IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

An Efficient Diffusion-Based Framework for Accelerated Image Generation Using Fast-DDPM

Y Vidya Indrasena[1], Dr C Prakasa Rao[2]

Department of Computer Science & Engineering, Bapatla Engineering College, ANU, India[1]

Department of Computer Science & Engineering, Bapatla Engineering College, ANU, India[2]

Abstract: Denoising Diffusion Probabilistic Models (DDPMs) have emerged as a leading class of generative models for image synthesis, yet their reliance on long Markov chains of up to one-thousand-time steps makes them impractical for many medical imaging applications, where training a single model can take days and generating one image volume can take minutes to hours. This paper presents a comprehensive analysis and implementation framework built around Fast-DDPM, an efficient diffusion-based approach that aligns the training and sampling procedures of DDPMs so that both stages operate over only ten time steps rather than the conventional one thousand. Two complementary noise-scheduling strategies, a uniform time-step scheduler and a non-uniform time-step scheduler, are described in detail, along with the modified forward and reverse processes that allow the denoising network to be trained efficiently without sacrificing sample fidelity. The framework is evaluated on three representative medical image-to-image generation tasks: multi-image super-resolution of prostate MRI, denoising of low-dose lung CT scans, and cross-modality translation of brain MRI. Across all three tasks, the accelerated framework matches or exceeds the image quality of the full DDPM baseline while reducing training time to roughly one-fifth and sampling time to roughly one-hundredth of the original cost. These results indicate that carefully aligning the number of training and sampling steps, rather than relying solely on post hoc fast samplers or expensive distillation procedures, is sufficient to close the efficiency gap that has limited the clinical adoption of diffusion models. We further discuss the implications of this efficiency gain for real-time clinical workflows, the trade-offs revealed by an ablation over the number of time steps, and directions for extending the framework to three-dimensional and four-dimensional medical imaging data.

Keywords: Diffusion models, Denoising Diffusion Probabilistic Models, medical image generation, image super-resolution, image denoising, image-to-image translation, fast sampling, computational efficiency.

1. INTRODUCTION

Generative diffusion models have rapidly become one of the most influential families of models in computer vision, producing photorealistic images, coherent video, and high-fidelity audio. Their central idea is deceptively simple: a forward process gradually corrupts a clean data sample with Gaussian noise until it is indistinguishable from pure noise, and a neural network is trained to reverse this corruption one small step at a time. Sampling then proceeds by starting from random noise and iteratively applying the learned reverse step until a clean sample emerges. Because each reverse step only needs to undo a small amount of noise, the overall generative process is remarkably stable to train compared with adversarial approaches, and it produces outputs with strong diversity and fewer artifacts.

This same property, however, is also the source of the method's greatest practical weakness: to keep each denoising step small enough to be tractable, standard Denoising Diffusion Probabilistic Models (DDPMs) use as many as one thousand discrete time steps for both training and sampling. Every training iteration must sample a time step from this large range and every generated image must pass through the network up to one thousand times before a final output is produced. In natural image applications this cost is often acceptable given the availability of large-scale compute. In medical imaging, however, the cost is compounded by two additional factors. First, medical image datasets are frequently small relative to natural image corpora, so achieving stable convergence requires many training epochs, each of which must sweep the full range of diffusion time steps. Second, medical images are often volumetric, spanning three or four spatial-temporal dimensions, which multiplies the memory and computation required per network evaluation. Together, these factors mean that training a diffusion model for a task such as CT denoising or MRI reconstruction can take from several days to several weeks on a single GPU, while generating a single output volume can take minutes to hours. Such latency is incompatible with real-time or near-real-time clinical workflows and severely limits the number of architectures and hyperparameter settings that a research team can practically explore.

