ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr391

# Multi-Modal Feature Extraction Network for Medical Image Fusion

V. Kowsalya[1]; D. Hariprasath[2]

[1] Master of Computer Application,
[2] Assistant Professor,

[1,2] Department of Computer Applications, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.

Publication Date: 2025/04/19

**Abstract: Medical image fusion is to synthesize multiple medical images from single or different imaging devices. This paper aims to improve imaging quality with accurate preserving for accurate diagnosis and treatment. This work plays an important role in the fields of surgical navigation, routine staging, and radio-therapy planning of malignant disease. Nowadays, computerized tomography (CT), magnetic resonance imaging (MRI), single-photo emission computed tomography (SPECT) modalities, and positron emission tomography (PET) are focused using medical image fusion. Bones and implants are clearly reflected by CT Image. High-resolution anatomical details for soft tissues are recorded using MRL images. However, the MRI image is not sensitive to the diagnosis of fractures compared to CT image. SPECT image is utilized to study the blood flow of tissues and organs by nuclear imaging technique. Our proposed work is Multi-Modal Based Medical Image fusion for directly learning image features from original images. Medical image fusion is a powerful tool that enhances the clinical value of individual imaging modalities, leading to better patient outcomes. As imaging technology advances and computational techniques evolve, the role of image fusion in modern medicine continues to grow.**

***Keywords:*** *Medical Image Fusion, Computed Tomography, Magnetic Resonance Imaging and Multi-Modal Image Fusion.*

**How to Cite:** V. Kowsalya; D. Hariprasath (2025) Multi-Modal Feature Extraction Network for Medical Image Fusion. *International Journal of Innovative Science and Research Technology*, 10(4), 577-582.https://doi.org/10.38124/ijisrt/25apr391

## I. INTRODUCTION

Medical Image Fusion refers to the process of combining information from multiple medical images, typically from different imaging modalities, to create a single composite image that provides a more comprehensive view of the anatomy or pathology. This fusion can help to improve diagnostic accuracy, treatment planning, and overall patient care. The common Imaging Modalities used in fusion are CT (Computed Tomography): Provides detailed cross-sectional images of the body's internal structures, offering high spatial resolution but limited soft tissue contrast.MRI (Magnetic Resonance Imaging): Offers superior soft tissue contrast, particularly useful for imaging the brain, spinal cord, and muscles, but with lower spatial resolution compared to CT.PET (Positron Emission Tomography): Provides functional information about metabolic activity and tissue function, commonly used for detecting cancer. SPECT (Single Photon Emission Computed Tomography): Similar to PET, but uses different tracers to provide functional information about organs and tissues. Ultrasound: Provides real-time imaging and is used for soft tissues, including during procedures like biopsies. The main benefits of Medical Image Fusion are Improved Diagnosis: Combining the anatomical detail from CT/MRI with the functional or metabolic data from PET/SPECT can lead to better understanding of the disease state. Enhanced Visualization: Increases the clarity and understanding of complex or ambiguous images, aiding clinicians in making more informed decisions. Precise Treatment Planning: In radiation therapy, image fusion helps accurately plan the targeting of tumors while minimizing exposure to surrounding healthy tissue. Surgical Planning and Navigation:[3] Provides 3D reconstructions that help in planning complex surgeries, like brain surgery or organ transplantation. The main techniques for Medical Image Fusion are Intensity-Based Fusion: Directly combines pixel intensities from different images. This method is often used in simple fusion tasks where the images are aligned spatially[5]. Feature-Based Fusion: Focuses on aligning specific features (e.g., edges, points, or contours) from the images before combining them. This technique is more robust in cases where the images may not align perfectly. Transform-Based Fusion: Uses transformations (such as rigid or non-rigid registration) to align images from different modalities before merging them into one composite image. Wavelet Transform: A more advanced technique that breaks down images into multiple

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr391

frequency components, which can then be combined at different scales for improved fusion. Deep Learning Approaches: Recent advancements in artificial intelligence and deep learning are being applied to image fusion, where convolutional neural networks (CNNs) or generative adversarial networks (GANs) are used to fuse images in a way that preserves important anatomical and functional details.

