Breaking the Continuum: Discrete Distribution Learning for Structural MRI Reconstruction

Supplementary Material

7. Mathematical Principle Interpretation of Discrete–Continuous Synthesis (*DiCoS*)

In this section, we provide a formal justification for why the proposed Discrete–Continuous Synthesis (*DiCoS*) constitutes a discrete generative modeling mechanism, rather than merely a multi-head predictor. We show that DPN implicitly models a discrete latent variable, produces samples from its conditional distribution, and performs approximate MAP inference through top-t hypothesis selection. In Algorithm 1, we described the workflow of *DiCoS*.

7.1. Idealized Discrete Generative Model

Consider an idealized *discrete generative prior* designed to capture the structural distribution of medical images. We introduce a discrete latent variable $z \in \mathcal{Z} = \{1, \dots, M\}$, which indexes different hypothesized anatomical structures (e.g., different templates or texture patterns). Its prior is

$$p(z) = \pi_z, \qquad \sum_{z=1}^{M} \pi_z = 1, \quad \pi_z \geq 0. \tag{12}$$

Step 1. Generation Model. Given a latent index z, a generator network G_θ produces an MR image:

$$x = G_\theta(z), \qquad x \in \mathbb{R}^N. \tag{13}$$

For accelerated MRI, the measurement model is

$$y = \mathcal{A}(x) + n, \qquad n \sim \mathcal{N}(0, \sigma^2 I), \tag{14}$$

where $\mathcal{A}$ denotes an undersampled Fourier operator.

Step 2. Posterior Distribution. Under this discrete generative formulation, the posterior over the latent index is

$$p(z \mid y) \propto p(y \mid z)p(z) = \exp\big(-E(z; y)\big), \tag{15}$$

with the energy defined as

$$E(z; y) = \frac{1}{2\sigma^2} \left\| \mathcal{A}\big(G_\theta(z)\big) - y \right\|_2^2 - \log \pi_z. \tag{16}$$

This energy quantifies how well the structural hypothesis indexed by z explains the measurement y.

Step 3. Ideal MAP Reconstruction. The ideal MAP estimate is obtained by searching over the discrete structural hypothesis space:

$$x^\star = G_\theta(z^\star), \qquad z^\star = \arg\min_{z \in \mathcal{Z}} E(z; y). \tag{17}$$

Algorithm 1 ***Discrete–Continuous Reconstruction***

Input: Undersampled k-space y, Operator $\mathcal{A}$, Pretrained score prior $s_\theta(x)$ and Parameters L, K, T, τ

Output: Reconstructed MR image $\hat{x}$

```
Initialize x_0 = A^H(y)
for stage ℓ = 1 ... L do
    Generate K candidates {x_ℓ^(k)} via DPN_ℓ(x_{ℓ−1})
    for each candidate x_ℓ^(k) do
        for t = 1 ... T do                    // MDC (Section 3.3)
            x ← x + η_t s_θ(x) + √(2η_t) ε_t          // PC
            x ← x + A^H(y − A(x))                     // DC
            update x_ℓ^(k) ← x
        end for
        Compute DBS Score(x_ℓ^(k))            // (Equation (8))
    end for
    Select top-t candidates
end for
Return x̂ = x_L
```

Complexity: $\mathcal{O}(LKTN)$, where N is image resolution

This means reconstruction requires ***searching over a discrete space of structural hypotheses***. However, brute-force search or MCMC sampling over a large number M of discrete states is computationally prohibitive and unsuitable for end-to-end learning.

7.2. Continuous Relaxation and Its Equivalence to Candidate Generation

To make the discrete structural search tractable, learnable, and differentiable, we apply a *continuous relaxation* to the discrete latent variable z. We show that the candidate-generation process in DPN is in fact a continuous relaxation of the discrete generative model.

Step 4. Embedding the Discrete State Into the Probability Simplex. Let $e_k \in \mathbb{R}^K$ denote the one-hot vector corresponding to the discrete state $z = k$. We embed the discrete space $\mathcal{Z}$ into the probability simplex:

$$\Delta^{K-1} = \left\{ s \in \mathbb{R}^K \,\middle|\, s_k \geq 0, \sum_{k=1}^{K} s_k = 1 \right\}, \tag{18}$$

whose vertices $\{e_1, \dots, e_K\}$ correspond exactly to the discrete states $\{1, \dots, K\}$. We define a continuous generative mapping:

$$G(h, s) = \sum_{k=1}^{K} s_k \, G_k(h), \tag{19}$$

where $h \in \mathbb{R}^d$ is the feature extracted from the previous reconstruction stage, $G_k(h)$ denotes the k-th image hypothesis (corresponding to the k-th 1×1 head in DPN), and $s \in \Delta^{K-1}$ is a continuous selection weight. When $s = e_k$, we exactly recover the discrete generative mapping:

$$G(h, e_k) = G_k(h), \tag{20}$$

which corresponds to the original discrete generation $x = G_\theta(z)$. During training and inference, however, we allow s to lie in the interior of the simplex, yielding a differentiable relaxation of the discrete latent variable. This provides the continuous foundation underlying the DPN candidate-generation mechanism.