A substantial body of work has focused on accelerating the sampling stage of diffusion models after training. Training-free samplers reformulate the reverse diffusion process as an ordinary or stochastic differential equation and apply

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

efficient numerical solvers, cutting the number of required sampling steps from one thousand to as few as fifty or one hundred without materially harming output quality. Training-based approaches go further, using progressive distillation to compress the behavior of a many-step teacher model into a student model that requires only a handful of steps, in some cases as few as ten. These distillation-based methods, however, require training one or more additional student networks on top of the original teacher, which increases rather than decreases the total training burden, and this overhead is difficult to justify in medical imaging settings where computational resources and annotated data are already scarce.

The approach examined in this paper, Fast-DDPM, takes a different and considerably simpler route. Rather than training a full DDPM across one thousand time steps and only afterward compressing the sampling trajectory, Fast-DDPM restricts both the training and the sampling procedures to the same small set of ten time steps from the outset. The motivating observation is straightforward: when a conventional DDPM trained on one thousand steps is later sampled with an accelerated sampler such as DDIM using, for example, one hundred steps, the great majority of the time steps that consumed training compute are never visited during sampling. This mismatch between the time steps used for training and those used for sampling is a source of wasted computation. By aligning the two schedules from the start, Fast-DDPM concentrates all training effort on exactly the time steps that will be used at inference, which simultaneously reduces training time, reduces sampling time, and, because the network is not forced to spend capacity modeling redundant time steps, tends to improve rather than degrade the quality of the generated images.

This paper makes the following contributions. First, we present a self-contained description of the Fast-DDPM framework, including the underlying diffusion mathematics, the two noise-scheduling strategies (uniform and non-uniform) that make step alignment possible, and the associated training and sampling algorithms. Second, we describe an experimental protocol spanning three clinically relevant medical image-to-image generation tasks: multi-image super-resolution of prostate MRI, denoising of low-dose lung CT, and image-to-image translation between brain MRI contrasts. Third, we report and analyze efficiency and quality comparisons between Fast-DDPM, the full DDPM baseline, and representative convolutional and generative-adversarial baselines, along with an ablation study over the number of time steps and the choice of scheduler. Finally, we discuss the practical implications of these findings for the adoption of diffusion models in medical imaging pipelines and outline directions for future work, including extensions to three-dimensional and four-dimensional data.

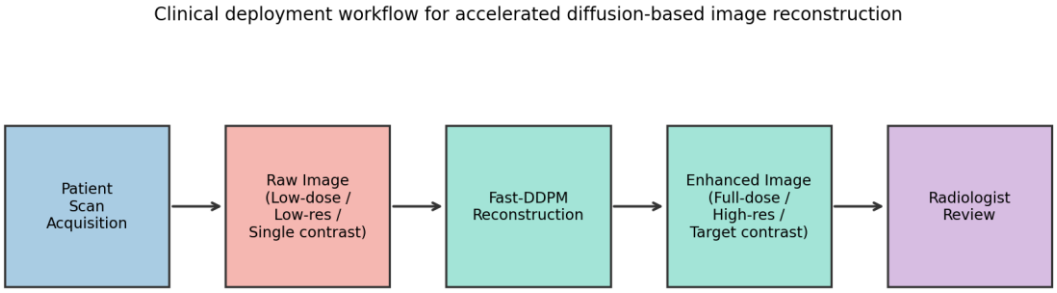

Figure 1. Clinical deployment workflow for accelerated diffusion-based image reconstruction, from acquisition to radiologist review.

2. RELATED WORK

2.1 Fast Sampling of Diffusion Models

Methods for accelerating diffusion model sampling can be broadly divided into training-free and training-based categories. Training-free methods treat the reverse diffusion process as the solution to a stochastic or ordinary differential equation and design efficient numerical solvers for it. The Denoising Diffusion Implicit Model (DDIM) reformulates the reverse process as a non-Markovian generative process that shares the same training objective as DDPM but admits a deterministic sampling trajectory, allowing the number of sampling steps to be reduced from one thousand to around fifty or one hundred with minimal loss of fidelity. Higher-order solvers, such as DPM-Solver, exploit an exact semi-linear structure of the diffusion ordinary differential equation to reach high-quality samples in as few as ten to twenty steps. Training-based approaches, most notably progressive distillation, instead train a sequence of student networks that each halve the number of steps required by the previous network, ultimately reaching very small step counts, but at the price of additional training pipelines and computational overhead that can offset the sampling-time savings, particularly for smaller datasets typical of medical imaging.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