## II. RELATED WORK

Medical image fusion using deep learning has revolutionized the ability to combine multiple imaging modalities, enabling the creation of high-quality fused images that provide comprehensive structural and functional information. This integration is crucial for diagnosis, treatment planning, and monitoring of diseases. Here are some of the main deep learning models and techniques applied in medical image fusion.

- *Convolutional Neural Networks (Cnns):*

CNNs are widely used in image fusion due to their power in extracting spatial features. In medical image fusion[4], CNNs help capture intricate details from different modalities, such as MRI and CT.

- *Single-Scale CNNs:*

These networks can be used to learn fusion rules and extract features at a single resolution, applying them across both modalities to create a fused image.

- *Multi-Scale CNNs:*

Multi-scale approaches capture details at different resolutions, which helps retain both high-frequency and low-frequency information. This is particularly useful in applications where different modalities, such as PET and CT, need to be merged to combine functional and anatomical details.

- *Residual Cnns:*

Using residual connections within CNNs, these networks can preserve critical features and retain essential information from each modality, reducing the risk of information loss during fusion.

- *Generative Adversarial Networks (Gans)*

GANs are a powerful tool in image synthesis and fusion, as they consist of a generator that creates fused images and a discriminator that learns to differentiate real and fused images, making the fused images more realistic and informative.

- *Fusiongan:*

This model is specifically tailored for image fusion. The generator combines multi-modality inputs, while the discriminator evaluates the fused image quality, guiding the generator to preserve critical information.

- *Conditional GANs (cGANs):*

These GANs use specific inputs as conditions to generate a fused output, allowing controlled fusion of images. For instance, an MRI can conditionally guide the fusion to focus on soft tissue, while a CT image emphasizes bone structure.

- *Cyclegan:*

CycleGANs can perform fusion when paired data (like MRI and CT scans of the same patient) are unavailable. By learning mappings between modalities, CycleGANs help in cross-modality translation while retaining the key details of each input.

- *Auto Encoders:*

Auto encoders, which are designed to learn compact representations of data, are often used in image fusion because they can reduce complex images to essential features and then reconstruct a fused image.

- *Stacked Auto encoders:*

These models stack multiple layers of encoders and decoders, learning low-dimensional representations from different modalities, which are then combined to form a comprehensive image.

- *Variational Auto Encoders (Vaes):*

By introducing a probabilistic layer, VAEs can capture uncertainty in the fused output. This is useful in cases where images have noise or missing information, such as ultrasound and MRI fusion, where ultrasound may lack detail or contain artifacts [10].

- *Deep Feature Encoders:*

These specialized autoencoders focus on extracting deep features from each modality, combining them in a way that enhances relevant details in the fused image.

- *Attention Mechanisms:*

Attention mechanisms improve fusion quality by allowing models to selectively focus on important regions within each modality, preserving the critical features in the fused image.

- *Self-Attention Networks:*

Self-attention allows the model to focus on relevant regions in each image, enhancing fusion in areas with high clinical interest, such as tumor sites in brain MRI and PET fusion.

- *Cross-Modality Attention:*

In cross-modality attention, features from each modality guide the network to emphasize areas of interest. For example, in MRI-PET fusion, the model can learn to emphasize functional areas from PET while highlighting structural areas in MRI.

- *Transformer-Based Attention:*

Transformer models, which are built around attention mechanisms, can be used for fusion by focusing on long-range dependencies within the image data,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr391

allowing for better contextual understanding of multi-modal information.

- *Hybrid Deep Learning Architectures:*

Combining different deep learning models allows for improved feature extraction and fusion capabilities.

- *Cnn-Rnn Architectures:*

This hybrid model combines CNNs for spatial feature extraction with RNNs (recurrent neural networks) to capture temporal changes or context in dynamic imaging, such as fusing real-time ultrasound with static MRI.

- *Cnn-Transformer Models:*

Combining CNNs' spatial feature extraction with transformers' ability to model long-range dependencies has proven effective in high-resolution fusion tasks, such as combining 3D MRI and PET images for detailed tumor characterization.