Step 5. DPN as a Continuous Relaxation of Discrete Posterior Inference. At the ℓ-th DPN stage, we first extract features from the previous reconstruction:

$$h_\ell = \phi_\ell(x_{\ell-1}), \tag{21}$$

and feed them into K linear heads to produce logits:

$$a_k = w_k^\top h_\ell + b_k, \qquad k = 1, \dots, K. \tag{22}$$

Applying a Boltzmann distribution with temperature yields a continuous distribution:

$$q_\ell(k \mid x_{\ell-1}) = \frac{\exp(a_k/\tau)}{\sum_{j=1}^K \exp(a_j/\tau)}, \tag{23}$$

where τ denotes the temperature. As $\tau \to 0$, the softmax distribution collapses to a simplex vertex:

$$q_\ell(k \mid x_{\ell-1}) \to e_{k^*}, \qquad k^* = \arg\max_k a_k, \tag{24}$$

that is the distribution over Δ^{K-1} approaches a discrete latent state. Thus, $q_\ell(k \mid x_{\ell-1})$ can be interpreted as a learnable approximation to the ideal discrete posterior $p(z \mid y)$.

Step 6. Two-Step Interpretation of DPN Candidate Generation. The first step is ***Continuous selection.*** Given $q_\ell(k \mid x_{\ell-1})$, we sample or select top-t:

$$z_\ell^{(k)} \sim q_\ell(\cdot \mid x_{\ell-1}) \quad \Longleftrightarrow \quad s_\ell^{(k)} \approx e_{z_\ell^{(k)}} \in \Delta^{K-1}. \tag{25}$$

The second step is ***Continuous generation.*** Using the selected continuous weight vector, the candidate is generated as:

$$x_\ell^{(k)} = G(h_\ell, s_\ell^{(k)}) = \sum_{j=1}^K s_{\ell,j}^{(k)}\, G_j(h_\ell) \approx G_{z_\ell^{(k)}}(h_\ell). \tag{26}$$

When the temperature τ is sufficiently small, this becomes equivalent to a hard top-k selection:

$$s_\ell^{(k)} \in \{e_1, \dots, e_K\}. \tag{27}$$

Hence, each candidate $x_\ell^{(k)}$ is effectively a sample produced from one discrete structural hypothesis z.

That is to say, under the continuous relaxation $s \in \Delta^{K-1}$, the DPN first learns an approximate posterior $q_\ell(s \mid x_{\ell-1})$, and then samples several points $s_\ell^{(k)}$ near the simplex vertices, thereby producing a set of candidates $\{x_\ell^{(k)}\}$ — which constitutes a Monte Carlo approximation of the discrete generative model $p(z)p(x \mid z)$.

Step 7. Connection to the Discrete Generative ELBO. If we associate the entire process with the discrete generative ELBO:

$$\max\ \mathbb{E}_{q_\ell(k)}[\log p_\theta(x_\ell^{(k)})] - \mathrm{KL}\big(q_\ell(z \mid y) \,\|\, p(z)\big), \tag{28}$$

and replace the discrete latent z with its relaxed counterpart $s_\ell^{(k)}$, we obtain exactly the training behaviour observed in DPN. Therefore, from the perspective of continuous relaxation, the DPN candidate-generation block implements:

- a learnable approximate posterior $q_\ell(k \mid x_{\ell-1})$,
- a continuous relaxation $s_\ell^{(k)}$ on the simplex,
- and a generator G_k that becomes discrete at the simplex vertices $x_\ell^{(k)}$.

In short, the DPN realizes a ***differentiable approximation to discrete generative modeling***: it replaces the true discrete variable z with a continuous, learnable relaxation.

8. Why Segmentation Results Reflect Semantic Performance

In this section, we justify why MedSAM-based [23] segmentation provides a valid proxy for evaluating the semantic performance of reconstruction models.

