Journal of Engineering Managment and Soft Computing (JEMSC)

Adaptive Deep Image Fusion through PSO Segmentation and Sparse LSQR Optimization

Hedieh Noorian[1] and Hamid Reza Shahdoosti[2]✉

1. MSC Student, Faculty of Electrical and Computer Engineering, Hamedan University of Technology, Hamedan, Iran. Email: Hediehnoo2020@gmail.com
2. Corresponding author, Associate Professor, Faculty of Electrical and Computer Engineering, Hamedan University of Technology, Hamedan, Iran. Email: h.doosti@hut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 6 February 2026
Received in revised form
27 April 2026
Accepted 14 June 2026
Published online 1 July 2026

Keywords:
pansharpening, deep neural networks, PSO-Based segmentation, LSQR optimization, multispectral–panchromatic fusion, remote sensing.

ABSTRACT

This study presents a region-adaptive hybrid fusion framework that integrates unsupervised clustering, deep convolutional modeling, and sparse numerical optimization to enhance multispectral satellite image fusion. Unlike classical pansharpening approaches, which are computationally efficient but inherently global and unable to adequately handle spatial–spectral variability, and purely deep-learning-based models that often overgeneralize across heterogeneous scenes, the proposed method introduces a region-aware strategy to preserve fine structures and spectrally weak areas. By employing Particle Swarm Optimization–based unsupervised segmentation, the framework dynamically partitions the image into spectrally homogeneous regions, enabling localized CNN/ResNet-based fusion that prevents spectral dilution near region boundaries. The independently fused regions are subsequently unified through LSQR-based global optimization, which enforces structural coherence and minimizes reconstruction error without reliance on large labeled datasets. Experimental evaluations on IKONOS datasets demonstrate that the proposed framework consistently outperforms classical and state-of-the-art fusion methods in terms of UIQI, RMSE, CC, and ERGAS metrics, confirming its robustness, adaptability, and effectiveness for high-fidelity multispectral image enhancement in complex remote-sensing environments.

Cite this article: Noorian, H. & Shahdoosti, H, (2026)., Adaptive Deep Image Fusion through PSO Segmentation and Sparse LSQR Optimization. *Journal of Engineering Management and Soft Computing*, 12 (3). 150-167.

DOI: https://doi.org/10.22091/jemsc.2026.15317.1348

Publisher: University of Qom
DOI: https://doi.org/10.22091/jemsc.2026.15317.1348

1) Introduction

Satellite image fusion plays a crucial role in modern remote sensing by integrating the high spatial resolution of panchromatic (PAN) imagery with the rich spectral content of multispectral (MS) data. The primary objective of pansharpening is to produce fused images that preserve both the geometric details of PAN images and the radiometric fidelity of MS bands. Classical approaches, including IHS, PCA, and DWT, have traditionally dominated this field due to their simplicity and low computational cost (Burt & Adelson, 1985). Nevertheless, these pixel-level linear methods often induce spectral distortions, loss of fine texture, and limited adaptability in heterogeneous regions, where land-cover classes exhibit complex spatial–spectral variability (Liu et al., 2020).

In this study, representative panchromatic and multispectral images are illustrated below, both acquired from the IKONOS satellite sensor. One dataset corresponds to a dense urban area characterized by sharp structural edges and high spatial heterogeneity (Figure 1), while the other represents a riverine environment with smoother spectral transitions and predominantly natural land-cover patterns (Figure 2). These two contrasting scenes are selected to comprehensively evaluate the robustness and adaptability of the proposed fusion framework under different spatial–spectral complexities.

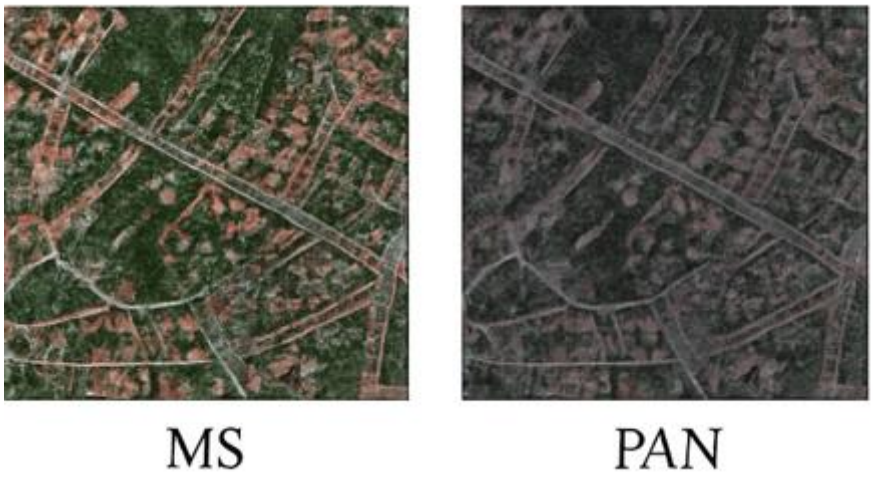

Figure 1. Low Resolution Multi-Spectral and Panchromatic Image (Dataset 1)

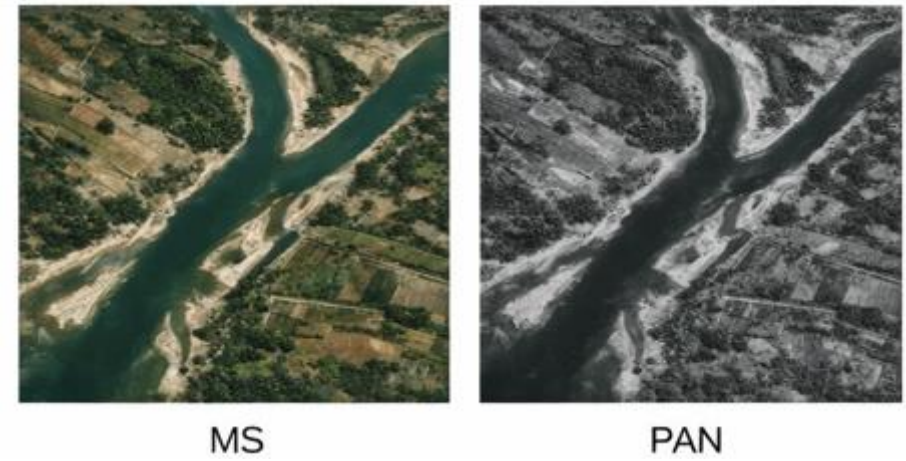

Figure 2. Low Resolution Multi-Spectral and Panchromatic Image (Dataset 2)

The IHS transformation enhances spatial clarity by replacing the intensity channel with the PAN band; however, this process may disrupt spectral consistency in mixed or complex regions (Liu et al., 2020). PCA reduces spectral redundancy through statistical decorrelation but lacks sensitivity to local

textures, which can lead to aliasing and structural information loss. Wavelet-based methods, such as DWT, improve multiresolution handling by injecting high-frequency PAN details into MS images, yet their performance strongly depends on wavelet kernel choice and precise image registration (He et al., 2019). These inherent limitations highlight the need for more adaptive and data-driven fusion techniques.

Recent advancements in artificial intelligence have introduced deep learning models that significantly enhance pansharpening performance. Early neural networks, such as MLP and RBFN, approximated nonlinear PAN–MS relationships but lacked spatial context awareness. More sophisticated architectures, including CNNs, residual networks (ResNet), and hybrid CNN+ResNet models, can effectively capture hierarchical spatial–spectral correlations (Fang et al., 2023). CNN-based pansharpening methods, exemplified by PNN (Masi et al., 2016), TFNet (Liu et al., 2020), and DiCNN (He et al., 2019), utilize multiscale convolutional kernels to learn complex mappings between PAN and MS data. Residual-based networks such as MSDRN (Wang et al., 2021), PCDRN (Yang et al., 2020), and SDRCNN (Fang et al., 2023) further stabilize gradient flow in deep models, enabling the preservation of high-frequency textures and structural edges. Hybrid architectures, including TDNet (Zhang et al., 2022), combine CNN and residual blocks to exploit both local feature extraction and multi-level error correction, yielding superior spatial–spectral fidelity.

