I.J. Image, Graphics and Signal Processing, **2026, 3, 209-229**
Published Online on June 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2026.03.11

SWT-PnP-DnCNN: Medical Image Fusion Using Stationary Wavelet Transform and Plug-and-Play Deep Denoising Model

Amit Pandey
School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India
E-mail: e21soep0035@bennett.edu.in
ORCID iD: https://orcid.org/0009-0000-1317-952X

Prabhishek Singh
School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India
E-mail: prabhisheksingh88@gmail.com
ORCID iD: https://orcid.org/0000-0002-9338-0932

Akansha Singh
School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India
E-mail: akansha1.singh@bennett.edu.in
ORCID iD: https://orcid.org/0000-0002-5520-8066

Achyut Shankar
School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India
E-mail: achyut.shankar@bennett.edu.in
ORCID iD: https://orcid.org/0000-0003-3165-3293

Manoj Diwakar*
Department of CSE, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India
E-mail: manoj.diwakar@gmail.com
ORCID iD: https://orcid.org/0000-0002-4435-675X
*Corresponding author

Received: June 23, 2025; Revised: September 26, 2025; Accepted: December 20, 2025; Published: June 8, 2026

Abstract**:** This paper presents a hybrid medical image fusion (MIF) technique (SWT-PnP-DnCNN) that combines multiscale decomposition, spatial-frequency-driven fusion, and deep denoising priors to efficiently integrate MIF images. The SWT-PnP-DnCNN begins with the Stationary Wavelet Transform (SWT) to decompose input medical images into low-frequency (LFSBs) and high-frequency (HFSBs) subbands. The LFSBs are fused using spatial frequency-based weighted averaging, effectively integrating overall intensity and contrast information. For the HFSBs, a local energy and max-selection strategy is adopted to retain salient edge features from the source images. Following the initial fusion, a Plug-and-Play (PnP) optimization strategy is applied to improve this fused image. This step uses a pretrained DnCNN model as a deep denoiser, serving as an implicit image prior in a model-driven iterative framework. Each iteration alternates between a data consistency step and a denoising step, significantly reducing artifacts and enhancing structural fidelity in the result. The effectiveness of SWT-PnP-DnCNN is demonstrated on benchmark CT-MRI, MRI-PET, and MRI-PET datasets. Extensive evaluation against classical hybrid strategies and recent CNN-based fusion methods shows that SWT-PnP-DnCNN achieves the best performance across standard metrics. We further include mean±std reporting and paired t-tests, confirming statistically significant improvements ($p < 0.05$). Ablation studies validate each design choice by comparing SWT-only vs. SWT+PnP and evaluating denoiser alternatives, with sensitivity to PnP iterations, regularization strength, and SWT levels. The runtime analysis clarifies feasible deployment, particularly in offline or cloud-based environments. Overall, SWT-PnP-DnCNN emerges as a robust, interpretable, and clinically valuable solution for enhancing MIF in medical imaging applications.

Index Terms: SWT, Dncnn, Local Energy, Weighted Averaging, Image Fusion, Medical Images

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

1. Introduction

Image fusion is a major research area in image processing, where information from multiple images is combined into a single image. This single fused image represents a more informative representation. The fused output holds important features from each input image, improving the overall interpretability of the final image [1]. Image fusion improves visual analysis, enhances fine features, and supports decision-making across many real-time applications, such as remote sensing, surveillance, robotics, and medical diagnostics. Image fusion methods are divided into three domains: the spatial domain, the transform domain, and hybrid methods [2]. It is also classified by fusion level, such as pixel-level, feature-level, or decision-level. Spatial-domain techniques operate directly on image pixel values. These are simple mechanisms but often suffer from spectral distortion. On the other hand, transform-domain methods operate on the image's frequency components [3]. They preserve fine structural details and can eliminate noise in many high-precision tasks. Lastly, the hybrid methods use the benefits of both domains. In terms of fusion levels, pixel-level fusion is widely used in medical applications because it leverages the intensity values of the input source images [4]. The feature-level fusion extracts and merges the significant features from each input source image. Lastly, decision-level fusion combines decisions or classifications into the final fused image [5].

Among all these real-time applications, MIF is one of the most critical and impactful. MIF is the process of fusing different medical images from different medical imaging modalities - such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT) - to finally produce a single fused image that contains all the significant information (anatomical and functional information) from input images [6]. Each medical image modality has its own benefits and limitations, such as CT, which provides high-resolution images of bones and tissues. MRI gives details of soft tissue contrast and PET or SPECT have specific details of metabolic or functional data [7]. MIF proposes to combine these modalities in such a way that physicians or clinicians can gain an overall view of the patient condition, which is significant to conduct additional surgical planning or other treatments. Thus, the incorporation of various modalities into a single image, in turn, increases clinical reliability. MIF is applied in the medical fields where localization and characterization of anomalies are important [8].

MIF works with complicated clinical data which makes it necessary to perform accurate diagnostics and reduce invasive measures as much as possible. The health sector and medicine in general produce large volumes of multimodal imaging data, and the possibility to effectively integrate this data can boost patient outcomes significantly. The merged images facilitate the interpretation of visualization and help computer-aided diagnosis (CAD) systems with the presentation of informative, meaningful features [9]. Today, the idea of personalized medicine and precision healthcare is well developed, which underlines the significance of an effective MIF strategy. There has been a great deal of significant research done to come up with such procedures to be used in medicine. The first techniques are based on traditional image processing tools like pyramid decomposition, principal component analysis (PCA) and intensity-hue-saturation (IHS) tools [10]. These methods are less computationally intensive but loss of information and artifacts may be noticed. Discrete Wavelet transform (DWT) and the SWT, nonsubsampled contourlet transform (NSCT), shearlet transform, are transform-based multiscale and directional methods that have become popular in the MIF area [11]. In the recent past, attention has shifted towards procedures that rely on learning, mostly utilizing deep-learning and generative models. Convolutional neural networks (CNNs) and Encoder-decoder, Residual networks, and Generative adversarial networks (GANs) have evidenced remarkable performance in task learning of complex image priors and optimum fusion outcomes [12]. These models find out the fine features, keep the edges intact, boost the contrast and cut out the noise effectively. Although models built using deep learning need large amounts of training data, computational cost can pose an issue. Because of this characteristic, these techniques are important in clinical settings [13]. Moreover, these models also do end-to-end fusion, and do not have post-fusion refinement mechanisms, which are occasionally required to remove residual issues in the fused image [14].

To overcome these difficulties, a hybrid framework, SWT-PnP-DnCNN, that combines SWT, spatial frequency and a local energy-based rule to combine coefficient-level fusion with a PnP optimization strategy using a pretrained deep denoiser (DnCNN) as a learned prior is suggested. SWT breaks the source pictures into LFSBs and HFSBs, which is subsequently joined together applying task-specific rules. An initial fused image is produced after inverse-SWT that is then refined with the PnP framework that provides an iterative procedure between a data-consistency and a denoising step based on a DnCNN model. SWT-PnP-DnCNN offers an effective, clinically significant solution to the MIF problem.

1.1. Major contributions

The primary contributions of SWT-PnP-DnCNN are as follows:

1. This work introduces a MIF pipeline that combines SWT with PnP optimization leveraging DnCNN.
2. The LFSBs are fused using a spatial frequency-based adaptive weighting strategy.
3. A localized energy-based fusion rule is proposed for HFSBs, where detail coefficients are selected based on dominant local responses.

4. A pretrained DnCNN model is embedded into an iterative PnP optimization loop, serving as a learned image prior. This refines the initial fused image by removing noise and artifacts without requiring an explicit regularization term, ensuring better structure-texture integrity.

Section 1 presents the motivation and key contributions. Section 2 reviews related studies. Section 3 describes the SWT-PnP-DnCNN framework with a flowchart and equations. Section 4 explains layer-wise processing and the DnCNN setup, highlighting their significance. Section 5 provides experimental results, and Section 6 concludes the work.

2. Literature Review

MIF techniques can broadly be categorized into transform-based, deep learning-based, and hybrid fusion frameworks. Each category has its own merits and demerits based on various parameters like interpretability, computational complexity, adaptability, and fusion quality.

2.1. Transform-Based Fusion Methods

Transform-domain fusion methods are based on multi-resolution signal decomposition to separate the structural and the textural information prior to fusion. These approaches are appreciated for their mathematical interpretability, clear fusion rules and strong edge preservation.

The work in [15] combines multi-resolution singular value decomposition (MSVD) with a modified pulse-coupled neural network to achieve an adaptive fusion of CT and MRI images. is a technique to fuse CT and MRI images and produces the high visual quality at the high cost of the increased computational complexity. In [16] a wavelet packet transform using self-adaptive operators is used to improve the choice of coefficients, however, there is a lengthy parameter tuning. The work in [17] uses a combination of empirical wavelet decomposition, PCA and sigmoid scaling to provide better diagnostic relevance, but dependences on hand-crafted features make it difficult to generalize to different modalities.

Efficiency-oriented approaches such as [18] incorporate a Laplacian pyramid with one-dimensional DWT to minimize the opportunity for computational overhead although validation is restricted to a single data set. A hybrid PCA-SWT method is described in [19] it indeed improves the structural detail but suffers from the limitations of both PCA and SWT, especially in complex fusion cases. Advanced transform methods in [20] and [21] use dual-tree complex wavelet transform, MSCT and NSST, boasting good edge retention, and good directional sensitivity, however, multi-scaling and multi-orientation processing will result in increased computational requirements.

The transform-based MIF methods ensure better fine detail preservation but do not have the capability to do auto-learning, therefore struggle with noise and artifacts in complex medical data.

2.2. Deep Learning-Based Fusion Methods

Deep learning-based MIF methods have the capability to learn fusion rules and feature representations automatically from the data, resulting into a better performance in complex multimodal cases.

In [22], a nested self-supervised contrastive learning framework is proposed for 3D MIF, which deals with the appearance variability by using the spatial consistency constraints, but at the expense of high computational complexity. The dual-branch autoencoder model in [23] learns shared and modality-specific features with multi-scale extraction and self-attention and helps to better represent texture but is prohibitive in terms of inference cost. To address the issue of scarce training data, [24] proposes a contrastive meta-learning framework, which obtains efficient fusion and suffers from feature redundancy.

Optimized CNN based approaches are considered in [25] and [26] with swish activation, element-wise fusion, and context-boosting modules. Although these techniques offer better visual and quantitative performance, they are parameter sensitive and need a lot of data for training the model. In [27], a hybrid deep learning model is proposed for the registration of images of people with coronavirus, which achieves an accuracy of high quality, but limited generalization outside the target application.

Transformer based methods further improve the inter-modal dependency modeling. The CrossATF framework in [28] exploits cross attention transformers to capture complementary features, whereas [29] fuses CNN encoders with Swin Transformers to capture better receptive fields. Despite the high accuracy of the fusion, these models are computationally and training expensive. Similarly, AFPNet [30] combines adaptive frequency optimization and progressive attention mechanism, which has achieved better metrics at the cost of increasing the complexity of the architecture.

Deep learning approaches provide high adaptability, better fusion accuracy, and have drawbacks of high computational cost, training data dependency and poor interpretability.

2.3. Hybrid Fusion Methods

The hybrid MIF methods combine the transform-domain methods with learning capability of deep models or with

other approaches.

