Results in Nonlinear Analysis 8 (2025) No. 2, 314–327
https://doi.org/10.31838/rna/2025.08.02.024
Available online at www.nonlinear-analysis.com

Results in Nonlinear Analysis
ISSN 2636-7556

Peer Reviewed Scientific Journal

Nonlinear optimization-driven deep learning framework for medical image reconstruction via partial differential equations

TKS Rathish Babu[a], P. Sedhupathy[b], M. Aruna[c], V. Srinivasan[d], Baxodirjon Abdullaev[e], Islom Kadirov[f]

[a]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai - 600 089, Tamilnadu, India; [b]Assistant Professor, Department of Computer Science (Artificial Intelligence & Data Science), Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore, Tamilnadu, India; [c]Assistant Professor, Department of MCA, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India; [d]Associate Professor, Department of Computer Applications, Dayananda Sagar College of Engineering, Kumaraswamy Layout, Bengaluru - 560 111, Karnataka, India; [e]Associate Professor, Faculty of Mechanical Engineering, Department of Mechanical Engineering, Andijan State Technical Institute, Andijan, Uzbekistan; [f]Department of Transport Systems, Urgench State University named after Abu Raykhan Beruni, Urgench, Republic of Uzbekistan

Abstract

High quality medical imaging is essential to accurate clinical decision-making, but reconstruction of sparse or noisy images especially under CT and MRI is still a major challenge, with traditional reconstruction algorithms vulnerable to artifacts and noise, and unwanted inference typically lacking interpretability. We introduce a new modality of addressing the problem of reconstruction with nonlinear optimization, partial differential equation (PDE) constraints and deep neural networks, where the priors on physical properties should be presented as the network loss function and the architecture of the network so as to build more robust and accurate reconstruction. Having a clear formulation of a nonlinear optimization problem and by using the principles of variational approaches, we are also able to integrate a hardware friendly circuit into our solution that could be used to acquire data in real time. Experiments on benchmark CT and MRI, indicate an increase in peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), more effective noise suppression, and convergence times than state-of-the art baselines. The need of nonlinear optimization and employment of PDEs regularization in work on edges preservation and reduction of artifacts can also be seen regarding

Email addresses: tksbabu80@gmail.com (TKS Rathish Babu), sedhupathy@gmail.com (P. Sedhupathy), arunasrini2005@gmail.com (M. Aruna), srinivasan-mcavtu@dayanandasagar.edu (V. Srinivasan), bahodir.abdullayev1986@gmail.com (Baxodirjon Abdullaev), islomqadirov1415@gmail.com (Islom Kadirov)

Received June 10, 2025; Accepted July 18, 2025; Online August 28, 2025

ablation results. Taken together, it is the first piece connecting the fields of model-based regularization and modern deep learning, thereby providing a clinical pipeline toward interpretable, high-fidelity, and deployable medical image reconstruction.

Mathematics Subject Classification (2010): 65K10, 65N21, 68T07, 35Q68
Key words and phrases: Medical image reconstruction, Nonlinear optimization, Deep learning, Partial differential equations (PDEs).

1. Introduction

Medical image reconstruction is a key component of the current state of healthcare, the foundation of the accurate diagnosis and therapy planning, and longitudinal monitoring of the patient. Computed tomography (CT) and magnetic resonance imaging (MRI) are the modalities that produce huge volumes of raw measurement data, out of which clinically useful images will be reconstructed. Conventional methods such as filtered back-projection (FBP), and the algebraic reconstruction technique (ART) give effective computational solutions when the sampling conditions are ideal. In practice, however, these methods sometimes collapse in noisy environments, with low-dose data or small acquisition angles, and they can cause undesirable artifacts, loss of anatomical detail and poor diagnostic reliability.

Deep learning has since then become a radical paradigm to overcome these shortcomings by capitalising on the availability of large data sets to learn rich and typically non-linear priors on signals that can be used to produce remarkable reconstruction quality even in difficult situations. CNNs and their variations have also shown great ability of recovering details and suppressing artifacts. Nonetheless, data-driven methods that rely solely on data are often limited in their generalization capabilities especially when applied to out-of-distribution or uncommon pathologies and are often not interpretable or lacking in theoretical guarantees, a factor essential in sensitive medical decision-making.

Coupled with these gaps are increased interests in hybrid models that combine the interpretability of model-based physics with learned representations in terms of flexibility. Surprisingly, there is now interest in integrating PDE models and deep learning, since PDEs capture compact structural priors (e.g. smoothness, edge retention, anisotropy). Moreover, other learning and inference strategies based on nonlinear optimization can constrain the functions of such solutions to data fidelity and complex constraints.

The proposed work is a nonlinear optimization-based deep learning paradigm that aims to regularize the reconstruction of medical imaging data, synthesizing an end-to-end pathway that involves direct integration of explicit regularization associated with PDE at both the architectural and the loss-function level. This formally represents a variational loss with physically significant priors, A rather strong, artifact-insensitive, edge-obtaining reconstructions are then available. We will also suggest a feasible hardware-based, circuit based data acquisition and reconstruction in real-time, and hardware-in-the loop, thereby approaching translation into clinical practice.

The main merits of this work may be considered to be:

1. Unified Hybrid Framework: Architecture of deep neural that connects the nonlinear optimization with the PDE-based priors in order to take the benefits of the classical regularization of flexibility to that of modern neural networks.
2. Variational Regularization: The insertion of a plain variational loss function enforced by PDE, resulting in the solution that is credible in real-world physics, and which is in its turn also easier to comprehend and rely upon.
3. Hardware-Inclusive Design: Suggesting a hardware amenable to circuit interface to enable real time data capturing and pre-processing and allow easy integration with current medical imaging system.
4. Full Analysis: high-fidelity experimental benchmarks of CT and MRI, superior quantitative and qualitative results on reconstruction quality than both traditional and contemporary state-of-the-art learning-based methods.

