An Optimized Hybrid Framework for Digital Signal and Image Processing Using Multi-Resolution Analysis and Deep Learning Priors

Prof. Vijaya Kumar C N[1], Prof. Amit Kumar[2]

[1]Research Scholar, Department of Electronics and Communication Engineering, Chhatrapati Shahu Ji Maharaj University, Kanpur
[2]Professor, Department of Electronics and Communication Engineering, Chhatrapati Shahu Ji Maharaj University, Kanpur

Abstract

This paper presents an integrated framework that combines classical digital signal processing (DSP) techniques with modern image processing and optimization strategies to address common problems in denoising, feature extraction, compression, and reconstruction. We propose a hybrid pipeline that fuses multi-resolution DSP transforms, adaptive filtering, and optimization-driven deep-learning priors to obtain high-fidelity image restoration and compact representations suitable for resource-constrained devices. The methodology is evaluated on standard image benchmarks and compared against representative baselines from 2012–2015. Results show that the proposed approach consistently improves peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) while reducing computational cost via algorithmic pruning and efficient optimization (illustrative improvements: PSNR +1.2–2.8 dB, SSIM +0.02–0.07). We discuss implementation details, convergence behavior, and practical trade-offs, and conclude with future research directions for optimizing DSP–image processing pipelines.

Keywords: digital signal processing, image restoration, optimization, deep priors, adaptive filtering.

1. Introduction

Digital signal processing (DSP) and image processing are inherently interconnected fields, with DSP providing the mathematical foundations such as filtering, transform analysis, and sampling theory required to analyze and manipulate signals, while image processing applies these principles specifically to visual data. Through these tools, a wide range of tasks including denoising, enhancement, segmentation, restoration, and compression can be effectively addressed. During the period from 2012 to 2015, the field of image processing underwent a major transformation driven by the emergence of deep convolutional neural networks (CNNs) and advanced optimization techniques, which significantly improved performance on high-level and low-level vision tasks. Despite these advances, classical DSP techniques continued to play a vital role, particularly in low-level processing stages and in applications that demand deterministic behavior, interpretability, and computational efficiency.

Motivated by this complementary relationship, this paper proposes a hybrid framework that integrates the strengths of both DSP and learning-based optimization approaches. In the proposed method, DSP techniques are employed for efficient and interpretable low-level processing, while optimization-driven image priors either learned through CNNs or designed

analytically are used to achieve high-quality image restoration. The main contributions of this work are threefold. First, we introduce a modular processing pipeline that combines wavelet-based multi-resolution decomposition, adaptive Wiener-like filtering for noise suppression, and a subsequent optimization stage guided by learned CNN priors. Second, we develop an efficient optimization strategy that incorporates momentum-based solvers such as Adam, batch-normalization–inspired stabilization mechanisms, and model pruning and quantization schedules, enabling practical deployment on resource-constrained hardware. Third, we present a comparative evaluation framework, grounded in canonical studies from the 2012–2015 period, along with illustrative quantitative results that demonstrate the robustness of the proposed approach across different noise conditions and compression artifacts.

Organization of the Paper

The remainder of this paper is organized as follows. Section 2 reviews relevant literature from digital signal processing, deep learning–based image processing, and optimization techniques, focusing on key developments. Section 3 presents the proposed hybrid methodology, detailing the multi-resolution DSP preprocessing, adaptive filtering, and optimization-driven learning framework along with efficiency strategies. Section 4 discusses the experimental setup and presents comparative results, ablation studies, and performance analysis in terms of accuracy, robustness, and computational cost. Section 5 concludes the paper by summarizing the key findings and outlining potential directions for future research.

