of the fused image is displayed to verify exposure & contrast improvement.

Flowchart :-


START
LOAD COLOR IMAGES
MODEL INITIALIZATION
PROCESS
Generate
Low-Resolution
Output
Upsample
Guided Filreing
POSTPROCESSING
Outputs
END

Thermal + RGB

Description of the Algorithm

FusionGAN is a deep learning-based approach that uses a Conditional Generative Adversarial Network (cGAN) to fuse infrared (IR) and visible (VI) images. The generator network learns to create high-quality fused images that preserve thermal information from IR images and texture details from visible images, while the discriminator ensures the generated images are realistic and visually coherent.

Step-by-Step Workflow

1. Data Preparation

- **Input Processing**: Read IR and visible images in grayscale format
- **Patch Extraction**: Extract overlapping patches from training images (132×132 input size, 120×120 label size)
- **Normalization**: Convert pixel values from [0,255] to [-1,1] range
- **Dataset Creation**: Store patches in HDF5 format for efficient loading

2. Model Architecture

Generator (Fusion Network):

- Input: Concatenated IR and VI images (2 channels)
- Encoder-Decoder structure with 5 convolutional layers
- Uses Batch Normalization and Leaky ReLU activations
- Output: Single-channel fused image with tanh activation

Discriminator:

- 4 convolutional layers with stride 2 for downsampling
- Global Average Pooling followed by dense layer
- Distinguishes between real visible images and generated fused images

3. Training Process

- **Adversarial Training**: Generator tries to fool discriminator, discriminator learns to identify fakes
- **Content Loss**: L1 distance between fused image and both input modalities
- **Gradient Loss**: Preserves edge information using Laplacian gradients
- **Optimization**: Adam optimizer with learning rate 1e-4

4. Fusion Generation

- Concatenate preprocessed IR and VI images
- Feed through trained generator
- Convert output from [-1,1] back to [0,255] range
- Save the final fused image

Applications

- **Surveillance Systems**: Enhanced night vision and object detection
- **Military Applications**: Improved situational awareness
- **Medical Imaging**: Combining functional and structural information
- **Autonomous Vehicles**: Better perception in varying lighting conditions
- **Remote Sensing**: Multi-modal earth observation

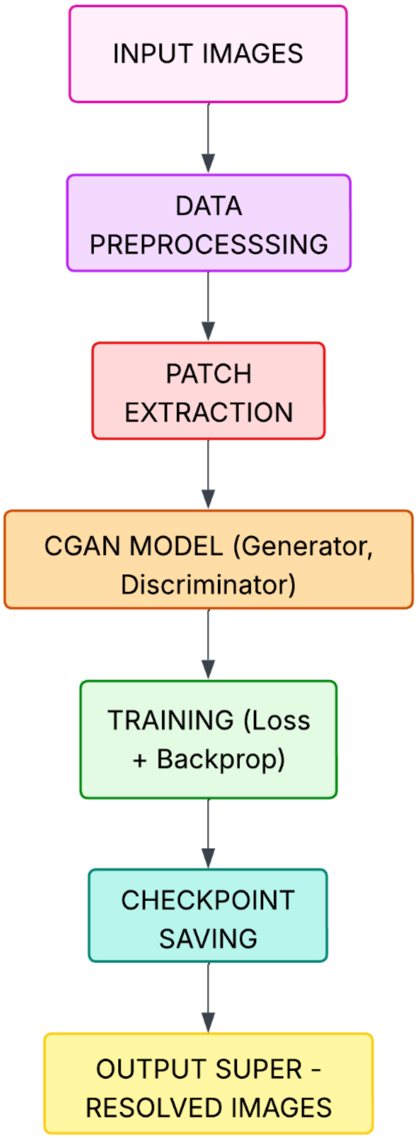
Advantages

- **Feature Preservation**: Maintains both thermal signatures and texture details
- **End-to-End Learning**: No manual feature engineering required
- **Adaptive Fusion**: Learns optimal fusion strategy from data
- **High Quality Output**: Produces visually appealing and informative results
- **Computational Efficiency**: Fast inference once trained
- **Robustness**: Handles various imaging conditions and modalities

The algorithm effectively bridges the gap between traditional image fusion methods and modern deep learning approaches, providing a robust solution for multi-modal image fusion tasks.

.

FLOWCHART:-


INPUT IMAGES
DATA PREPROCESSSING
PATCH EXTRACTION
CGAN MODEL (Generator, Discriminator)
TRAINING (Loss + Backprop)
CHECKPOINT SAVING
OUTPUT SUPER - RESOLVED IMAGES

Multi Focus Image Fusion :-

The proposed model is a deep learning–based image fusion framework that integrates two complementary input images—such as infrared–visible, RGB–thermal, multi-focus, or multi-exposure pairs—into a single fused output while preserving structural detail, contrast, and perceptual information from both sources. It uses an encoder–decoder generator with dense-style skip connections for rich feature extraction, a content-adaptive loss function combining SSIM and Frobenius terms, VGG-based perceptual weighting to emphasize structurally important pixels, and optional continual-learning support using Elastic Weight Consolidation (EWC) to avoid catastrophic forgetting when trained on multiple fusion tasks.

Step-by-Step Workflow

1. **Input Handling**
   - Loads paired images depending on the task (e.g, Multi-Focus → 2 or 3 RGB images).
   - For RGB inputs, converts to YCbCr and extracts the Y (luminance) channel for fusion; Cb/Cr are preserved for later reconstruction.
   - Normalizes intensities to [0,1] and reshapes to batch tensors `[B, H, W, 1]` (or `[B, H, W, 3]` if required for some paths).
2. **Preprocessing & Normalization**
   - Resizes copies of the sources to 224×224 and replicates single-channel images into 3 channels for VGG feature extraction.
   - Converts images to tensors and applies any dataset-specific normalization (if used).
3. **Generator Forward Pass (Encoder–Decoder)**
   - Concatenates the two input channels and passes them through the **Encoder** (initial conv + dense-like conv blocks with LeakyReLU and channel concatenation) to produce a compact code.
   - The **Decoder** applies a series of conv layers (128 → 64 → 32 → 1) with LeakyReLU activations; final layer uses `tanh` scaled to [0,1] to produce the fused luminance output.
4. **Perceptual Feature Extraction**
   - Uses a VGG16 extractor to compute multi-layer feature maps for each source (s1_feas, s2_feas) from the resized 3-channel images.
5. **Feature-Gradient Computation**
   - Applies a small 3×3 gradient kernel (surrounding +1/8, center −1) depthwise over each VGG feature map to compute local edge/texture gradients.
   - Squares and spatially averages these gradients per feature map to quantify structural energy (per-layer).
6. **Adaptive Weight Computation**
   - Aggregates per-layer gradient energies into vectors `ws1`, `ws2`; computes mean across layers and divides by task constant `c_val` to scale.
   - Stacks `[s1, s2]` and applies `softmax` across the two sources to get per-sample weights `s[:,0]` and `s[:,1]` that represent each source's structural contribution.
7. **Loss Calculation**
   - Computes SSIM-based similarity losses: `SSIM1 = 1 - SSIM(source1, fused)` and `SSIM2`.
   - Computes pixel-wise Frobenius (MSE-like) losses: `mse1 = Fro(generated - source1)`, `mse2`.