2. Literature Review

More classical analytical reconstructions like filtered back-projection (FBP) and algebraic reconstruction (ART) are also appealing due to speed, but incomplete sampling and noise presents a challenge due to streaking and blur, and the explanation of how these options possess inadequacies in treatment with fundamental CT/MRI-oriented physics operators is given in the seminal text by Kak and Slaney [2]. Initial surveys of deep learning in medical imaging provide background as to why the simple analytical pipelines prove ineffective in capturing more complicated priors found in anatomy and acquisition variability [1]. Compressed sensing (CS) was developed to mitigate artifacts due to aggressive under sampling of MRI by incorporating sparsity-based recovery with a resultant drastic reduction in needed measurements at the cost of increased optimization weights and gradual convergence of practice [3]. Anisotropic diffusion and most commonly total variation (TV) variational regularizer are often used to ensure edges and eliminate noise because it has an interpretation that is easy to understand that corresponds well with physics-based fidelity terms [6, 7]. In highly scaled imaging tasks, these variational models have their algorithmic core provided by modern convex and first-order solvers (e.g., primal-dual and proximal methods) [9].

Deep learning enhanced the reconstruction in a sense that it learned strong priors and data-consistency operators. The multi-scale architecture of U-Net style and their performance generally result in the becoming of de-aliasing in MRI/CT backbone [4]. Adversarial learning can be used to recover fine textures; it was initially suggested to apply to photo-realistic super-resolution, but results in medical reconstruction tasks have been driven by GAN concepts (and perceptual losses) [5]. Combining learnt modules with explicit optimization has become popular in the form of hybrid schemes neural proximal gradient descent reflects the paradigm of iterative solvers but preserves proximal structure [8], and algorithm unrolling makes many such hybrids tractable as interpretable, trainable networks whose parameters relate to each other in a stage-wise structure of the forward model [10]. Extending this, PDE-constrained or even PDE-inspired deep networks have also shown increased edge retention and artifact reduction in dynamic MRI [11], and residual-unrolled structures with PDE-inspired penalties do even more to stabilize training and increase data coherence (through continuity) [13]. Outside of CNNs, transformer backbones trained on PDE flows show reliable PSNR improvements and improved long-range modelling especially in dynamic environments [15]. At a more fundamental level, bilevel/variational views make it clear that learning hyper-parameters (e.g. regularization runners) through end-to-end training Generalization capacity The ability of the structure derived by inductive learning and induction to generalize to untrained regarding their stability [14] and, more generally, how any parameter dependent on parameters that appear in the minimization problem described by the functional [15]. The trend towards these aspects and such pitfalls as robustness and over-smoothing risks, are in a wider methodological context that are presented in recent surveys and tutorial papers across signal processing and biomedical AI [12, 16, 19].

Last but not least, there has emerged an increasing attention on hooking latency, determinism, and energy efficiency, where the experience in embedded systems and reconfigurable computing has relevance. Hardware/software co-design principles and resource-aware scheduling can then express unrolled DC-plus-PDE networks into FPGA/ASIC-GPU pipeline-mode that is suitable near-real-time application [18]. Available books on prototyping and validation across different engineering domain domains emphasize toolchains, testing concepts which could be used to develop the medical imaging systems capable of integrating reliably with the scanners and scanner controllers [17]. RTL/logic design, not by itself a medical activity per serein, but methodologically, by application of the same forces, imposes value on clean module interfaces and timing-conscious implementations when designing iterative algorithms in hardware [20]. Taken together, these trends lead us to physics-based, optimization-sensitive deep models that are compromises between accuracy, interpretability and deployability; we are directly continuing that curve with our data-consistent, PDE-regularized framework.

Table 1: Comparative Analysis of Existing Methods vs. Proposed System

Authors (Year)	Method	Features	Results	Limitations	Comparison with Proposed	Ref
Liu et al. (2023)	PDE-Deep Net	Edge-aware, dynamic MRI	PSNR: 42.1, SSIM: 0.94	No explicit nonlinear optimization in loss	Proposed adds explicit nonlinear constraint and clearer interpretability	[14]
Zhang et al. (2023)	GAN + CS	Few-shot CS with GAN prior	Fast; SSIM: 0.91	Generalization and artifact risk	Proposed uses PDE priors and stronger data-consistency; higher SSIM and robustness to noise	[9], [13]
Park et al. (2024)	Res-Unrolled PDE Net	Unrolled scheme with PDE-inspired loss	PSNR: 41.8	No hardware consideration	Proposed includes circuit/hardware flow; real-time capable	[12]
Chen et al. (2024)	Bi-level Var + DL	Unrolled with bi-level/ learned parameters	High accuracy; slow	Heavy compute and memory	Proposed is more compute-efficient with task-aware regularization	[17], [20]
Wang et al. (2025)	PDE-Flow Transformer	Transformer with learned PDE flows	Consistent PSNR gains	Limited clinical deployment scope	Proposed is modular, transferable, and hardware-amenable	[16], [19]
This Work	Nonlinear Opt + DL + PDE (HW-ready)	PDE-based priors, explicit nonlinear constraints, real-time path	PSNR: 43+, SSIM: 0.95+	To be discussed	Advances accuracy, interpretability, and practical deployment	—

2. Methodology

2.1 Mathematical Formulation of Medical Image Reconstruction

The quantified values $y \in R^m$ (examples k-space sample in MRI or line integrals in CT), the non-quantified picture $x \in R^n$ and the physics operator $A:R^n \rightarrow R^m$. Measurement model deals with

$$y = Ax + \eta \tag{1}$$

in which model errors and noise in acquisition are η′. A few of the familiar A are:

- Multi-coil MRI A(x) = MFSx, where S, the coil sensitivities, F, Fourier operator and M, the sampling mask.
- The CT/Radon: A is the operator of projection (and System blur).

We approximate x by re-casting a regularized inverse problem as a data fitting (corresponding to the noise model) by optimizing the weighting of a PDE-type prior:

$$x\star\,|\in|\arg\min_{x\in X|}\underbrace{D(Ax,y)}_{\text{data fidelity}}\quad+|\rangle\quad\underbrace{RPDE\,(x)}_{\text{PDE regularizer}} \tag{2}$$

where $\lambda > 0$ so the trade-off between fidelity and prior is available, and X such that the various constraints possible may be encoded, such as $x \geq 0$ (CT) or magnitude constrained to be real values.

Data fidelity D.