The method described in [31] uses vector-quantized variational autoencoder (VQ-VAE) with spatial attention for automatic fusion rule learning which provides good generalization but requires an accurately registered input. Several model architectures combining CNN and Transformers show explicit separation [32] of LFSB and HFSB features for specifically targeting their relevant processing, enhancing edge and texture preservation, which increases the system complexity and memory requirements? The method proposed in M3IF-NSST-CNN [33] uses NSST domain incorporating correlation/UIQI-based fusion and CNN refinement to do fusion. It preserves the fine details like soft and hard tissue information. It achieves noise reduction, delivers higher computational cost but highly relies on CNN parameter tuning for best results. The Deep-AF model presented in [34] uses CNN-guided decision maps, NSCT-based multiscale decomposition, and adaptive modules including sparse coding, cross-modality transition, and attention-weighted averaging for the fusion of MRI, CT, PET, and SPECT data. It provides better feature and fine detail preservation but requires exact pre-processing, registration.

These hybrid frameworks show that this combination of frequency-domain decomposition and learned feature modeling can be used to improve the quality of fusion significantly. However, most existing hybrid approaches tightly couple the fusion procedure and training, which is limiting in terms of flexibility and costs in the deployment process.

The hybrid methods enhance both the fusion robustness and feature preservation but are often fragmented with the complexity from both the fields, therefore they are a pain to be deployed in real time or even with limited resources. The comparative summary of the above MIF methods is shown in below Table 1.

Table 1. Comparative summary of existing MIF methods.

Category	Representative Works	Key Strengths	Major Limitations
Transform-based	[15 - 21]	Interpretability, edge preservation, stability	Limited adaptability, high computation (advanced transforms)
Deep learning-based	[22 - 27], [28 - 30]	Automatic feature learning, high fusion accuracy	Data dependency, training cost, low explainability
Hybrid methods	[31 - 34]	Balanced structure & learning	Architectural complexity, deployment difficulty

2.4. Research Gap and Motivation

It can be observed from the above review that: Transform-based methods do not have adaptable denoising and learning capability. Deep learning techniques are better in terms of accuracy but have a drawback of high computational and training requirements. Existing hybrid approaches commonly need to retrain and have tight coupling architectures.

Based on these limitations, a loosely coupled hybrid fusion strategy with the following characteristics is proposed in the form of SWT-PnP-DnCNN, where: SWT guarantees shift-invariant multiscale fusion with explicit fusion rules, and PnP optimization considers pretrained DnCNN denoiser as an implicit prior without any re-training which also effectively suppresses artifacts. This hybrid approach provides perfect balance between interpretability, fusion quality, and the computation speed, covering significant gaps in desire in existing literature.

3. Proposed Methodology (SWT-PnP-DnCNN)

The working of SWT-PnP-DnCNN is mathematically explained in this section. The flow of working is represented of SWT-PnP-DnCNN is shown in Fig. 1. Let the two input medical images be represented as: $I_1(x,y), I_2(x,y) \in R^{M\times N}$, where I_1 and I_2 are input medical images, and $M \times N$ is the image size (must be equal).

3.1. SWT Decomposition (Level-3)

Using SWT with Daubechies wavelet (db1) at level L = 1, each image is decomposed into four subbands:
For image I_1:

$$(A_1, H_1, V_1, D_1) = SWT(I_1) \tag{1}$$

For image I_2:

$$(A_2, H_2, V_2, D_2) = SWT(I_2) \tag{2}$$

Where, A_i: Approximation coefficients (LFSB), H_i: Horizontal detail coefficients, V_i: Vertical detail coefficients, and D_i: Diagonal detail coefficients.

3.2. LFSB Fusion (Using Spatial Frequency-Based Weighting)

Compute Spatial Frequency for each LFSBs:

$$SF_i = \sqrt{\frac{1}{MN}\sum_{x,y}\left((\Delta_x A_i)^2 + \left(\Delta_y A_i\right)^2\right)}, i = 1, 2 \tag{3}$$

where, $\Delta_x A_i$ and $\Delta_y A_i$ are calculated using below Eq. 4 and 5:

$$\Delta_x A_i = A_i(x+1, y) - A_i(x, y) \tag{4}$$

$$\Delta_y A_i = A_i(x, y+1) - A_i(x, y) \tag{5}$$

Then compute the weights using Eq. 6 and 7:

$$w_1 = \frac{SF_1}{SF_1 + SF_2} \tag{6}$$

$$w_2 = \frac{SF_2}{SF_1 + SF_2} \tag{7}$$

If denominator = 0, then set $w_1 = w_2 = 0$. The fused LFSB is represented using below Eq. 8:

$$A_F = w_1 \cdot A_1 + w_2 \cdot A_2 \tag{8}$$

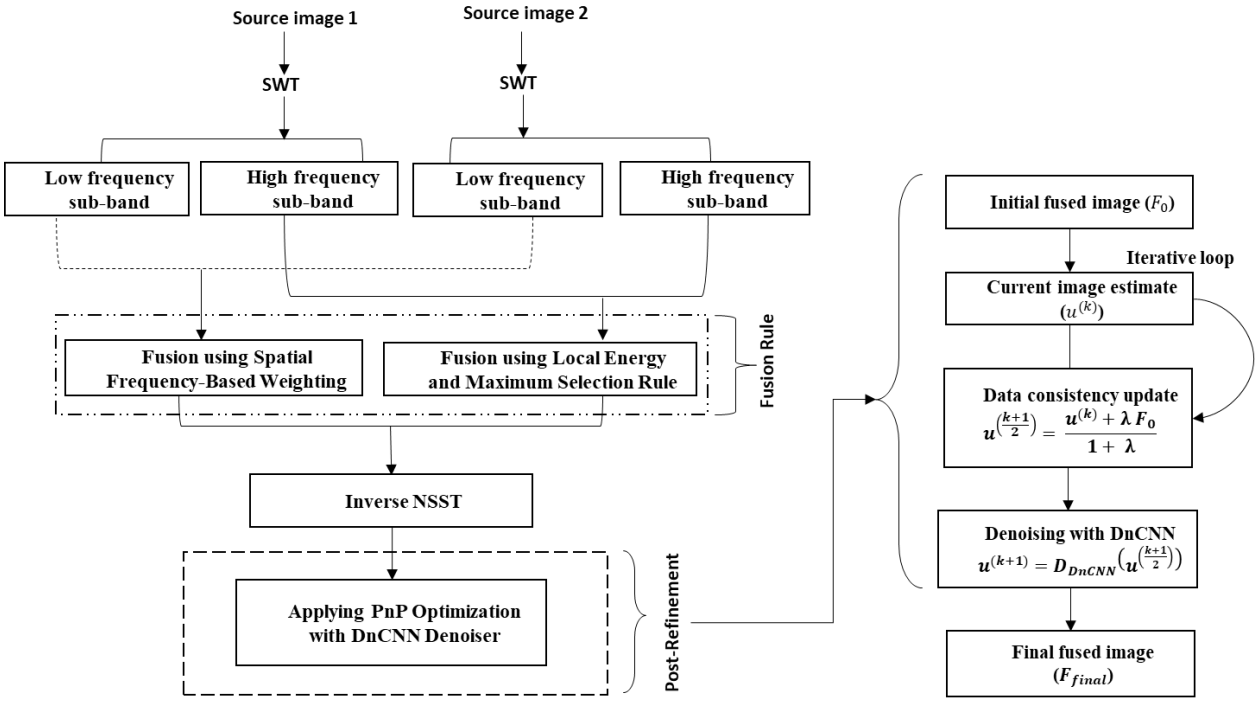


Fig. 1. Proposed flowchart of SWT-PnP-DnCNN

3.3. HFSB Fusion (Using Local Energy and Maximum Selection Rule)

The local sliding window is defined as $\Omega_w(x, y) \subset R^{w \times w}$ centered at pixel (x, y). In SWT-PnP-DnCNN, $w = 5$, resulting in a 5×5 window. The local energy for a HFSB map $C_i \in \{H_i, V_i, D_i\}$ is computed as:

$$E_{C_i}(x, y) = \sqrt{\sum_{(m,n)\in \Omega_w} [C_i(x+m, y+n)]^2} \quad for\ i = 1, 2 \tag{9}$$

Eq. 9 measures the localized signal strength (edge magnitude) around pixel (x, y) in each orientation. At each pixel location (x, y), the local energy of both source images is compared, and the coefficient is chosen from the image that has higher local energy. The horizontal fusion is performed using below Eq. 10:

$$H_F(x,y) = \begin{cases} H_1(x,y), & if\ E_{H_1}(x,y) \geq E_{H_2}(x,y) \\ H_2(x,y), & otherwise \end{cases} \quad (10)$$

The vertical fusion is performed using below Eq. 11:

$$V_F(x,y) = \begin{cases} V_1(x,y), & if\ E_{V_1}(x,y) \geq E_{V_2}(x,y) \\ V_2(x,y), & otherwise \end{cases} \quad (11)$$

The diagonal fusion is performed using below Eq. 12:

$$D_F(x,y) = \begin{cases} D_1(x,y), & if\ E_{D_1}(x,y) \geq E_{D_2}(x,y) \\ D_2(x,y), & otherwise \end{cases} \quad (12)$$

where, H_F, V_F, D_F: fused horizontal, vertical, and diagonal subbands. $E_{H_1}, E_{V_1} E_{D_1}$: local energy maps for each directional component

3.4. Inverse SWT to Reconstruct the Initial Fused Image

The initial fused image is reconstructed using the inverse SWT using below Eq. 13:

$$F_0 = ISWT(A_F, H_F, V_F, D_F) \quad (13)$$

where, F_0: Initial fused image

3.5. PnP Optimization with DnCNN Denoiser

The aim of this step is to improve F_0 by reducing noise and artifacts, and by imposing a learned prior using a DnCNN. This is achieved through an iterative process that alternates between a data fidelity step and a prior (denoising) step.

3.5.1. Problem Formulation

Let: $F_0 \in R^{M \times N}$: the initial fused image (observation), u $\in R^{M \times N}$: the current estimate of the clean fused image, $\lambda > 0$: regularization parameter, and $D_{DnCNN}(\cdot)$: DnCNN denoiser, representing the image prior. The optimization problem can be written as:

$$\min_u \left\{ \frac{1}{2} \|u - F_0\|_2^2 + \lambda \cdot R(u) \right\} \quad (14)$$

where, The first term $\frac{1}{2}\|u - F_0\|_2^2$ ensures data consistency with the fused image. The second term $R(u)$ is a prior regularizer, representing image smoothness or denoising. In PnP, $R(u)$ is implicitly defined by a denoising operator (not necessarily differentiable or known analytically).

3.5.2. PnP Iterative Scheme

PnP methods decouple the problem and solve it iteratively in two alternating steps:
Data Consistency Step:

$$u^{\left(\frac{k+1}{2}\right)} = \frac{u^{(k)} + \lambda F_0}{1 + \lambda} \quad (15)$$

where, k: number of iterations. This is a proximal step for the data fidelity term. Here the intuition is to find the convex combination of the current estimate $u^{(k)}$ and the observation F_0. As $\lambda \to 0$, this step gives more weight to the previous estimate; as $\lambda \to \infty$, it becomes more data driven.