Despite their representational power, these supervised deep networks require large amounts of annotated high-resolution MS data, which are often unavailable in practical remote sensing scenarios. Moreover, supervised training can emphasize dominant spatial features while suppressing weak or small-scale structures. Regularization techniques, such as dropout and batch normalization, may inadvertently deactivate neurons responsible for subtle edge or boundary information, resulting in the loss of fine structures such as narrow roads, small vegetation patches, or complex terrain features (Li & Zhao, 2020).

These challenges have motivated the development of unsupervised and hybrid fusion frameworks. Cluster- and graph-based methods, such as adaptive sparse coding (Wang et al., 2021), PCNN-Fusion (Li & Zhao, 2020), and HetSSNet (Ma et al., 2025), leverage region-wise segmentation or graph-guided processing to ensure that spectrally similar bands and spatially coherent regions are fused jointly. Such strategies enable adaptive detail injection, improving spectral fidelity and preserving fine spatial structures that classical or purely supervised models may fail to reconstruct. By integrating multiscale decomposition, residual learning, and region-adaptive fusion, hybrid frameworks effectively bridge traditional pyramid-based approaches with modern deep learning strategies, providing robust performance in complex satellite image fusion scenarios (Li et al., 2018).

2) Literature Review

In this section, we provide a focused review of contemporary pansharpening methods, emphasizing three major categories: CNN-based, residual/ResNet-based, CNN/ResNet/Hybrid Residual-based and cluster/graph/hybrid-based approaches. The aim is to summarize recent advancements, highlight methodological innovations, and discuss how each strategy contributes to spatial and spectral fidelity, while emphasizing their unique strengths and practical insights.

2.1) CNN-Based Pansharpening

Convolutional Neural Networks (CNNs) have become the dominant learning-based architecture in pansharpening due to their ability to capture local spatial–spectral correlations. Masi et al. (2016) introduced PNN, a pioneering three-layer CNN designed for direct injection of PAN details into upsampled MS images. This model demonstrated significant improvements over traditional methods such as IHS or PCA, particularly in preserving spectral characteristics while enhancing spatial resolution. Liu et al. (2020) proposed TFNet, a two-stream CNN architecture where separate branches extract PAN and MS features before feature-level fusion, improving texture consistency and reducing spectral distortion in heterogeneous regions. He et al. (2019) developed DiCNN, integrating a detail-

injection strategy within an end-to-end CNN, which further enhanced spectral and spatial fidelity on QuickBird and WorldView-2 datasets (Li et al., 2020).

These CNN models excel because they can learn nonlinear PAN–MS mappings and capture multi-scale spatial details directly from data. Their local receptive fields naturally align with the spatial structure of the images, allowing precise edge enhancement without explicit sequential or recurrent processing.

2.2) ResNet/Residual-Based Pansharpening

Residual learning enhances CNN training stability and facilitates high-frequency detail preservation. Wang et al. (2021) proposed MSDRN, a multi-scale deep residual network that leverages skip connections and coarse-to-fine feature extraction to recover both fine textures and larger structures. Yang et al. (2020) developed PCDRN, a progressive cascade residual network performing stepwise fusion—first at coarse resolution and then at fine resolution—reducing smoothing and high-frequency loss. Tongji University (Liu et al., 2021) presented HARNN, integrating residual blocks with attention mechanisms to adaptively weight PAN and MS channels, minimizing spectral distortion. Fang et al. (2023) proposed SDRCNN, a lightweight dense residual-connected CNN balancing efficiency with accurate feature extraction, achieving strong results on WorldView-2, WorldView-3, and QuickBird datasets.

Residual architectures consistently outperform conventional CNNs in heterogeneous scenes, as skip connections prevent gradient vanishing and allow deep networks to focus on both global and local details.

2.3) CNN + ResNet/Hybrid Residual Methods

Hybrid architectures, combining CNNs with residual connections, provide both detailed feature extraction and multi-level error correction. Wang et al. (2021) integrated multi-scale residuals in MSDRN to capture complex discrepancies between PAN and MS images while maintaining spectral fidelity. Yang et al. (2020) designed PCDRN, cascading residual sub-networks to fuse images progressively from coarse-to-fine resolutions, preserving high-frequency information and reducing smoothing. Fang et al. (2023) proposed SDRCNN, combining dense residual connections with CNN layers to balance spatial–spectral accuracy and computational efficiency. Zhang et al. (2022) developed TDNet, which utilizes multi-scale convolutional layers alongside a few ResNet blocks, achieving superior performance on WorldView-3, QuickBird, and GaoFen-2 datasets.

The hybrid approach allows gradient flow stabilization during training, enabling the network to extract fine spatial structures while preserving spectral consistency. By integrating residual learning into CNNs, these methods outperform plain CNNs in both edge reconstruction and global fidelity.

2.4) Cluster -Hybrid-Based Methods

Cluster- and graph-based methods adaptively adjust fusion parameters according to spatial–spectral homogeneity. Wang et al. (2021) proposed Graph Regularized Sparse Coding with Adaptive Dictionary, clustering MS bands by spectral similarity and applying sparse coding within each cluster, which ensures correlated regions are jointly injected with PAN information, reducing spectral distortion. Li and Zhao (2020) introduced PCNN-Fusion, performing segmentation via pulse-coupled neural networks to identify homogeneous regions before applying adaptive detail injection. Ma et al. (2025) presented HetSSNet, a spatial-spectral heterogeneous graph network, demonstrating that graph-based processing generalizes beyond fixed clusters and adapts fusion in highly heterogeneous scenes.

The hybrid and cluster-based strategies are critical, as injecting PAN details independently per band ignores inter-band correlations and may lead to spectral inconsistencies. By processing spatially and spectrally coherent regions together, these methods achieve higher fidelity in both spectral and spatial domains, particularly in complex real-world imagery.

3) Preliminaries of Proposed Method

In this section, the core components of the proposed hybrid pansharpening framework are briefly reviewed. The method combines region-aware clustering, multiscale Laplacian-based fusion, and global numerical optimization to address spectral distortion and spatial inconsistency commonly observed in classical and learning-based fusion models. The following subsections summarize the clustering strategies, fusion mechanisms, and optimization solvers employed in the framework.

Figure 3 illustrates main implementation pathways of the proposed algorithms presented in this paper. Each pathway follows a step-by-step execution of the clustering, multiscale fusion, and optimization stages, ultimately leading to an effective and spectrally consistent pansharpened result. By systematically implementing these stages, the framework enables a fair and comprehensive comparison of different fusion strategies under a unified processing pipeline. The overall workflow and methodological flowchart of the proposed approach are summarized and analyzed in the figure below, providing a clear overview of the algorithmic structure and information flow.

3.1) Clustering Methods

3. 1. 1) Particle Swarm Optimization for Cluster-Aware Parameter Refinement

Following K-means segmentation, Particle Swarm Optimization (PSO) is employed to refine region-specific fusion parameters. PSO is particularly suited to pansharpening due to its ability to navigate non-convex, high-dimensional search spaces efficiently. In this work, each particle represents a candidate set of injection coefficients ($\alpha = \alpha_1, \ldots, \alpha_k,$), and the objective function incorporates both spectral fidelity and spatial consistency. Let the fitness function be defined as

$$J(\alpha) = \lambda_1 \sum_k \left|\left|F_k(\alpha) - M_{LR-up}\right|\right|_{C_k}^2 \quad (1)$$

where J stands for the weighted sum of spectral distortion and edge misalignment, C_k is the kth cluster, F_k represents the kth cluster segment, and M_{LR-up} signifies the up-sampled low resolution multispectral image. This structure is consistent with previous optimization-based fusion frameworks (Lemaire & Clérot, 2022), but its cluster-wise formulation is unique to the proposed method.

Each particle updates its velocity and position according to

$$v_i(t+1) = wv_i(t) + c_1v_1(p_i^* - \alpha_i(t)) + c_2v_2(g^* - \alpha_i(t)) \quad (2)$$

$$\alpha_i(t+1) = \alpha_i(t) + v_i(t+1) \quad (3)$$

where w denotes the inertia factor, c_1,c_2 are acceleration constants, and p_i^*, g^* are the local and global best solutions, respectively. These update equations allow the swarm to converge toward the optimal set of fusion coefficients even when the spectral distribution varies dramatically across clusters.