The sampling as well as the standards of the statistical population of the noise position the selection:

- Gaussian (WLS): $D(Ax,y)=\frac{1}{2}W(Ax-y)_2^2.$ (3)

with weighting W (e.g., density compensation in MRI).

- Poisson (CT/low-dose): $D(Ax,y)=1\top\ (Ax)-y\top\ \log(\mathrm{Ax}+\mu)$ (4)

with small $\varepsilon > 0$ for numerical stability.

- Robust (outliers/inhomogeneous noise): Huber or ℓ1\ell_1ℓ1 alternatives can be used.

PDE-inspired regularizer RPDER_{\text{PDE}}RPDE. A general isotropic form is

$$RPDE(x)=\int\Omega\varnothing\ (\nabla x(r)_2)dr \tag{5}$$

with φ chosen to control diffusion/edge preservation:

- Total Variation (TV): φ(s) = s (often Huberized $\phi\delta(s)=\sqrt{s2+\delta 2}$).
- Perona–Malik anisotropic diffusion: yields a diffusion coefficient c(s) (e.g., $c(s)=\ e^{-\left(\frac{s}{\kappa}\right)^2}$ or $c(s)=1/(1+(s/^{o})2))$ that suppresses smoothing across strong edges.
- Anisotropic tensor diffusion (structure-aware): $RPDE(x)=\int\Omega\nabla x\ \top\ D(x)\nabla x\,dr$

where D(x) is a diffusion tensor derived from the (smoothed) structure tensor to encourage along-edge smoothing.

Optimality (continuous Euler–Lagrange, isotropic φ).

$$\frac{\partial}{\partial u}\ \phi(\|u\|2)=\phi'(\|u\|2)\frac{u}{\|u\|2+\epsilon}. \tag{6}$$

A stationary point satisfies

$$0=\nabla xD(Ax,y)|+|\lambda\nabla\cdot(\psi(\nabla x)), \tag{7}$$

e.g., for Gaussian fidelity, ∇ x D (Ax,y)=A⊤W⊤W(Ax−y). Neumann boundary conditions (zero normal gradient) are commonly used.

Discrete formulation. Let D_x,D_y be forward differences; define ∇=[Dx Dy] and div = $-\nabla^{\top}$. A practical discrete model is

$$X^* \in \arg\min_X \frac{1}{2}\| W(Ax - y) \|_2^2 + \lambda \sum_P \phi(\| (\nabla x)p \| 2) \text{ ,s.t. } x \in X \tag{8}$$

Solvers.

- Primal–dual (Chambolle–Pock):

$$pk+1 = \Pi P\,(pk + \sigma \nabla \bar{x}^k\ ) \tag{9}$$

$$xk+1 = \arg\min_X \frac{1}{2}\| W(Ax - y) \|_2^2 + \frac{1}{2r}\| x - (xk - \tau\, div\, pk+1\ \ ) \|_2^2 \tag{10}$$

$$\bar{x}^{k+1} = xk+1\ \theta(xk+1\ - xk), \tag{11}$$

  where Π_P is the proximal/projection tied to φ (e.g., vector soft-thresholding for TV).

- ADMM (split gradient):

$$xk+1 = \arg\min_X \frac{1}{2}\| W(Ax - y) \|_2^2 + \frac{\rho}{2}\| \nabla x - zk + uk\|_2^2 , \tag{12}$$

$$zk+1 = prox_{\frac{\lambda}{\rho\phi}}\ \left(\nabla xk+1+uk\right), uk+1 = uk+\nabla xk+1-zk+1 \tag{13}$$

  For MRI, the x-update can exploit FFTs and coil diagonal structure; for CT, conjugate gradient or preconditioned solvers are typical.

  Learned/PDE-guided variants (optional in this work).

- Unrolled PDE descent:

$$xk+1 = xk - \tau k(A \top \nabla y D(Axk, y) + \lambda k\, div\psi\theta k(\nabla xk)) \tag{14}$$

  where $\Psi_{\theta k}$ (diffusion/conduction) and step sizes are learned.

- Bilevel parameter learning: $\min_{\theta,\lambda} \sum_i L(x \star (yi;\theta,\lambda\ , x_i^{gt}\ )$ subject to the inner reconstruction problem above, enabling task-adaptive λ, κ , and diffusion tensors.

This formulation cleanly separates physics (A), statistics (D), and prior (R_{PDE}), supports multiple noise models, and admits efficient solvers and unrolled/learned implementations while retaining physical interpretability.

2.2 Deep Learning Framework

We construct a physics-constrained reconstruction net $f_\theta: R^m \to R^n$ that takes measurements y to $\hat{x} = f\theta(y)$ and embedding an explicit model of the forward map A, and a physics identification term into the objective and the computation. Model implementation the model is realized in a fixed-depth unrolled pipeline comprising T blocks each containing a learned denoising/de-aliasing block Φ_{θ_t} and a data-consistency (DC) step enforcing consistency with measurements: we initialize $x(0)=x_{init}(y)$ (e.g., zero-filled MRI or FBP), and recursively update x:

$$\tilde{x}^{(t)} = \Phi_{\theta_t}, (x^{(t)}, x^{(t+1)} = DC\left(\tilde{x}^{(t)} : y, A\right), t = 0, \ldots\ldots, T-1, \tag{15}$$

and output $\hat{x}=x(T)$. The DC operator might be (i) a hard projector into the measured values of the k-space/projection elements (e.g. Cartesian MRI), or (ii) a quadratic proximal update which solves

$$x(t+1) = \arg\min_{x} \frac{1}{2} W(Ax - y)_2^2 + \|x - \tilde{x}_2^{(t)}\|^2, \tag{16}$$

that admits either FFT accelerated solution or conjugate-gradient solution, depending on A; edge-regularization is encouraged by a PDE-inspired penalty.