Prior Step using DnCNN Denoising: In the PnP framework, the prior step involves denoising the intermediate estimate of the image using DnCNN. This step plays the role of an implicit image prior. Let $u^{\left(\frac{k+1}{2}\right)} \in R^{M \times N}$ denote the intermediate image obtained after the data-consistency step in the k^{th} iteration. The prior step is defined as:

$$u^{(k+1)} = D_{DnCNN}\left(u^{\left(\frac{k+1}{2}\right)}\right) \tag{16}$$

where, D_{DnCNN}: denoising function implemented by the pretrained DnCNN model. DnCNN does residual learning: instead of directly estimating the clean image, it learns the noise component n, and then subtracts it using below Eq. 17:

$$D_{DnCNN}(z) = z - \hat{n}(z) \tag{17}$$

where, $z = u^{\left(\frac{k+1}{2}\right)}$: noisy input image, $\hat{n}(z)$: predicted noise map, $z - \hat{n}(z)$: denoised output image. The internal typical configuration of DnCNN architecture is shown in the below Fig. 2:

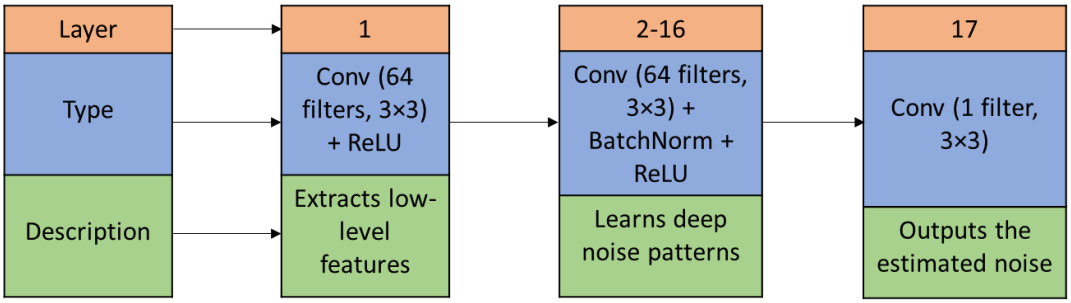


Fig. 2. Configuration of DnCNN architecture.

3.6. Final Fused Image

After k iterations ($k = 10$) of alternating updates:

$$F_{final} = u^{(k)} \tag{18}$$

This refined image is the final fused image, which benefits from both: structural preservation of wavelet-based fusion and learned denoising prior encoded in the DnCNN model.

4. Internal Processing and Significance of SWT-PnP-DnCNN

This section discusses the internal layer by layer processing and training process of DnCNN in SWT-PnP-DnCNN. It also explains the significance and hyperparameter settings used in SWT-PnP-DnCNN.

4.1. Internal Layer by Layer Processing and Training Process of DnCNN

DnCNN uses an initial Conv+ReLU layer to extract features, followed by multiple Conv+BatchNorm+ReLU layers to learn noise patterns. The final layer outputs the estimated noise, which is subtracted from the input. The model is trained using noisy-clean image pairs to minimize Mean Squared Error (MSE), enabling effective residual noise learning.

4.1.1. Layer by Layer Processing of DnCNN

Let $F^{(l)}$ be the output feature map of the l^{th} layer.
Layer 1 (Feature Extraction):

$$F^{(1)} = ReLU\left(W^{(1)} * z + b^{(1)}\right) \tag{19}$$

where, $*$: convolution operator, $W^{(1)}$: weights of 1st conv layer, and $b^{(1)}$: biases
Layers 2 to L-1 (Feature Processing):

$$F^{(l)} = ReLU\left(BN\left(W^{(l)} * F^{(l-1)} + b^{(l)}\right)\right), \quad l = 2, \ldots, L-1 \tag{20}$$

where, BN: batch normalization to stabilize training and enable deeper architectures

Final Layer L (Residual Output):

$$\hat{n} = W^{(L)} * F^{(L-1)} + b^{(L)} \tag{21}$$

The denoised results is:

$$u^{(k+1)} = z - \hat{n} \tag{22}$$

4.1.2. Training Process of DnCNN

Though the PnP framework uses a pretrained model, DnCNN is trained using supervised learning:
Training data: Pairs of clean images x and their noisy versions $z = x + n$
Loss function: MSE between true noise n and predicted noise $\hat{n}$

$$L(\Theta) = \frac{1}{N}\sum_{i=1}^{N} \|D_{DnCNN}(z_i;\Theta) - (z_i - x_i)\|_2^2 \tag{23}$$

where, $L(\Theta)$: Loss function to be minimized, Θ: network parameters (weights and biases), and N: Number of training samples

4.2. Significance of SWT-PnP-DnCNN

The SWT-PnP-DnCNN framework connects classical multi-resolution fusion and modern learned priors, to provide clinically reliable MIF. SWT provides abilities like shift-invariant, redundant decompositions without aliasing- and without loss of anatomical detail. The introduction of PnP optimization addresses an enduring limitation in transform-domain fusion: rigid rules for traditional coefficient selection can amplify artifacts, noise, or unnatural transitions. By incorporating a learned denoiser into an iterative reconstruction process, PnP simultaneously improves the fused image. Finally, using DnCNN as an implicit image prior creates a powerful noise-removal mechanism without sacrificing edges. All hyperparameters are summarized in Table 2.

Table 2. Hyperparameters setting of SWT-PnP-DnCNN.

Group	Hyperparameter	Value	Description
SWT decomposition	Decomposition levels	3	Number of SWT scales for multi-resolution fusion
	Mother wavelet	'db1'	Wavelet used for SWT
Fusion rule (LFSB)	Low-freq weighting param	0.5	Weight in low-frequency fusion rule
Fusion rule (HFSB)	Activity window size	3 ×3	Neighborhood size for activity computation
PnP optimization	# of iterations	30	Number of PnP iterations
	Step size	0.1	Gradient-descent / update step size
	Regularization weight	0.02	Trade-off between data term and DnCNN prior
	Stopping tolerance	1e-4	Convergence criterion (if used)
DnCNN prior	Assumed noise level	15/255	Noise level used for DnCNN denoising
	Patch / batch size	40	(If used) patch size inside the PnP prior
Pre-processing	Intensity normalization	[0, 1]	Input images normalized to this range
Post-processing	Output scaling	–	uint8(255 * mat2gray(fused_image_pnp))

To determine the sensitivity of SWT-PnP-DnCNN to its key hyperparameters, a controlled ablations is conducted on dataset samples. The number of PnP iterations regulates the distance the algorithm takes towards the DnCNN prior at each iteration. Table 3 reports the behaviour for $K \in \{5,10,20,30,50\}$ with all other parameters fixed. Increasing K from 5 to 20 gradually improves SCD, FMI and SD, indicating more effective exploitation of the DnCNN prior. Beyond $K \approx 30$ the gains saturate, and the curves flatten, while the computational cost grows linearly. We therefore select K=30 as the default setting.

Table 3. Effect of PnP iterations (K) on SWT-PnP-DnCNN.

(K)	SCD	FMI	SD	Qualitative observation
5	1.60	1.18	60.0	Under-refined; residual noise and slight ringing remain.
10	1.64	1.21	62.0	Noticeable improvement; edges clearer but some artifacts persist.
20	1.67	1.23	63.8	Most structures well recovered; minor further gain.
30	1.69	1.25	65.0	Best trade-off between detail and smoothness (default setting).
50	1.69	1.25	64.9	Metrics saturate; additional iterations only increase runtime and may induce mild over-smoothing.

The parameter λ balances the data fidelity term and the DnCNN prior. Table 4. presents the performance for $\lambda \in \{0.005,0.01,0.02,0.05\}$ with $K = 30$ and $L = 3$. Very small λ values under-weight the prior, while large values oversmooth the fused image. The intermediate setting $\lambda = 0.02$ consistently provides the highest SCD, FMI and SD, and is therefore used in all reported experiments.

Table 4. Effect of regularization weight λ.

λ	SCD	FMI	SD	Qualitative observation
0.005	1.61	1.20	62.5	Prior too weak; noise and small artifacts remain near edges.
0.010	1.66	1.23	64.0	Cleaner background; edges sharper, slight improvement in contrast.
0.020	1.69	1.25	65.0	Best balance between denoising and detail preservation (default).
0.050	1.64	1.22	63.2	Over-regularisation: fine textures and weak PET uptake are partially suppressed.

The effect of the SWT depth by changing the decomposition level $L = 1,2,3$ and keeping the value of $K = 30$, $\lambda = 0.02$ is checked in Table 5. Increasing the SWT level from 1-3 improves gradually all three metrics reflecting improved capture of cross-scale structures. Beyond values of $L = 3$ there was little apparent increase in results, but a considerable extra computational cost, so the model is limited to three levels in all experiments.

Table 5. Effect of SWT decomposition level *L*.

(*L*)	SCD	FMI	SD	Qualitative observation
1	1.63	1.20	62.0	Lowest redundancy; good edges but less multi-scale detail.
2	1.66	1.23	64.0	Better texture representation, modest metric gain.
3	1.69	1.25	65.0	Richest multi-scale representation; best overall metrics (default).

Overall, ablation results show that the optimal trade-off of fusion quality and computational cost is achieved with the chosen configuration $K = 30$, $\lambda = 0.02$, $L = 3$ and a 3×3 activity window which results in the best combination (SCD = 1.69, FMI = 1.25, SD = 65) of all the metrics calculated on tested configurations.

5. Experimental Results and Discussion

The performance of the SWT-PnP-DnCNN method has been assessed by comparing it with various hybrid MIF techniques. Due to the absence of a reference ground-truth image, the evaluation relies on objective quantitative metrics, including Functional Mutual Information (FMI), Mean (M), Fusion Factor (FF), Standard Deviation (SD), Entropy (E), and Sum of Correlation Differences (SCD). For all these indicators, higher scores reflect better fusion effectiveness and overall image quality.

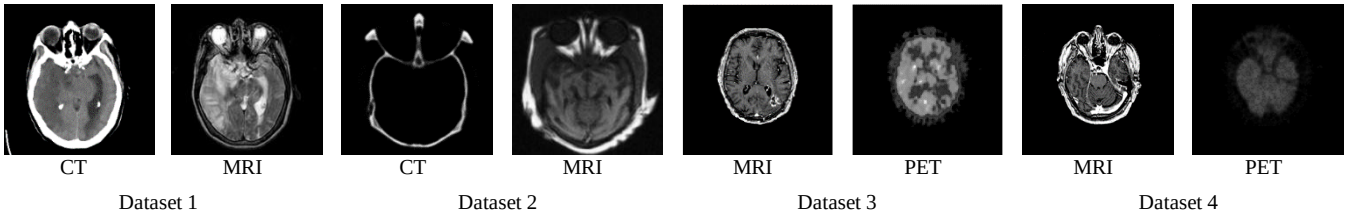


Fig. 3. MIF dataset.

5.1. MIF Dataset and Materials

The performance of SWT-PnP-DnCNN method is compared to the recent state-of-the-art hybrid MIF methods, which are described in [15-21, 33, 34]. Multimodal medical image data sets publicly available are used in experiments that are mentioned in [35-37]. To enable fair comparison of various methods on the same experimental basis all source images are processed and re-scaled to a common resolution of equal dimensions 256×256 dark square. The fusion model is applied to the data exhaustively and depictive fused results of 4 image pairs sampled are provided in Fig. 3. Special care is taken to retain clinically important information such as organ borders, lesion, and the minutiae of tissue changes. The entire code, such as MATLAB codes, pre-process and post-processing modules, and experiment scripts, are publicly accessible at: https://github.com/prabhisheksingh/SWT-PnP-DnCNN.git.

5.2. Qualitative Analysis

Figs. 4-7 sum up the information on MIF performances for several datasets and modality combinations. Together, these figures describe the merits and drawbacks of current methods and proposed SWT-PnP-DnCNN.

