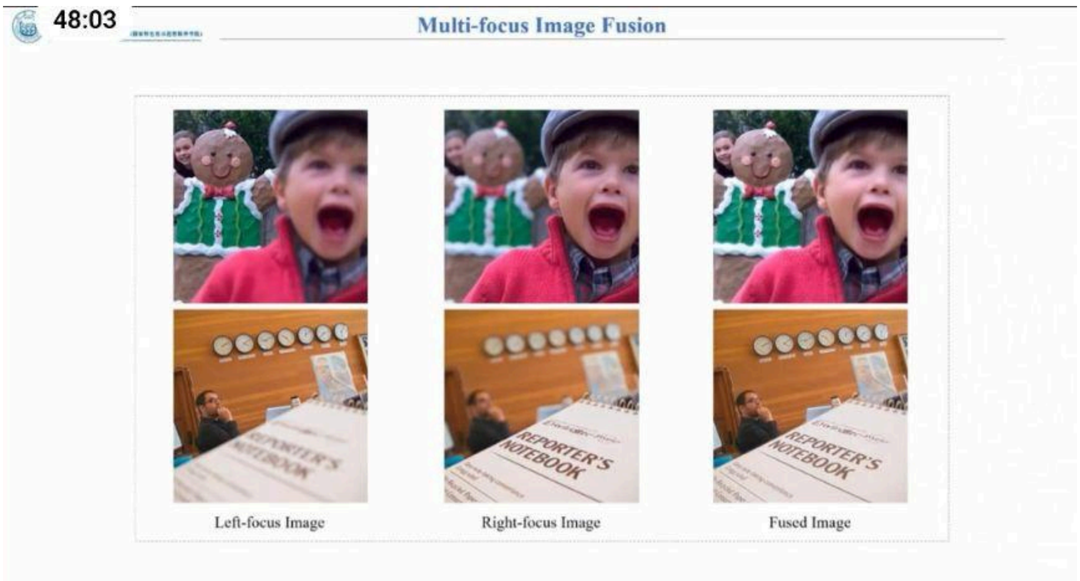
Image Fusion

- **Definition**: A technique where multiple images are combined into a single image to enhance overall information or quality.
- **Purpose**: To create a composite image with more detailed or comprehensive information than any single input image.
- **Types of Image Fusion**:
  - **Pixel-level fusion**:
    - Merges pixel values from different images.
    - Techniques include averaging, PCA, and wavelet transforms.
  - **Feature-level fusion**:
    - Combines extracted features (e.g., edges, textures, shapes).
    - Focuses on combining relevant features from the images.
  - **Decision-level fusion**:
    - Merges decisions made on each image (e.g., classifications or object detections).
- **Goal**:
  - Improve clarity, information content, or visualization in the combined image.
  - Useful when working with images taken under different conditions or sensors.
- **RGB-IR Image Fusion**:
  - process of combining **RGB** images with **Infrared (IR)** images to create a new image that enhances both visible and infrared information.
  - This fusion technique leverages the strengths of both image types for applications where both visible and thermal details are required.

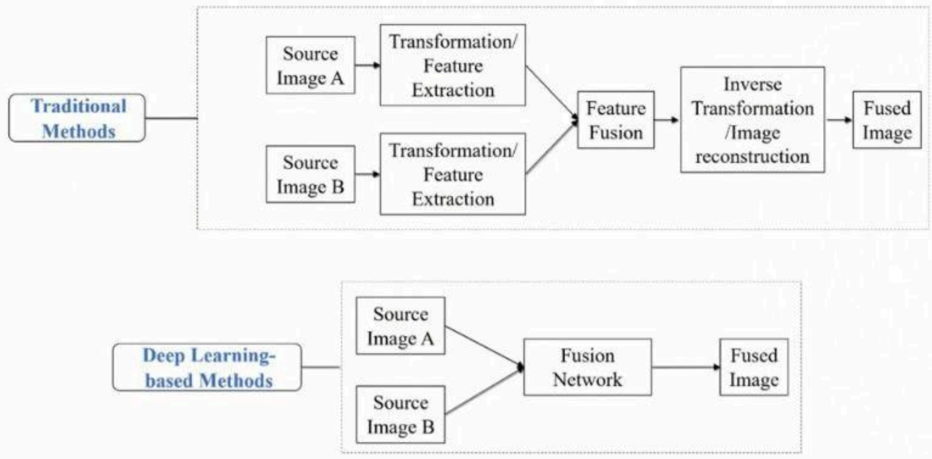

49:16
Multi-exposure Image Fusion
Under-exposure Image
Over-Exposure Image
Fused Image