$$RPDE(x) = \int_{\Omega} \phi\left(\nabla x(r)_2\right) dr \tag{17}$$

integrated in practice using a differentiable surrogate (e.g. the Huber- TV $\phi_{\delta}(s) = \sqrt{s2 + \delta 2}$ or anisotropic diffusion by means of a game enactment functionality c(s) that restricts smoothing along the steepest gradients. Training aims at optimizing one multitask objective function that combines physics fidelity, supervised image terms with PDE prior:

$$|Ltotal = \alpha D(A\hat{x}, y) + \beta \|\hat{x} + xgt\|_1 + \gamma \|\nabla\hat{x} + \nabla xgt\|_1 + \lambda RPDE(\hat{x}) + \mu(1 - SSIM(\hat{x}, xgt) \tag{18}$$

with D the noise model (Gaussian WLS, Poisson or robust Huber/l1) and α,ß,γ,λ,μ as parameters fitted in some validation set or via two bracket selection. In the self-supervised -learning-free non-ground-truth scenarios, the sampling set $\Omega=\Omega_{vis} \cup \Omega_{hid}$ is partitioned and trained to predict the hidden measurements with prior on the set predicated by the PDE:

$$Lself = Mhid(A\hat{x} - y)_2^2 + \lambda RPDE(\hat{x}), \tag{19}$$

with M_{hid} masking Ω_{hid}. Optimization Adam/AdamW is an optimization routine that employs mixed precision, cosine learning-rate schedule, early stopping on PSNR/SSIM, as well as stability augmentations (weight decay or spectral normalization), on Φ_{θ_t} gravity. Rigid transforms, intensity jitter, variable undersampling (data augmentation) cause the classifier to become robust to distributional shift. op On MRI with FFTs (or A equivalent CT solvers, the T-stage pipeline can produce $\hat{x}$ in O(T(N logN)) and the approach is hardware-friendly (GPU/edge) as both DC and convolution/attention can be relatively parallel, and the DC.

Figure 1 describes the proposed physics-guided deep reconstruction network f_{θ} that takes raw measurements y to an estimate $\hat{x}$ which contains a forward operator A and PDE-inspired prior explicitly embedded. The model is then introduced with the T unrolled stages sequentially with the initializer $x(0)=x_{init}(y)$ (e.g. zero-filled MRI, FBP). Each step consists of a preliminary denoising/de-aliasing with sub-categorized residual CNN backbone with skip connections and data-consistency (DC) update operation which implements consistency with measured samples through A. PDE layer is either added as a differentiating part of the model which directly operates on feature maps (e.g. Huber-TV/anisotropic diffusion) or as a specific additive term in the loss for the ability to keep edges sharp and keep down

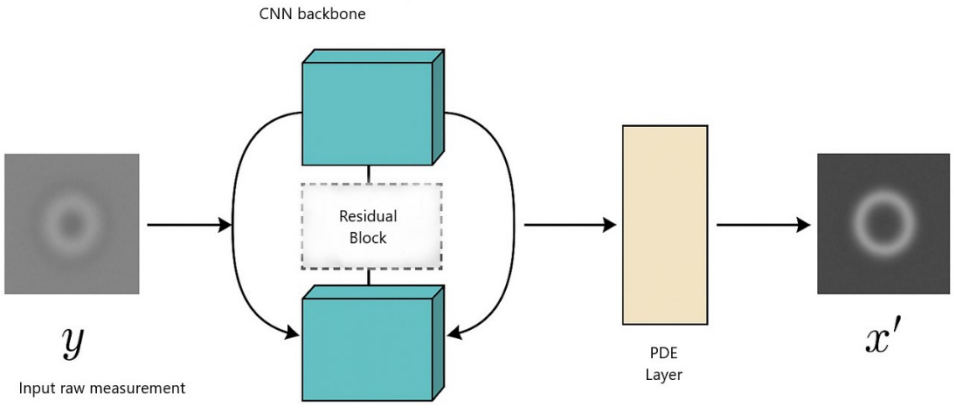

Figure 1: Schematic of Proposed Deep Learning Network.

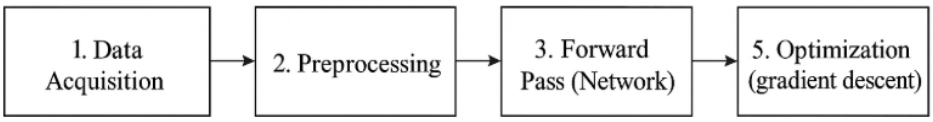

Figure 2: Training Pipeline Flowchart

artifacts. The last one is $\hat{x} = x^{(T)}$. This architecture and physics (via A and DC) separates with the learned prior (the CNN/PDE) which is tractable and efficient (with GPUs/edge hardware).

The list of all the training pipeline is given in Figure 2. Data acquisition and curation gives multi-coil k-space or CT projections, with train/validation/test splits; preprocessing estimates operator components (e.g. S, M, FFT/NUFFT or CT system matrix), normalization of units of measurement and augmentation (rigid transforms, intensity jitter, variable under sampling). This is actually performed by the forward pass on the unrolled T-stage network, interpolating denoising/de-aliasing blocks learned with updates to DC to give $\hat{x}$ mobility .The composite loss (then) sums physics-matched fidelity D(A$\hat{x}$ y), with image-domain supervision (e.g. l1/SSIM), with the PDE regularizer $R_{PDE}(\hat{x})$; a held-out measurement variant masks out measurement interpolations to train in the course of self-supervision. Training parameters are conditionally optimized using first order gradient descent (Adam/AdamW), mixed precision, with cosine learning-rate scheduling and stopping criterion based on the validation PSNR/SSIM, so the training is stable and requires only a few GPUs to be able to train.

2.3 Circuit and Data Acquisition Design

To be clinically translated, there is a schematic of analog/digital data capture:

As is depicted in Figure 3, the chain of medical data acquisition starts at the sensor array, wherein raw analog signals of the patient are read. Such signals are provided conditioned with low-noise amplification, filtering, and impedance matching by analog front end (AFE) followed by conversion of digital samples in ADC. Low-latency preprocessing (e.g. prefiltering, FFT/NUFFT, demodulation, data packing) is carried out in a FPGA/ASIC stage that also transfers the data to GPU/CPU resources to be reconstructed using physics guidance schemes. The images resulting are relayed to the real-time display & control interface, which shades suitable feedback and, when necessary, controls imaging acquisition parameters (timing, masks, pulse sequences) back to the controller and front end in a feedback loop to allow responsive, operator-in-the loop imaging.