Problem setup. Let $x_{\mathrm{GT}} \in \mathbb{R}^{H\times W}$ denote the fully-sampled ground-truth MR image, and $\hat{x} = \mathcal{R}(y)$ be the reconstruction produced by a model $\mathcal{R}$ from undersampled measurements y. Let $y_{\mathrm{seg}} \in \{1, \dots, C\}^{H\times W}$ denote the ground-truth anatomical segmentation labels (organs, tissues, lesions, etc.). We consider a hypothetical semantic oracle $f^\star : \mathbb{R}^{H\times W} \to \{1, \dots, C\}^{H\times W}$ which maps an image to the correct anatomical labels, i.e.

$$f^\star(x_{\mathrm{GT}}) = y_{\mathrm{seg}}. \tag{29}$$

Ideally, a reconstruction $\hat{x}$ is *semantically consistent* with x_{GT} if it induces the same semantic interpretation:

$$f^\star(\hat{x}) = f^\star(x_{\mathrm{GT}}) = y_{\mathrm{seg}}. \tag{30}$$

MedSAM as a semantic oracle. In practice, the oracle $f^\star$ is unavailable. Instead, we use a pre-trained segmentation network f_{M} (MedSAM), kept *frozen* during all experiments. We assume that, on fully-sampled clinical MR images, f_{M} approximates the oracle in the sense that

$$\mathbb{P}\big[f_{\mathrm{M}}(x_{\mathrm{GT}}) = y_{\mathrm{seg}}\big] \approx 1, \tag{31}$$

That is, MedSAM [23] produces near-correct anatomical labels for high-quality inputs. Under this assumption, we can use f_{M} as a surrogate for $f^\star$:

$$f^\star(\hat{x}) \approx f_{\mathrm{M}}(\hat{x}), \quad f^\star(x_{\mathrm{GT}}) \approx f_{\mathrm{M}}(x_{\mathrm{GT}}). \tag{32}$$

Segmentation agreement as semantic consistency. Define the MedSAM-generated masks:

$$\hat{y} = f_{\mathrm{M}}(\hat{x}), \qquad \tilde{y} = f_{\mathrm{M}}(x_{\mathrm{GT}}). \tag{33}$$

If the reconstruction $\hat{x}$ preserves the semantic content of x_{GT}, then Equation (30) implies

$$\hat{y} = f_{\mathrm{M}}(\hat{x}) \approx f^\star(\hat{x}) = f^\star(x_{\mathrm{GT}}) \approx f_{\mathrm{M}}(x_{\mathrm{GT}}) = \tilde{y}. \tag{34}$$

Conversely, any semantic distortion in $\hat{x}$ (e.g., shifted boundaries, missing structures, spurious lesions) changes the input to the fixed MedSAM [23] model and leads to discrepancies between $\hat{y}$ and $\tilde{y}$. Let $\mathcal{D}(\cdot,\cdot)$ denote a segmentation similarity metric such as Dice or IoU. Then

$$\mathcal{D}\big(\hat{y}, \tilde{y}\big) = \mathcal{D}\big(f_{\mathrm{M}}(\hat{x}), f_{\mathrm{M}}(x_{\mathrm{GT}})\big) \tag{35}$$

is high if and only if the two segmentations agree on most pixels. Under the above approximation of $f^\star$ by f_{M}, this is equivalent to saying that $\hat{x}$ and x_{GT} induce the same anatomical interpretation:

$$\begin{aligned} & \mathcal{D}\big(f_{\mathrm{M}}(\hat{x}), f_{\mathrm{M}}(x_{\mathrm{GT}})\big) \uparrow \\ \iff \quad & \text{Semantic consistency} \uparrow \text{ between } \hat{x} \text{ and } x_{\mathrm{GT}}. \end{aligned} \tag{36}$$

Conclusion. Because segmentation requires assigning a semantic label to each anatomical region, MedSAM [23] implicitly encodes high-level anatomical knowledge. Keeping f_{M} fixed and comparing $\hat{y} = f_{\mathrm{M}}(\hat{x})$ with $\tilde{y} = f_{\mathrm{M}}(x_{\mathrm{GT}})$ therefore measures how well the reconstruction preserves those semantics. Dice and IoU between MedSAM masks on reconstructed and ground-truth images thus serve as a mathematically well-defined proxy for the semantic performance of MRI reconstruction models.

9. Continuous Score-based SDE Pretraining and Its Role in DiCoS

Why a continuous score model is needed. The discrete hypothesis generator in *DiCoS* captures anatomical structures but lacks the fine-grained textures and local continuity required for high-fidelity MRI reconstruction. To compensate, we pretrain a continuous score-based diffusion model that provides a smooth gradient field $\nabla_x \log p_t(x)$ over the entire image domain. This continuous field naturally complements the discrete hypotheses by injecting fine-scale regularity and local geometric detail.