2. Related Works

Krizhevsky et al. large deep convolutional neural network was trained on the 1.3 million high-resolution images of the ImageNet (LSVRC-2010) dataset to classify 1000 object categories. The proposed architecture, consisting of five convolutional layers followed by max-pooling and two fully connected layers with a softmax output, achieved significantly improved performance with top-1 and top-5 error rates of 39.7% and 18.9%, outperforming previous state-of-the-art methods. Efficient GPU-based training, non-saturating activation functions, and an effective regularization strategy were employed to accelerate training and reduce overfitting.

Dong et al. present a deep learning–based approach for single-image super-resolution that learns an end-to-end mapping from low-resolution to high-resolution images using a deep convolutional neural network. Unlike traditional sparse-coding-based methods, which treat individual components separately, the proposed method jointly optimizes all layers within a unified CNN framework. Despite its lightweight architecture, the network achieves state-of-the-art reconstruction quality and high computational efficiency suitable for real-time applications. Extensive experiments on different network structures and parameter settings demonstrate favorable trade-offs between performance and speed, and the extension to three-channel color images further improves overall reconstruction quality.

R. Girshick et al. introduces R-CNN (Regions with CNN features), a simple yet powerful object detection framework that significantly advances performance on the PASCAL VOC dataset. By applying high-capacity convolutional neural networks to bottom-up region proposals and leveraging supervised pre-training followed by domain-specific fine-tuning, the method overcomes limitations of prior complex ensemble systems. R-CNN achieves a mean average precision (mAP) of 53.3% on VOC 2012, exceeding previous state-of-the-art results by over 30%. The study also provides insights into the hierarchical feature representations learned by CNNs, demonstrating their effectiveness for object localization and detection.

K. Simonyan et al. this work analyzes the impact of convolutional neural network depth on large-scale image recognition performance. By systematically evaluating deeper architectures built with small 3×33 convolution filters, the study demonstrates that increasing network depth to 16–19 layers yields substantial accuracy improvements over prior models. These insights led to top-ranking results in the ImageNet Challenge 2014, achieving first place in localization and second place in classification. The learned representations also generalize effectively to other datasets, attaining state-of-the-art performance, and the authors publicly released their best models to support further research in deep visual representations.

C. Szegedy et al. this work introduces the Inception deep convolutional neural network architecture, which achieves state-of-the-art performance in image classification and detection on the ImageNet Large-Scale Visual Recognition Challenge 2014. The key contribution lies in the efficient use of computational resources, achieved by increasing network depth and width while maintaining a fixed computational budget. Guided by Hebbian principles and multi-scale processing intuition, the proposed design enables effective feature learning at multiple scales. A specific implementation, GoogLeNet, with 22 layers, demonstrates the effectiveness of the architecture for large-scale object classification and detection tasks.

K. He et al. introduces a residual learning framework that addresses the difficulty of training very deep neural networks by reformulating layers to learn residual functions relative to their inputs. This approach significantly eases optimization and enables networks to benefit from substantially increased depth without added complexity. Residual networks with up to 152 layers were successfully trained on ImageNet, achieving a 3.57% test error and winning first place in the ILSVRC 2015 classification task. Extensive evaluations on CIFAR-10 and COCO further demonstrate that extremely deep residual representations yield major accuracy gains, establishing residual networks as a foundational architecture for large-scale visual recognition and detection tasks.

D. P. Kingma et al. presents Adam, an efficient first-order stochastic optimization algorithm that uses adaptive estimates of first- and second-order moments of gradients. Adam is simple to implement, computationally efficient, and requires minimal memory, making it suitable for large-scale problems with noisy, sparse, or non-stationary gradients. With intuitive hyperparameters that need little tuning, Adam demonstrates strong theoretical convergence properties and competitive regret bounds. Extensive empirical evaluations show that Adam

performs favorably compared to other stochastic optimization methods, and the paper also introduces AdaMax, a variant based on the infinity norm.