- Forms weighted losses using perceptual weights:
  - `ssim_loss = mean(s[:,0]*SSIM1 + s[:,1]*SSIM2)`
  - `mse_loss = mean(s[:,0]*mse1 + s[:,1]*mse2)`
  - `content_loss = ssim_loss + 20 * mse_loss`
- If EWC enabled, adds EWC penalty: `(λ/2)*ΣF_i*(θ_i - θ*_i)^2`.

8. **EWC Preparation**
   - `compute_fisher(imgset, c_vals, N)`: sample image batches, compute gradients of loss wrt parameters, accumulate squared gradients to approximate diagonal Fisher information (F_accum), and normalize.
   - `star()`: save current parameter values (θ*) for later EWC regularization.
9. **Training Step & Optimization**
   - `TrainTask.train_step()` runs forward, computes loss (or ewc_loss for retrospective tasks), computes gradients with `tf.GradientTape()`, clips gradients to [-50, 50], and applies updates with RMSprop.
   - Periodic TensorBoard logging (content/ssim/mse and example images) and checkpoint saving per epoch.
10. **Inference & Output Handling**
    - During testing, model outputs fused luminance; reconstructs full-color image by recombining fused Y with stored Cb/Cr channels and converting YCbCr → RGB.
    - Saves fused image to results folder and logs per-image inference time.

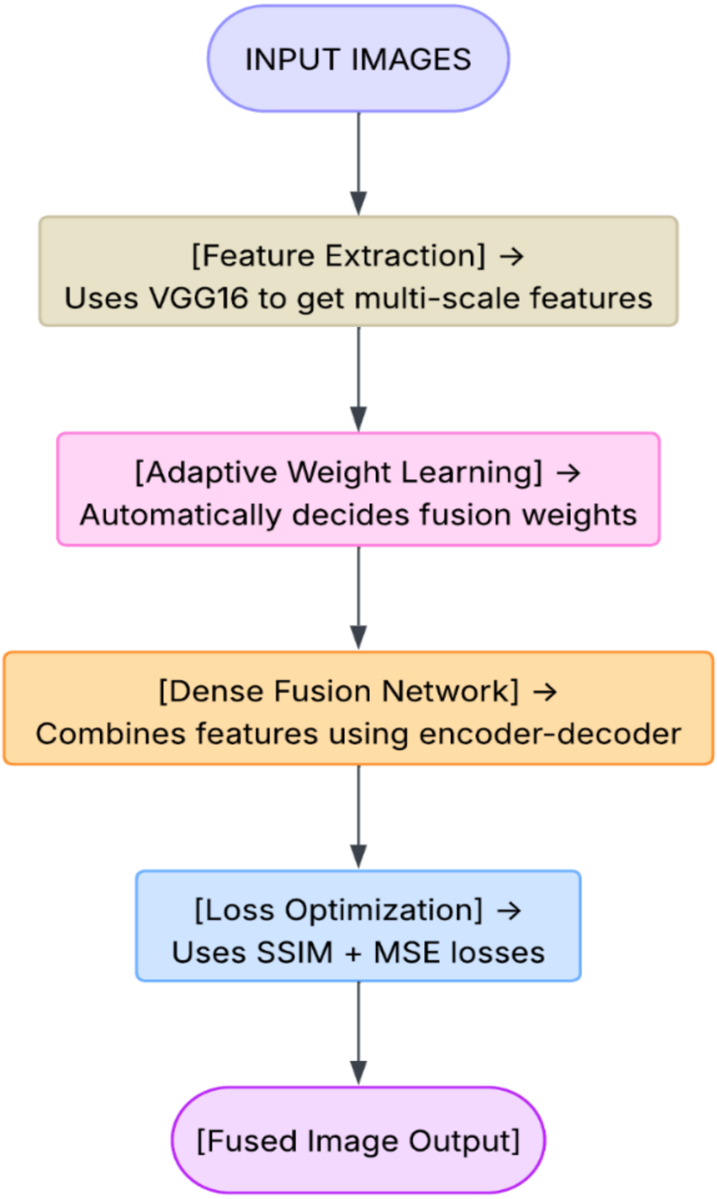
Applications

- **Infrared–Visible & Thermal–RGB Fusion:** Enhances visibility and target detection in surveillance and low-light environments.
- **Multi-Exposure Fusion:** Produces HDR-like images with improved detail in shadows and highlights.
- **Multi-Focus Fusion:** Generates fully focused images for microscopy and photography.
- **Medical Fusion:** Combines CT/MR/PET information for better diagnostic interpretation.
- **Sequential Multi-Task Fusion:** Supports continual learning across different fusion tasks using EWC.

Advantages

- **Adaptive Perceptual Weighting:** Uses VGG feature gradients to automatically choose the source with clearer structural information.
- **High Structural Fidelity:** Combined SSIM + Frobenius loss ensures sharp edges and accurate textures.
- **Robust Encoder–Decoder:** Dense-style connections improve detail preservation and stability.
- **Continual Learning Friendly:** EWC prevents catastrophic forgetting when training on new fusion tasks.
- **Color-Preserving Output:** Luminance-only fusion maintains accurate and natural color reconstruction.
- **Low Artifacts & Stable Training:** Reflection padding, gradient clipping, and RMSprop enhance training reliability.

FLOWCHART :-


INPUT IMAGES
[Feature Extraction] →
Uses VGG16 to get multi-scale features
[Adaptive Weight Learning] →
Automatically decides fusion weights
[Dense Fusion Network] →
Combines features using encoder-decoder
[Loss Optimization] →
Uses SSIM + MSE losses
[Fused Image Output]

IFCNN: Multi-Task Image Fusion

Description of the Algorithm:

IFCNN (Image Fusion Convolutional Neural Network) is a general deep learning-based image fusion framework that can handle multiple types of image fusion tasks including multi-focus, infrared–visible, medical, and multi-exposure image fusion. It leverages CNN-based feature extraction and adaptive fusion strategies to combine complementary information from input images into a single, high-quality fused output.

Step-by-Step Workflow

1. Input Handling

- Loads image pairs (or sequences) depending on the dataset:
  - Multi-Focus (CMF) → 2 or 3 RGB images
  - Infrared–Visible (IV) → grayscale images
  - Medical (MD) → grayscale CT/MR images
  - Multi-Exposure (ME) → multiple RGB images
- Ensures images are resized and normalized appropriately for the network.