52:40
Infrared and Visible Image Fusion
Infrared Image
Visible Image
Fused Image


48:03
Multi-focus Image Fusion
Left-focus Image
Right-focus Image
Fused Image

46:49

Literature Review



45:08

Literature Review

Traditional Methods

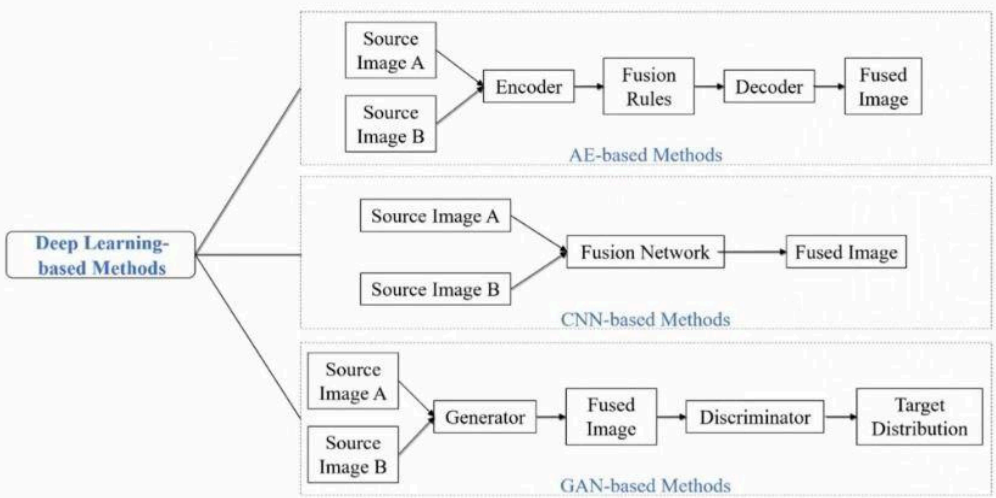
- **Decomposition** — Decompose the source image with a specific image transformation method
- **Fusion** — Fuse the decomposed coefficients with the pre-designed fusion rules
- **Reconstruction** — Reconstruct the fusion result is by inverting the fused coefficients