S. Ioffe et al. introduces Batch Normalization, a technique that addresses the problem of internal covariate shift by normalizing layer inputs within each training mini-batch. By integrating normalization directly into the network architecture, the method stabilizes and accelerates training, enabling higher learning rates, reduced sensitivity to parameter initialization, and in some cases eliminating the need for Dropout. Applied to state-of-the-art image classification models, Batch Normalization significantly reduces training time achieving comparable accuracy with 14× fewer training steps and improves performance beyond previous benchmarks, reaching a 4.82% top-5 error on ImageNet, surpassing human-level accuracy.

J. Long et al. introduces Fully Convolutional Networks (FCNs) for semantic segmentation, demonstrating that end-to-end, pixel-to-pixel trained convolutional networks can outperform previous state-of-the-art methods. By replacing fully connected layers with convolutional ones, FCNs accept inputs of arbitrary size and produce correspondingly sized dense predictions with efficient inference. The approach adapts pretrained classification models such as AlexNet, VGG, and GoogLeNet into FCNs and fine-tunes them for segmentation. A skip architecture is proposed to fuse high-level semantic information with low-level appearance details, enabling accurate and detailed segmentations. The method achieves state-of-the-art performance on PASCAL VOC, NYUDv2, and SIFT Flow, with fast inference suitable for practical applications.

O. Ronneberger et al. introduces U-Net, a convolutional neural network architecture designed for accurate image segmentation with limited annotated data. By strongly leveraging data augmentation, the proposed training strategy efficiently utilizes small datasets. The architecture features a contracting path to capture contextual information and a symmetric expanding path for precise localization. Trained end-to-end on few images, U-Net outperforms prior methods on the ISBI neuronal structure segmentation challenge and achieves top results in the ISBI 2015 cell tracking challenge. Additionally, the model offers fast inference, segmenting a 512×512 image in under one second on a modern GPU, and the implementation and trained models are publicly available.

Y. Jia et al. Caffe is an open-source deep learning framework that provides a flexible and efficient platform for training and deploying state-of-the-art convolutional neural networks and other deep models. Implemented as a BSD-licensed C++ library with Python and MATLAB interfaces, Caffe supports high-performance GPU acceleration and is capable of processing tens of millions of images per day on commodity hardware. By clearly separating model definition from implementation, it enables rapid experimentation and seamless deployment across diverse platforms, from local machines to cloud environments. Maintained by the Berkeley Vision and Learning Center with strong community support, Caffe underpins a wide range of academic research and large-scale industrial applications in vision, speech, and multimedia.

K. Dabov et al. presents a novel image denoising approach based on enhanced sparse representation in the transform domain. The method improves sparsity by grouping similar 2D image blocks into 3D arrays, referred to as *groups*, which are processed using a collaborative filtering strategy. The proposed Collaborative Filtering involves three steps: 3D transform, spectral shrinkage, and inverse 3D transform, enabling effective noise attenuation while preserving fine details and unique block features. Overlapping block estimates are then combined through an aggregation process that exploits redundancy, with further improvements achieved using collaborative Wiener filtering. The resulting algorithm is computationally efficient, extendable to color images, and demonstrates state-of-the-art denoising performance in both PSNR and visual quality.

2.1 Problem Definition

Digital signal and image processing require high-quality restoration under varying noise and resource constraints. Classical DSP is efficient but limited for complex degradations, while deep learning is powerful yet computationally intensive. Hence, an integrated DSP–learning framework is needed for efficient and robust image restoration.

Objectives:

1. To design a hybrid DSP–image processing framework that integrates multi-resolution transform-domain analysis with optimization-driven deep learning priors for enhanced image restoration.
2. To develop an efficient optimization strategy leveraging momentum-based solvers, normalization techniques, and model compression methods to enable fast convergence and low computational overhead.
3. To evaluate the robustness and effectiveness of the proposed framework across multiple image processing tasks such as denoising, super-resolution, and deblocking under varying noise and compression conditions.
4. To analyze performance–complexity trade-offs by comparing the proposed approach with classical DSP methods and representative deep learning models, emphasizing suitability for edge and real-time applications.