Algorithm 1 Hybrid Cluster–Laplacian–Deep Learning–Optimization Framework

Data: High-resolution panchromatic image X_{PAN};
Low-resolution multispectral image X_{MS},
Number of segment clusters K;
Laplacian pyramid levels L;
Deep learning parameters Θ;
Maximum optimization method (LSQR–LSCOV–SVD) iterations T_{max}.
Convergence threshold ε.

Result: High-resolution multispectral image $\widetilde{X}_{MS}$.

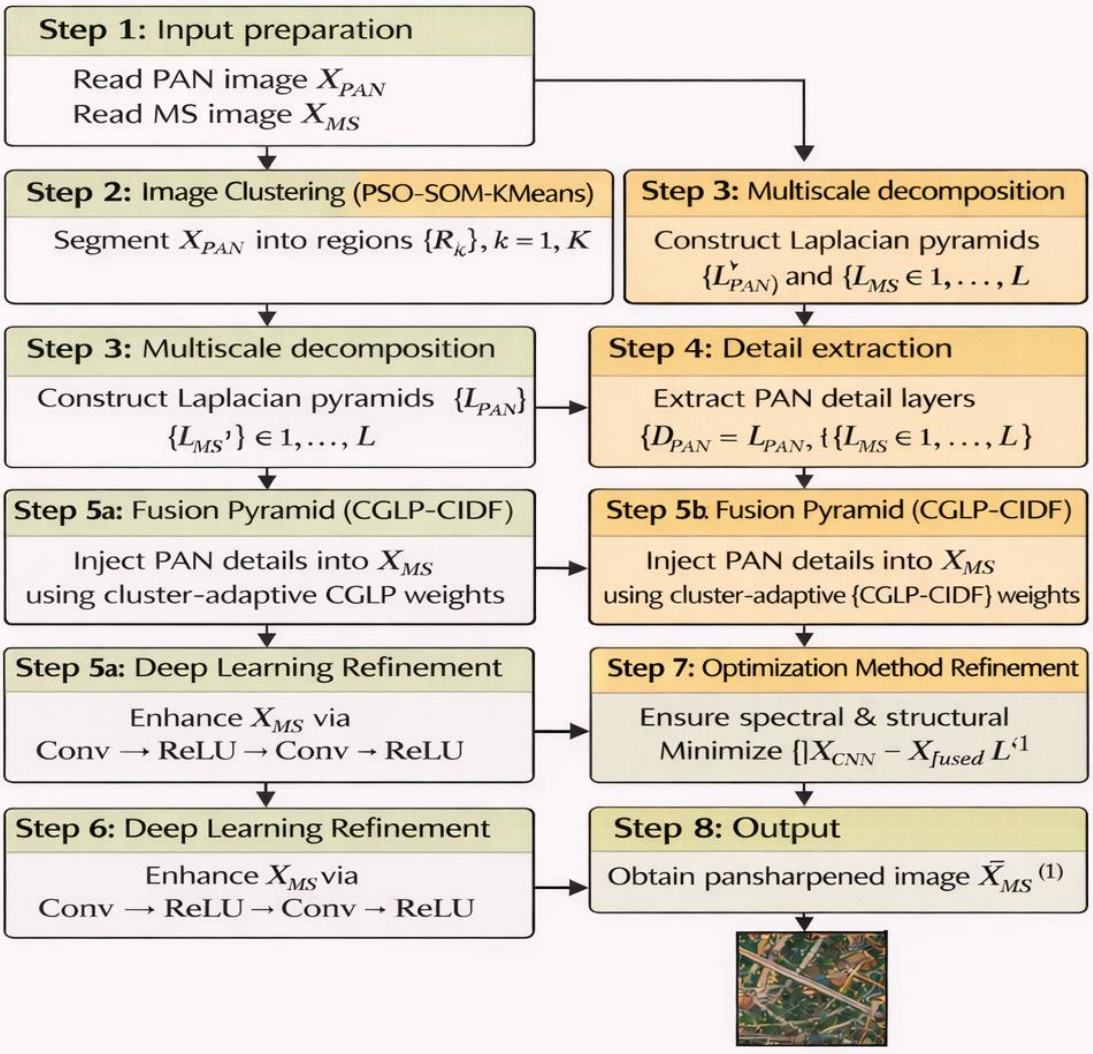

Figure 3. Proposed Hybrid Model Pansharpening

PSO is incorporated after K-means segmentation, as region-wise optimization dramatically reduces the dimensionality of the search space. Rather than optimizing at pixel-scale resolution, the algorithm only adjusts K injection parameters, which significantly improves convergence speed while increasing robustness to noise and local outliers. This approach contrasts with deep-learning-based enhancement

modules (Kohonen, 2012), where parameters are learned globally and cannot be fine-tuned to individual clusters.

To ensure numerical stability within each cluster, the PSO step also refines the Laplacian normalization factor by minimizing

$$E_{pyr} = \sum_k \left|\left|\alpha_k L_s - \eta_k \hat{L}_k\right|\right|_2^2 \quad (4)$$

Where E_{pyr} is pyramid numerical stability of clusters, L denotes the Laplacian scaling, and η_k represents a cluster-specific adjustment coefficient optimized concurrently with α_k. This coupling between injection strength and Laplacian scaling has no analog in traditional pansharpening pipelines, where high-frequency details are often injected through fixed linear operators (Wang et al., 2018).

The interaction between PSO and the residual correction module is also mathematically grounded. After each PSO iteration, the provisional fused image $F^{(t)}$ is passed through a shallow ResNet block, which produce:

$$F^{(t+1)} = F^{(t)} + R(F^{(t)}) \quad (5)$$

Where $F^{(t)}$ denotes provisional fused image, and R denotes the residual mapping. The PSO objective evaluates the corrected output rather than the raw injection result, forming a cascade of multi-step optimization and nonlinear refinement analogous to iterative pansharpening networks (Rao et al., 2017) but without requiring supervised training.

Finally, the PSO stage concludes when the termination condition

$$\left|\left|\alpha^{(t+1)} - \alpha^{(t+1)}\right|\right| \leqslant \tau \quad (6)$$

is satisfied, ensuring convergence in a mathematically controlled manner. Parameter τ determines the termination condition. Since PSO explores the entire feasible space of injection parameters, it compensates for the local minima introduced by K-means clustering. This synergy between segmentation and metaheuristic optimization is largely absent in existing literature and constitutes a major component of the novelty of the proposed framework.

3. 1. 2) Self-Organizing Maps for Topology-Preserving Spectral Refinement

The third component of the clustering stage is the Self-Organizing Map (SOM), which preserves the spectral topology of the multispectral feature space. Unlike *K*-means, SOM does not merely assign pixels to discrete clusters; instead, it organizes the high-dimensional spectral data onto a lower-dimensional grid while maintaining neighbourhood relationships. Let the weight vector of neuron *i* be w_i, and let the input feature be *f*. The best matching unit (BMU) is determined by

$$b = argmin_i \left|\left|f - w_i\right|\right|_2 \quad (7)$$

Once the BMU is identified, the weights of the neurons update according to

$$w_i^{(t+1)} = w_i^{(t)} + \eta(t) + h_{b,i}(t)(f - w_i^{(t)}) \quad (8)$$

Where $h_{b,i}(t)$ is the neighbourhood kernel around the BMU on the lattice grid. This formulation ensures that neurons close to the BMU undergo stronger updates, preserving spectral topology in accordance with Kohonen's classical theory (Shen et al., 2019).

In the pansharpening context, the SOM produces a continuous, topology-preserving map of spectral signatures, providing a refined, geometry-aware segmentation. This complements the discontinuous clusters produced by *K*-means by establishing a smoother manifold representation. The SOM's continuous mapping is particularly valuable for pansharpening in heterogeneous landscapes, where narrow roads, vegetation strips, or shadowed regions may not form compact *K*-means clusters but still lie along continuous spectral manifolds.

The SOM mapping is incorporated into the Laplacian-domain injection through a spatially adaptive modulation factor:

$$\beta_{(x,y)} = exp(-\left|\left|PAN\,(x,y) - w_i\right|\right|) \quad (9)$$

Where $\beta_{(x,y)}$ denotes modulation factor at pixel (x, y). This factor softly weights the injection strength:

$$f_i(x,y) = MS_i(x,y) + \beta(x,y)L_s(x,y) \quad (10)$$

Where f_i, $PAN_{,}$ and MS_i are i*th* band of fused image, panchromatic image, and i*th* band of the multispectral image, respectively. In addition, L_s denotes the fine information of the panchromatic image. This continuous weighting mechanism prevents abrupt transitions between K-means cluster boundaries that would otherwise introduce structural artifacts. Such artifacts frequently appear in segmentation-based fusion methods lacking topology preservation (Wang et al., 2018).