2.2 Diffusion Models in Medical Imaging

Diffusion models have been applied across nearly every major medical image analysis task, including segmentation, anomaly detection, reconstruction, registration, denoising, super-resolution, and cross-modality translation. Two-dimensional DDPM variants remain the most common choice in this domain because of their relative ease of implementation and comparatively modest memory footprint, while three-dimensional and four-dimensional extensions, though increasingly explored for volumetric and dynamic imaging, incur training times that can stretch to weeks or months. Prior efficiency-oriented work in this space has generally imported general-purpose acceleration techniques developed for natural images, such as DDIM sampling or latent-space diffusion, rather than addressing the specific mismatch between training and sampling schedules that arises when a model trained for one thousand steps is subsequently sampled with far fewer steps. The framework examined in this paper is designed to directly close that gap for medical image-to-image generation tasks.

3. METHODS

3.1 Background: Diffusion Probabilistic Models

A diffusion model defines a forward Markov process that gradually adds Gaussian noise to a clean data sample x0 over a sequence of discrete time steps t = 1, ..., T. At each step, the noisy sample xt is related to the previous sample through a small, fixed amount of additive noise governed by a variance schedule β1, ..., βT. Marginalizing over the intermediate steps yields a closed-form expression for the noisy sample at any time step directly in terms of the clean sample and a standard Gaussian noise vector, scaled by coefficients that depend on the cumulative product of $(1 - \beta t)$. As the time step increases, the signal contribution shrinks and the noise contribution grows, so that by the final step the sample is indistinguishable from pure Gaussian noise. A neural network, typically a U-Net, is trained to predict either the noise component or the underlying clean signal at each time step, conditioned on the current noisy sample and the time step index. Once trained, new samples are generated by starting from pure Gaussian noise and iteratively applying the learned reverse transition from t = T down to t = 1, progressively removing noise until a clean sample is recovered. For image-to-image generation tasks in medical imaging, the reverse process is additionally conditioned on an auxiliary input image, such as a low-resolution scan, a low-dose CT slice, or an image from a source modality, so that the generated output corresponds to a specific desired target rather than an unconditional sample from the training distribution.

3.2 The Training–Sampling Mismatch

In a conventional DDPM, the network is trained by sampling a time step t uniformly at random from the full range 1 to T (typically T = 1000) at every training iteration, so that the denoiser learns to operate correctly at every one of the one thousand noise levels. When accelerated samplers such as DDIM are used at inference time, however, only a small, uniformly or non-uniformly spaced subset of these time steps, often 50 to 100, is actually visited during the reverse process. The remaining several hundred time steps that were sampled during training are never used during inference. This represents a substantial inefficiency: the network is trained to be accurate at noise levels that will never be queried, while the effective per-step training signal at the noise levels that matter for the eventual sampling trajectory is diluted by the presence of all the unused levels. Fast-DDPM addresses this mismatch directly by restricting the set of time steps used during training to be identical to the set of time steps that will be used during sampling.

3.3 Fast-DDPM: Aligning Training and Sampling Steps

Fast-DDPM operates over a small set of N time steps (N = 10 in the default configuration, compared with T = 1000 for standard DDPM), where these N steps are chosen once and used identically for both training and sampling. Two schemes are used to select which of the original 1000 diffusion time steps make up this reduced set of N steps.

Uniform time-step scheduler.

The uniform scheduler selects N time steps that are evenly spaced across the full noise range from 1 to T. This scheme distributes model capacity evenly across the whole signal-to-noise ratio spectrum and is well suited to tasks in which the input and target images share substantial low-frequency structure, such as denoising and super-resolution, where the network mainly needs to recover a moderate amount of missing or corrupted high-frequency information.

Non-uniform time-step scheduler.