- *GAN with Attention:*

Adding attention layers to GANs can enhance their fusion abilities by helping the generator focus on important regions from each modality, especially in applications like fusing high-resolution MRI and PET for better visualization of brain structures.

- *Multi-Modal Feature Extraction Networks*

Multi-modal feature extraction networks use separate branches to extract specialized features from each modality before combining them, preserving each modality's unique attributes.

- *Feature Fusion Networks:*

These networks apply different branches for each modality, which are later merged through fusion layers. This architecture is particularly useful for modalities with stark differences, like fusing MRI and CT, where MRI provides soft tissue contrast and CT highlights bone structure.

- *Deep Fusion Modules:*

Deep fusion modules integrate feature maps from each modality at multiple layers within the network. For example, in fusing CT and PET for oncology, deep fusion modules can repeatedly merge and refine feature maps, producing high-quality fused outputs that retain essential functional and structural information.

- *Self Supervised Learning :*

Self-supervised learning is an effective way to train fusion models with minimal labeled data, which can be advantageous in medical imaging, where labeled datasets are often scarce.

- *Contrastive Learning:*

In contrastive learning, pairs of similar and dissimilar images are used, where pairs from the same modality are marked as similar, and different modality pairs are marked as dissimilar. This enables the model to learn significant features that can later enhance fusion tasks.

- *Pretext Tasks:*

Tasks like image rotation prediction or image patch jigsaw puzzles enable the model to learn robust representations without explicit labels. Once trained on these tasks, models can be fine-tuned for fusion to leverage learned spatial relationships. The Applications of Deep Learning in Medical Image Fusion are Deep learning-based image fusion has found applications across various clinical areas, including:

- *Neuroscience:*

Fusion of MRI and PET for brain imaging provides detailed views of brain structures alongside functional activity, helping with disease localization and assessment of neurological disorders like Alzheimer's and tumors.

- *Oncology:*

Fusing CT and PET images enables precise tumor localization, staging, and monitoring, allowing clinicians to assess structural and metabolic data concurrently.

- *Cardiology:*

Combining real-time ultrasound with MRI or CT can help visualize cardiac motion and structure simultaneously, aiding in the diagnosis and treatment of heart diseases.

- *Orthopedics:*

Fusing MRI and CT images provides a detailed view of bone and soft tissue, which can be critical for pre-surgical planning and treatment of musculoskeletal injuries.

## III. PROPOSED WORK

In our proposed work deep learning has significantly advanced medical image fusion by creating high-quality, informative fused images from different modalities. These methods allow for a more holistic understanding of complex medical data and are poised to play an increasingly prominent role in personalized medicine, diagnostics, and treatment planning. Multi-modal feature extraction networks for medical image fusion utilize distinct feature extraction layers or "branches" for each imaging modality. Each branch is specifically trained to capture important features unique to that modality, such as high-contrast areas in MRI for soft tissue or high-intensity regions in PET for metabolic activity. Once features from each modality are extracted, they are fused through a series of fusion layers that combine the extracted information to create a single, comprehensive output image. The Key Components of Multi-Modal Feature Extraction Networks are Feature Extraction Branches: Each modality (e.g., MRI, CT, PET) has a dedicated branch in the network, which processes the input image and extracts modality-specific features. For instance: An MRI branch could focus on soft tissue and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr391

anatomical details. A CT branch could emphasize bone structures and high-contrast areas. A PET branch might highlight regions with high metabolic activity. Feature Fusion Layers: After feature extraction, fusion layers combine features from each modality. Fusion techniques may include concatenation, summation, or more complex strategies like attention-based weighting or learned fusion strategies. Multi-Scale Feature Integration: Many networks include multi-scale fusion, where feature maps from each modality are fused at different levels of granularity, enabling the final output to retain both high-level (contextual) and low-level (fine-grained) details. Output Layer: The fused features are passed through a final output layer to create a comprehensive fused image that provides a clearer representation of the combined information.

➢ *Input Images*

- *MRI Image:*
  High spatial resolution, highlighting soft tissue structures.

- *PET Image:*
  Lower spatial resolution, but highlights regions with high metabolic activity (e.g., potential tumor areas).