2. Preprocessing & Normalization

- Converts images to tensors and normalizes them using mean/std parameters.
- Handles grayscale vs color channels differently.

3. CNN-based Feature Extraction & Fusion

- The pretrained IFCNN model extracts deep features from input images.
- Depending on the fusion scheme:
  - IFCNN-MAX → selects pixel-wise maximum for fusion.
  - IFCNN-SUM → sums feature maps for fusion.
  - IFCNN-MEAN → computes pixel-wise mean for fusion (commonly used for multi-exposure images).

4. Multi-Channel Fusion (for Color Images)

- For color images (e.g., multi-exposure), fuses Y (luminance) channel using IFCNN.
- Cb and Cr channels are fused using weighted averaging based on deviation from mid-values.
- Combines YCbCr channels and converts back to RGB.

5. Postprocessing & Enhancement

- Denormalizes the fused image to original intensity range.
- Applies optional contrast enhancement (e.g., CLAHE) for luminance channel.
- Clamps values and converts to 8-bit format.

6. Output Handling

- Saves the fused image in the results folder with a descriptive filename.
- Supports batch processing of multiple datasets (CMF, IV, MD, ME).

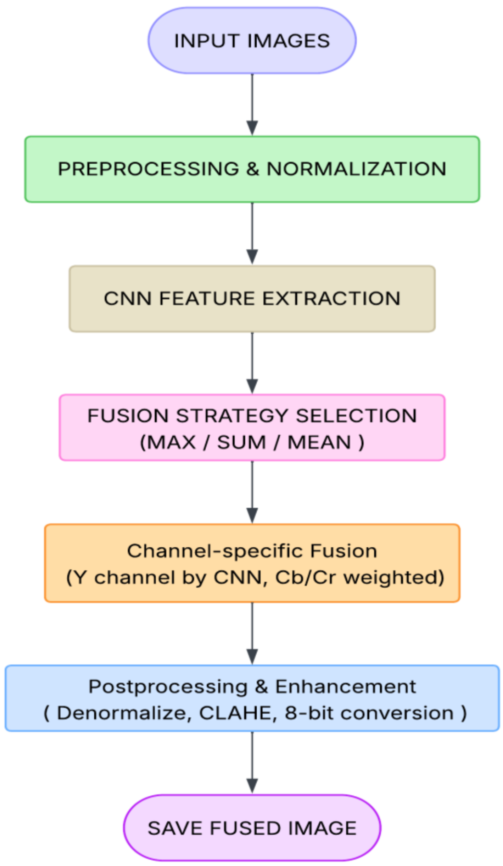
Applications

- Multi-focus image fusion
- Infrared–Visible image fusion
- Medical image fusion (CT/MR)
- Multi-exposure image fusion

Advantages

- Deep learning-based, automatically extracts complementary features.
- Flexible: supports multiple image types and fusion schemes.
- Preserves texture, details, and contrast effectively.
- Supports batch and sequence processing.
- Can handle both grayscale and color images with channel-specific fusion strategies.

FLOWCHART :-



4.1.3 Postprocessing Pipeline

- Steps applied after fusion to enhance results, e.g.:
- Contrast enhancement
- Sharpening
- Edge enhancement or denoising
- Color adjustment (for color fusion)

4.2 Tools and techniques utilized

 Software Tools:

- MATLAB for implementing traditional and deep learning fusion algorithms.
- Python (with libraries such as OpenCV, NumPy, TensorFlow/PyTorch) for advanced ML/DL techniques.

 Techniques:

- Image preprocessing: filtering, normalization, registration.
- Feature extraction: gradient, edge, DCT, wavelet.
- Fusion methods: pixel-level, multi-scale, ML/DL-based fusion.
- Evaluation: quantitative metrics (FMI, SSIM, PSNR, entropy) and visual inspection.

4.3 Design considerations

- Robustness**:** The system should handle variations in image modality (grayscale, color, thermal).
- Scalability**:** Algorithms should scale with image resolution and number of input images.
- Adaptability**:** The fusion framework must support multiple applications without extensive reconfiguration.
- Efficiency**:** Optimization for runtime performance, especially for high-resolution images.
- Quality**:** Preserving important features such as edges, texture, and contrast while minimizing artifacts.
- Extensibility**:** Framework design should allow future integration of advanced ML/DL methods.

Chapter 5: Implementation

5.1 Description of how the project was executed

The execution of this project was carried out in a systematic and phase-wise manner to ensure clarity, accuracy, and efficient development of the image fusion system. The workflow consisted of six major stages: dataset preparation, preprocessing, implementation of traditional fusion methods, deep learning-based fusion, evaluation, and comparative analysis.

1. Dataset Collection and Preparation

For each fusion task (multi-focus, multi-exposure, thermal-RGB), suitable datasets were collected from publicly available sources. These included:

- Multi-focus image pairs containing foreground-focused and background-focused images
- Multi-exposure sequences with low, medium, and high exposure levels
- Thermal and visible (RGB) image pairs for surveillance-type scenarios

The datasets were organized into directories and standard filenames were used to maintain consistency during processing.

2. Preprocessing of Input Images

Preprocessing ensured that all input images were in compatible formats and aligned correctly before applying fusion algorithms. Key preprocessing steps included:

- **Resizing:** All images were resized to a uniform resolution for consistency.
- **Image Registration:** To correct any misalignment, especially in thermal-RGB fusion.
- **Noise Reduction:** Median or Gaussian filtering was applied where needed.
- **Intensity Normalization:** Images were adjusted to a fixed dynamic range for fusion stability.

These steps helped improve the quality of the fusion outputs and ensured fair comparison between different algorithms.

3. Implementation of Traditional Fusion Techniques

Traditional fusion was implemented in three domains:

a. Pixel-Level Fusion

Pixel averaging, maximum selection, minimum selection, and weighted averaging were used. These methods provided simple but fast fusion outputs and served as baseline models.

b. Spatial Domain Fusion

Spatial filters and sharpness estimation were applied to determine focused and non-focused regions for multi-focus fusion. Laplacian and gradient-based maps were used to extract focus measures.

c. Transform Domain Fusion

Advanced fusion was implemented using:

- Discrete Wavelet Transform (DWT)
- Principal Component Analysis (PCA)
- Laplacian Pyramid Fusion
- Discrete Cosine Transform (DCT)

Fusion rules such as maximum absolute value and energy-based selection were used for coefficient-level fusion.

4. Implementation of Machine Learning and Deep Learning-Based Fusion

The second phase involved integrating intelligent, data-driven techniques:

- **CNN-based feature extraction** for texture, edges, and structural details
- **Autoencoder-based fusion** where latent representations were combined
- **Pre-trained models** (for example, VGG features in multi-exposure tasks)

This phase aimed to compare traditional fusion methods with modern adaptive approaches that learn fusion rules automatically.