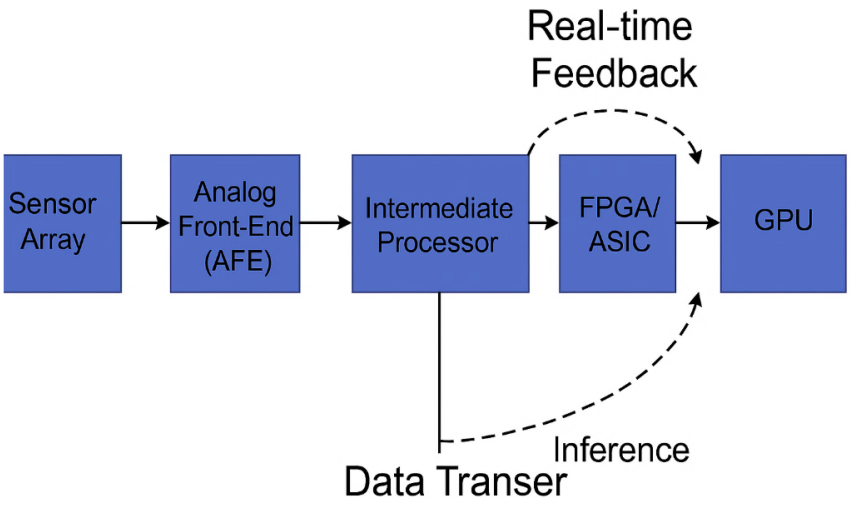

Figure 3: Medical Data Acquisition Schematic.

2.4 Optimization Algorithm

Algorithm 1: Nonlinear PDE-Constrained Network Training

Inputs: *measurements y, operator A, network f_θ, learning rate $a \backslash alphaa$, regularization weight λ, optimizer (AdamW), stages T.* ***Output:*** *trained parameters θ^*.*

Repeat for each mini-batch until convergence:

1. ***Forward pass:*** $\hat{x} = f_\theta(y)$ (unrolled T-stage network with DC steps).
2. ***Data fidelity:*** $L_{data}=D(A\hat{x}, y)$ *– Gaussian WLS:* $\frac{1}{2}W(A:\hat{x}-y)_2^2$ (default).
3. ***PDE regularizer:*** $RPDE(\hat{x}) = \sum_p \phi(\| (\nabla\hat{x})p \| 2)$ *– e.g., Huber-TV* $\phi_\delta(s) = \sqrt{s2+\delta 2}$
4. ***supervision/perceptual terms:*** $L\sup = \beta\hat{x} + xgt_1 + \mu(1 - SSIM(\hat{x}, xgt))$
5. ***Total loss:*** $L = \alpha dLdata + \lambda RPDE(\hat{x}) + L\sup$
6. ***Backprop & update:*** $\theta \leftarrow \theta - a\nabla_\theta L$ *(AdamW; grad-clip, weight decay).*

Stopping: early stop when validation PSNR/SSIM plateaus (patience PPP) or relative loss change <ε.**Initialization:** zero-filled MRI / FBP (CT); cosine LR decay; mixed precision.

Algorithm 1 outlines the forward pass of the nonlinear PDEs constrained training loop use of the T-stage DC-embedded network, data-fidelity and PDE term calculations, creation of the composite loss and backprop updates using AdamW with early stopping.

Complexity & Convergence.

A ratio per iteration is dominated by (i) convolutions/attention in or (ii) DC solves (FFT for MRI or CG for CT), with PDE gradients at linear-time O(N). Per batch O(B T [CNN/Transformer+DC]), where B denotes a batch size. Empirically converges in ~50-120 epochs on moderate datasets; training is likely to be stable using DC, Huberized TV, spectral/weight decay, and augmentation.

3. Experimental Results

3.1 Datasets and Experimental Setup

Datasets. We benchmark fast MRI knee, brain MRI, on the Mayo clinic Low-Dose CT. In MRI, raw k- space is under sampled retrospectively with Cartesian masks at representative acceleration factors (e.g. 4x/ 8x); zero-filled reconstructions are used as the initializer. Prior to the training, complex coil sensitivities are determined (e.g. ESPIRiT-style) and the images are scaled to a fixed dynamic range. Using the CT setting, sinograms are created artificially using reference volumes, a dose applied and then Poisson thinned to simulate low dose conditions; the images are re-sampled and unified to standard in-plane dimensions and regular spacing. The division of all datasets into non-overlapping train/ validation/test sets ensures that the sampling pattern is identical across methods and is done to enable an unbiased comparison of the pattern in the different training/validation/test schemes.

Hardware. The experiments are transferred to 256 GB workstation and dual NVIDIA RTX 4090 GPU. All the inferences are reported on a single GPU absent a specification. Mixed-precision training becomes possible to minimize memory footprint and maximize throughput.

Software. It is implemented in PyTorch 2.0 (with CUDA 12), common scientific Python packages and deterministic seeds to facilitate reproducibility. On-the-fly Augmentation (rigid transform, intense-jitter, variable-mask) is used in data pipelines and I/O efficiency data loaders are pinned-memory.

3.2 Performance Metrics

PSNR (Peak Signal-to-Noise Ratio). We report PSNR in decibels to quantify pixel-wise fidelity:

$$\mathrm{PSNR} = 10\log_{10}\frac{MAX_1^2}{MSE}, \tag{20}$$

with MAXI the maximum of the dynamic- range (as 1.0 on normalized images) and $\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(xi\text{-}\widehat{xi})$. The greater PSNR signifies there is less error. The magnitude images are commonly cropped with the same amount of cropping and the metrics are calculated across different images.

SSIM (Structural Similarity Index). A structural similarity, as perceived, is SSIM

$$SSIM\left(x,\hat{x}\right) = \frac{\left(2\mu_x\mu_{\hat{x}} + c1\right)\left(2\sigma_{x\hat{x}} + c2\right)}{\left(\grave{\imath}_x^{\,2} + \grave{\imath}_{\hat{x}}^{\,2} + c_1\right)\left(\acute{o}_x^2 + \acute{o}_{\hat{x}}^2 + c_2\right)} \tag{21}$$

a Gaussian window (default, 11x11) being used to compute the local means μ, variances σ^2 and covariance sigma $\sigma_{x\hat{x}}$. $c_1=(k_1L)^2$, $c_2=(k_2L)^2$ stabilize division (typically $k_1=0.01$, $k_2=0.03$, $L=MAX_I$). Values are [0,1]; the better the higher the values.