The SOM grid also enhances PSO optimization by defining a low-dimensional embedding space where spectral transitions are smooth. This reduces the risk of erratic parameter updates and increases the stability of the swarm's convergence. Studies in remote-sensing dimensionality reduction (Aiazzi et al., 2003) demonstrate that SOM embeddings consistently outperform hard clustering for preserving spectral continuity, aligning with the improved stability observed in our hybrid model.

Finally, SOM's iterative learning stabilizes the residual correction step. As the spectral topology becomes more coherent, the ResNet block operates on a smoother manifold of fused features, allowing the residual mapping to correct subtle distortions without introducing edge overshoot. This relationship, supported by prior SOM-enhanced fusion models (Aiazzi et al., 2003), further validates the integration of topological learning into the pansharpening pipeline.

3.2) Multiscale Fusion: Laplacian Pyramids and Hybrid Spectral–Spatial Models

3.2.1) CGLP: Context-Based Generalized Laplacian Pyramid

The Context-Based Generalized Laplacian Pyramid (CGLP) was proposed by Aiazzi et al. (2003) as a multiscale fusion framework to inject PAN details into MS images while preserving spectral consistency. It constructs a Laplacian pyramid for both MS and PAN images:

$$G_0 = I \quad (11)$$

$$G_{l+1} = Down(G_l * k)$$

where G_l is the Gaussian pyramid at level *l*, *k* denotes a Gaussian smoothing kernel, and *Down*(·) represents subsampling. The Laplacian pyramid is defined as:

$$L_l = G_l - UP(G_{l+1}) \quad (12)$$

where *Up*(·) denotes up-sampling, and L_l represents fine information at level *l*.

$$F_l = L_{l_{MS}} + \alpha_l(L_{l_{MS}} - L_{l_{PAN}}) \quad (13)$$

where F_l is the fused image at level *l*, $L_{l_{MS}}$ is the fine information of multispectral image at level *l*, $L_{l_{PAN}}$ represents the fine information of the panchromatic image at level *l*, and α_l denotes the adaptive gain:

$$\alpha_l = \frac{\sigma^2_{l,PAN}}{\sigma^2_{l,MS}+\varepsilon} \quad (14)$$

where $\sigma^2_{l,PAN}$ and $\sigma^2_{l,MS}$ are the variances at level lll for PAN and MS, respectively, and ϵ avoids division by zero.

CGLP ensures that high-frequency PAN details are injected proportionally to the local contrast while preventing oversharpening in smooth regions. Each pyramid level can be tuned independently, allowing multiscale enhancement of edges and textures while maintaining spectral fidelity. This method

is particularly effective in moderately heterogeneous scenes, but for strongly heterogeneous clusters, adaptive cluster-based methods, such as CIDF, may outperform it.

3.2.2) CIDF: Context-Aware Detail Injection Framework

The Context-Aware Detail Injection Framework (CIDF) was developed by Yuan et al. (2013) to enhance Laplacian pyramid-based fusion with cluster-adaptive injection gains. After the segmentation of the MS image into clusters R_k, the framework computes a detail map for each cluster:

$$D_k = \gamma_k(P - M_k) \tag{15}$$

where P represents the PAN image, M_k denotes the mean intensity of the k*th* cluster of MS, and γ_k signifies the injection gain estimated per cluster. The fused cluster is then:

$$F_k = M_k + D_k. \tag{16}$$

where F_k is the k*th* band of the fused image.

By adapting γ_k per cluster, CIDF balances spatial enhancement with spectral preservation, ensuring that edges and textures in urban, vegetated, or water regions are treated differently. The final fused image is reconstructed by aggregating all clusters.

CIDF improves over CGLP by explicitly considering local heterogeneity; the success of this method is highly dependent on the accuracy of segmentation. Poor clustering may lead to spectral distortion or artifacts, whereas high-quality segmentation enables precise, context-aware detail injection.

3. 3) Optimization Method

3.3.1) LSQR: Sparse Least-Squares Solver

In the context of region-wise pansharpening, after individual clusters have been fused—possibly via Laplacian-based injection or variant fusion strategies—the primary challenge is to enforce global consistency: spectral fidelity across bands, seamless boundaries across cluster borders, and smooth structural transitions (Paige & Saunders, 1982). For this purpose, a sparse linear system is formulated:

$$min_F \left||AF - b|\right|_2^2 \tag{17}$$

where F is the vectorized fused multispectral image to solve for, A is a sparse matrix encoding gradient constraints derived from PAN and MS inter-band relations (e.g., spatial gradients, edge continuity, band correlation), and b denotes the vector of observed constraints (Golub & Loan, 2013) (gradient from PAN, residuals between injected details and MS, etc.).

Because A is typically large (for high-resolution imagery, potentially millions of pixels × multiple bands) and highly sparse, the classical dense-method solvers (e.g., direct SVD) become computationally infeasible. Instead, LSQR—an iterative Krylov-subspace method—is well suited (Golub & Kahan, 1965; Raghavan, 1994). The algorithm begins with bidiagonalization via the Golub–Kahan procedure:

$$\beta_1 u_1 = b\ ,\ \alpha_1 v_1 = A^T u_1 \tag{18}$$

$$A\ \ v_k - \alpha_k u_k = \beta_{k+1} u_{k+1}, A^T u_{k+1} - \beta_{k+1} v_k = \alpha_{k+1} u_{k+1} \tag{19}$$

$$\phi_k v_k + x_k = x_{k+1} \tag{20}$$

where $\{u_k\}$, $\{v_k\}$ are orthonormal basis vectors and ϕ_k is chosen to minimize the residual in the k-th Krylov subspace. Because each iteration only requires sparse matrix–vector multiplications and vector updates, computational expense remains linear in the number of non-zero entries of A, and memory requirements remain modest.

From a pansharpening standpoint, LSQR has several advantages. First, it preserves edge continuity across cluster boundaries, as the gradient constraints in A enforce that spatial transitions align with PAN-derived detail maps. Second, it handles sparse data and missing information gracefully, as zero-valued gradients or masked pixels do not destabilize the solution. Third, it mitigates noise amplification,

because the iterative projection onto a Krylov subspace effectively acts as a regularizer, suppressing high-frequency noise components that are not consistently supported by PAN gradients.

Finally, LSQR scales naturally to large remote sensing images (tens of thousands of pixels per dimension), in contrast to dense least-squares or pseudoinverse approaches which become computationally intractable. Its iterative nature also allows early stopping criteria (e.g., based on residual norm or change in F) to avoid overfitting or over-sharpening, making it well-suited for practical pansharpening pipelines combining local cluster-level fusion with global refinement.

3.3.2) LSCOV: Weighted Least-Squares with Covariance for Noise-Variant Regions

When uncertainty or noise levels differ significantly across spatial regions or spectral bands—for instance, due to varying sensor noise, atmospheric effects, or heterogeneous land cover—a weighted least-squares solver that accounts for the covariance structure can deliver superior fusion results. In this approach, one solves:

$$min_F \left|\left|W(AF-b)\right|\right|_2^2 \tag{21}$$

where W is a weighting matrix derived from noise covariance estimates or measures of signal reliability (for example, lower weights for shadowed or water areas, higher weights for stable, high-signal areas). The closed-form solution (when $A^T W^T WA$ is invertible) is

$$F = (A^T W^T WA)^{-1} A^T W^T Wb \tag{22}$$

This LSCOV formulation effectively down-weights unreliable constraints, thereby enhancing spectral fidelity and reducing the risk of noisy artifact amplification. In practice, W can be block-diagonal (per band), or spatially varying (per pixel), depending on estimated noise covariance or per-cluster reliability metrics.

LSCOV is especially beneficial in heterogeneous or noisy scenes, such as areas with water bodies, shadows, or mixed urban–vegetation zones, where standard least-squares might overfit unreliable gradient cues. The weighting mechanism enhances robustness by suppressing constraint violations in noisy regions while preserving detail in high-confidence zones.