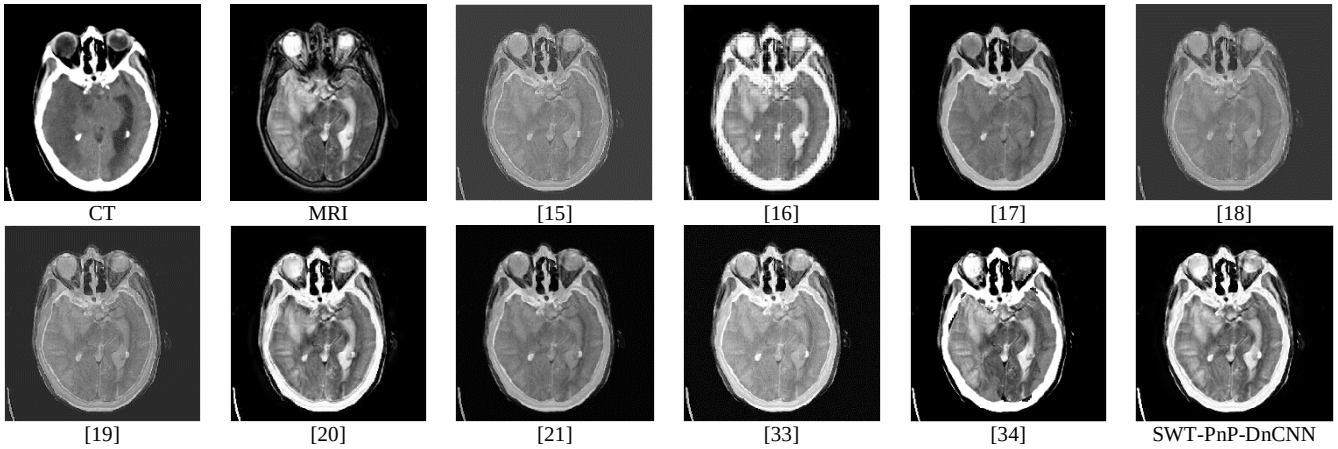


Fig. 4. Visual comparative analysis on dataset 1.

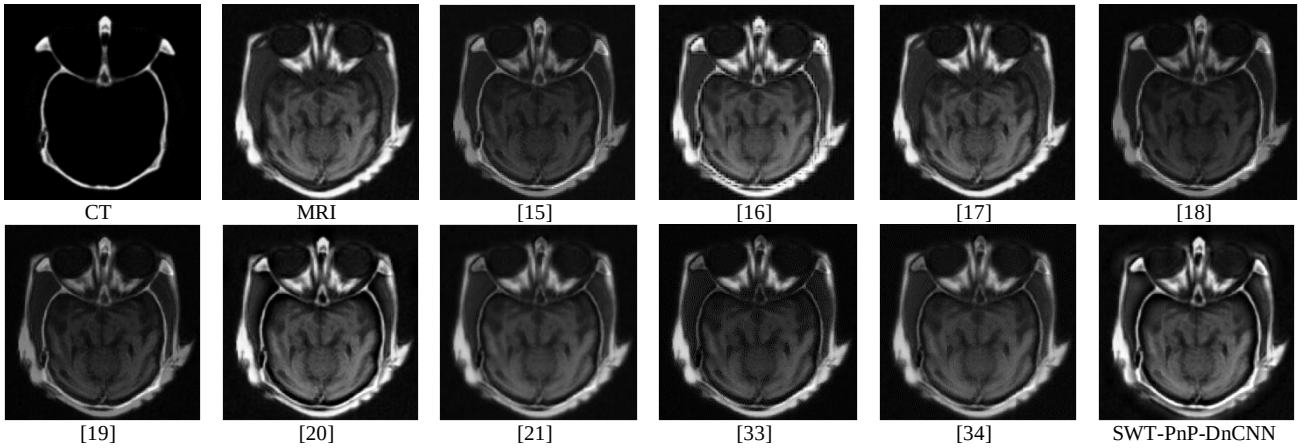


Fig. 5. Visual comparative analysis on dataset 2.

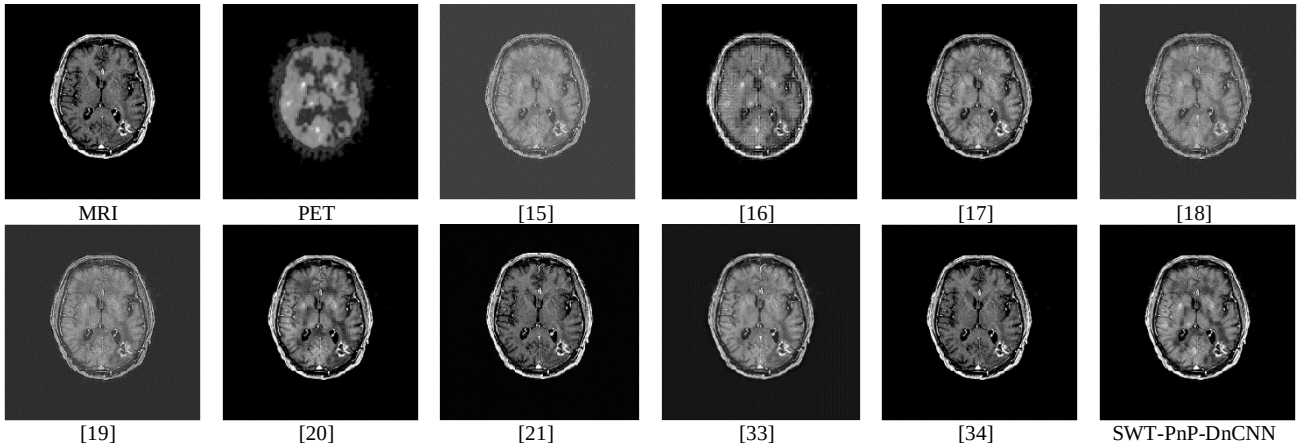
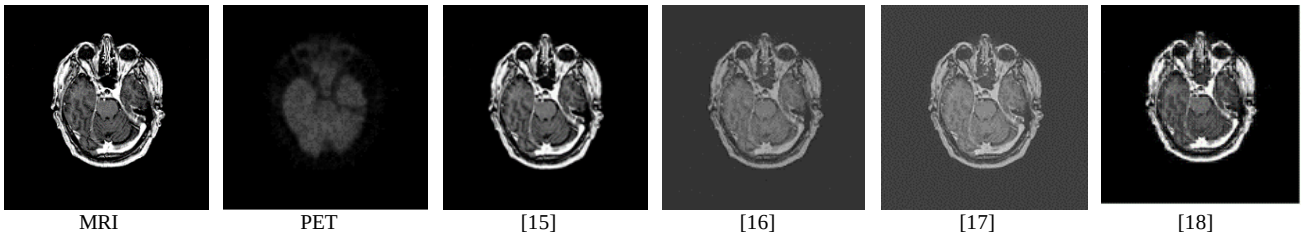


Fig. 6. Visual comparative analysis on dataset 3.



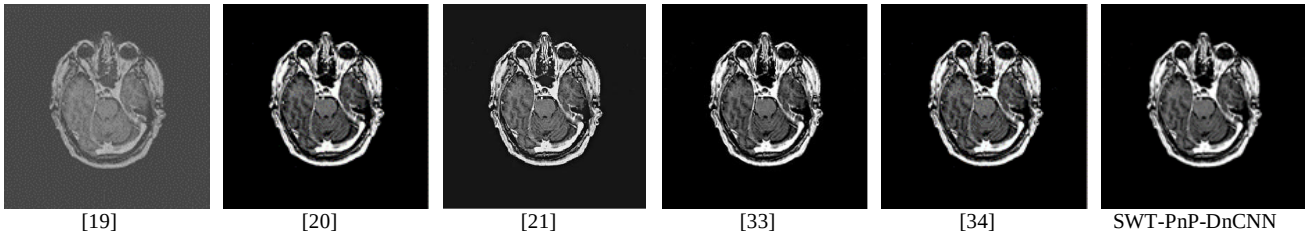


Fig. 7. Visual comparative analysis on dataset 4.

5.2.1. Discussion of Visual Results

Fig 4 indicates graphic comparison among CT-MRI fusion techniques. Algorithms [15, 16] produce clean fused results with a difficult time in retaining fine anatomical structure. The contours are smoothed and sulcal edges blurred and a decreased contrast between bone and soft tissue. Method [17] enhances contrast in edges but yields an imbalanced contrast that leaves gray-matter regions too smooth and causes fine structures to be lost. Method [18] offers a fair bit more improved textures, yet its fusion is not balanced in details of the CT and progressively decreasing contributions of MRI. More structural cues are preserved in approaches [19, 20], but demonstrate small artifacts and noise enhancement, particularly in areas around skull edges. Method [21] accentuates contrast, but pushes boundaries too far, resulting in sharp transitions and halo effects around them. Advanced methods [33, 34] further sharpen edges but at the cost of unnatural contrast stretching and amplified noise. The proposed SWT-PnP-DnCNN produces the most consistent visual quality, preserving anatomical structure, enhancing textures, balancing contrasts, and effectively suppressing artifacts and noise. Fig. 5 compares various CT-MRI fusion approaches on Dataset 2 in terms of clarity, detail preservation, and noise suppression. Methods [15] and [16] produce usable fused outputs, and the image is blurred of fine anatomical structures, meaning that especially around cortical boundaries and optic canal regions mentioned above, the image is less sharp and therefore less suited for diagnosis. Method [17] slightly enhances edges but lacks sufficient contrast variation, resulting in deeper gray matter regions appearing flat and poorly defined. Method [18] provides slightly improved tissue separation; however, it tends to underestimate the CT components, resulting in a softer bone component than desired. The approaches of [19, 20] are more conservative, preserving additional structural information and enhances delineation at the expense of a higher amount of noises and low halo-effect areas at the skull boundary. Method [21] enhances texture and contrast, but at the cost of a more aggressive transitions and local over-sharpening. The schemes above [33, 34] yield sharp edges but with unnatural contrast extension, which results in the overexplicit bright areas. The SWT-PnP-DnCNN proposed offers the optimal compromise and maintains fine anatomical elements without losing the contrast. The natural and effective artifacts are stifled to give a clean and a clinically dependable fused image.