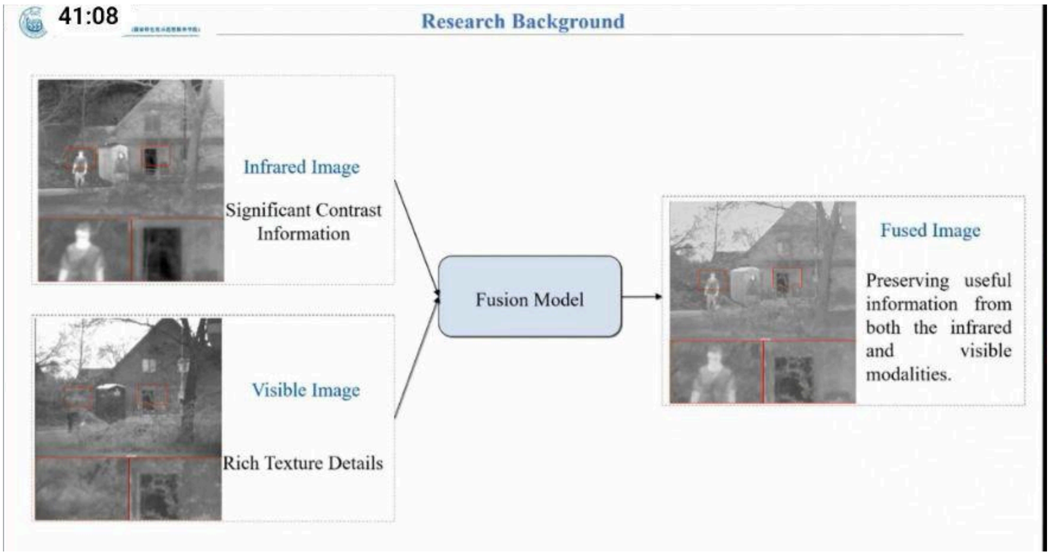
Limitations

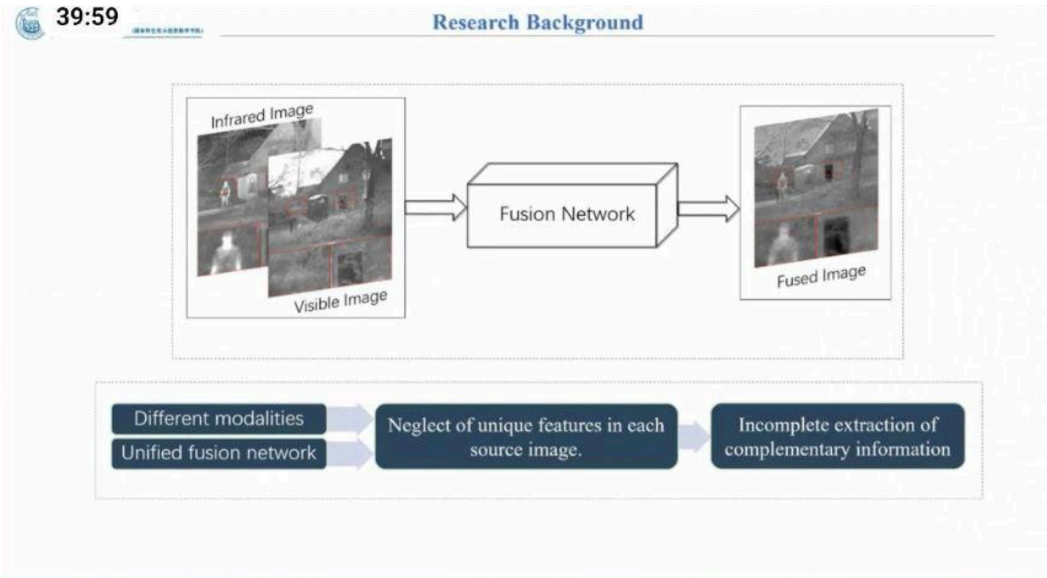
- ➢ Manually design decomposition methods and fusion rules
- ➢ It is difficult to fully account for all the features of the input image
- ➢ Time-consuming and laborious

43:47

Literature Review




41:08
Research Background
Infrared Image
Significant Contrast Information
Visible Image
Rich Texture Details
Fusion Model
Fused Image
Preserving useful information from both the infrared and visible modalities.


39:59
Research Background
Infrared Image
Visible Image
Fusion Network
Fused Image
Different modalities
Unified fusion network
Neglect of unique features in each source image.
Incomplete extraction of complementary information

Y-shape Dynamic Transformer (YDTR)

- **Goal**: Develop a novel infrared and visible image fusion method
- **Objective**: Generate a composite image that combines:
  - Salient target in the infrared image.
  - Texture details from the visible image.

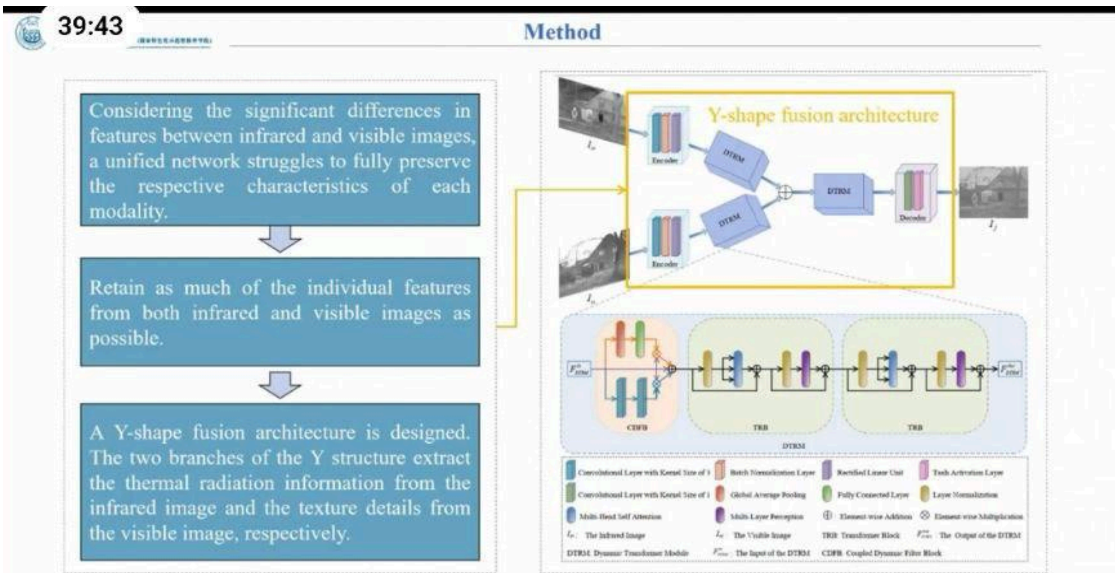
- **Problem with existing methods**: Current deep learning based methods rely on convolutional operations, which limit global feature preservation.