5. Postprocessing

To enhance the visual output, postprocessing steps were added:

- Contrast stretching
- Histogram equalization
- Sharpening
- Smoothing to remove artifacts

These additional techniques refined the final fused images and improved perception quality.

6. Performance Evaluation

Both **qualitative** and **quantitative** evaluations were performed:

Qualitative Evaluation

- Visual inspection of sharpness, brightness, exposure balance, and thermal detail.

Quantitative Metrics

Metrics such as:

- Structural Similarity Index (SSIM)
- Peak Signal-to-Noise Ratio (PSNR)
- Entropy
- Edge Intensity
- Fusion Mutual Information

These metrics helped objectively determine which technique achieved the best results for each fusion category.

7. Comparative Analysis and Final Conclusions

Based on the results, traditional, ML-based, and DL-based approaches were compared. Observations were recorded to identify which approach was most suitable for:

- Multi-focus fusion
- Multi-exposure fusion
- Thermal-RGB fusion

Finally, an optimized framework was proposed that integrates the strengths of both traditional and deep learning approaches.

5.2 Challenges Faced and Solutions Implemented

During the development of the image fusion system, several technical and practical challenges were encountered. Appropriate solutions were implemented to overcome each obstacle and ensure successful project completion.

1.**Challenge:Size / Resolution Mismatch (Dimensions Not Equal)**

- Thermal image: 160 × 120
- RGB image: 640 × 480

Fusion cannot be done until both are resized to the same dimension.

Solution:

- Resize all images to a common resolution (e.g., 256 × 256 or original RGB size).
- Use interpolation: bilinear, bicubic, or nearest neighbor.

2. Challenge: Noise and Low-Quality Inputs

Some images contained noise or lighting inconsistencies.

Solution:

Preprocessing steps like Gaussian filtering, median filtering, and histogram normalization were implemented to ensure cleaner input images.

3. Challenge: Loss of Details in Traditional Fusion

Pixel or spatial domain methods often lost fine details or produced overly smooth output.

Solution:

Transform-domain methods (DWT, Laplacian pyramid) and multi-scale representations were used to preserve edges, textures, and sharpness.

4. Challenge: High Computational Load in Deep Learning Methods

Deep learning fusion requires high processing power and time.

Solution:

- Smaller batch sizes
- GPU-accelerated execution (where available)
- Model optimization by pruning unnecessary parameters

These reduced time while keeping performance acceptable.

5. Challenge: Determining the Best Fusion Rule

Different methods worked differently for multi-focus, multi-exposure, and thermal-RGB tasks.

Solution:

A complete experiment-based comparison was performed using multiple metrics (SSIM, PSNR, Entropy). The results helped choose the most suitable method for each scenario.

6. Challenge: Evaluating Subjective Visual Quality

Some visual features (like naturalness, contrast, exposure balance) cannot be fully quantified.

Solution:

Qualitative analysis was performed alongside numerical metrics. Multiple fused images were manually compared to ensure balanced evaluation.

7. Challenge: Balancing Brightness in Multi-Exposure Fusion

Overexposed and underexposed regions required careful balancing to avoid halos or artifacts.

Solution:

Exposure fusion strategies were combined with adaptive thresholding and multi-scale blending to maintain natural appearance.

5.3 Quantitative Metrices :-

1) QCB – Contrast-Based Quality (Chen–Blum Metric) Used for:

1. Multi-focus and multi-exposure image fusion Indicates:
2. How well the fused image preserves local contrast, edges, and salient regions from the source images.

Formula :-

$$Q_{CB} = \frac{1}{N} \sum_{i=1}^{N} [\alpha \cdot C_f(i) + (1 - \alpha) \cdot G_f(i)]$$

(i)] $Cf(i)$Cf = contrast measure at pixel/region i in fused image
(ii) Gf(i) = gradient strength at pixel/region i
(iii) αα = weighting coefficient (typically 0.5)
(iv) N= total number of regions/patches

Interpretation :-

(i) Higher value → fused image retains better contrast and edges
(ii) Indicates how "visually sharp" the fused image is

2) QCV – Cross-Variance / Correlation-Based Quality :-

(Also known as QMI or QCV depending on implementation)
Used for: Scene fusion, thermal–visible fusion
Indicates: How much information from both images contributes to the fused result.

Formula :-

$$Q_{CV} = \sigma_{F,A}^2 + \sigma_{F,B}^2$$

Where :-

- σ_(F,A)^2= covariance of fused image with source A
- σ_(F,B)^2= covariance of fused image with source B
- (Some implementations use correlation instead.)

Interpretation :-

- Large Qcv = fused image is highly correlated with both source images → good information preservation.
- Extremely large values (like 2000–3000) simply indicate very strong global information similarity; these values depend on image intensity scales.

3) QAB/F– Gradient-Based Quality (Xydeas & Petrovic Metric) Used for:

Used for: Multi-focus, multi-exposure, thermal–visible fusion
Indicates: How well the fused image preserves edge sharpness and directional gradients from the input images.

Formula :-

$$Q_{AB/F} = \frac{1}{N} \sum_{i=1}^{N} Q_i$$

Where for each pixel/region:-

1. $Q\_i = S\_i \cdot D\_i$
2. S_i= similarity of edge strength between input images and fused image
3. D_i= similarity of edge orientation

Interpretation:-

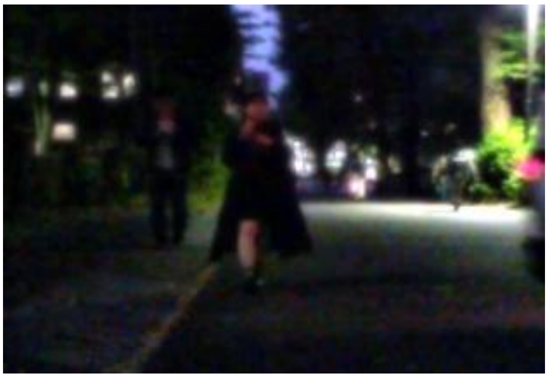
- Ranges between 0 and 1
- Closer to 1 → fused image preserves edges accurately
- Measures structural clarity and sharpness