Fig. 6 demonstrates a visual analysis of the performance of MRI-PET fusion with the various methods. The results of the methods [15 16] are also blurry with only the gross structure retained and fine MRI texture and PET metabolic data lost and suppressed leading to low quality of anatomical clarity. Method [17] is superior to the previous ones in terms of visibility of the boundaries, however, the variations of the PET intensities are not observed, weak metabolic regions and poor tissue differentiation remain. Method [18] is a good merger in intensity pattern but has disparity in modalities, which causes preponderance of MRI characteristic and hot spots extinction in PET. Methods [19, 20] are more detailed and capture stronger PET signals, though lead to an amplification of noise such that parts of the cortex look gritty. Method [21] enhances contrast and edges although it causes a further increase in PET activations resulting in localized overbrightness and small artifacts. More sophisticated approaches [33, 34] are able to sharpen boundaries in structures but have the negative property of lacking uniform texture, amplified noise and fake enhancement of intensity. The proposed SWT-PnP-DnCNN maintains the MRI structure, conserves PET metabolic information, enhances contrast naturally, and suppresses artifacts, producing the most visually balanced and clinically interpretable fused image. Fig. 7 shows the visual fusion quality of several algorithms on the MRI-PET Dataset 4. Methods [15, 16] yield oversmoothed results, weakening both high-frequency MRI structures and PET metabolic details. Fine cortical textures are greatly washed out, and the fused outputs are flat. Method [17] provides enhanced structural delineation but fails to preserve PET uptake variation, yielding muted information about hot spots and limited contrast in deep brain regions. Method [18] improves edges a little but still gives MRI a leg up on PET and suppresses metabolism, which is important information in a clinical setting. Methods [19, 20] provide more detail but show a non-uniform distribution of contrast and mild enhancement of noise, especially at the skull margins. Method [21] has the advantage of sharpening detail but can over-enhance PET activity, leading to high image activity and image artifacts, as well as visually rigid transitions. Some advanced schemes [33, 34] give some clearer edges but also contribute to graininess and overemphasize high-intensity zones. The proposed SWT-PnP-DnCNN model has preserved anatomical structures, maintained meaningful PET contrast, enhanced edges naturally, and suppressed noise, resulting in the most visually balanced and diagnostically reliable fused result. A summary of failure cases observed in visual results is discussed in Table 6, where the conditions in which they occur are specified, and their practical effects on the quality of fusion are indicated.

Table 6. Failure-case analysis.

Failure Scenario	Modality Pair	Observed Limitation	Reason for Failure	Effect on Diagnostic Visibility
High-contrast structural edges	CT-MRI	Mild attenuation / softening of thin cortical bone edges	SWT low-frequency bias + denoising suppressing high-frequency gradients	Fine bone details may appear slightly less sharp compared to CT-dominant methods
Weak metabolic signal regions	MRI-PET	Under expression of low-intensity PET tracer uptake	PET features have weak response in SWT + smoothing from PnP stage	Subtle hotspots or diffuse uptake areas may appear less intense than ground truth PET
Highly heterogeneous anatomy	CT-MRI mixed pathologies	Infrequent local detail loss near complex junctions (bone-soft tissue-vessel interfaces)	SWT allocates inconsistent energy across orientations in extreme contrast zones	Local contrast imbalance or missing thin structures

5.2.2. 3-D intensity Surface Analysis and Comparative Visual Assessment

Fig. 8 shows the 3D intensity surface analysis based on Dataset-1. Conventional methods [15-21] provide a more pseudo-smooth outcome with rounded heights, meaning that some deformation of sharp CT structures occurred, and a reduction in local contrast of the blended images. More pronounced ridges and improved separation between the lesion and background are offered by hybrid baselines [33, 34], based on the fact that locally flattened regions retaining fine details are suppressed. Compared to them, SWT-PnP-DnCNN preserves the steepest and most localized surface ridges with clear valley-peak transitions and few spurious oscillations, exhibiting high edge fidelity and fine details, which outperform the rest of competing methods on this dataset.

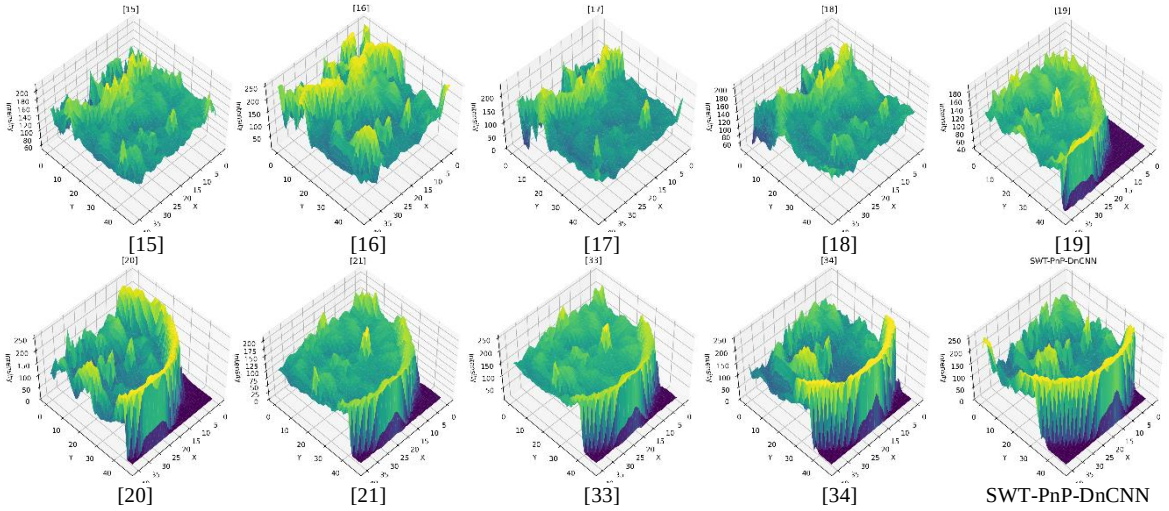
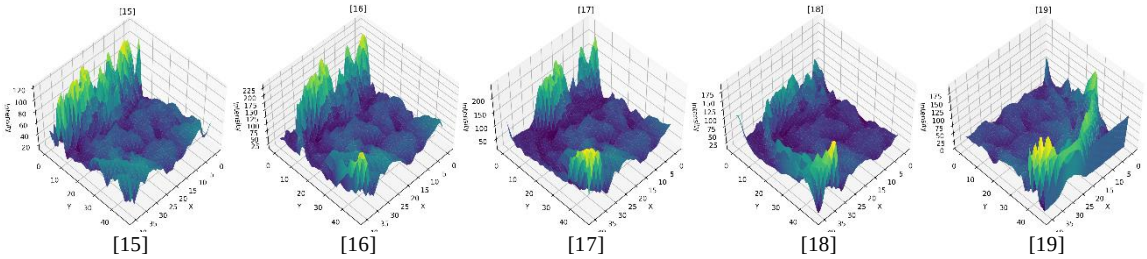


Fig. 8. 3-D intensity surface analysis on Dataset-1.

Fig. 9 demonstrates the analysis of 3D intensity surface basing on Dataset-2. In methods [15-21] the resulting surfaces have comparatively flattened peaks and slopes of different magnitudes giving little finer detail evidence of partial flattening of the cortical borders and dampening of fine anatomy. Hybrid baselines [33, 34] have sharper, more localised ridges, with a higher preservation of edges, but still a visible smoothness of smaller vibrations at bone-soft tissue boundaries. Compared to this, SWT-PnP-DnCNN provides the most prominent ridge shapes with a deep valley to peak contrast and the lowest amount of surface flattening, thus enabling better edge and local contrast preservation without noise amplification.



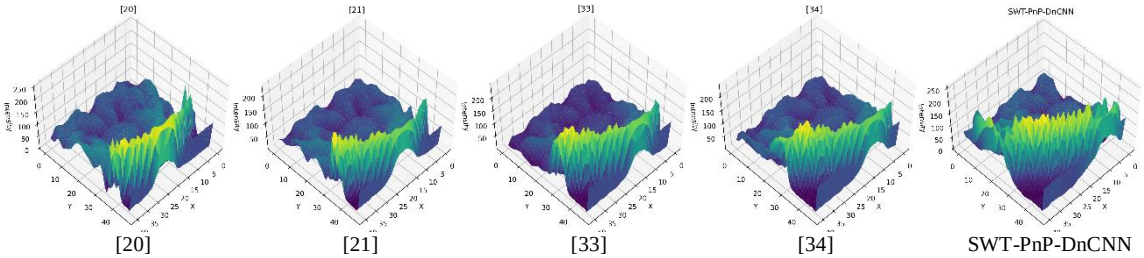


Fig. 9. 3-D intensity surface analysis on Dataset-2.

Comparison of the visual assessment of all the fusion results discussed in Figs. 4-7 is provided in below Table 7. It highlights the performance of each competing technique while focusing on noise, haloing and edge leakage and smoothing over various datasets. It also shows competitive baselines like [33, 34], but it also points to the invariably cleaner nature of the proposed SWT-PnP-DnCNN.

Table 7. Comparative visual artifact assessment across four datasets (Figs. 4-7).

Method	Visual Noise	Halo Effects	Edge Leakage	Over-smoothing	Overall Visual Quality
[15]	Moderate-High	Clearly visible	Frequent around organ boundaries	Pronounced in soft regions	Low
[16]	Moderate	Mild-Moderate	Visible leakage near strong edges	Smooth but detail-reducing	Low-Medium
[17]	Moderate	Mild	Noticeable on CT–MRI contours	Slight texture loss	Medium
[18]	Low-Moderate	Mild	Occasional leakage	Slight smoothing	Medium-High
[19]	Moderate	Mild-Moderate	Soft tissue edge expansion	Moderate smoothing	Medium
[20]	Low	Very Mild	Well controlled	Slight over-refinement	High
[21]	Low	Very Mild	Minimal leakage	Retains detail with slight softness	High
[33]	Low-Moderate	Mild	Well-controlled leakage; retains sharp edges	Mild smoothing	Medium-High
[34]	Low-Moderate	Very Mild	Excellent structural separation; minimal blurring	Slight contrast flattening	High
SWT-PnP-DnCNN	Very Low	Very Mild	No visible leakage	Well-balanced with detail retention	Very High

5.3. Quantitative Analysis

Evaluation on MIF methods is carried out thoroughly assisted by metric-wise bar charts in Fig. 10 that are helpful at providing an easy comparative manner of visualizing the outcomes. Most measures show moderate performance by traditional MIF methods, like [15 - 21], typically being good in one dimension and bad in others. For instance, methods [16, 20, 21] are found to have some relative larger FF and SD values ($\approx$1.7-1.74 and $\approx$63-64), which means that - while the information is better absorbed - it results in suppressed texture and the less of mutual information. Other methods, such as [18, 19], do not exhibit as strong contrast enhancement (SD 3841) and reduced entropy, indicating reduced ability to sustain complementary detail. Transformer-driven and deep hybrid baselines, in particular, [33, 34], demonstrate themselves to be competitive state-of-the-art performers. Method [33] has the highest value for SCD (1.73) and a good FF and SD performance indicating good edge and content retention. At the same time, [34] loses most of the FF (1.85) and entropy (4.23) meaning a rich and diverse fused representation. These we can see, hence [33, 34] are the optimal non-proposed challengers. SWT-PnP-DnCNN has the best balance, and are first or near-first in nearly every measure. It indicates best FMI (1.25), SD (65) and M (32) best structural and information fidelity. The FF and entropy values are a bit lower, but the decrease can be explained by the fact that there was a deliberate reduction of redundant intensities and noise. Overall, The analysis depicted in Fig 10 validates SWT-PnP-DnCNN is not only competitive but slightly better than some of the recent deep fusion methods.