- **Proposed solution**:
  - YDTR uses a dynamic Transformer module (DTRM) to capture both local features and significant context information
- **YDTR architecture** consists of:
  - **Two Y-shaped branches**: One branch extracts thermal information from the IR image, and the other extracts texture details from the visible image.
  - Each branch uses an **encoder** to capture shallow features and a **dynamic Transformer module (DTRM)** to model long-range relationships.
  - The **main path** combines these features through a DTRM and a decoder to reduce dimensions and integrate the information.

- **Loss function**: Combines two terms to enhance fusion quality:
  - Structural similarity (SSIM)
  - Spatial frequency (SF)

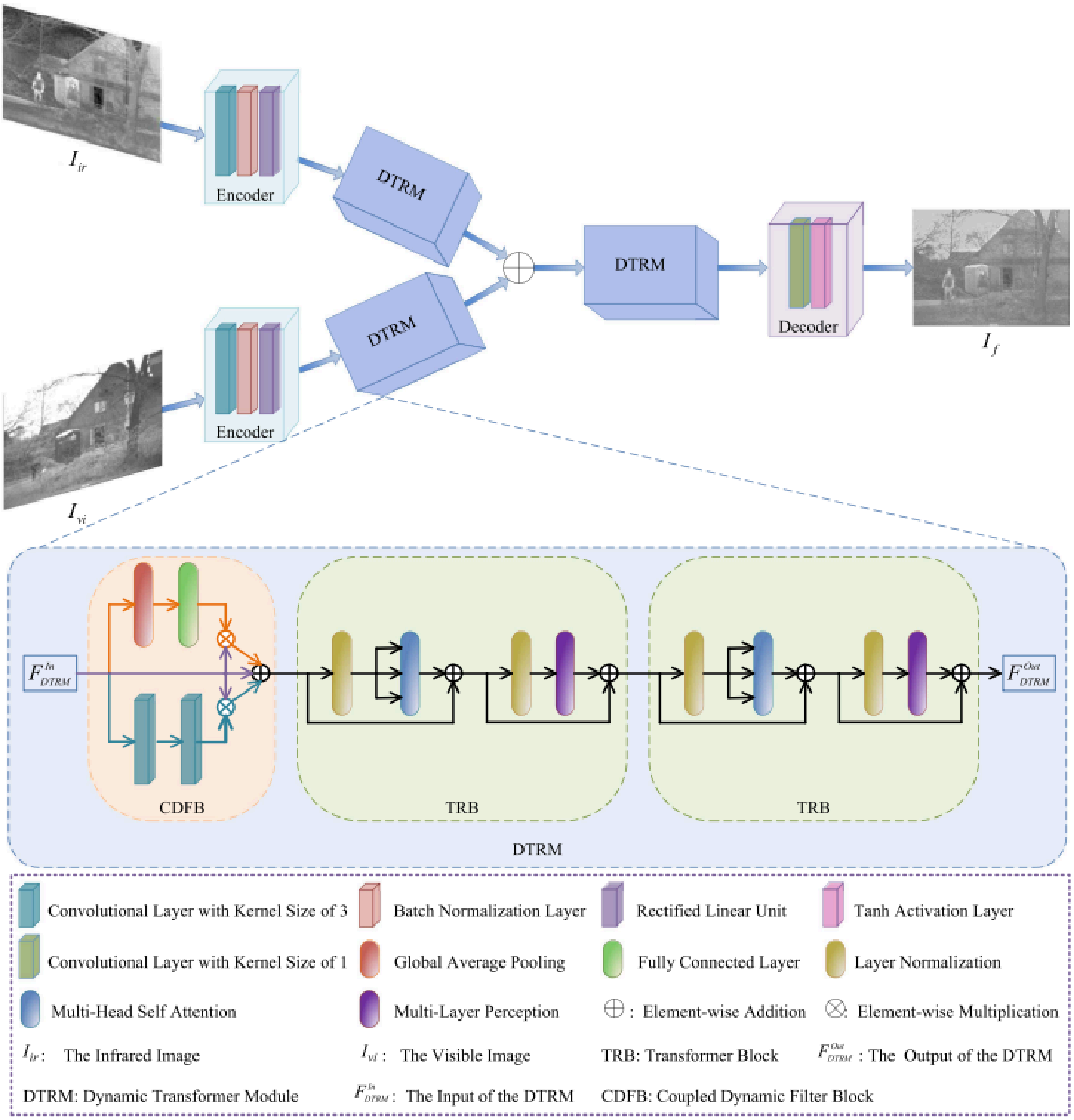
- **Extension**: YDTR can be extended to handle:
  - Infrared and RGB-visible images.
  - Multi-focus images.