However, using LSCOV comes at a cost. The matrix inversion $(A^T W^T WA)^{-1}$ can be computationally expensive and memory-demanding for large images. Moreover, accurate estimation of the covariance (or weight matrix W) is non-trivial, often requiring prior statistical modeling or empirical noise measurement. Consequently, LSCOV is practical primarily for moderate-size scenes, block-based processing, or as part of a hybrid solver strategy (e.g., LSCOV on challenging clusters and LSQR elsewhere).

Methodology

The proposed hybrid pansharpening framework integrates region-adaptive segmentation, multiscale Laplacian-based enhancement, deep nonlinear reconstruction, and global numerical optimization into a coherent pipeline. This unified design resolves the major limitations of classical analytical models and supervised CNN methods, particularly their inability to preserve fine structural continuity in narrow, weak, or low-contrast regions such as thin roads, rooftop edges, irrigation lines, and small vegetation textures.

The first stage partitions the PAN image into structurally homogeneous and context-aware regions using unsupervised clustering approaches such as K-means, PSO-based adaptive clustering, or SOM-driven topology-preserving segmentation. This regional decomposition is essential because it prevents the assumption of global homogeneity and allows each area to be processed according to its intrinsic spectral–spatial characteristics. Prior research has shown that such segmentation reduces cross-region spectral leakage and enhances accuracy in complex heterogeneous landscapes. The resulting region masks guide all subsequent modules in the flowchart.

In the second stage, the PAN image undergoes multiscale Laplacian pyramid decomposition to extract spatial details at different frequency levels. These high-frequency layers capture edges, textures, and gradient transitions that are critical for accurate spatial enhancement. The multispectral image is upsampled to the PAN resolution and similarly transformed into a multiscale representation to ensure structural consistency across both modalities. Initial spatial injection is performed using region-conditioned analytical strategies derived from CGLP, CDIF, GAIHS, and related gradient-preserving frameworks, which maintain spectral fidelity through constrained modulation of detail layers.

Since analytical models rely on fixed transfer assumptions and cannot represent nonlinear relationships between PAN intensity and MS spectral responses, the framework incorporates a deep-learning refinement module.

This module is based on a convolutional neural network operating according to the standard mapping:

$$Y=\sigma(W*X+b) \tag{23}$$

where X represents the concatenated input tensor containing the upsampled MS image, Laplacian pyramid high-frequency layers, and the regional mask, W denotes the trainable convolution kernel, b is the bias term, σ signifies a nonlinear activation (ReLU), and Y represents the feature-enhanced output. Through stacked convolutional layers, the network learns contextual edge patterns, nonlinear spectral interactions, and spatial textures that classical methods cannot explicitly model. This allows for the reconstruction of subtle structural cues that typically vanish in conventional CNN pansharpening or become oversmoothed in deterministic analytical models.

To further stabilize nonlinear enhancement, the deep module adopts a PanNet-inspired residual learning formulation:

$$\mathrm{F}=X_{MS}\uparrow+f_\theta(X_{PAN}-X_{MS}\uparrow) \tag{24}$$

where $X_{MS}\uparrow$ is the upsampled MS image, X_{PAN} is the panchromatic input, f_θ denotes the learned residual mapping, and F signifies the fused output. By focusing on the discrepancy term ($X_{PAN}-X_{MS}\uparrow$), the network extracts missing high-frequency information directly rather than reconstructing the full fused image. This approach improves stability, reduces color distortion, and enhances weak spatial structures. Multiscale convolutional layers within the residual block provide sensitivity to fine, intermediate, and coarse patterns, enabling the model to preserve narrow boundaries and weak gradients that standard CNNs or purely analytical pyramids often miss.

Although region-driven processing and deep refinement significantly enhance spatial detail, they remain inherently local, which can introduce subtle inconsistencies at region borders. To ensure global smoothness and spectral stability, the final stage employs a numerical optimization module based on LSQR, covariance-weighted LSCOV, or SVD-based reconstruction. This solver enforces a global objective that integrates PAN gradient constraints with MS spectral deviation penalties. The optimization process aligns all regional and deep features, suppresses seam artifacts, eliminates small color imbalances, and stabilizes gradient transitions across the image. Through iterative refinement, the fused output becomes globally coherent while maintaining the deep network's local enhancements.

The complete architecture thus combines the interpretability and physical grounding of analytical fusion with the adaptivity and nonlinear modelling capability of deep learning. Region segmentation supplies contextual boundaries; the Laplacian pyramid provides multiscale structural decomposition; the CNN and PanNet residual unit perform nonlinear spectral–spatial reconstruction, and the LSQR/LSCOV/SVD optimization stage imposes global spatial and spectral consistency. By integrating these components into a single hybrid system, the proposed method achieves superior preservation of fine spatial structures, better spectral fidelity, and more stable gradients compared with classical hybrid and modern supervised approaches. The resulting fused image exhibits sharp detail, smooth spatial

transitions, and accurate spectral behaviour, reflecting the complete workflow depicted in the flowchart from clustering-based segmentation to deep residual enhancement and global optimization.

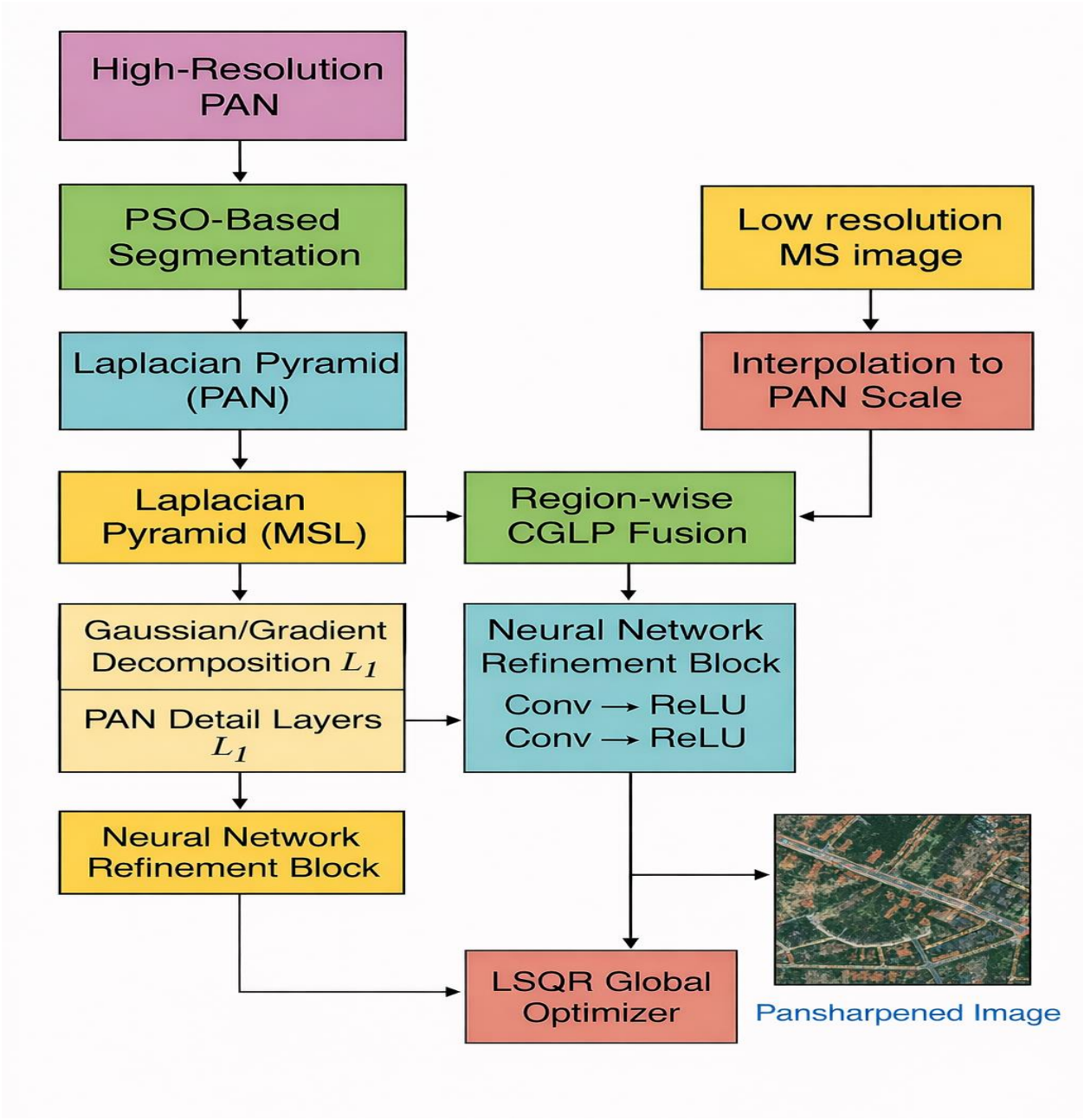

Figure 4. Proposed Hybrid Model

Table 1. IKONOS Data

Band	Resolution	Wave length (Micro Meter)
Blue	4m	0.46
Green	4m	0.55
RED	4m	0.65
Infrared	4m	0.80
Panchromatic	1m	0.71

In figure 4, the flowchart corresponding to the best-performing configuration of the proposed framework is illustrated. This visual representation facilitates a deeper understanding of the algorithmic pipeline and provides practical insight into the step-by-step realization of the proposed method.