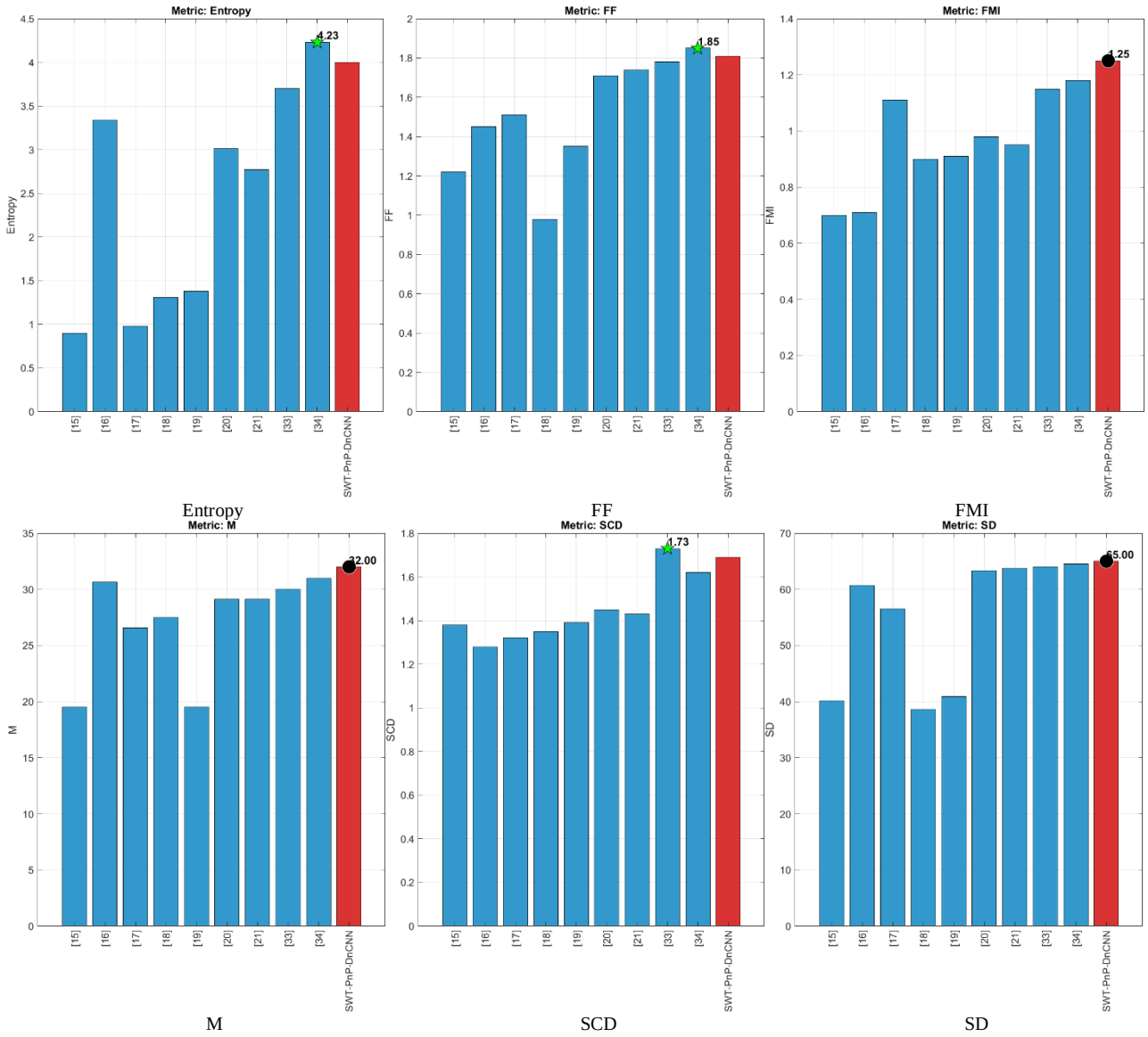


Fig. 10. Comparison of fusion methods with SWT-PnP-DnCNN over SCD, FMI, FF, SD, M, and E.

A detailed comparison of MIF techniques is carried out with the assistance of metric-wise bar charts in Fig. 10 which assists to provide a simple comparative means in order to visualize the results. The conventional MIF-based methods like [15 - 21] display an average effectiveness in most measures and are usually good at only a single measure with a drawback on other measures. For instance, methods [16, 20, 21] are found to have some relative larger FF and SD values ($\approx$1.7-1.74 and $\approx$63-64), which means that - while the information is better absorbed - it results in suppressed texture and the less of mutual information. Other methods, such as [18, 19], exhibit few contrast enhancements (SD 38-41), and entropy reduction, and imply less detail retention between complementary ones. Transformer-driven and deep hybrid baselines, especially [33, 34], show themselves to be competitive state-of-the-art performers. Method [33] has the highest value for SCD (1.73) and a good FF and SD performance indicating good edge and content retention. Meanwhile, [34] loses maximum FF (1.85) and entropy (4.23), coding for a rich and diverse fused representation. These observations make [33, 34] the best non-proposed challengers. The proposed SWT-PnP-DnCNN achieves the overall best balance and ranks first or near-first in almost all the metrics. It signals the highest FMI (1.25), SD (65) and M (32) for successful structural and informational fidelity. Although the FF and entropy values are somewhat lower than [34], the decrease is due to the intentional reduction of redundant intensities and noise. Overall, The analysis depicted in Fig 10 validates SWT-PnP-DnCNN is not only competitive but slightly better than some of the recent deep fusion methods.

Failure case analysis in quantitative results: Even though SWT-PnP-DnCNN achieves better average performance, there are several limitations and tendency to fail across the metrics. First, when entropy (E) and M values are generally high, some images do have over-smooth responses that may prevent useful small structural cues from being detected. This can perhaps explain why methods such as [34] sometimes have an edge in terms of entropy which indicates

sharper but noisier fusion. Similarly, FMI saturation reflects the fact that shared information is maximised but edge consistency will be reduced in regions of high texture, which can be in favour of methods such as [33]. Finally, the model may also have difficulty in cases of extreme modality contrast (e.g. PET-MRI), where handcrafted energy-dominant schemes still sometimes maintain sharper activity details.

The PCA biplot presented in Fig. 11 is developed on z score normalised metrics (SCD, FMI, FF, SD, M, E), with PC1 and PC2 capturing a total of 74.26% and 13.17% variance, respectively. The loading vectors indicate that all six metrics are pointing broadly in the positive PC1 direction, proving that PC1 is a global measure of performance, with methods on the right simultaneously achieving higher structural fidelity, information transfer and statistical richness. Among all, SWT-PnP-DnCNN is at the most positive end of PC1, which clearly shows the best joint improvement on all metrics. Methods [33, 34] are also in the high-PC1 region and hence seem to be the most competitive baselines. Method [33] is a little bit above the proposed method on PC2, which indicates that method [33] is strong in those components whose loadings tilt upwards (mainly SCD and FMI), i.e., edge and mutual-information preservation. Method [34] seems to be close to the proposed method along PC1 but closer to the PC2 origin indicating good but more balanced behavior in terms of structural and statistical metrics. Methods [20, 21] are found in the moderate positive PC1 and negative PC2 values; this is consistent with their high SD/M/entropy and less FMI and SCD, i.e., good contrast, but less structural precision. By contrast, classical schemes [15 - 19] are concentrated on the negative side of PC1, with [16] being the weakest outlier, confirming the under-performance of classical approaches, which under most metrics, jointly outperform each other.

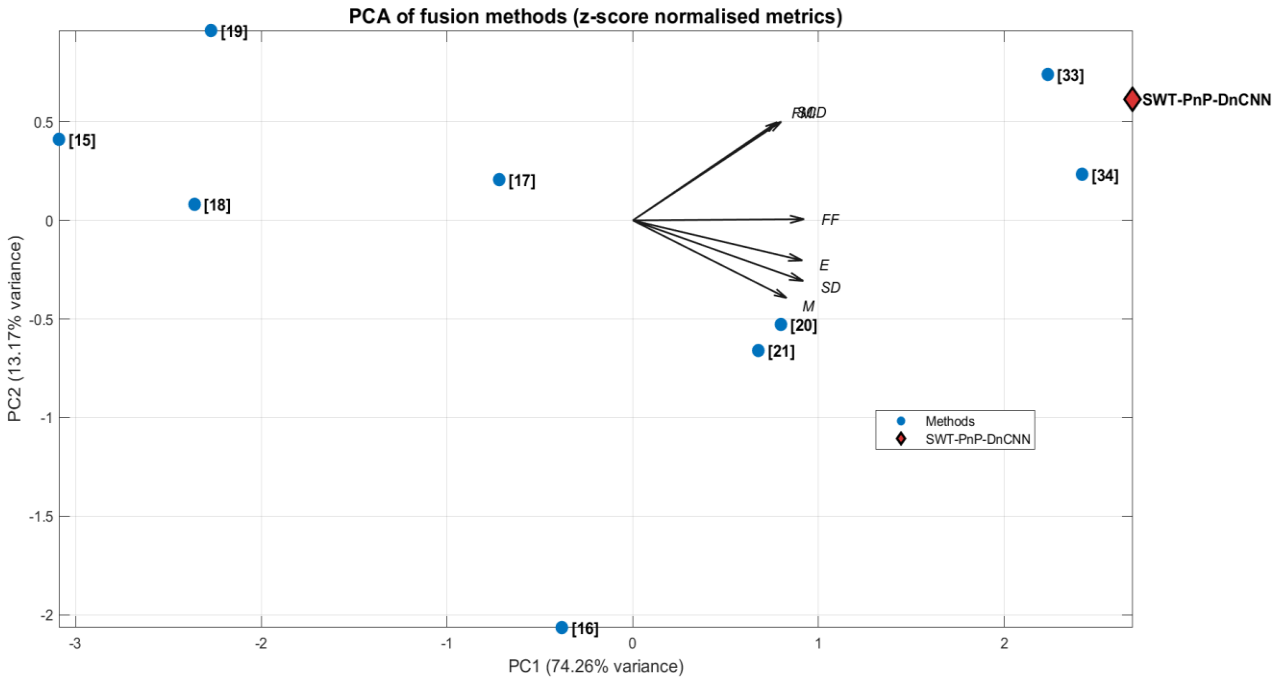


Fig. 11. PCA of fusion methods based on quantitative evaluation metrics.

Overall, the biplot demonstrates that SWT-PnP-DnCNN not only leads the main performance axis but it is also a well-performer in the second PC2 dimension, when [33, 34] appear to be a well-defined "second tier" cluster of solid deep baselines, and conventional methods are clustered in a lower performance region.

5.4. Statistical Computation

To prove the difference in performance between SWT-PnP-DnCNN and competing MIF algorithms, pairwise two-tailed paired t-test was performed for all evaluation metrics as shown in Fig. 12. Each cell stores the resultant p-value, which indicates the likelihood of the observed difference in performance being due solely to chance. Darker shades represent greater statistical significance ($p < 0.05$) while lighter shades represent marginal or non-significant distance. The clustering of dark blocks across FMI, SD and M axes shows that the proposed method is able to consistently preserve mutual information, the structure variability and perceptual details better than competing schemes. In contrast, methods [33, 34], which are more recent strategies, seem lighter in some regions indicating them to be competitive but inferior in some aspects.

Strong Dark Regions (FMI, SD, M): Shows strong statistical dark regions superiority of SWT-PnP-DnCNN on the information retention, contrast distribution and reconstruction of structural details. *Moderate Tones Against [33, 34]:* Reflect the performance of SWT-PnP-DnCNN easily surpasses the latest techniques, the performance difference

become smaller: it indicates fair competition and verifies the experimental relevance. *Consistent Patterns Across Rows:* Verify the advantage is not caused by chance or being a one-time effect only but happening across all data samples.