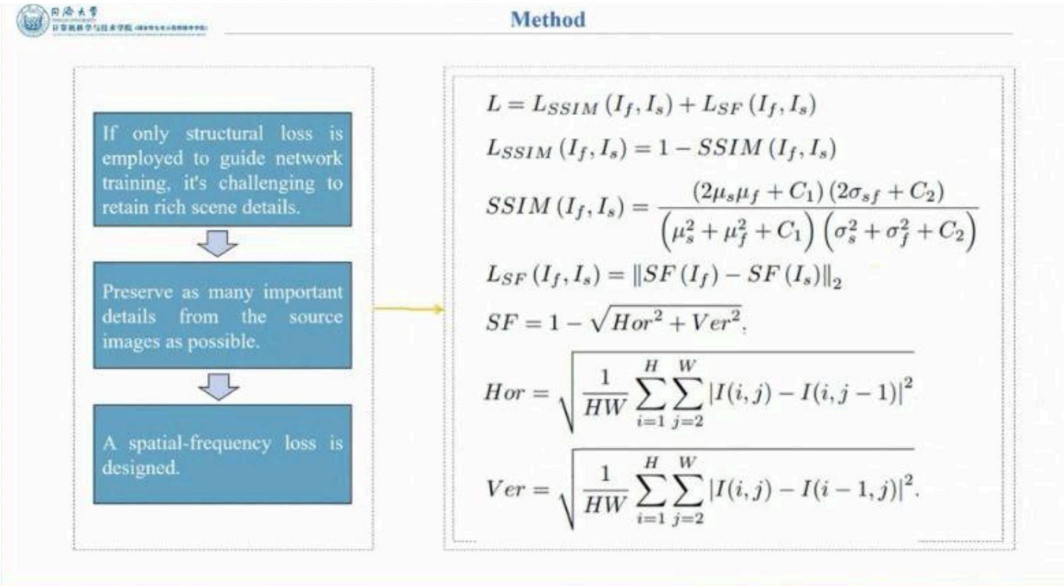
- **Generalization**: The method demonstrates strong generalization capability without requiring fine-tuning.