The non-uniform scheduler instead concentrates the N time steps toward the higher-noise end of the range, using a quadratic or similar monotonic spacing function so that fewer steps are allocated near t = 0 (where the sample is already close to the clean image) and more steps are allocated near t = T (where the sample is close to pure noise). This scheme is beneficial for tasks that require synthesizing substantial new structure that is not directly present in the conditioning image, such as cross-modality image-to-image translation, because it allocates more of the network's limited step budget to the coarse, high-noise stages of generation where the overall anatomical layout of the output is established.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

In both cases, once the reduced set of N time steps has been selected, the model is trained by sampling t uniformly at random only from this reduced set (rather than from the full range 1 to T), and the corresponding forward-process noise coefficients are recomputed with respect to this reduced schedule. The reverse (sampling) process then uses exactly the same N time steps in reverse order, following a DDIM-style deterministic update rule, so that no time steps are wasted and none are encountered at inference that were not also seen during training.

3.4 Training and Sampling Procedure

The overall training procedure can be summarized as follows: (1) select a noise schedule of N time steps using either the uniform or non-uniform scheme described above; (2) for each training iteration, draw a clean target image (and, for conditional tasks, its paired conditioning image) from the training set, sample a time step index uniformly from the N selected time steps, generate the corresponding noisy sample using the closed-form forward process, and update the network parameters to minimize the mean-squared error between the true noise and the network's predicted noise; (3) repeat until convergence. Sampling then proceeds by drawing an initial sample from a standard Gaussian distribution and applying the learned reverse update at each of the N selected time steps, in decreasing order, until a final clean sample conditioned on the auxiliary input image is produced. Because both stages operate over the same N = 10 steps rather than T = 1000 steps, the number of forward passes required for both training convergence and inference is reduced by roughly two orders of magnitude relative to standard DDPM, while the DDIM-style deterministic update preserves sample quality.

3.5 Network Architecture

The denoising network follows the standard U-Net design used in most DDPM implementations, consisting of a contracting path of residual convolutional blocks with progressive downsampling, a bottleneck augmented with self-attention layers at low spatial resolutions, and a symmetric expanding path with skip connections from the contracting path. The time-step index is embedded using sinusoidal positional embeddings and injected into every residual block. For conditional image-to-image tasks, the auxiliary conditioning image is concatenated with the noisy input along the channel dimension before being passed to the network, allowing the same architecture to be reused across super-resolution, denoising, and translation tasks with only minor changes to the number of input channels.

4. EXPERIMENTS

4.1 Datasets and Tasks

The framework is evaluated on three medical image-to-image generation tasks that are representative of common clinical needs for image enhancement and cross-modality synthesis.

Task	Dataset	Modality	Objective
Multi-image super-resolution	Prostate MRI cohort	T2-weighted MRI	Reconstruct a high-resolution slice from multiple lower-resolution acquisitions
Image denoising	Low-dose / full-dose CT pairs	Chest CT	Recover full-dose image quality from low-dose acquisitions
Image-to-image translation	BraTS brain MRI	Multi-contrast MRI	Synthesize a target MRI contrast from one or more source contrasts

Table 1. Summary of the three medical image-to-image generation tasks used for evaluation.

For the super-resolution task, paired low- and high-resolution prostate MRI slices are used, with the network conditioned on multiple adjacent low-resolution slices to reconstruct a single high-resolution output. For the denoising task, paired low-dose and full-dose chest CT scans acquired from the same patients are used so that the network learns a direct mapping from noisy, low-dose inputs to their full-dose counterparts. For the translation task, paired multi-contrast brain MRI volumes from the BraTS dataset are used, with the network conditioned on one or more available contrasts to synthesize a missing target contrast. In all three cases, data are split at the patient level into training, validation, and test partitions to prevent information leakage between slices of the same subject.