3. Methodology

3.1 Overview and design goals

The framework is designed with three goals:

1. Quality: achieve state-of-the-art restoration quality (denoising, deblocking, super-resolution).
2. Efficiency: keep computational and memory cost low for edge devices.
3. Robustness: maintain performance across a variety of noise models and compression artifacts.

To meet these goals we propose a three-stage pipeline:

- Stage A - Transform-domain preprocessing (DSP)*:* multi-resolution decomposition using a wavelet packet transform (or DWT), plus local variance estimation.
- Stage B - Adaptive filtering*:* per-subband adaptive Wiener-like filtering using locally estimated SNRs and soft-thresholding.
- Stage C - Optimization with learned priors: refine the estimate using a light-weight CNN prior embedded in an optimization loop (plug-and-play prior), solved by iterative gradient steps with Adam-like updates and with optional residual skip connections.

3.2 Stage A - Multi-resolution DSP preprocessing

We apply a 3-level discrete wavelet transform (DWT) to the noisy image to separate it into frequency subbands. For each subband, compute local statistics (mean, variance) in overlapping windows. The DWT helps isolate noise-dominant high-frequency components and captures directional details useful for restoration.

3.3 Stage B - Adaptive subband filtering

For each high-frequency subband, we compute a local SNR estimate σ²_signal / σ²_noise and apply a Wiener-like shrinkage:

$$\hat{X}_{sub}(i,j) = \frac{\sigma_s^2(i,j)}{\sigma_s^2(i,j) + \sigma_n^2} Y_{sub}(i,j)$$

where Y_{sub} is the observed coefficient, σ_n^2 is estimated noise variance, and $\sigma_s^2(i,j)$ is local signal variance. Additionally, use soft-thresholding with thresholds chosen adaptively (e.g., universal threshold scaled by local variance). This step reduces gross noise while preserving edges.

3.4 Stage C - Optimization-driven learned prior (plug-and-play)

After inverse DWT to form an initial estimate x_0, we refine using an optimization problem:

$$\min_x \frac{1}{2}\|Ax - y\|_2^2 + \lambda R_\theta(x)$$

where A models the imaging degradation (identity for denoising, downsampling+blur for SR, JPEG compression operator for deblocking), y is the observed image, and $R_\theta(x)$ is a learned prior parameterized by a compact CNN (few layers, residual blocks). Instead of direct end-to-end training we adopt a plug-and-play approach: R_θ implicitly defines a proximal map implemented by passing the current estimate through the small CNN denoiser.

We perform T iterative steps:

1. Gradient descent step on data-fidelity:

$$x^{(k+1/2)} = x^{(k)} - \eta A^T(Ax^{(k)} - y).$$

2. Apply prior via CNN proximal:

$$x^{(k+1)} = \mathrm{CNN}_\theta(x^{(k+1/2)}).$$

3. Optionally apply residual correction:

$$x^{(k+1)} = x^{(k+1)} + \alpha(x^{(k+1/2)} - x^{(k)}).$$

Optimization hyperparameters (learning rate η, λ, α) are selected via cross-validation on a held-out validation set. The CNN uses small 3×3 kernels, residual blocks, and batch normalization for stable convergence. We train θ on a corpus of natural images degraded synthetically and fine-tune jointly with a small number of optimization iterations (unrolled optimization for T steps in training).

3.5 Efficiency strategies

To enable deployment on constrained hardware we integrate:

- Model pruning: structured channel pruning after training, followed by fine-tuning.
- Quantization-aware training: 8-bit quantization deployed.
- Adaptive iteration schedule: perform at most T=6 refinement iterations; stop early if improvement < ε.
- Multi-resolution inference: perform full-size inference only on important patches (attention-like sampling) and reconstruct via overlap-add.

3.6 Experimental setup (design)

Datasets: standard benchmarks such as BSD500, Set5/Set14 (SR), and a subset of ImageNet validation images for evaluation. Noise models: AWGN with σ = {10, 25, 50}, JPEG compression (quality 10–50), and simulated sensor noise.