➢ *Feature Extraction Branches*

- *MRI Branch:*
  Processes the MRI input through several convolutional layers, extracting features related to anatomical structures.

- *PET Branch:*
  Processes the PET input through its convolutional layers, emphasizing areas with metabolic activity.

➢ *Feature Fusion*

- *Concatenation Fusion:*
  The extracted features from the MRI and PET branches are concatenated to form a joint feature map.

- *Attention-Based Fusion:*
  An attention mechanism weighs features from each modality, emphasizing PET data in regions with high metabolic activity and MRI features in regions with detailed anatomy.

➢ *Multi-Scale Fusion*
  Fused feature maps are merged at different resolutions, allowing the network to combine both high-resolution structural and functional details into a single output.

➢ *Output Layer*
  A final convolutional layer processes the fused features to produce a single fused image, retaining essential information from both MRI and PET.

The output of a multi-modal feature extraction network for MRI and PET fusion would be a fused image that provides the following insights:

- *Enhanced Anatomical Detail:*
  Retains the soft tissue detail from MRI.

- *Metabolic Hotspots:*
  Clearly highlights regions of high metabolic activity from PET, which might suggest areas of concern such as tumors.

➢ *Visual Representation of Sample Output*

- *MRI Image (Input):*
  Shows clear soft tissue structures, with limited functional data.

- *PET Image (Input):*
  Displays regions of high metabolic activity but lacks structural clarity.

- *Fused Image (Output):*
  Combines the anatomical detail from MRI with the functional hotspots from PET, providing a single, clear image that highlights potential tumor locations with accurate anatomical context.

In practice, such fused outputs are highly beneficial in clinical settings, as they allow clinicians to see structural and functional data within a single image, enhancing diagnostic accuracy and aiding in treatment planning.

For a multi-modal image fusion process, here's what typical input images might look like for MRI and PET scans of the brain:

➢ *MRI Image (Input):*

- *Description:*
  This input image is grayscale, showing high-resolution structural details of the brain. There is a clear differentiation between gray matter, white matter, and cerebrospinal fluid (CSF) spaces. It provides excellent anatomical information but lacks functional data, which is where the PET scan becomes valuable.

- *Characteristics:*
  High spatial resolution, excellent for observing tissue contrasts but lacks metabolic information.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr391

➢ *PET Image (Input):*

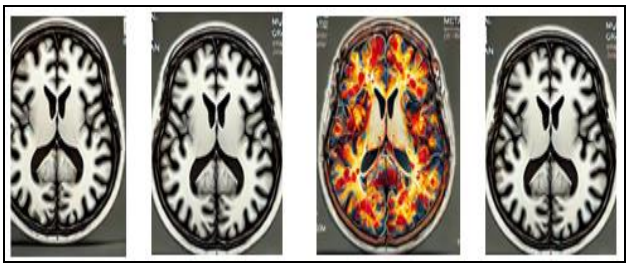
- *Description:*

The PET scan is usually in color, often with warm colors (e.g., reds, yellows) representing areas of high metabolic activity, which can indicate regions of increased glucose uptake—often associated with tumor activity or other areas of interest.

- *Characteristics:*

Lower spatial resolution compared to MRI, but highlights functional metabolic processes in the brain.

MRI – IMAGES



PET IMAGES

Fig 1 Input Image

These two inputs as shown in figure 1, with MRI providing structural clarity and PET highlighting functional hotspots, are then processed by the multi-modal feature extraction network. This fusion provides clinicians with an enhanced single output image combining both types of data, assisting in diagnosis and treatment planning. Here are sample input images typically used in multi-modal medical image fusion. The MRI image is in grayscale, showing detailed anatomical structures, while the PET image displays functional metabolic activity highlighted in warm colors, ideal for combining structural and functional insights in a fused output.