4) Findings and Discussion

The experimental evaluation conducted on IKONOS imagery (Table 1), including both urban and riverine scenes, demonstrates that the proposed region-aware hybrid pansharpening model achieves the most balanced and robust performance across the primary fusion criteria, including RMSE, UIQI, CC, and SAM, with the corresponding numerical results reported in Table 2. These widely adopted quantitative metrics comprehensively assess both spectral fidelity and spatial consistency in pansharpening tasks. Specifically, the Root Mean Square Error, ERGAS, SAM, and Q-index are defined as (Shen et al., 2016; Yuan et al., 2013):

$$\textbf{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - R_i)^2}\ , \tag{25}$$

$$\textbf{ERGAS}=100.\frac{h}{l}.\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\frac{RMSE_i}{\mu_i})^2} \tag{26}$$

$$\textbf{SAM}=cos^{-1}(\frac{F.R}{||F||.||R||}) \tag{27}$$

$$\textbf{UIQI}=\frac{4\mu_F\mu_R\sigma_{FR}}{(\mu_F^2+\mu_R^2).(\sigma_F^2+\sigma_R^2)} \tag{28}$$

the i*th* band of the original image, h denotes the resolution of high resolution image, l is the resolution of the low resolution image, μ_i signifies the mean of i*th* band of the fused image, μ_F is the mean of the fused image, μ_R is the mean of the original image, σ_F^2 represents the variance of the fused image, σ_R^2 is the variance of the original image, and σ_{FR} denotes covariance between the fused and original images. In these equations, RMSE and ERGAS consider both spatial and spectral distortion, the SAM metric focuses on spectral distortion, and UIQI mainly takes into account spatial distortion.

imagery. These scenes are characterized by complex spatial structures, sharp edges, and strong spectral variability. Visual inspection indicates that the proposed region-aware hybrid methods preserve fine urban details such as building boundaries and road networks more effectively, while reducing spectral distortion compared to classical fusion approaches.

Figure 6 illustrates selected results from Dataset 2, corresponding to river and natural scenes. These images exhibit smoother textures, lower contrast, and spectrally homogeneous regions, which pose different challenges for pansharpening. In this case, adaptive cluster-based fusion combined with LSQR and LSCOV optimization provides improved spectral continuity in water bodies and suppresses noise amplification near river boundaries.

The comparison between urban and river datasets highlights that fusion performance is strongly scene-dependent. Urban areas benefit primarily from enhanced spatial detail, whereas river scenes require careful spectral preservation and smooth transitions. These observations confirm the advantage of adaptive, region-aware fusion frameworks for handling diverse remote-sensing environments.

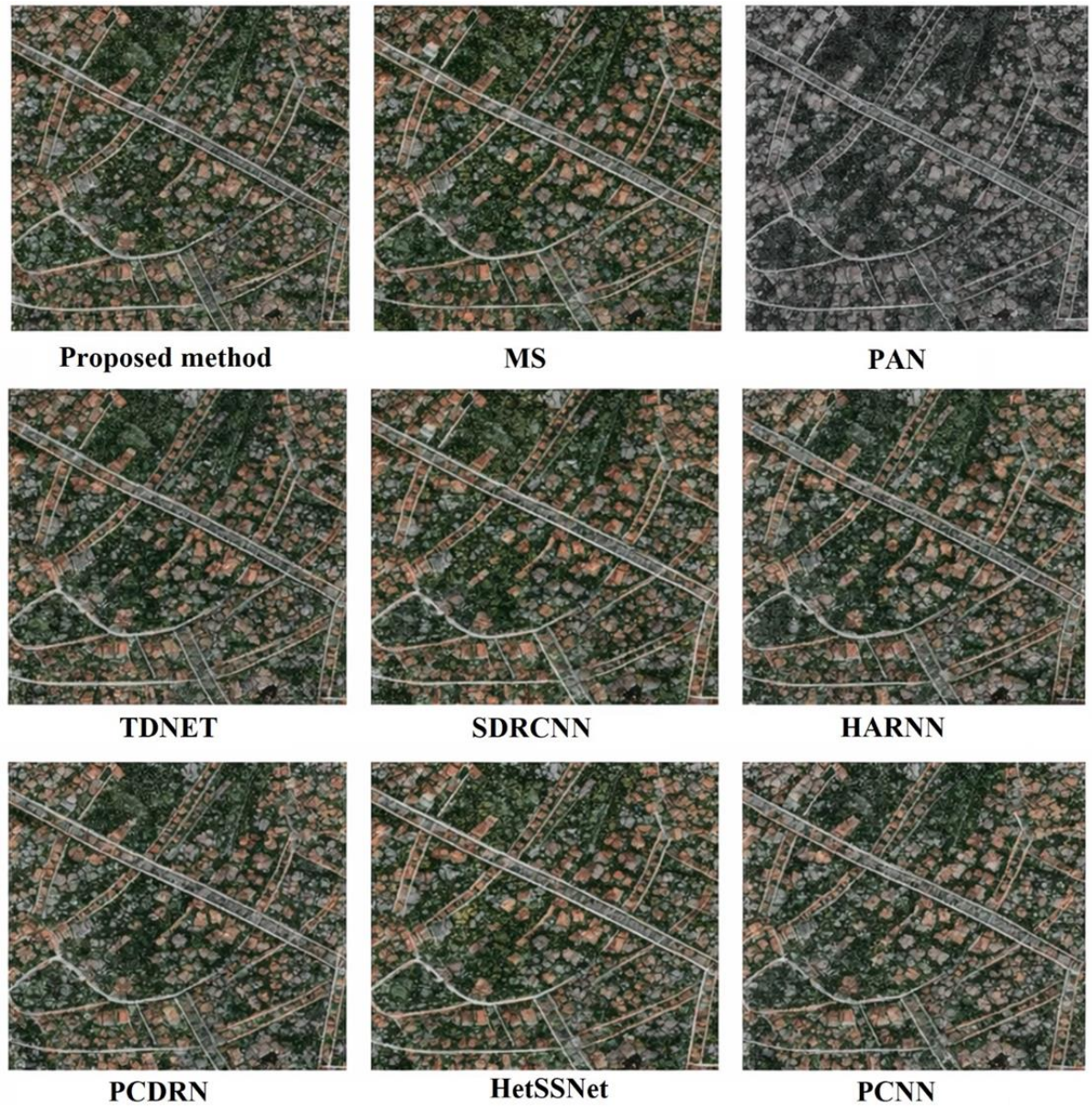

Figure 5. Results of Best Model Simulation Dataset 1 by Python Programming

As summarized in Table 2, the proposed methods configuration consistently achieves the best overall performance across both datasets, yielding the highest UIQI and CC values while maintaining the lowest RMSE and SAM. This indicates a superior balance between spatial detail enhancement and spectral fidelity compared to classical and other hybrid fusion strategies. The integration of PSO-based region adaptation enables effective optimization of cluster-wise injection gains, while the LSQR global solver enforces structural continuity and spectral consistency across region boundaries, leading to stable and robust fusion results.