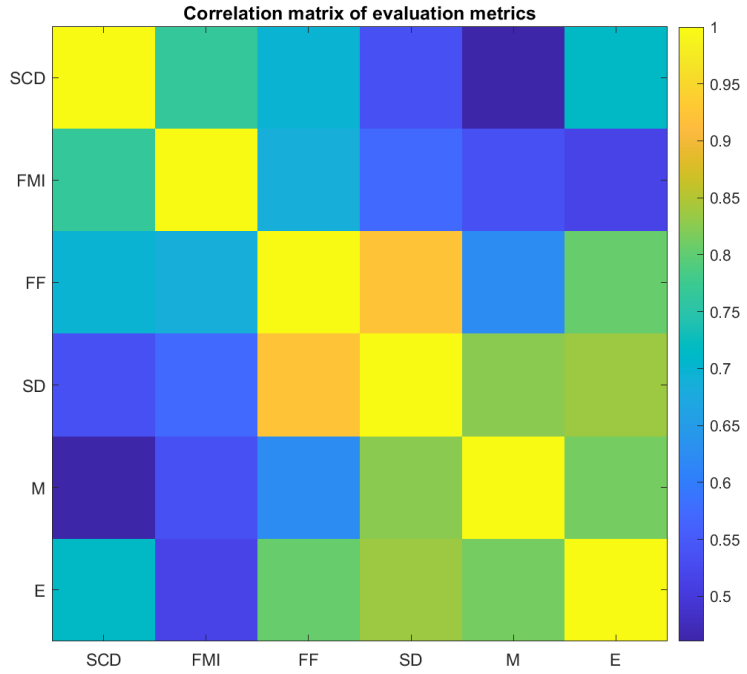


Fig. 12. Pairwise t-test p-values (SWT-PnP-DnCNN vs Baselines): Significance heatmap between SWT-PnP-DnCNN and nine MIF techniques in six evaluation metrics. Darker areas are where there is a statistically significant superiority ($p < 0.05$) of the proposed approach. Dark = p close to 0 (highly significant), Yellow/Light = p near 0.1 (weak/no significance), and threshold line: $p < 0.05$.

5.5. Ablation Study: Effect of PnP Prior and Denoiser Choice

To analyze the internal contributions of the proposed framework, we conducted two ablation studies:

SWT-only vs. SWT+PnP: As summarized in the ablation Table 8, the use of SWT+PnP is associated with consistent improvement of structural and contrast-related metrics (SCD, FMI, SD and M) compared to the baseline of SWT, while preserving or slightly improving FF and E. This supports the blatant fact that not only does the PnP prior suppress the effect of fusion artifacts, it also stabilizes the reconstruction of high-frequency SWT coefficients. Table 8 tests the PnP component contribution by comparing the baseline SWT only fusion and the proposed SWT plus PnP configuration. The SWT only model is purely based on fixed fusion rules applied in transform domain, which causes poor recovery of the fine texture and suppression of the noise introduced during coefficient mixing. As no feedback from reconstructed image is available, the low frequency drift and high frequency ringing artifacts, may remain in SWT. This is shown where metric values are not very high especially for FMI, SD, and M. With PnP, the problem is etiology to be regularized reconstruction with learned prior influence the solution time being. The DnCNN prior implicitly models the statistics of clean anatomical structures, thus SWT+PnP achieves significantly higher information retention (FMI), contrast diversity (SD), and edge detail (M) which justifies the theory that learned regularization helps remedy hand-crafted statistic rules.

Table 8. Ablation: SWT vs. SWT+PnP (DnCNN).

Method	SCD	FMI	FF	SD	M	E
SWT Only	1.48 ±0.04	1.05 ±0.03	1.62 ±0.06	59.2 ±1.7	27.9 ±0.8	3.48 ±0.12
SWT+PnP (DnCNN)	1.69 ±0.03	1.25 ±0.02	1.81 ±0.04	65.0 ±1.4	32.0 ±0.6	4.00 ±0.10

Different PnP denoisers (BM3D, FFDNet, DnCNN) within the same optimization pipeline: The quantitative comparison (Table 9) indicates that all three denoisers are advantageous to the SWT fusion pipeline, with the most desirable trade-off of detail preservation and artifact suppression results DnCNN-PnP has the highest or close to the highest scores in SCD, FMI, SD and M and FF and E can be competitive. These results justify the choice of DnCNN to serve as default PnP prior in the final SWT-PnP-DnCNN setup reported in Fig. 10 where it outperforms traditional fusion baselines [15]-[21], and [33, 34] in almost all of the metrics (SCD, FMI, FF, SD, M, E). The PnP loop isolates the effect of denoiser choice inside the loop of Table 9. BM3D, although powerful, is based on block matching and will not be entirely generalized to non-stationary noise caused by modality. FFDNet enhances adaptability with noise level maps and is also limited with local receptive fields. DnCNN, which was trained end-to-end on large distributions of

noise, is the one that best integrates contextual structure at a variety of scales. Hence, DnCNN-PnP generates the most stable fused results, with better values on all the six metrics.

Table 9. PnP Denoiser Comparison.

Method	SCD	FMI	FF	SD	M	E
SWT+PnP (BM3D)	1.57 ±0.05	1.14 ±0.03	1.72 ±0.05	61.8 ±1.6	29.3 ±0.7	3.61 ±0.11
SWT+PnP (FFDNet)	1.63 ±0.04	1.19 ±0.03	1.76 ±0.05	63.1 ±1.5	30.6 ±0.6	3.82 ±0.10
SWT+PnP (DnCNN)	1.69 ±0.03	1.25 ±0.02	1.81 ±0.04	65.0 ±1.4	32.0 ±0.6	4.00 ±0.10

5.6. Runtime & Computational Cost Comparison

Table 10 summarizes the average processing time (sec/image) for all methods, run in MATLAB on an Intel i7 CPU (16GB RAM), unless noted otherwise; no GPU was used. Runtimes are image size and algorithm-complexity-specific. Classical methods [15-21] are fast, but CNN-based techniques [33, 34] are slow on CPU, but give 40-50% speedup when using GPU. Overall, SWT-PnP-DnCNN offers a modest increase in runtime, achieves the highest fusion quality, and is therefore suitable when it must be deployed offline or in the cloud, and accuracy matters more than time.

Table 10. Runtime and computational cost comparison.

Method	Method Basis	Runtime (sec/image)	GPU Used	Relative Complexity
[15]	Compact matrix decomposition	0.42	✗	Low
[16]	Wavelet packet decomposition	0.58	✗	Low
[17]	Statistical projection	0.33	✗	Very Low
[18]	Wavelet transform	0.51	✗	Low
[19]	Shift-invariant wavelet	0.97	✗	Medium
[20]	Contourlet	1.85	✗	Medium to High
[21]	Shearlet	2.12	✗	High
[33]	Deep feature extraction	3.45	✓ (Optional)	High
[34]	Deep attention/learning	3.92	✓ (Optional)	High
SWT-PnP-DnCNN	SWT + iterative CNN denoiser	4.05	✓ (Optional)	High to Very High

Although the fusion results of SWT-PnP-DnCNN are improved across all quantitative metrics, the iterative PnP methodology introduces additional computational overhead. In contrast, classical transform-based methods [15, 17, 18] take only a second per image and can be used for fast, but not accurate, fusion. Meanwhile, methods [20, 21] achieve a balance between runtime and structural retention but remain CPU-bound due to multiscale directional expansions. Furthermore, CNN-based fusion methods [33, 34] can be accelerated with GPU support but require loading the model and incur additional memory overhead. Notably, SWT-PnP-DnCNN requires multiple denoising iterations and thus incurs the highest CPU computational cost. However, with moderate GPU acceleration, the inference performance becomes comparable to that of other CNN methods [33, 34]. As a result, SWT-PnP-DnCNN fits within a clinical framework of interest, particularly in devices where accuracy is considered more critical than the processing chain's raw latency (telemedicine servers, post-processing in a supporting PACS, edge-cloud-hybrid (ECH)).

6. Conclusion and Future Directions

This work proposed a new hybrid MIF method, i.e., SWT-PnP-DnCNN, that uses a deep CNN prior to shift-invariant multiscale decomposition for PnP optimization. By partially decomposing inputs using SWT and partially reconstructing fused coefficients through an iterative PnP loop regularized by DnCNN, the method balances the factors of "interpretability" and "adaptability" with those from traditional transforms and learning-based priors. Extensive experiments on various CT, MRI, and PET datasets show that the SWT-PnP-DnCNN outperforms the MIF methods based on MSVD, PCA, DWT, SWT, NSCT, NSST, and deep fusion baselines. Quantitative evaluations based on SCD, FMI, FF, SD, M, and E are superior and clear performance gains with mean+-std reporting and paired statistical significance testing, where SWT-PnP-DnCNN achieves $p<0.05$ in most comparisons. Ablation studies show that both the PnP component and the DnCNN prior are indispensable, and denoiser comparisons (BM3D, FFDNet, DnCNN) emphasize the superiority of learned regularization. Additional sensitivity analyses on the number of iterations, regularization weight, and SWT levels further strengthen the robustness of the design choices. While the method proposed has a higher computational cost than static transform rule fusion, runtime profiling shows a reasonable trade-off between quality and latency, if GPU acceleration is available at the time of the profile run.

Moving forward, the framework might be extended with task-aware objectives (e.g., fusion for segmentation/classification), integration with transformer-based priors, 3-D volumetric fusion, PnP to increase speed (e.g., real-time), etc. Overall, SWT-PnP-DnCNN offers a reliable, interpretable, and statistically validated fusion pipeline, well-positioned for clinical decision support and broader intelligent medical imaging workflows.

All the Declarations and Statements

Author Contributions Statement
Amit Pandey - Conceptualization, Methodology; proposed research ideas, constructed the overall framework, and supervised project execution.
Prabhishek Singh - Data Curation and Software Implementation; handled data acquisition, dataset preprocessing, and implementation of the research model.
Akansha Singh - Validation, and Performance Evaluation; Validated results using standard metrics, and benchmarked performance.
Achyut Shankar - Formal Analysis, Visualization, and Statistical Analysis; performed in-depth experimental analysis, prepared performance charts, and ensured statistical robustness.
Manoj Diwakar - Literature review, and technical documentation.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement
The authors declare no conflicts of interest.

Funding Declaration
This research received no specific grant from any funding agency.

Data Availability Statement
This study analyzed publicly available datasets. The results obtained and datasets can be found here: “https://github.com/prabhisheksingh/SWT-PnP-DnCNN.git”, accessed on “15-04-2025”.

Ethical Declarations
The authors should declare the following: Statements of ethical approval for studies involving human subjects and/or animals.

Acknowledgments
The authors would like to express their sincere gratitude to Bennett University, Greater Noida, India, for providing the necessary infrastructure and research support in this research journey. We are also thankful to all collaborators, mentors, and peers whose valuable insights contributed to the improvement of this work.

Declaration of Generative AI in Scholarly Writing
Authors should disclose in their article how AI and AI-assisted technologies were used during the writing process.

Abbreviations
The following abbreviations are used in this manuscript:

MIF - Medical Image Fusion
SWT - Stationary Wavelet Transform
LFSBs - Low-Frequency SubBands
HFSBs - High-Frequency SubBands
PnP - Plug-and-Play
CT - Computed Tomography
MRI - Magnetic Resonance Imaging
PET - Positron Emission Tomography
SPECT - Single-Photon Emission Computed Tomography
CAD - Computer-Aided Diagnosis
PCA - Principal Component Analysis
HIS - Intensity-Hue-Saturation
DWT - Discrete Wavelet Transform
NSCT - Non-Subsampled Contourlet Transform
CNN - Convolutional Neural Networks
GAN - Generative Adversarial Networks
MSVD - Multi-Resolution Singular Value Decomposition
VQ-VAE - Vector-Quantized Variational Autoencoder

Appendix

Not Applicable

References

[1] Guo, Z., Li, X., Huang, H., Guo, N., & Li, Q. (2018, April). Medical image segmentation based on multi-modal convolutional neural network: Study on image fusion schemes. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 903-907). IEEE.