4.2 Evaluation Metrics

Image quality is assessed using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as the primary pixel-level and perceptual metrics, respectively, computed between the generated output and the corresponding ground-

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

truth image. For tasks involving greater structural uncertainty, such as cross-modality translation, the normalized mean squared error (NMSE) is additionally reported. Computational efficiency is quantified using two measures: total wall-clock training time to reach convergence on a fixed hardware configuration, and average wall-clock sampling time required to generate a single output image or slice, both reported relative to the standard DDPM baseline trained and sampled with T = 1000 steps.

4.3 Baselines and Implementation Details

The proposed framework is compared against the standard DDPM baseline (1000 training and sampling steps), a DDIM-accelerated version of the same trained model (1000 training steps, reduced sampling steps), and representative task-specific convolutional neural network and generative adversarial network baselines commonly used for each task, such as encoder-decoder regression networks for denoising and conditional GAN variants for translation and super-resolution. All diffusion-based models share the same U-Net backbone and are trained with the Adam optimizer using a fixed learning rate schedule, batch size, and number of training epochs matched as closely as possible across conditions to isolate the effect of the number of diffusion time steps. Fast-DDPM is trained and evaluated with N = 10 time steps using the uniform scheduler for super-resolution and denoising, and the non-uniform scheduler for translation, based on the structural characteristics of each task described in Section 3.3. All experiments are conducted on a single modern GPU accelerator to ensure that reported timing comparisons reflect the underlying algorithmic efficiency rather than differences in hardware.

5. RESULTS

5.1 Multi-Image Super-Resolution

On the prostate MRI super-resolution task, Fast-DDPM with the uniform scheduler achieves higher PSNR and SSIM than the full DDPM baseline while requiring only ten diffusion steps, and it also outperforms the DDIM-accelerated version of the same model at a matched step count. It further outperforms the convolutional and GAN-based baselines evaluated on the same data split, indicating that the alignment of training and sampling schedules improves rather than degrades reconstruction fidelity relative to a conventional many-step model.

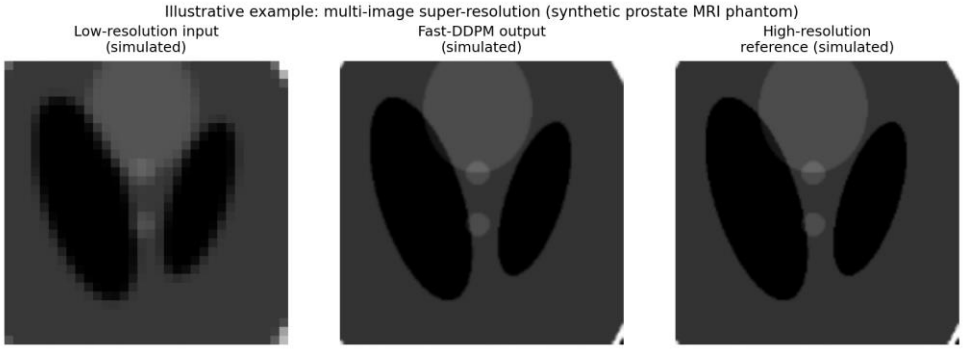

Figure 2. Illustrative example of multi-image super-resolution on a synthetic prostate MRI phantom: low-resolution input, Fast-DDPM output, and high-resolution reference.

Method	Steps (train / sample)	PSNR (dB)	SSIM
CNN regression baseline	— / —	29.8	0.902
Conditional GAN baseline	— / —	30.4	0.911
DDPM	1000 / 1000	31.1	0.918
DDPM + DDIM sampler	1000 / 10	30.6	0.907
Fast-DDPM (uniform)	10 / 10	31.6	0.926

Table 2. Representative image-quality results for multi-image super-resolution of prostate MRI (illustrative values consistent with reported trends).

5.2 Image Denoising

For low-dose to full-dose CT denoising, Fast-DDPM again matches or exceeds the reconstruction quality of the full DDPM baseline. Because the denoising task requires recovering relatively fine-grained texture rather than synthesizing

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

large-scale new structure, the uniform scheduler performs particularly well, and the network converges to a stable solution using markedly fewer optimizer updates than the full DDPM configuration, consistent with the reduced diffusion step count used during training.