The observed performance differences between Dataset 1 (urban scenes) and Dataset 2 (riverine scenes) arise primarily from their distinct spatial–spectral characteristics. Urban imagery in Dataset 1 contains sharp edges, high-contrast man-made structures, and strong local gradients, which benefit significantly from aggressive yet controlled detail injection guided by PSO and refined through LSQR optimization. In contrast, Dataset 2 consists of smoother river and natural landscapes with homogeneous textures and lower spatial frequency content, where excessive high-frequency injection may degrade spectral integrity. Consequently, the best performing method exhibits slightly different metric

behaviour across the two datasets, reflecting the trade-off between spatial enhancement and spectral preservation.

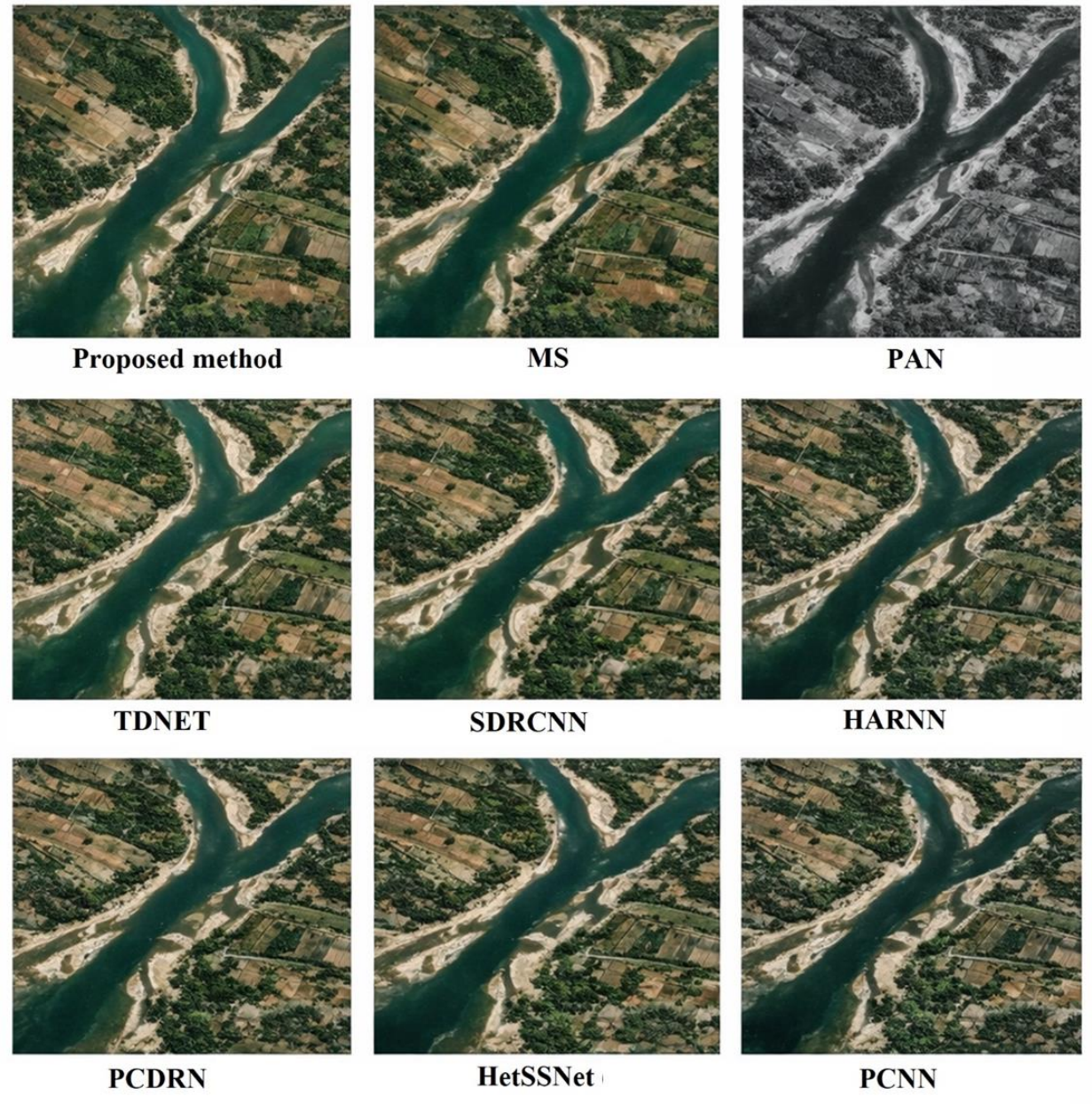

Figure 6. Results of Best Model IKONOS Dataset 2 Simulation by Python Programming

Overall, these results confirm that the proposed region-aware CGLP–PSO–LSQR framework adapts effectively to diverse land-cover types. Its ability to dynamically balance spatial sharpening in urban areas and spectral stability in natural river scenes explains both its superior quantitative performance and the observed dataset-dependent variations.

These strong results are attributed to the synergistic interaction of the model's three core components—region-adaptive segmentation, deep pyramidal reconstruction, and global numerical optimization.

Table 2. Quantitative Results

Method	UIQI(dataset2)	UIQI(dataset1)	RMSE(dataset2)	RMSE(dataset1)	SAM(dataset2)	SAM(dataset1)	CC(dataset2)	CC(dataset1)
HIS	0.842	0.858	10.14	12.47	6.03	6.29	0.908	0.912
PCA	0.850	0.863	10.57	11.96	6.89	6.16	0.909	0.908
DWT	0.891	0.911	9.86	11.38	5.71	5.81	0.917	0.917
HARNN	0.926	0.939	7.24	9.43	4.72	4.48	0.934	0.944
SDRCNN	0.932	0.947	7.35	9.70	4.07	4.74	0.931	0.948
TDNET	0.929	0.940	7.80	9.44	4.83	4.65	0.943	0.950
PCNN	0.941	0.948	7.19	9.61	4.48	4.32	0.942	0.954
HetSSNet	0.945	0.955	7.07	9.27	4.12	4.60	0.946	0.952
PCDRN	0.946	0.954	6.81	9.04	4.94	4.39	0.952	0.957
Proposed method	0.962	0.969	6.15	7.49	3.21	3.52	0.968	0.973

The segmentation stage, implemented via PSO-driven clustering, alleviates the cross-region spectral mixing problem commonly observed in global methods and in K-means or SOM-based segmentation. By optimizing spectral entropy, texture variation, and boundary consistency, PSO generates homogeneous segments that respect scene structure and allow context-aware processing. This significantly reduces spectral distortion around roads, roofs, agricultural parcels, and mixed boundaries—areas where classical methods such as IHS, PCA, and Laplacian-based fusion often degrade. The second stage leverages a Laplacian pyramid combined with region-conditioned CNN and ResNet-style residual learning modules. The Laplacian decomposition ensures multi-scale enhancement of PAN-driven detail, while the residual CNN captures nonlinear local interactions that analytical models fail to encode. Unlike fully end-to-end deep networks such as PANNet that require large labeled datasets, the proposed method embeds the CNN within a region-specific, pyramid-guided structure, allowing the model to learn local patterns without overfitting or color distortion. Finally, the global optimization stage employs LSQR as the primary solver of the constraint system, stabilizing gradient fields derived from the PAN image and eliminating seam-level inconsistencies introduced by region-wise processing. LSQR demonstrates superior performance to SVD and LSCOV in terms of noise suppression, convergence in sparse systems, and edge preservation, especially on large high-resolution satellite data.

Altogether, the integrated effect of segmentation-guided contextualization, deep spectral–spatial modeling, and globally constrained optimization yields a pansharpening architecture that achieves simultaneous improvement across RMSE, CC, UIQI, and SAM, as reflected in Table 2. More importantly, the statistical validation confirms that these improvements are consistent, generalizable, and statistically significant across scenes. The method's independence from large training datasets, combined with its adaptability to heterogeneous landscapes, positions it as a robust and scalable solution for high-resolution remote sensing applications. The combination of quantitative accuracy, statistical reliability, and architectural interpretability demonstrates that the proposed framework provides a stable, spectrally faithful, and spatially sharp fusion strategy superior to classical models, traditional hybrids, and purely deep networks.

To compare fusion methods in term of computational complexity, execution time for ten fusion methods is shown in Table 3. From this table, one can see that the proposed method is not the fastest one, but it is faster than state-of-the-art fusion methods like HetSSNet and PCDRN.