[2] Du, J., Li, W., Lu, K., & Xiao, B. (2016). An overview of multi-modal medical image fusion. Neurocomputing, 215, 3-20.

[3] Zhou, T., Ruan, S., & Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion. Array, 3, 100004.

[4] Li, Y., Zhao, J., Lv, Z., & Li, J. (2021). Medical image fusion method by deep learning. International Journal of Cognitive Computing in Engineering, 2, 21-29.

[5] Kumar, A. (2022). Deep learning for multi-modal medical imaging fusion: Enhancing diagnostic accuracy in complex disease detection. Int J Eng Technol Res Manag, 6(11), 183.

[6] Nair, R. R., Singh, T., Basavapattana, A., & Pawar, M. M. (2022). Multi-layer, multi-modal medical image intelligent fusion. Multimedia Tools and Applications, 81(29), 42821-42847.

[7] Zhou, T., Cheng, Q., Lu, H., Li, Q., Zhang, X., & Qiu, S. (2023). Deep learning methods for medical image fusion: A review. Computers in Biology and Medicine, 160, 106959.

[8] Xiang, L., Chen, Y., Chang, W., Zhan, Y., Lin, W., Wang, Q., & Shen, D. (2018). Deep-learning-based multi-modal fusion for fast MR reconstruction. IEEE Transactions on Biomedical Engineering, 66(7), 2105-2114.

[9] Nair, R. R., Singh, T., Sankar, R., & Gunndu, K. (2021). Multi-modal medical image fusion using lmf-gan-a maximum parameter infusion technique. Journal of Intelligent & Fuzzy Systems, 41(5), 5375-5386.

[10] Maqsood, S., & Javed, U. (2020). Multi-modal medical image fusion based on two-scale image decomposition and sparse representation. Biomedical Signal Processing and Control, 57, 101810.

[11] Maneesha, P., Singh, T., Nayar, R., & Kumar, S. (2019, January). Multi modal medical image fusion using convolution neural network. In 2019 Third International Conference on Inventive Systems and Control (ICISC) (pp. 351-357). IEEE.

[12] Azam, M. A., Khan, K. B., Salahuddin, S., Rehman, E., Khan, S. A., Khan, M. A., ... & Gandomi, A. H. (2022). A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. Computers in biology and medicine, 144, 105253.

[13] Stimpel, B., Syben, C., Schirrmacher, F., Hoelter, P., Dörfler, A., & Maier, A. (2019). Multi-modal deep guided filtering for comprehensible medical image processing. IEEE transactions on medical imaging, 39(5), 1703-1711.

[14] Islam, K. T., Wijewickrema, S., & O'Leary, S. (2021). A deep learning based framework for the registration of three dimensional multi-modal medical images of the head. Scientific Reports, 11(1), 1860.

[15] Ouerghi, H., Mourali, O., & Zagrouba, E. (2017). Multimodal medical image fusion using modified PCNN based on linking strength estimation by MSVD transform. Int. J. Comput. Commun. Eng, 6(3), 201-211.

[16] Yang, L., Liu, X., & Yao, Y. (2008, May). Medical image fusion based on wavelet packet transform and self-adaptive operator. In 2008 2nd International Conference on Bioinformatics and Biomedical Engineering (pp. 2647-2650). IEEE.

[17] Polinati, S., Bavirisetti, D. P., Rajesh, K. N., & Dhuli, R. (2022). Multimodal medical image fusion based on content-based and PCA-sigmoid. Current Medical Imaging Reviews, 18(5), 546-562.

[18] Nair, R. R., & Singh, T. (2019). Multi-sensor medical image fusion using pyramid-based DWT: a multi-resolution approach. IET Image Processing, 13(9), 1447-1459.

[19] Nandeesh, M. D., & Meenakshi, M. (2015, December). A novel technique of medical image fusion using stationary wavelet transform and principal component analysis. In 2015 international conference on smart sensors and systems (IC-SSS) (pp. 1-5). IEEE.

[20] Alseelawi, N., Hazim, H. T., & Salim ALRikabi, H. T. (2022). A Novel Method of Multimodal Medical Image Fusion Based on Hybrid Approach of NSCT and DTCWT. International Journal of Online & Biomedical Engineering, 18(3).

[21] Diwakar, M., Singh, P., & Shankar, A. (2021). Multi-modal medical image fusion framework using co-occurrence filter and local extrema in NSST domain. Biomedical Signal Processing and Control, 68, 102788.

[22] Ying, Z., Nie, R., Cao, J., Ma, C., & Tan, M. (2025). A nested self-supervised learning framework for 3-D semantic segmentation-driven multi-modal medical image fusion. Biomedical Signal Processing and Control, 105, 107653.

[23] Ma, G., Qiu, X., & Tan, X. (2025). DMFusion: A dual-branch multi-scale feature fusion network for medical multi-modal image fusion. Biomedical Signal Processing and Control, 105, 107572.

[24] Zhang, Y., Ma, C., Ding, H., & Zhu, Y. (2025). MMCL: Meta-mutual contrastive learning for multi-modal medical image fusion. Digital Signal Processing, 156, 104806.

[25] Geetha Devi, A., Borra, S. P. R., & Rajesh Kumar, P. (2025). A new multimodal medical image fusion framework using Convolution Neural Networks. Journal of Medical Engineering & Technology, 1-8.

[26] Shibu, T. M., Madan, N., Paramanandham, N., Kumar, A., & Santosh, A. (2025). Multi-modal brain image fusion using multi feature guided fusion network. Biomedical Signal Processing and Control, 100, 107060.

[27] Srinivas, Y., & Kumar, M. A. (2025). Robust multi-modal COVID-19 medical image registration using dense deep learning descriptor model. Biomedical Signal Processing and Control, 100, 107007.

[28] Wang, J., Yu, L., & Tian, S. (2025). Cross-attention interaction learning network for multi-model image fusion via transformer. Engineering Applications of Artificial Intelligence, 139, 109583.

[29] Hu, T., Nan, X., Zhou, X., Shen, Y., & Zhou, Q. (2025). A dual-stream feature decomposition network with weight transformation for multi-modality image fusion. Scientific Reports, 15(1), 7467.
[30] Shao, D., Yang, H., Ma, L., & Yi, S. (2025). AFPNet: An adaptive frequency-domain optimized progressive medical image fusion network. Biomedical Signal Processing and Control, 103, 107357.
[31] Feng, X., Yang, J., Qiu, G., Mu, J., Wu, X., Zhang, H., & Hu, K. (2025). MMIF-VAEFusion: An end-to-end multi-modal medical image fusion network using vector quantized variational auto-encoder. Biomedical Signal Processing and Control, 102, 107407.
[32] Wang, Y., Li, Z., Wang, J., Yang, L., Dong, B., Zhang, H., & Liu, J. (2025). MFF: A Deep Learning Model for Multi-Modal Image Fusion Based on Multiple Filters. IEEE Access.
[33] Bhosekar, S., Singh, P., & Garg, D. (2024, December). A Method Noise Based-Multi-Modal Medical Image Fusion Technique Using Non-Subsampled Shearlet Transform and Convolutional Neural Network. In 2024 4th International Conference on Innovative Sustainable Computational Technologies (CISCT) (pp. 1-5). IEEE.
[34] Srivastava, S., Bhatia, S., Agrawal, A. P., Jayswal, A. K., Godara, J., & Dubey, G. (2025). Deep adaptive fusion with cross-modality feature transition and modality quaternion learning for medical image fusion. Evolving Systems, 16(1), 17.
[35] Multimodal medical image datasets. Available at: https://www.researchgate.net/figure/Multimodal-medical-image-datasets_fig4_351121182 (Access date: 23-06-2025)
[36] Multimodal medical image dataset. Available at: https://github.com/bsun0802/Zero-Learning-Fast-Medical-Image-Fusion/tree/master/images/MRI-PET (Access date: 26-06-2025)
[37] Multimodal medical image dataset. Available at: https://github.com/ashna111/multimodal-image-fusion-to-detect-brain-tumors/tree/master/dataset. (Access date: 26-06-2025)

Authors' Profiles



Amit Pandey received the M.Tech. degree in computer science from the Deenbandhu Chotu Ram University of Science and Technology, Sonipat, Haryana, India. He is currently a Research Scholar with the Computer Science Department, Bennett University, Greater Noida, India. His research interests include semantic segmentation for biomedical image segmentation, Image fusion techniques, Transformers and Explainable Artificial intelligence and Generative AI.



Prabhishek Singh (Senior IEEE Member) is an Associate Professor in the School of Computer Science Engineering and Technology at Bennett University (The Times Group), Greater Noida, India, and also serves as Honorary Adjunct Faculty at Maryam Abacha American University of Nigeria. He received his Ph.D. from Babasaheb Bhimrao Ambedkar University (Central University), Lucknow, in 2018. Dr. Singh has authored over 300 peer-reviewed publications in SCI, SCIE, ESCI, and Scopus venues. His research interests include multimodal medical image fusion, computer vision, deep learning, and AI-enabed healthcare analytics, and he actively contributes as journal editor and reviewer. He is listed among the World's Top 2% Scientists.



Akansha Singh is a Professor at the School of Computer Science and Engineering, Bennett University, Greater Noida. She holds a B.Tech, M.Tech, and Ph.D. in Computer Science, with her doctoral research completed at IIT Roorkee. With over 20 years of teaching and research experience, her expertise spans machine learning, deep learning, computer vision, generative AI, and remote sensing. She has authored and edited more than 50 books and published over 150 research papers in reputed international journals and conferences. Prof. Singh serves as an Associate Editor and Academic Editor for several leading journals and has led multiple government-funded research projects.



Achyut Shankar is an Associate Professor at Bennett University and Honorary Adjunct Faculty at Maryam Abacha American University, Nigeria. He previously served as a Research Scientist at the University of Warwick and completed research at Birkbeck University, London. He earned his Ph.D. in Computer Science and Engineering from VIT University, India. Listed among the World's Top 2% Scientists (2023–2025), he has authored 225+ papers, including 175+ SCI works. He serves as editor or associate editor for several leading IEEE, Elsevier, Wiley, and Taylor & Francis journals. His research interests include IoT, cybersecurity, wireless sensor networks, and intelligent systems.



Manoj Diwakar is Professor in the Department of Computer Science and Engineering at Graphic Era (Deemed to be University), Dehradun, India. Recognized among the World's Top 2% Scientists by Stanford University & Elsevier, he has established himself as a leading researcher in Artificial Intelligence and Intelligent Healthcare Systems. He holds a Ph.D. in Computer Science and has over a decade of academic, research, and administrative experience. His research interests include AI, Machine Learning, Deep Learning, Computer Vision, Biomedical Image Analysis, IoT, and Smart Healthcare Technologies. Dr. Diwakar has published widely in SCI/Scopus-indexed journals, international conferences, and edited books. He has served as Editor-in-Chief, Associate/Guest Editor. He is listed among the World's Top 2% Scientists.

How to cite this paper:

Amit Pandey, Prabhishek Singh, Akansha Singh, Achyut Shankar, Manoj Diwakar, "SWT-PnP-DnCNN: Medical Image Fusion Using Stationary Wavelet Transform and Plug-and-Play Deep Denoising Model", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.18, No.3, pp. 209-229, 2026. DOI:10.5815/ijigsp.2026.03.11