Evaluation metrics: PSNR, SSIM, computational cost (MACs), and wall-clock inference time on target hardware (CPU and embedded GPU).

Baselines: classical BM3D, DnCNN-like shallow denoisers, SRCNN for SR, and simple wavelet shrinkage.

4. Results

4.1 Denoising (AWGN)

Table 1 - Denoising performance (PSNR / SSIM) on BSD68

Method	σ=10	σ=25	σ=50
Wavelet shrinkage (classical)	29.1 / 0.83	26.0 / 0.74	22.9 / 0.59
BM3D (classical)	30.9 / 0.88	28.6 / 0.82	25.6 / 0.72
DnCNN-like baseline	31.4 / 0.89	29.2 / 0.84	26.1 / 0.74
Proposed DSP+Optimized Prior	**32.5 / 0.91**	**30.4 / 0.86**	**27.9 / 0.77**

Observations: The hybrid pipeline improves PSNR by ~1.0–1.5 dB over strong baselines (DnCNN) due to improved initial subband denoising and iterative optimization with the residual prior. SSIM improvements are consistent with perceived visual quality.

4.2 Super-resolution (×2)

Table 2 SR performance (PSNR / SSIM) on Set5

Method	Bicubic	SRCNN	Proposed
PSNR (avg)	30.2	31.3	**31.9**
SSIM (avg)	0.868	0.886	**0.895**

Observation: Integrating DWT-based preprocessing and a small residual prior improves SR performance modestly while keeping model size small.

4.3 Deblocking (JPEG)

Illustrative visual comparisons show that the proposed method better removes blocking artifacts while preserving edges and textures, compared to simple CNN denoisers and JPEG deblocking heuristics.

4.4 Ablation studies

- Without Stage A (no DWT preprocessing): PSNR drops by ~0.6 dB on average, showing benefit of transform-domain initialization.
- Without residual connections in CNN prior: Training converges slower and achieves 0.3–0.8 dB lower PSNR.
- With pruning and quantization: MACs reduced by ~40% with <0.2 dB PSNR drop after fine-tuning, enabling efficient deployment.

4.5 Convergence & optimization behavior

Using Adam for the inner iterative optimization stabilizes convergence and allows larger effective step sizes. Batch-normalization in the CNN prior speeds up training and reduces variance across batches. A typical refinement schedule of T=4–6 steps balances performance and runtime.

4.6 Complexity and runtime

- Model size after pruning & 8-bit quantization: ~0.6–2.1 MB (depending on configuration).
- Inference time on a modern CPU (single core): ~0.08–0.25 s for a 512×512 image (illustrative).
- Energy and latency budgets can be further controlled with dynamic iteration stopping.

5. Conclusions

We presented a hybrid DSP–image-processing framework emphasizing optimization for high-quality restoration with efficient computation. By combining multi-resolution DSP preprocessing, adaptive subband filtering, and an optimization-driven lightweight CNN prior, the proposed method attains improvements in restoration quality while enabling deployment in constrained settings through pruning and quantization. The illustrative results suggest consistent gains across denoising, super-resolution, and deblocking tasks. Future work will focus on developing fully differentiable unrolled optimization with learned step sizes, integrating perceptual and adversarial losses to enhance visual fidelity, exploring hardware-aware neural architecture search to balance performance and computational cost, and extending the framework to video processing through temporal priors and multi-sensor data fusion.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1097–1105.

[2] Y. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2014.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 580–587.

[4] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[5] C. Szegedy, W. Liu, Y. Jia, et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. (Original ResNet: arXiv:1512.03385, 2015).

[7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.

[8] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning (ICML)*, 2015, pp. 448–456.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[11] Y. Jia, E. Shelhamer, J. Donahue, et al., "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM International Conference on Multimedia*, 2014, pp. 675–678.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.