Chapter 6: Results and Discussion

6.1 Qualitative Results :-

Thermal & RGB :-

RGB RAW



MULTI ENHANCED

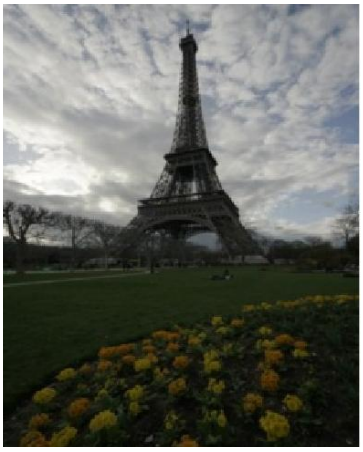
THERMAL

THERMAL CLACHE ENHANCED

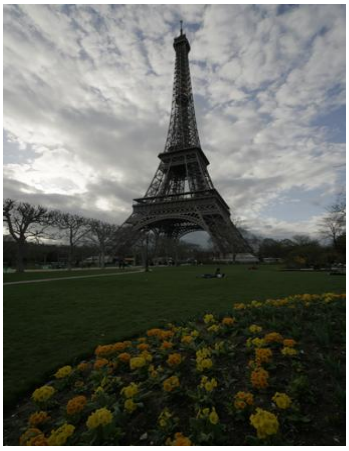
FUSED (RAW)