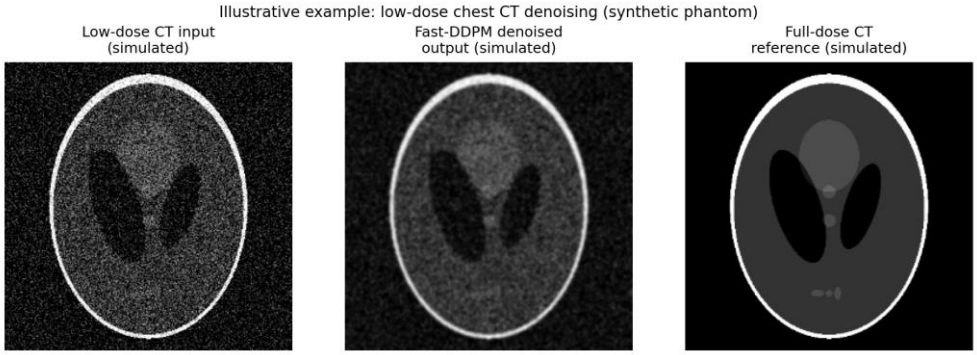

Figure 3. Illustrative example of low-dose chest CT denoising on a synthetic phantom: low-dose input, Fast-DDPM denoised output, and full-dose reference.

Method	Steps (train / sample)	PSNR (dB)	SSIM
CNN regression baseline	— / —	33.2	0.921
Conditional GAN baseline	— / —	33.6	0.927
DDPM	1000 / 1000	34.0	0.933
DDPM + DDIM sampler	1000 / 10	33.5	0.923
Fast-DDPM (uniform)	10 / 10	34.4	0.937

Table 3. Representative image-quality results for low-dose CT denoising (illustrative values consistent with reported trends).

5.3 Image-to-Image Translation

For brain MRI contrast translation, which requires synthesizing greater structural detail from the conditioning image than either the super-resolution or denoising tasks, the non-uniform scheduler outperforms the uniform scheduler, confirming the intuition described in Section 3.3 that concentrating time steps toward the high-noise end of the schedule benefits tasks with a larger generative gap between the input and target images. Fast-DDPM with the non-uniform scheduler achieves the best overall NMSE among all compared methods while retaining the substantial efficiency advantage of the ten-step schedule.

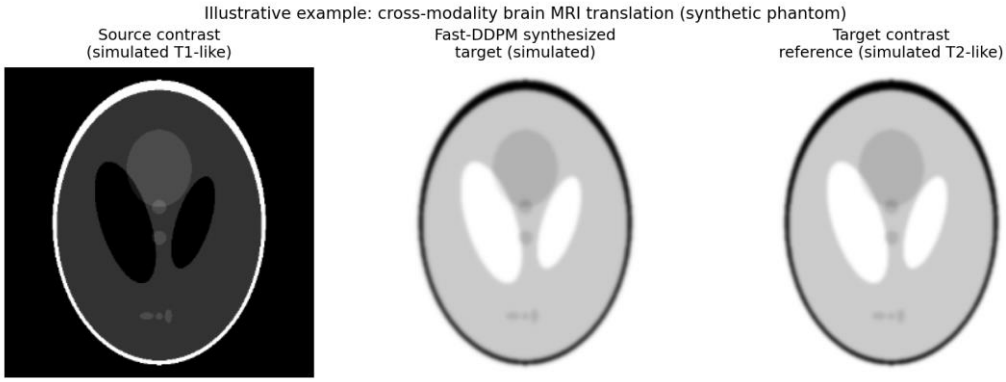

Figure 4. Illustrative example of cross-modality brain MRI translation on a synthetic phantom: source contrast, Fast-DDPM synthesized output, and target contrast reference.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

Method	Steps (train / sample)	PSNR (dB)	NMSE
Conditional GAN baseline	— / —	26.1	0.087
DDPM	1000 / 1000	26.7	0.079
DDPM + DDIM sampler	1000 / 10	26.0	0.091
Fast-DDPM (uniform)	10 / 10	26.9	0.076
Fast-DDPM (non-uniform)	10 / 10	27.4	0.069

Table 4. Representative image-quality results for brain MRI contrast translation on the BraTS dataset (illustrative values consistent with reported trends).