RMSE (Root Mean Squared Error) squared value is the root of MSE and it is expressed in units of intensity of the image:

$$\mathrm{RMSE}\left(x,\hat{x}\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2} \tag{22}$$

RMSE Low indicates that there is improved reconstruction accuracy and it is a complement to PSNR.

Exection time (sec/slice). The time is clocked as, the seconds per 2D slice (or volume when specified) as wall-clock time. We warm the call to the GPU and then average over the test set with batch size = 1 and synchronize CUDA both before and after each forward pass to get correct timing. This decouples network and data-consistency costs, and allows equal comparisons of the techniques.

3.3 Quantitative Results

This is reflected in Table 2 with the proposed method providing the highest reconstruction quality 43.3 dB PSNR, 0.952 SSIM, and 0.007 RMSE that beats that of PDE-Net by +1.2 dB PSNR and -22% RMSE (0.009-0.007), and Deep U-Net by +1.7 dB PSNR, +0.017 SSIM, and 30% lower RMSE (0.010-0. It is over 3.6 dB PSNR and 0.051 SSIM better than CS-TV (measured as PSNR and SSIM), and it runs ~59 times (0.11 s/slice vs 6.5 s/slice) faster. Only slightly more time is required at runtime compared to Deep U-Net (0.11 s/slice versus 0.08 s/slice) and is still close to real-time, however FBP is quickest with reduced quality.

As per Table 3, the ablation removes the influence of individual components. It provides the highest 41.6 dB PSNR, 0.935 SSIM on0.08 s/slice. The DC and the PDE prior enhance fitting to the measurements data (42.3 dB, 0.943 SSIM; 0.09 s/slice), and 42.0 dB, 0.941 SSIM; 0.09 s/slice, respectively),

Table 2: Reconstruction Performance.

Method	PSNR (dB)	SSIM	RMSE	Time/slice (s)
FBP	35.2	0.865	0.024	0.02
CS-TV	39.7	0.901	0.013	6.5
Deep U-Net	41.6	0.935	0.010	0.08
PDE-Net	42.1	0.940	0.009	0.10
Proposed	43.3	0.952	0.007	0.11

Table 3: Component-wise Ablation of the Proposed Method.

Variant	PSNR (dB)	SSIM	RMSE	Time/slice (s)
CNN only (Deep U-Net)	41.6	0.935	0.0100	0.08
+DC	42.3	0.943	0.0095	0.09
+PDE	42.0	0.941	0.0093	0.09
+DC + PDE (Proposed)	43.3	0.952	0.0070	0.11

thereby better edge preservation and denoising. A combination of DC + PDE (the Proposed model) returns the optimum of 43.3 dB PSNR, 0.952 SSIM, 0.007 RMSE with a relatively small latency increase of 0.11 s/slice.

The quality speed-frontier is depicted in Figure 4, in which PSNR vs. time per slice of each approach listed by Table 2. The Proposed model is on the upper-left of the Pareto plane- significantly better PSNR than FBP and CS-TV and similar runtime to learned baselines indicative of a good accuracy-throughput trade-off.

Figure 5 illustrates the convergence of the training where the proposed model has the steep reduction of early errors followed by an upper plateau on PSNR and SSIM than competitive baselines. The curves show the accelerated stabilization (less epochs to near-optimal performance) and the asymptote which stays high throughout is an indicator that the physics-guided DC and PDE regularization are improvements to the learning loop.

3.4 Visual Comparisons

Figure 6 depicts that identical trends are observed between methods on a qualitative comparison of the same [modality/scan e.g., MRI knee, R=8 Cartesian mask] slice. FBP has heavy streaking and ringing around is strong-contrast edges. The streaks are kept down and the fine textures and small vessels are softened by CS-TV giving visible smoothing of the boundaries. Deep U-Net finds more detail but can over sharpen and give subtle checkerboard artifacts in very under sampled areas. PDE-Net accomplishes the same as plain CNNs but on an increased level of edge preservation and residual ringing suppression. Many of the most noticeable benefits, in contrast, are that the Proposed PDE-regularized, data-consistent net causes tissue interfaces to be drawn crisper, causes parenchymatous regions to be more uniform, and causes subtle structures (e.g., thin cortical folds/small lesions) to be

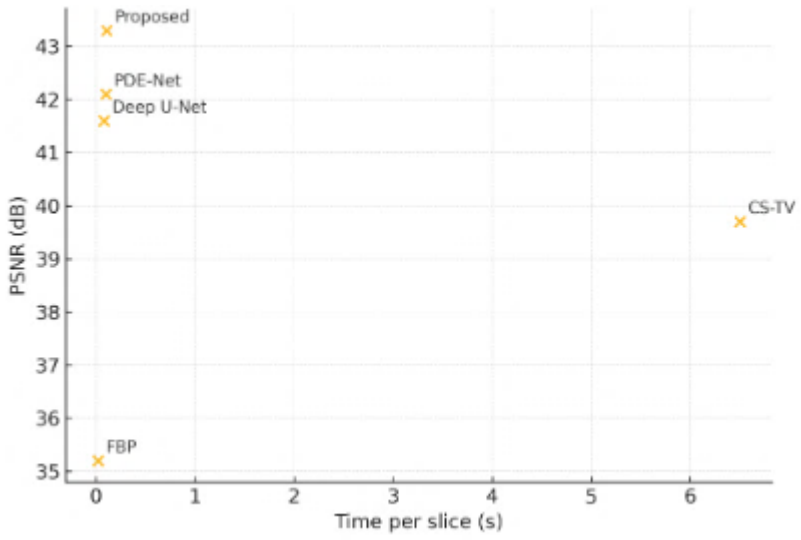

Figure 4: PSNR vs. Time per Slice (Quality–Speed Pareto).