Architecture and resolution generalization. The score model $s_\theta(x_t, t)$ adopts a lightweight U-Net with four resolution levels, FiLM-modulated time embeddings, and fully convolutional residual blocks. While the continuous score model is trained under a VE-SDE, which is resolution-independent: it only defines the noise scale $\sigma(t)$ and does not constrain the spatial size. Because no component depends on absolute spatial size, the learned score field $s_\theta(\cdot, t)$ is *resolution-equivariant*: its output is consistent across 40, 80, 160, and 320 resolutions. This property allows a single pretrained model to refine candidates at all stages:

$$s_\theta : \mathbb{R}^{H\times W} \to \mathbb{R}^{H\times W}, \quad \text{holds for any } (H, W). \tag{37}$$

Thus, refinement quality is shared across the entire coarse-to-fine hierarchy.

Training objective. Following Song et al.[34], we adopt denoising score matching (DSM) under the VE-SDE:

$$\begin{aligned} \mathcal{L}_{\mathrm{DSM}} &= \mathbb{E}_{t,x_0,\varepsilon} \left\| s_\theta(x_t, t) + \frac{\varepsilon}{\sigma(t)} \right\|_2^2, \\ x_t &= x_0 + \sigma(t)\varepsilon, \end{aligned} \tag{38}$$

with a log-linear noise schedule

$$\sigma(t) = \sigma_{\min} \left(\frac{\sigma_{\max}}{\sigma_{\min}} \right)^t, \tag{39}$$

Where $\sigma_{\min} = 2 \times 10^{-3}$ and $\sigma_{\max} = 5 \times 10^{-2}$. Training uses AdamW with a learning rate 10^{-4}, cosine decay, gradient clipping (1.0), and EMA decay (0.9999). The result is a converged smooth score field that models the continuous MRI manifold.

In summary, the continuous score model complements discrete hypothesis generation by supplying resolution-agnostic, texture-preserving, and physically consistent refinement, acting as a universal continuous teacher throughout the entire *DiCoS* pipeline.

10. Visualization of Discrete Hypothesis Space

To more clearly illustrate the behavior of our discrete hypothesis generator, Figure 8 visualizes one full stage of *DiCoS*. The central image is the current-stage input passed to the DPN. Surrounding it are the K generated candidates, each representing a structurally plausible hypothesis consistent with the same undersampled measurement. Across candidates, one can observe variations in boundary sharpness, tissue transitions, and fine-scale texture patterns while maintaining global anatomical consistency.

To further interpret these patterns, we note that the observed variations primarily concentrate in regions where the undersampling introduces ambiguity. For knee MRI, candidates differ in cartilage–meniscus boundaries, soft-tissue contrast, and subtle high-frequency textures; for brain MRI, variations appear in ventricle boundary sharpness, cortical–subcortical contrast, and local texture recovery. Despite these localized differences, all hypotheses preserve the global anatomical structure, illustrating that *DiCoS* explores a constrained and semantically meaningful

hypothesis manifold rather than generating unstable noise. The DBS-selected hypotheses consistently achieve a balanced recovery of structural details, confirming that our discrete–continuous reasoning pipeline effectively identifies the most reliable candidate for downstream refinement.

11. Fairness of MDC's Pretrained Prior

To test dependence on VE-SDE [6], we replace MDC's prior with HFS-SDE [4]. As shown in Table 2, performance is slightly improved, indicating our gains are not tied to a specific pretrained prior. This robust performance across different priors confirms that our gains are inherent to the DPN architecture rather than being tied to a specific pretrained model, ensuring a fair comparison with existing methods.

Table 5. Comparison in pretrained model in MDC module

Pretrained	NMSE ↓	PSNR ↑	SSIM ↑	Dice ↑	IoU ↑
VE-SDE	**$1.43^{\pm 0.74}$**	**$35.32^{\pm 2.26}$**	**$86.13^{\pm 3.59}$**	**0.921**	**0.842**
HFS-SDE	$1.41^{\pm 1.03}$ ▲	$35.37^{\pm 2.35}$ ▲	$86.22^{\pm 2.89}$ ▲	0.919 ▼	0.845 ▲

12. Limitations and Future Work

Limitations. *DiCoS* currently models the discrete structural manifold with a fixed number K of hypotheses at each stage. Although this formulation successfully captures diverse anatomical modes, it may not fully adapt to variations in structural complexity across slices or modalities. In addition, our evaluation focuses on 2D reconstructions for consistency with prior work, and the extension to full 3D volumetric reasoning remains to be explored.