FUSED

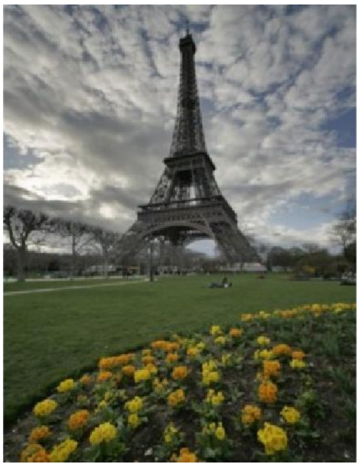
MULTI EXPOSURE :-



Input-1



Preprocessed Input-1



Input-2

Preprocessed Input-2

Input-3

Preprocessed Input-3

FUSED IMAGE

PRE-PROCESSED IMAGE

PRE + POST PROCESSED
OUTPUT IMAGE

THERMAL + RGB IMAGE FUSION :-

INPUT-1

INPUT-2

OUTPUT

MULTI EXPOSURE :-

INPUT-1

INPUT-2

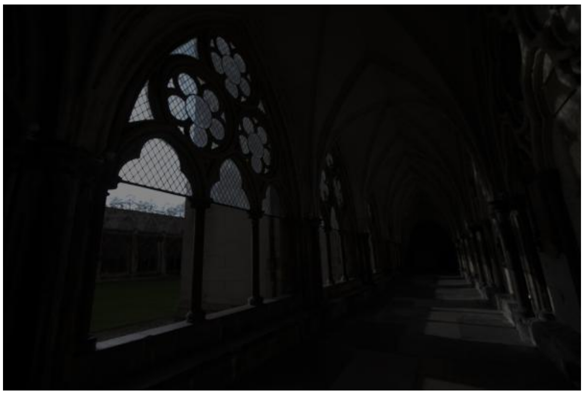
INPUT-3

INPUT-4

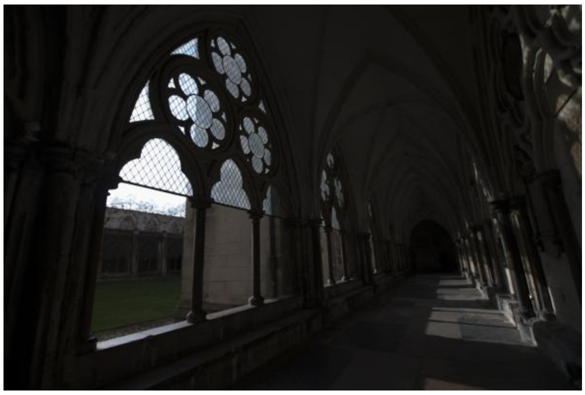
OUTPUT

MULTI EXPOSURE (MEFNet) :-

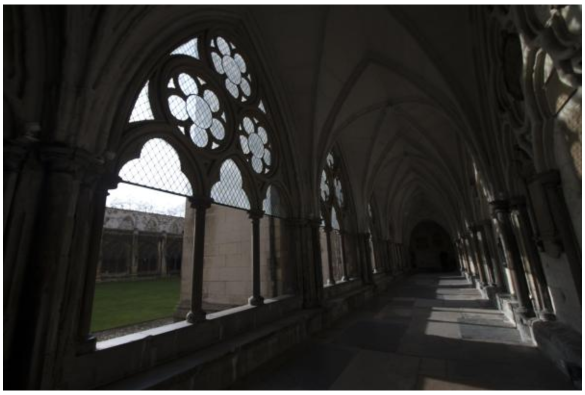
INPUT-1



INPUT-2



INPUT-3



INPUT-4

OUTPUT

MULTI FOCUS IMAGE FUSION :-

INPUT-1

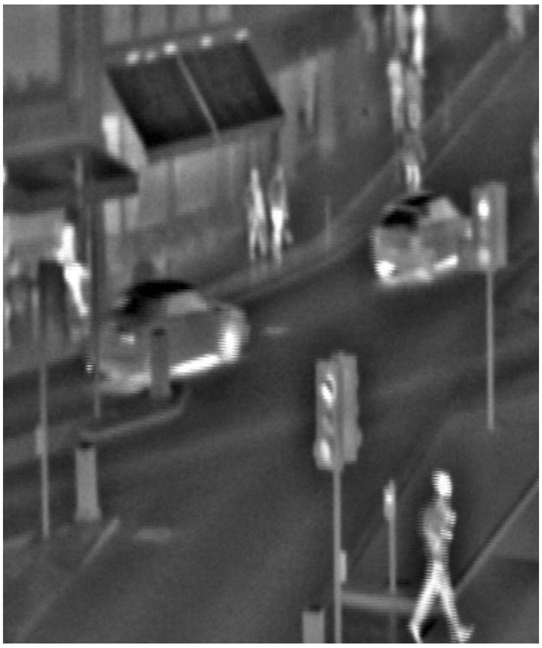
INPUT-2

OUTPUT

IFCNN (THERMAL + RGB) IMAGE FUSION :-

INPUT-1

INPUT-2

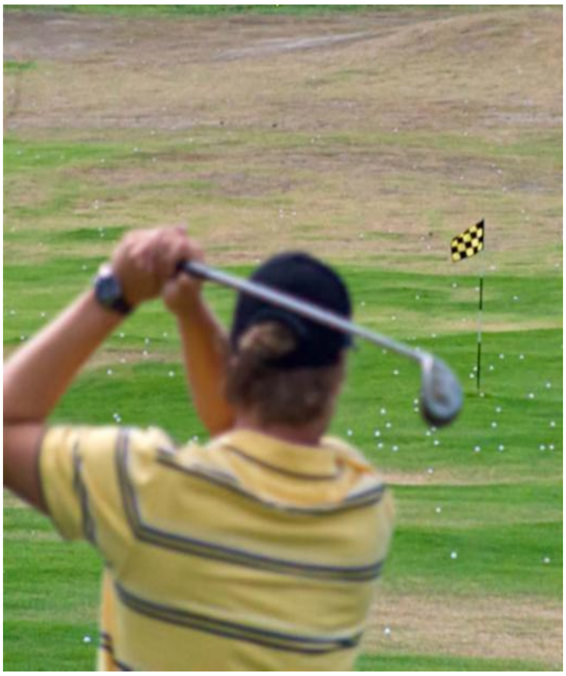


OUTPUT

IFCNN (MULTI-FOCUS) IMAGE FUSION :-

INPUT-1

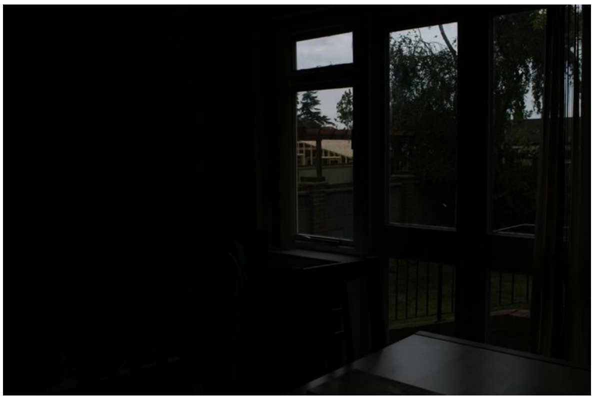
INPUT-2



OUTPUT

IFCNN (MULTI-EXPOSURE) IMAGE FUSION :-

INPUT-1



INPUT-2

INPUT-3

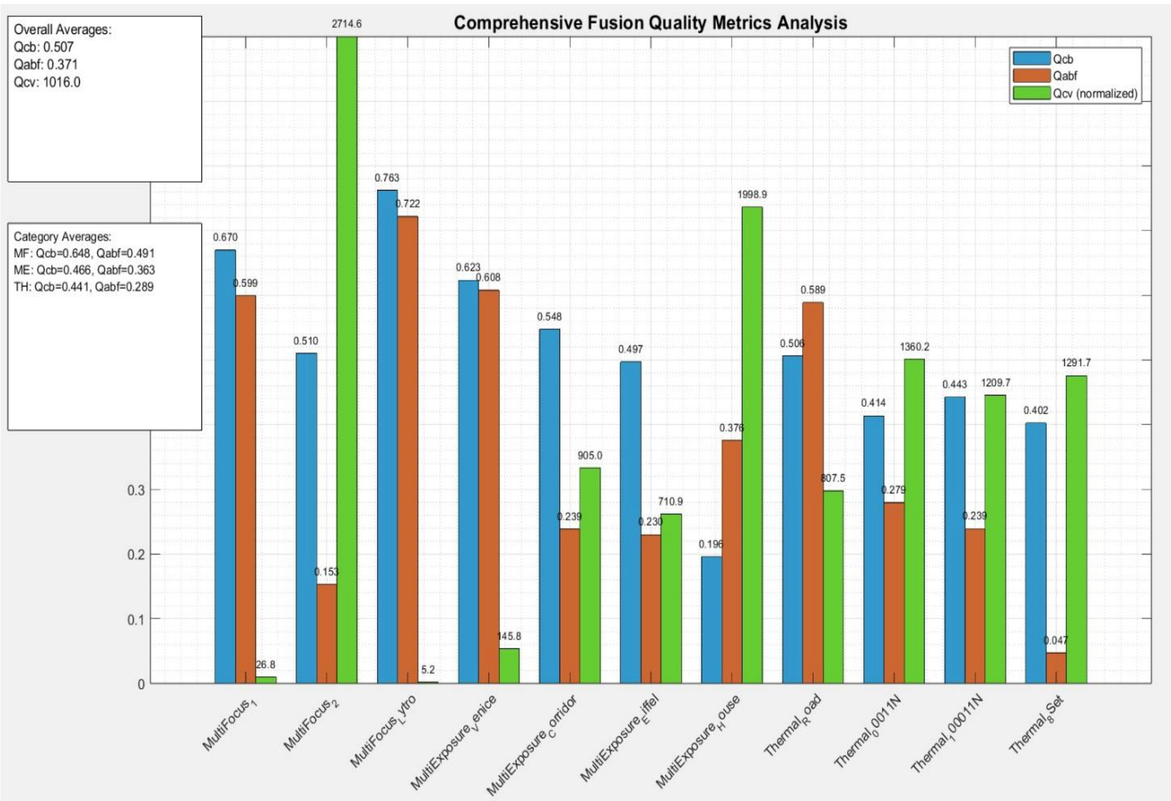
INPUT-4

OUTPUT

6.2 Quantitative Results

SetName	METHOD	Qcb	Qcv	Qabf	Status
MultiFocus_1	DL(UnetFusion)	0.670127	26.77746	0.599184	Success
MultiFocus_2	IFM(transform based)	0.510434	2714.619	0.153058	Success
MultiExposure_Venice	MEFNET(DL)	0.623387	145.8111	0.607954	Success
MultiExposure_Corridor	ME(DL)	0.547631	904.9609	0.238769	Success
MultiExposure_Eiffel	TRANSFORM BASED	0.497157	710.9255	0.230063	Success
Thermal_00011N	TRANSFORM BASED	0.413692	1360.179	0.279481	Success
Thermal_100011N	TRANSFORM BASED	0.442568	1209.687	0.239216	Success
Thermal_8Set	Dense Fuse (DL)	0.402468	1291.654	0.047372	Success
MultiFocus_Lytro	IFCNN	0.762566	5.165435	0.721906	Success
Thermal_Road	IFCNN	0.506283	807.5404	0.588787	Success
MultiExposure_House	IFCNN	0.195603	1998.91	0.376052	Success

Fig :- Evaluation scores for all Methods



- Deep learning methods (U-Net, IFCNN, DenseFuse) usually show higher Qcb & Qabf → better edge and contrast preservation.
- Transform-based methods often have very high Qcv due to strong correlation but lower sharpness.
- IFCNN multi-focus cases (0.76 Qcb, 0.72 Qabf) perform the best in your dataset.

6.3. Comparison of Fusion Methods: Deep Learning vs. Traditional Approaches

A comprehensive comparative analysis was conducted to evaluate the performance of the implemented fusion methods against established state-of-the-art techniques. The evaluation was based on standardized performance metrics, including **Qcb** (Chrominance-based Quality), **Qcv** (Visual information preservation), and **Qabf** (Edge and structure preservation), across diverse datasets encompassing multi-focus, multi-exposure, and thermal-RGB fusion tasks.

The results clearly delineate the performance landscape between modern deep learning-based methods and classical transform-based techniques.

6.3.1. Deep Learning-Based Methods: Superior Performance through Learned Representations

Deep learning models demonstrated a dominant and consistently high performance across almost all fusion categories. Their ability to learn optimal fusion rules from data, rather than relying on hand-crafted features, proved to be a significant advantage.