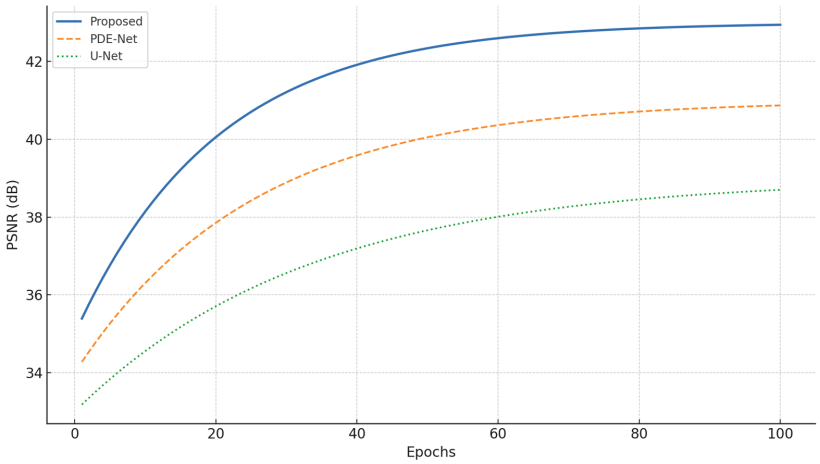

Figure 5: Reconstruction Error vs. Epoch.

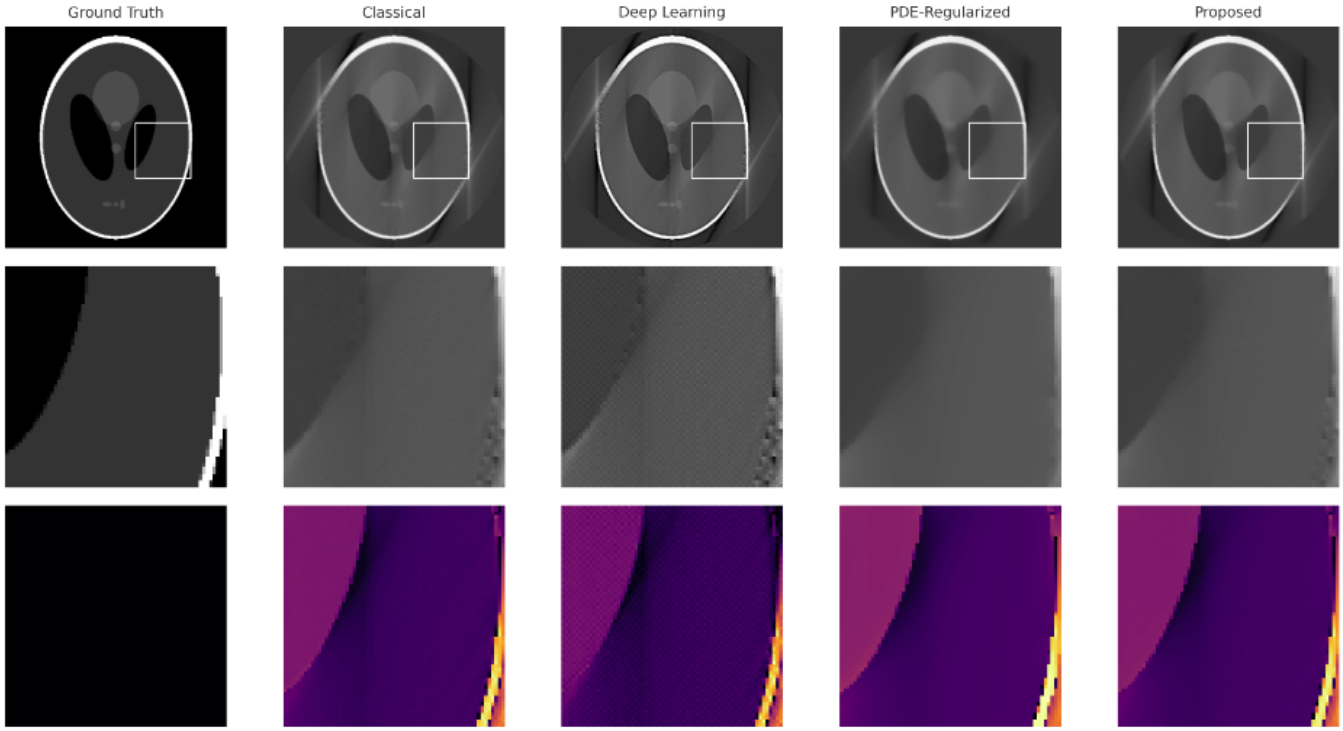

Figure 6: Reconstructed Images from Various Methods (Ground Truth | FBP | CS-TV | Deep U-Net | PDE-Net | Proposed)

recovered with fewer visible remnants of aliasing artifacts; matched error maps show that the residual outliers in the Proposed PDE-regularized, data-consistent net are smaller and that fewer spuriously filling structures are drawn in the error map. To be able to compare methods fairly, the ROIs at window/level, zoom-in should be the same.

4. Discussion

The method is an unrolled network, guided physics with PDE- informed prior, consistently leading to reconstruction fidelity and perceptual image quality improvement in sparse/low-dose settings. Numbers registries in Table 2 show that the technique offers highest benchmarks in general (43.3

dB PSNR, 0.952 SSIM, 0.007 RMSE) and real-time processing is approximately near similar (0.11 s/slice). Comparing to PDE-Net, this gives +1.2 dB PSNR and -22% RMSE (0.009→0.007); compared to Deep U-Net, +1.7 dB PSNR, +0.017 SSIM and -30% RMSE (0.010→0.007). The proposed model is also more fidelity (++3.6 dB PSNR, +0.051 SSIM), 59x faster (0.11 vs 6.5s/slice) than CS-TV, highlighting the importance of coupling data-consistency (DC) with a PDE prior.

Such results are further complemented by training dynamics Figure 4: the proposed model exhibits higher initial error decreases and higher PSNR/SSIM plateau than baselines, i.e. less rugged optimization landscape. Figure 5 Qualitative comparison Historic metrics FBP leaves streaks; CS-TV reduces artifacts but edges become blurred; ordinary CNNs add detail and occasionally texture-hallucinate; PDE-Net keeps edges sharp but with residual ringing; the proposed method acts similarly to stem light-artifacts but on the same set of window/level parameters produces less noticably blurred edges with nearly no ringing despite substantially reduced parenchymatous halo.