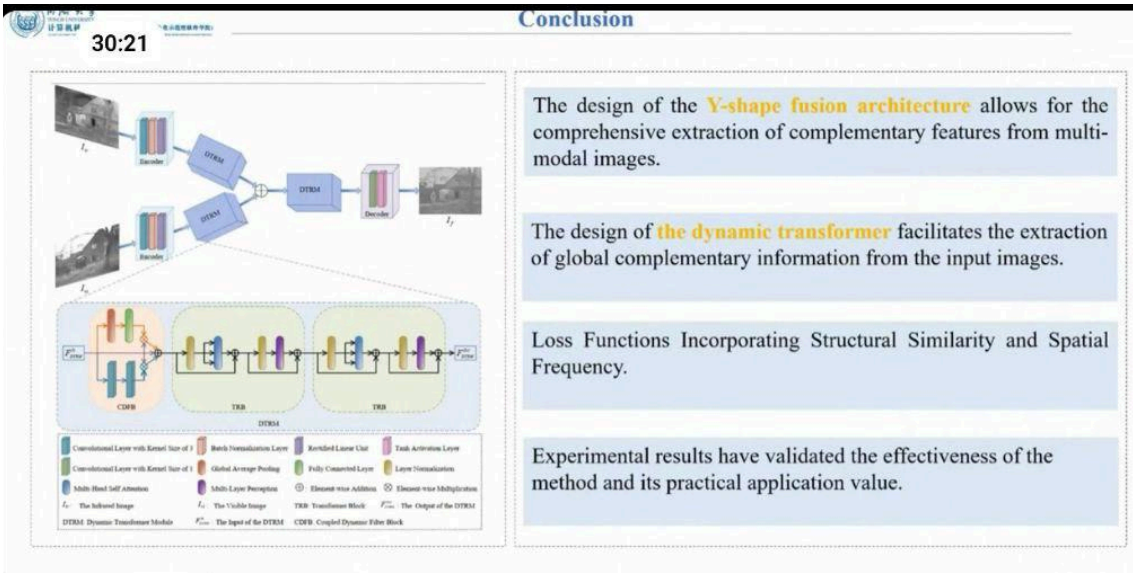


IR images - Thermal radiation information

RGB images - texture details







COMPARISON WITH STATE-OF-THE-ART IMAGE FUSION ALGORITHMS

Methods	End-to-End	Convolutional Operation	Transformer	Y-shape	SSIM Loss	SF Loss	Unsupervised	Generalization Ability
CNN [25]	×	✓	×	×	×	×	×	×
AUIF [26]	×	✓	×	×	✓	×	✓	×
DenseFuse [18]	×	✓	×	×	✓	×	✓	×
FusionGAN [19]	✓	✓	×	×	×	×	✓	×
AttentionFGAN [2]	✓	✓	×	✓	×	×	✓	×
GANMcC [24]	✓	✓	×	×	×	×	✓	✓
MgAN-Fuse [27]	✓	✓	×	×	×	×	✓	×
U2Fusion [20]	✓	✓	×	×	✓	×	✓	×
RFN-Nest [23]	✓	✓	×	×	✓	×	×	✓
MFE-EAG [28]	✓	✓	×	×	✓	×	×	✓
CSF [29]	×	✓	×	×	✓	×	✓	×
IFT [30]	×	✓	✓	×	✓	×	✓	×
DNDT [31]	✓	✓	✓	×	✓	×	✓	×
PPT Fusion [32]	×	×	✓	×	×	×	×	×
TGFuse [33]	✓	✓	✓	×	✓	×	✓	×
YDTR	✓	✓	✓	✓	✓	✓	✓	✓

DL-Based Fusion

Methods:

- **CNN-based Methods**: Liu et al. [25] introduced CNNs for infrared and visible image fusion, using a Siamese network for activity level measurement. Following this, methods like **DenseFuse** [18] and **FusionGAN** [19] leveraged dense connections and GANs to enhance fusion quality.

- **Attention Mechanisms**: FusionGAN was enhanced with a **multi-scale attention mechanism** [2] to focus on discriminative regions.

- **Multi-classification and Attention**: Methods like **GANMcC** [24] and **MgAN-Fuse** [27] incorporated **multi-classification constraints** and **multi-grained attention modules** for improved fusion.

Challenges:

- **Limited Long-Range Context**: Convolutional operations capture local features but fail to model global context, leading to a loss of significant global features.

- **Feature Extraction Approach**: Many methods use single or parallel networks without tailoring to the specific characteristics of infrared and visible images.

Transformer in Image Fusion:

- The **Transformer** architecture [37], introduced for NLP, addresses CNN's limited receptive field. The **Vision Transformer (ViT)** [41] applies this to image classification, improving global feature handling.

- **Transformer-based Fusion Methods**:

  - **Multi-Scale Fusion** [30] uses a two-stage training approach.
  - **DenseNet-Transformer** [31] combines DenseNet for encoding and dual-Transformers for fusion.
  - **Patch Pyramid Transformer (PPT)** [32] uses a patch Transformer for sequence transformation and a Pyramid Transformer for feature extraction.

Proposed Approach:

- The paper proposes a **hybrid CNN-Transformer fusion method** to preserve both **local and global features**. This approach overcomes the limitations of existing DL-based methods, enhancing the overall fusion quality for infrared and visible images.