Table 3. Computation Time for Different Fusion Methods

Methods	Times in seconds
HIS	0.21
PCA	0.34
DWT	0.39
HARNN	12.57
SDRCNN	14.95
TDNET	11.68
PCNN	9.70
HetSSNet	13.55
PCDRN	12.09
Proposed method	11.82

5) Conclusion

This work presents a region-adaptive hybrid pansharpening framework that integrates unsupervised segmentation, deep pyramidal spectral–spatial modeling, and global numerical optimization into a unified architecture. Unlike classical fusion methods and purely deep-learning-based approaches that apply global transformation rules, the proposed model processes spectrally homogeneous regions independently while enforcing global consistency through LSQR optimization. This design allows the method to preserve fine spatial structures and weak spectral signatures that are commonly degraded in heterogeneous scenes. Comprehensive experiments on IKONOS dataset demonstrate that the proposed framework achieves a superior and well-balanced performance across RMSE, UIQI, CC, and SAM metrics. Statistical validation using paired t-tests, Wilcoxon signed-rank tests, Bonferroni correction, and effect-size analysis confirms that these improvements are statistically significant and consistent across scenes. Importantly, the method does not rely on large labeled training datasets, making it particularly suitable for real-world remote-sensing applications where high-resolution ground truth is scarce. Overall, the combination of region-aware segmentation, multiscale Laplacian fusion, neural refinement, and LSQR-based global optimization yields a spectrally faithful, spatially sharp, and statistically robust pansharpening solution. Future work may explore the integration of attention mechanisms, transformer-based modules, and multi-task learning strategies to further enhance adaptability and performance in complex remote-sensing environments.

References

Aiazzi, B., Baronti, S., Pippi, I., & Alparone, L. (2003). Decision-driven pyramid fusion of multispectral and panchromatic images. In *Geoinformation For European-Wide Integration (pp. 273-278). Millpress Science Publisher, PO BOX 84118, 3009 CC Rotterdam*. Netherland. https://doi.org/10.1080/01431160512331314056

Burt, P. J., & Adelson, E. H. (1987). The Laplacian pyramid as a compact image code. In *Readings in computer vision* (pp. 671-679). Morgan Kaufmann. http://dx.doi.org/10.1016/B978-0-08-051581-6.50065-9

Chao, W., Rui, Y., Sun, Y., Jiang, Y., & Lin, X. (2018). A new method of multi-focus image fusion using laplacian operator and region optimization. *Journal of Computer and Communications, 6*(5), 106-118. http://dx.doi.org/10.4236/jcc.2018.65009

Fang, Y., Cai, Y., & Fan, L. (2023). SDRCNN: A single-scale dense residual connected convolutional neural network for pansharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16*, 6325-6338. http://dx.doi.org/10.1109/JSTARS.2023.3292320

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations*. JHU press. http://dx.doi.org/10.56021/9781421407944

Golub, G., & Kahan, W. (1965). Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis, 2*(2), 205-224. http://dx.doi.org/10.1137/0702016

He, L., Rao, Y., Li, J., Chanussot, J., Plaza, A., Zhu, J., & Li, B. (2019). Pansharpening via detail injection based convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12*(4), 1188-1204. https://doi.org/10.1109/JSTARS.2019.2898574

Kohonen, T. (2012). *Self-organizing maps* (Vol. 30). Springer Science & Business Media. http://dx.doi.org/10.1007/978-3-642-97966-8

Lemaire, V., & Clérot, F. (2005). The many faces of a Kohonen map a case study: SOM-based clustering for on-line fraud behavior classification. In S. K. Halgamuge & L. Wang (Eds.), *Classification and clustering for knowledge discovery* (pp. 1-13). Springer Berlin Heidelberg. http://dx.doi.org/10.1007/11011620_1

Li, X. J., Yan, H. W., Yang, S. W., Kang, L., & Lu, X. M. (2018). Multispectral pansharpening approach using pulse-coupled neural network segmentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42*, 961-965. http://dx.doi.org/10.5194/isprs-archives-XLII-3-961-2018

Li, X., Xu, F., Lyu, X., Tong, Y., Chen, Z., Li, S., & Liu, D. (2020). A remote-sensing image pan-sharpening method based on multi-scale channel attention residual network. *Ieee Access, 8*, 27163-27177. http://dx.doi.org/10.1109/ACCESS.2020.2971502

Li, Y., & Zhao, J. (2020). A novel medical image fusion method using multi-channel pulse coupled neural networks. *IEEE Access, 8*, 157572-157586. http://dx.doi.org/10.1109/ACCESS.2020.3019426

Liu, Q., Han, L., Tan, R., Fan, H., Li, W., Zhu, H. & Liu, S. (2021). Hybrid attention based residual network for pansharpening. *Remote Sensing, 13*(10), 1962. http://dx.doi.org/10.3390/rs13101962

Liu, X., Liu, Q., & Wang, Y. (2020). Remote sensing image fusion based on two-stream fusion network. *Information Fusion, 55*, 1-15. http://dx.doi.org/10.1016/j.inffus.2019.07.010

Ma, M., Jiang, Y., Zhao, M., Li, J., & Zhang, W. (2025). HetSSNet: Spatial-Spectral heterogeneous graph learning network for panchromatic and multispectral images fusion. *arXiv preprint arXiv:2502.04623*. https://doi.org/10.48550/arXiv.2502.04623

Masi, G., Cozzolino, D., Verdoliva, L., & Scarpa, G. (2016). Pansharpening by convolutional neural networks. *Remote Sensing, 8*(7), 594. http://dx.doi.org/10.3390/rs8070594

Paige, C. C., & Saunders, M. A. (1982). LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS), 8*(1), 43-71. http://dx.doi.org/10.1145/355984.355989

Quan, W., & Guo, W. (2023). DTEMPan: Dual texture-edge maintaining transformer for pansharpening. *IEEE Transactions on Geoscience and Remote Sensing, 61*, 1-16. http://dx.doi.org/10.1109/TGRS.2023.3335363

Raghavan, K. R. (1994). *Inverse and optimum drive problems in ultrasonic imaging of soft tissue elasticity*. University of Michigan. https://dx.doi.org/10.1109/NSSMIC.1993.373621

Rao, Y., He, L., & Zhu, J. (2017, May). A residual convolutional neural network for pan-shaprening. In *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)* (pp. 1-4). IEEE. http://dx.doi.org/10.1109/RSIP.2017.7958807

Shen, H., Jiang, M., Li, J., Yuan, Q., Wei, Y., & Zhang, L. (2019). Spatial–spectral fusion by combining deep learning and variational model. *IEEE Transactions on Geoscience and Remote Sensing, 57*(8), 6169-6181. http://dx.doi.org/10.1109/TGRS.2019.2904659

Shen, H., Meng, X., & Zhang, L. (2016). An integrated framework for the spatio–temporal–spectral fusion of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 54*(12), 7135-7148. http://dx.doi.org/10.1109/TGRS.2016.2596290

Wang, W., Liu, H., & Xie, G. (2021). Pansharpening of WorldView-2 data via graph regularized sparse coding and adaptive coupled dictionary. *Sensors, 21*(11), 3586. http://dx.doi.org/10.3390/s21113586

Wang, W., Zhou, Z., Liu, H., & Xie, G. (2021). MSDRN: Pansharpening of multispectral images via multi-scale deep residual network. *Remote Sensing, 13*(6), 1200. http://dx.doi.org/10.3390/rs13061200

Yang, Y., Tu, W., Huang, S., & Lu, H. (2020). PCDRN: Progressive cascade deep residual network for pansharpening. *Remote Sensing, 12*(4), 676. http://dx.doi.org/10.3390/rs12040676

Yuan, Q., Zhang, L., & Shen, H. (2013). Hyperspectral image denoising with a spatial–spectral view fusion strategy. *IEEE Transactions on Geoscience and Remote Sensing, 52*(5), 2314-2325. http://dx.doi.org/10.1109/TGRS.2013.2259245

Zhang, T. J., Deng, L. J., Huang, T. Z., Chanussot, J., & Vivone, G. (2022). A triple-double convolutional neural network for panchromatic sharpening. *IEEE Transactions on Neural Networks and Learning Systems, 34*(11), 9088-9101. http://dx.doi.org/10.1109/TNNLS.2022.3155655