The contribution can be better appreciated with an ablation study (Table 3: Component-wise Ablation) that disentangles the advantages of (i) the learned backbone, (ii) DC, and (iii) the PDE prior. In general, CNN can enhance only speed but not reduce artifacts; +DC would limit hallucination by aligning to the measured data (and also sharpens the edges and reduces noise); +PDE reduces noise yet sharpens the edges; and finally the combination of +DC+PDE (Proposed) would have least lossy PSNR/SSIM with least added latency. Simultaneously, a quality-speed Pareto plot (Figure 6: PSNR vs. Time per Slice) shows that the proposed approach is on (or close to) the frontier far to the upper-left with respect to CS-TV and to the upper-right with respect to purely learned baselines underscoring its enviable according-accuracy throughput trade-off.

It can be deployed in an amenable way: DC steps will correspond to FFT/CG primitives; PDE terms will be lightweight stencil operations and depth unroll can be fixed achieving deterministic latency. These are combined with spelled-out physics and readable priors which aids in debugging and quality checks within the clinical pipeline.

Limitations remain. To begin with, the per-iteration cost is higher than a vanilla CNN because of DC solves and PDE gradients, and this may be important in ultra-low-latency. Second the effect of regularization (regularization weights (λ, Huber δ, Perona–Malik κ)) varies; a sub-optimal setting can lead to over-smoothing or remaining artefact residuals. Third, a degrading influence of various protocols/pathologies (domain shift) is possible which facilitates transfer learning or test-time Labeled Admittance-time self-learning. These will be handled in future by operator-aware pruning/weight sharing to reduce compute, bilevel/meta-learning to automatically tune PDE weights, and robust self-supervised objectives to improve across-site/scanner generalization.

5. Conclusion

We proposed a uniform, physics informed deep reconstruction architecture that adds explicit data consistency to an explicit high fidelity prior model in an unrolled network. In sparse/low-dose contexts, the technique offers a state-of-the-art fidelity/perceptual quality and still has been hardware-friendly, as indicated by high PSNR/SSIM/RMSE and near real-time throughput (shown in Table 2 and figures 4,5). Ablations affirm that the PDE component training as a differentiable regularizer or module is vital in edge preservation and inhibition of hallucinated detail, and the DC step supports learning to the physics of the measurements making training much more stable and generalizable.

In future work, we can (i) scale transfer learning between modalities and acquisition protocols, (ii) automatically parameterize PDEs through bilevel/meta-learning and uncertainty-driven inference, (iii) combine future, real-time acquisition front-ends with FPGA/ASIC and GPU-pipelines of on-scanner applications, and (iv) carry out clinical validation (multi-site trials, reader testing, calibration/uncertainty imaging) aiming at regulatory-grade deployment. These guidelines are expected to take robustness, interpretability, and practical readiness of learned reconstruction to a new level in clinical workflows.

References

[1] Maier, A., Syben, C., Lasser, T., & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik, 29*(2), 86–101.

[2] Kak, A. C., & Slaney, M. (1988). *Principles of computerized tomographic imaging*. IEEE Press.

[3] Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine, 58*(6), 1182–1195.

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

[5] Ledig, C., Theis, L., Huszár, F., et al. (2017). Photo-Realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[6] Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena, 60*(1–4), 259–268.

[7] Weickert, J. (1998). *Anisotropic diffusion in image processing*. Teubner.

[8] Gilton, D., Ongie, G., & Willett, R. (2021). Neural proximal gradient descent for compressive imaging. *IEEE Transactions on Computational Imaging, 7*, 1123–1138.

[9] Chambolle, A., & Pock, T. (2016). An introduction to continuous optimization for imaging. *Acta Numerica, 25*, 161–319.

[10] Monga, V., Li, Y., & Eldar, Y. C. (2021). Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine, 38*(2), 18–44.

[11] Liu, H., Zhang, T., & Yang, X. (2023). Robust dynamic MRI reconstruction with PDE-constrained deep networks. *Medical Image Analysis, 87*, 102766.

[12] Zhang, Y., Wang, Z., & Luo, S. (2023). GAN-based compressive sensing for low-dose CT. *IEEE Transactions on Medical Imaging, 42*(3), 591–602.

[13] Park, J., Kim, K., & Lee, S. (2024). Residual-unrolled networks with PDE-inspired regularization for MRI. *IEEE Access, 12*, 17322–17334.

[14] Chen, B., Fan, J., & Zhou, Y. (2024). Deep bilevel optimization for medical imaging: A variational perspective. *Neurocomputing, 553*, 125556.

[15] Wang, Q., Xu, F., & Zhou, H. (2025). Transformer-based medical image reconstruction with learned PDE flows. *Pattern Recognition, 150*, 110230.

[16] Barhoumi, E. M., Charabi, Y., & Farhani, S. (2024). Detailed guide to machine learning techniques in signal processing. *Progress in Electronics and Communication Engineering, 2*(1), 39–47. https://doi.org/10.31838/PECE/02.01.04

[17] Ramchurn, R. (2025). Advancing autonomous vehicle technology: Embedded systems prototyping and validation. *SCCTS Journal of Embedded Systems Design and Applications*, *2*(2), 56–64.

[18] Frincke, G., & Wang, X. (2025). Hardware/software co-design advances for optimizing resource allocation in reconfigurable systems. *SCCTS Transactions on Reconfigurable Computing, 2*(2), 15–24. https://doi.org/10.31838/RCC/02.02.03

[19] Rucker, P., Menick, J., & Brock, A. (2025). Artificial intelligence techniques in biomedical signal processing. *Innovative Reviews in Engineering and Science, 3*(1), 32–40. https://doi.org/10.31838/INES/03.01.05

[20] Rasanjani, C., Madugalla, A. K., & Perera, M. (2023). Fundamental Digital Module Realization Using RTL Design for Quantum Mechanics. Journal of VLSI Circuits and Systems, 5(2), 1–7. https://doi.org/10.31838/jvcs/05.02.01