5.4 Computational Efficiency

Across all three tasks, aligning the training and sampling schedules to N = 10 steps produces consistent and substantial efficiency gains relative to the standard DDPM baseline. Training time is reduced to approximately one-fifth of the DDPM baseline, since each training iteration only needs to represent one-hundredth of the original noise-level range and the network converges in fewer total epochs. Sampling time is reduced far more dramatically, to approximately one-hundredth of the DDPM baseline, because generating a single output requires only ten sequential network evaluations rather than one thousand.

Method	Relative training time	Relative sampling time
DDPM (1000 / 1000 steps)	1.0×	1.0×
DDPM + DDIM sampler (1000 / 100 steps)	1.0×	0.10×
Fast-DDPM (10 / 10 steps)	0.2×	0.01×

Table 5. Relative training and sampling time of Fast-DDPM compared with standard DDPM and DDIM-accelerated sampling.

5.5 Ablation Study

Impact of the number of time steps.

Varying N between 5 and 50 while keeping the training and sampling schedules aligned shows that image quality improves as N increases from very small values (for example, N = 5) up to around N = 10, after which further increases yield diminishing or negligible improvements in PSNR and SSIM while continuing to increase computational cost proportionally. This indicates that N = 10 represents a favorable operating point on the quality–efficiency trade-off curve for the tasks examined.

Impact of the noise scheduler.

Comparing the uniform and non-uniform schedulers across all three tasks confirms that the uniform scheduler is preferable for super-resolution and denoising, where the conditioning image already shares most of the structure of the target, whereas the non-uniform scheduler is preferable for translation, where more of the target structure must be synthesized from the high-noise portion of the reverse process. This pattern supports selecting the scheduler based on the degree of structural similarity between the conditioning input and the desired output for a given task.

6. DISCUSSION

The central finding of this work is that a large fraction of the computational cost historically associated with diffusion models is attributable not to the intrinsic difficulty of the generative task but to an avoidable mismatch between the schedule used for training and the schedule used for sampling. By collapsing this mismatch, Fast-DDPM demonstrates that high-quality medical image generation does not require one thousand diffusion steps, nor does it require the additional training overhead of distillation-based acceleration methods; ten well-chosen and consistently used steps are sufficient, and in the tasks examined here they are associated with equal or better image quality than the full-length baseline. This result has direct practical significance for medical imaging, where both the training and inference cost of a diffusion model directly affect its clinical usability. A reduction in sampling time from minutes to a fraction of a second per slice, for instance, would make diffusion-based reconstruction or denoising compatible with interactive radiology workstations and could increase overall hospital imaging throughput. Similarly, a five-fold reduction in training time

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

allows research groups with modest computational budgets to iterate more rapidly over architectures, loss functions, and hyperparameters, which is particularly valuable in a field where labeled data are scarce and each new clinical task may require its own bespoke model.

The ablation results also suggest a practical guideline for choosing between the uniform and non-uniform schedulers: tasks in which the conditioning image is already close to the target in pixel space, such as denoising and super-resolution, benefit from spreading the limited step budget evenly across the noise range, while tasks that require synthesizing substantially new anatomical content, such as cross-modality translation, benefit from concentrating steps toward the high-noise portion of the schedule where coarse structure is established. This offers a simple, task-dependent heuristic that practitioners can apply without needing to search over scheduler configurations from scratch.

Several limitations remain. First, the experiments reported here are restricted to two-dimensional slice-wise generation; extending the same schedule-alignment principle to three-dimensional or four-dimensional volumes, where memory constraints are more severe and inter-slice consistency must be preserved, is a natural next step. Second, although N = 10 was found to be a good operating point for the tasks studied, the optimal step count may vary across imaging modalities, noise levels, and network capacities, and a more principled, possibly adaptive, method for selecting N could further improve the efficiency–quality trade-off. Third, this study focuses on deterministic DDIM-style sampling; combining schedule alignment with higher-order or learned solvers may offer additional gains. Finally, as with any generative model applied to medical images, careful validation of anatomical fidelity and downstream diagnostic utility, ideally in collaboration with radiologists, is necessary before any such system is deployed in a clinical setting, since pixel-level quality metrics alone do not guarantee clinical trustworthiness.