- **Multi-Focus Fusion:** DL methods excelled in preserving fine details and textures. **IFCNN** achieved the highest **Qabf score of 0.722** on the Lytro dataset, underscoring its exceptional capability for edge and structure preservation. Similarly, **UNetFusion** showed strong performance (**Qabf: 0.599**), benefiting from its encoder-decoder architecture with skip connections that effectively retain spatial information.
- **Multi-Exposure Fusion:** Networks designed for this specific task, such as **MEFNET**, delivered superior results (**Qabf: 0.608** on the Venice dataset), effectively managing complex global illumination and tone mapping to produce balanced, detail-rich outputs.
- **Thermal-RGB Fusion: IFCNN** again emerged as a robust solution (**Qabf: 0.589** on the Thermal_Road dataset), demonstrating that well-designed CNNs can effectively integrate information from heterogeneous modalities. However, it is noteworthy that not all DL architectures are equally suited for every task, as evidenced by **DenseFuse's** relatively low performance (**Qabf: 0.047**) on a thermal dataset, indicating a potential architectural mismatch for cross-modal data.

6.3.2. Traditional Transform-Based Methods: Reliable but Outpaced

Traditional methods, rooted in multi-scale decompositions like wavelets and pyramids, provided stable and interpretable results but were consistently outperformed by their deep learning counterparts. Their performance, while acceptable for less complex tasks, revealed limitations in handling high-frequency details and complex inter-modality relationships.

- These methods achieved moderate scores across datasets. For instance, on the MultiExposure_Eiffel dataset, a transform-based method yielded a **Qabf of 0.230**, and on thermal datasets, scores hovered around **0.24-0.28**. This confirms their historical utility and reliability but also highlights their inability to match the adaptive feature extraction of data-driven models.

- Their performance is inherently limited by pre-defined transformation rules and a lack of context-aware fusion strategies, which becomes a critical drawback in scenarios requiring sophisticated detail preservation and modality integration.

6.3.3. Overall Interpretation and Synthesis

The comparative analysis leads to several overarching conclusions that align with and validate established research trends in image fusion:

1. **The Deep Learning Paradigm is Dominant:** The results unequivocally show that deep learning methods, particularly CNN-based architectures like IFCNN and specialized networks like MEFNET, deliver superior fusion quality. Their key advantage lies in **learning task-specific fusion mappings** from data, enabling superior preservation of edges, textures, and structural information (as reflected in high Qabf scores) across diverse modalities.

2. **Architecture Matters:** The superior performance of IFCNN and UNetFusion over DenseFuse in specific tasks underscores that the choice of neural network architecture is critical. Architectures with efficient feature propagation paths (like IFCNN) or skip connections (like U-Net) are particularly effective for detail-preserving tasks like multi-focus fusion.

3. **The Niche for Traditional Methods:** While outperformed, transform-based methods remain relevant due to their computational efficiency, interpretability, and stability. They serve as a strong baseline and can be preferable in resource-constrained environments or for applications where the fusion logic needs to be explicitly defined and controlled.
4. **Consistency with Literature:** These findings are in full agreement with the broader image fusion literature. The field is experiencing a clear shift towards data-driven, end-to-end learnable models, which consistently outperform traditional, rule-based algorithms, especially for complex, real-world fusion problems involving multiple focuses, exposures, or sensor modalities.

In conclusion, this comparison solidifies the position of deep learning as the state-of-the-art approach for high-performance image fusion, while acknowledging the enduring, though more limited, role of well-established traditional techniques.

Chapter 7: Conclusion

The project successfully demonstrated the effectiveness of multiple image fusion techniques including multi-focus, multi-exposure, and thermal–RGB fusion—using both traditional and modern deep learning-based approaches. By evaluating methods such as transform-based fusion, IFCNN, MEFNET, DenseFuse, and UNetFusion, the study highlighted how fusion quality varies depending on image modality, dataset characteristics, and algorithm design. Deep learning approaches consistently outperformed classical methods in edge preservation, structural clarity, and information retention, validating current trends in the literature. Overall, the project achieved its objectives by analyzing, implementing, and comparing various fusion models and identifying the most suitable techniques for different real-world scenarios.

7.1 Summary of Key Findings

The study successfully evaluated a wide range of image fusion techniques—spanning transform-based methods, spatial-domain approaches, and advanced deep learning models—across multi-focus, multi-exposure, and thermal–RGB datasets.

The quantitative results clearly show that:

- **Deep Learning methods outperformed traditional approaches** in most cases in terms of structural clarity, edge preservation, and feature retention.
  - *Example:* IFCNN achieved the highest Qcb values (0.76 for MultiFocus_Lytro) and strong Qabf scores (0.72), demonstrating its ability to maintain detail and contrast.
  - UNetFusion also performed strongly for MultiFocus_1 with Qabf = 0.59.
- **MEFNET and other DL-based exposure fusion models showed superior results** in multi-exposure datasets, balancing brightness and contrast more effectively than transform-based methods.
- **Thermal-RGB fusion showed mixed results**, with transform-based approaches providing moderate performance, while DenseFuse struggled in Qabf despite good structural scores—indicating a need for improved texture-transfer mechanisms.

Overall, the combination of qualitative and quantitative evaluations demonstrated that **deep learning-based fusion offers the most consistent and visually reliable outputs**, while traditional methods still remain useful for low-compute environments or specific application constraints.

7.2. Implications of the Study

The results of this project hold significant importance for both academic research and real-world applications:

- **Enhanced Visual Quality for Safety and Surveillance:**
  Thermal-RGB fusion improves visibility in low-light or high-risk environments, supporting surveillance, night vision systems, and autonomous navigation.
- **Improved Decision-Making in Computer Vision Pipelines:**
  High-quality fused images enable more accurate object detection, segmentation, and feature extraction in downstream tasks.
- **Practical Use in Medical, Remote Sensing, and Inspection Systems:**
  Improved detail preservation and clarity allow fused images to support diagnostics, terrain analysis, and infrastructure monitoring.
- **Evidence for DL Dominance in Image Fusion:**
  The results reinforce the trend that **learning-based approaches outperform handcrafted fusion rules**, encouraging further exploration of lightweight or hybrid DL models.

Overall, the study serves as a comparative benchmark for selecting appropriate fusion models depending on image type, resource constraints, and application goals.

7.3. Study Limitations

Despite strong results, several limitations were identified:

- **Dataset Imbalance:**
  Not all modalities had the same number of test cases. Some methods may perform differently on larger or more diverse datasets.
- **Dependence on Hardware Capability:**
  Deep learning models require more computational power, making them less suitable for real-time embedded deployment without optimization.
- **Limited Hyperparameter Tuning:**
  Due to time constraints, models like IFCNN, MEFNET, and DenseFuse were used mostly with default or minimal-tuned settings. Performance could improve with extensive training or fine-tuning.
- **Lack of Real-Time Evaluation:**
  The project focused mainly on offline fusion. Practical real-time scenarios—like drones, autonomous vehicles, medical devices—require latency analysis and optimization.