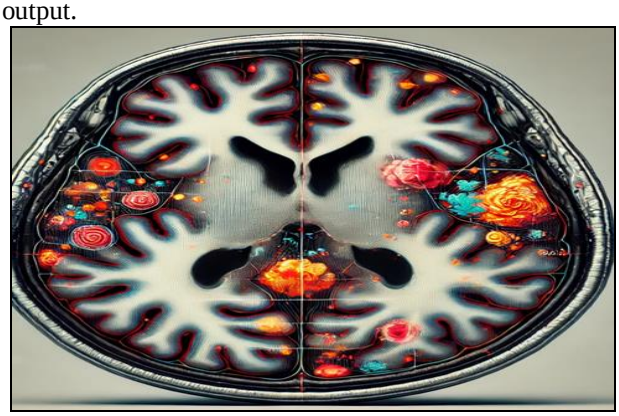


Fig 2 Output Image

The fused medical image as shown in figure 2 combining MRI and PET data. It displays clear anatomical structures from the MRI with PET's metabolic activity areas overlaid, providing a comprehensive view often used in diagnostic imaging to locate and assess areas of concern such as tumors.

## IV. CONCLUSION

Image Fusion is an essential technique to combine the input received from various imaging modality . In this work, Brin images are taken as a input for computation. MRI images and PET images are given as input for the Machine learning techniques. Machine learning techniques with various kernel function are applied and the suitable for Gaussian kernel has been concluded for this fusion technique due to nonlinear data points.

## REFERENCES:

[1]. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 2004, 13, 600–612.

[2]. Wang, Q.; Shen, Y.; Zhang, J.Q. A nonlinear correlation measure for multivariable data set. Phys. D Nonlinear Phenom. 2005, 200, 287–295.

[3]. Koroleva, O.A.; Tomlinson, M.L.; Leader, D.; Shaw, P.; Doonan, J.H. High-throughput protein localization in Arabidopsis using Agrobacterium-mediated transient expression of GFP-ORF fusions. Plant J. 2005, 41, 162–174.

[4]. He, C.; Liu, Q.; Li, H.; Wang, H. Multimodal medical image fusion based on IHS and PCA. Procedia Eng. 2010, 7, 280–285.

[5]. Ismail, W.Z.W.; Sim, K.S. Contrast enhancement dynamic histogram equalization for medical image processing application. Int. J. Imaging Syst. Technol. 2011, 21, 280–289.

[6]. Han, Y.; Cai, Y.; Cao, Y.; Xu, X. A new image fusion performance metric based on visual information fidelity. Inf. Fusion 2013, 14, 127–135.

[7]. Du, J.; Li,W.; Xiao, B.; Nawaz, Q. Union Laplacian pyramid with multiple features for medical image fusion. Neurocomputing 2016, 194, 326–339.

[8]. Du, J.; Li,W.; Xiao, B. Anatomical-functional image fusion by information of interest in local Laplacian filtering domain. IEEE Trans. Image Process. 2017, 26, 5855–5866.

[9]. Jiang, W.; Yang, X.; Wu, W.; Liu, K.; Ahmad, A.; Sangaiah, A.K.; Jeon, G. Medical images fusion by using weighted least squares filter and sparse representation. Comput. Electr. Eng. 2018, 67, 252–266.

[10]. Ma, J.; Xu, H.; Jiang, J.; Mei, X.; Zhang, X.P. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Trans. Image Process. 2020, 29, 4980–4995.

[11]. Dong, X.; Bao, J.; Chen, D.; Zhang, W.; Yu, N.; Yuan, L.; Chen, D.; Guo, B. Cswin transformer: A general vision transformer backbone with cross-shaped windows.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr391

In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 12124–12134.

[12]. Wu, S.; Wu, T.; Tan, H.; Guo, G. Pale transformer: A general vision transformer backbone with pale-shaped attention. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; pp. 2731–2739.

[13]. Li, C.; Zhou, A.; Yao, A. Omni-Dimensional Dynamic Convolution. In Proceedings of the International Conference on Learning Representations, Virtual, 25–29 April 2022.

[14]. Li, W.; Peng, X.; Fu, J.; Wang, G.; Huang, Y.; Chao, F. A multiscale double-branch residual attention network for anatomical– functional medical image fusion. Comput. Biol. Med. 2022, 141, 105005.

[15]. Lamba, K.; Rani, S. A novel approach of brain-computer interfacing (BCI) and Grad-CAM based explainable artificial intelligence: Use case scenario for smart healthcare. J. Neurosci. Methods 2024, 408, 110159.