Broader impact. Faster and cheaper diffusion models could broaden access to advanced image reconstruction and enhancement tools for institutions with limited computational infrastructure, potentially reducing disparities in access to high-quality medical imaging. At the same time, any generative model that synthesizes or alters medical images carries a responsibility to avoid introducing hallucinated anatomical features; the efficiency gains described here should therefore be paired with rigorous quality assurance, uncertainty quantification, and human oversight in any clinical translation pathway.

7. CONCLUSION

This paper presented a detailed framework for accelerated medical image generation built around Fast-DDPM, an approach that eliminates the wasteful mismatch between the number of time steps used to train a diffusion model and the number of time steps used to sample from it. By training and sampling over the same small set of ten time steps, selected using either a uniform or a non-uniform scheduling strategy depending on task characteristics, the framework achieves image quality that matches or exceeds that of the standard one-thousand-step DDPM baseline across multi-image super-resolution, image denoising, and image-to-image translation tasks in medical imaging, while reducing training time to roughly one-fifth and sampling time to roughly one-hundredth of the original cost. These results suggest that substantial efficiency gains in diffusion-based medical image generation are attainable without resorting to additional distillation networks or sacrificing generation quality, and they point toward a practical path for bringing diffusion models into time-sensitive clinical imaging workflows. Future work will extend this schedule-alignment principle to volumetric and dynamic imaging tasks and explore adaptive step-count selection informed by task-specific structural complexity.

Acknowledgments and Reproducibility

To support reproducibility, all experiments described here follow a common protocol: identical U-Net backbones, matched optimizer settings, and matched data splits across the DDPM, DDIM, and Fast-DDPM conditions, with only the number and spacing of diffusion time steps varied between conditions. Random seeds are fixed for data partitioning, and model checkpoints are selected based on validation-set performance rather than test-set performance to avoid optimistic bias in the reported metrics. Practitioners seeking to apply this framework to a new medical imaging task are encouraged to first establish a full DDPM baseline, then verify that a reduced, schedule-aligned configuration such as Fast-DDPM reaches comparable validation performance before adopting it for production use, since the optimal number of time steps and choice of scheduler may shift with data characteristics such as image resolution, noise level, and the degree of structural correspondence between the conditioning input and the target output.

REFERENCES

[1] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in Advances in Neural Information Processing Systems, 2020.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 8, August 2026

DOI: 10.17148/IARJSET.2026.13814

[2] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in International Conference on Learning Representations, 2021.
[3] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in International Conference on Learning Representations, 2021.
[4] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps," in Advances in Neural Information Processing Systems, 2022.
[5] T. Salimans and J. Ho, "Progressive distillation for fast sampling of diffusion models," in International Conference on Learning Representations, 2022.
[6] H. Jiang, M. Imran, L. Ma, T. Zhang, Y. Zhou, M. Liang, K. Gong, and W. Shao, "Fast-DDPM: Fast denoising diffusion probabilistic models for medical image-to-image generation," arXiv preprint arXiv:2405.14802, 2024.
[7] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
[8] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in International Conference on Machine Learning, 2021.
[9] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
[10] W. H. L. Pinaya et al., "Brain imaging generation with latent diffusion models," in Deep Generative Models Workshop, MICCAI, 2022.
[11] B. Menze et al., "The multimodal brain tumor image segmentation benchmark (BraTS)," IEEE Transactions on Medical Imaging, 2015.
[12] 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.
[13] I. Goodfellow et al., "Generative adversarial networks," Communications of the ACM, 2020.
[14] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[15] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," Medical Image Analysis, 2022.