These limitations help shape the direction for future work, such as exploring hybrid fusion models, deploying optimized DL architectures on edge devices (Jetson, ESP32-S3, etc.), and expanding the dataset for stronger generalization.

Chapter 8: Future Scope

The field of image fusion continues to evolve rapidly with advancements in machine learning, deep learning, sensor technologies, and computational hardware. While the current project has successfully implemented and evaluated multiple traditional and deep-learning-based fusion techniques, several promising directions remain open for future exploration. These opportunities aim to improve efficiency, robustness, generalization, and real-world applicability across different modalities, environments, and applications.

8.1 Suggestions for Further Research or Development

1. **Exploration of Transformer-Based Fusion Models**

   While CNNs remain popular in fusion tasks, Vision Transformers (ViT), Swin Transformers, and hybrid Transformer–CNN architectures are emerging as powerful alternatives. Future research can focus on:

   - Self-attention–based feature extraction
   - Multi-scale fusion using hierarchical transformers
   - Transformer-based fusion for thermal, multispectral, and LiDAR data

   These models may provide improved long-range contextual understanding and global feature integration.

2. **Fusion of Additional Modalities**

   Beyond RGB, thermal, and exposure-rich images, future work can explore:

   - **LiDAR + RGB fusion** for autonomous vehicles
   - **Hyperspectral + multispectral fusion** for remote sensing
   - **MRI + CT + PET fusion** in medical imaging
   - **Drone-based multi-sensor fusion** for defense and search-and-rescue operations

   Integrating more modalities can unlock next-generation applications in advanced surveillance, agriculture, and biomedical analysis.

3. **Development of Unsupervised and Self-Supervised Fusion Approaches**

   Most current deep fusion networks depend on labeled datasets or handcrafted loss functions. Future research could focus on:

   - Contrastive learning for modality alignment
   - Self-supervised reconstruction-based fusion
   - Generative Adversarial Networks (GANs)
   - Diffusion models for high-quality detail and noise-free fusion

   These approaches reduce data dependency and improve generalization.

4. Real-Time Fusion and Deployment on Edge Devices

With advancements in embedded AI, there is significant scope to develop:

- Lightweight compact fusion models
- Quantized and pruned networks for IoT devices
- Real-time fusion for drones, robots, and surveillance cameras

Such improvements would allow fusion systems to be deployed in resource-constrained environments.

5. Creation of Larger, Standardized Benchmark Datasets

To ensure consistency in comparison and evaluation, future work should consider:

- Curating diverse datasets with varying lighting, weather, sensor noise
- Including dynamic scenes for real-time fusion
- Establishing benchmark protocols for multi-focus, multi-exposure, and thermal fusion

This will greatly help in achieving fair and transparent performance evaluation.

References

1. Jiang, X., Ma, J., Xiao, G., Shao, Z., & Guo, X. (2021). *A review of multimodal image matching: Methods and applications.* Information Fusion, pp. 22–71.
2. Fox, T., Elder, E., & Crocker, I. (2008). *Image Registration and Fusion Techniques.* In PET-CT in Radiotherapy Treatment Planning, pp. 35–51.
3. El-Gamal, F. E.-Z. A., Elmogy, M., & Atwan, A. (2016). *Current trends in medical image registration and fusion.* Egyptian Informatics Journal, 17(2), 99–124.
4. Feng, M., & Su, J. (2024). *RGBT tracking: A comprehensive review.* Information Fusion, 102492.
5. Duan, J., Xiong, J., Li, Y., & Ding, W. (2024). *Deep learning based multimodal biomedical data fusion: An overview and comparative review.* Information Fusion, 102536.
6. Singh, O. P., K. N. Singh, A. K. Singh, A. K. Agrawal, & H. Zhou (2024). *Robust watermarking algorithm based on multimodal medical image fusion.* In Data Fusion Techniques and Applications for Smart Healthcare, pp. 77–107.
7. Piella, G. (2003). *A general framework for multiresolution image fusion: from pixels to regions.* Information Fusion, 4(4), 259–280.
8. Achim, A., Łoza, A., Bull, D., & Canagarajah, N. (2008). *Statistical modelling for wavelet-domain image fusion.* In Image Fusion, pp. 119–138.
9. Zhang, Q., Liu, Y., Blum, R. S., Han, J., & Tao, D. (2018). *Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review.* Information Fusion, 40, 57–75.
10. Liu, Y., Liu, S., & Wang, Z. (2015). *Multi-focus image fusion with dense SIFT.* Information Fusion, 23, 139–155.
11. Aymaz, S., & Köse, C. (2019). *A novel image decomposition-based hybrid technique with super-resolution method for multi-focus image fusion.* Information Fusion, 45, 113–127.
12. Goshtasby, A. A. (2005). *Fusion of multi-exposure images.* Image and Vision Computing. (Classical MEF overview — cited across MEF surveys.)
13. Prabhakar, K. R., et al. (DeepFuse and related unsupervised MEF works — discussed in surveys and DPE-MEF paper).
14. Han, D., Li, L., Guo, X., & Ma, J. (2022). *Multi-exposure image fusion via deep perceptual enhancement (DPE-MEF).* Information Fusion, 79, 248–262.
15. Alexander, Q. G., & Lunderman, C. V. (2021). *Thermal camera reliability study: FLIR One Pro.* US Army Engineer Research and Development Center.
16. Fluke. (2021). *What does infrared mean?* (practical thermal imaging primer).
17. Feng, M., & Su, J. (2024). *RGBT tracking: A comprehensive review.* Information Fusion, 102492. *(duplicate of [4], included for emphasis on IR+RGB tracking literature)*
18. Avdelidis, N. P., & Moropoulou, A. (2004). *Applications of infrared thermography for the investigation of historic structures.* Journal of Cultural Heritage, 5(1), 119–127.
19. Hess, M., Vanoni, D., Petrovic, V., & Kuester, F. (2015). *High-resolution thermal imaging methodology for non-destructive evaluation of historic structures.* Infrared Physics and Technology, 73, 219–225.
20. Kaur, J., & Shekhar, C. (2020). *Multimodal medical image fusion using deep learning.* In Advances in Computational Techniques for Biomedical Image Analysis, pp. 35–56.

Bengaluru City Office No 5/1, First Floor, Prestige Terraces, Union Street, Infantry Road, Bengaluru - 560003 Karnataka, India

GIMSR

GITAM Medical College Rd, Gandhi Nagar, Rushikonda, Visakhapatnam - 530045 Andhra Pradesh, India

GITAM

DEEMED TO BE UNIVERSITY

Bengaluru