Future Work. Future work includes developing adaptive hypothesis allocation mechanisms, extending DiCoS to 3D volumetric MRI, and integrating semantics-aware priors such as segmentation-conditioned diffusion models.

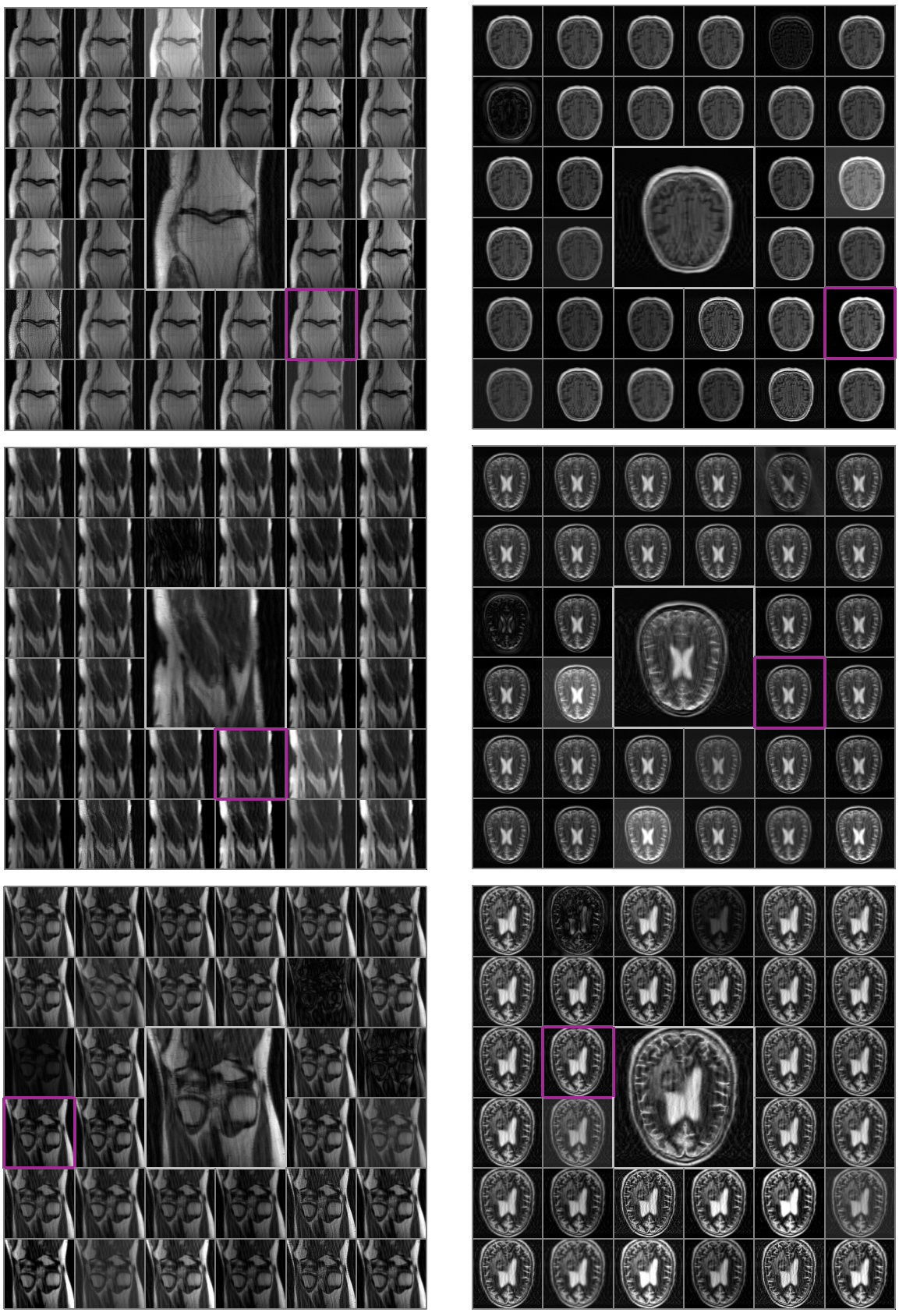

Figure 8. Visualization of the current-stage input and the K candidate reconstructions generated by *DiCoS*. The purple box marks the candidate selected by our Dual-domain Balanced Scoring (DBS) module, which is subsequently propagated to the next stage for refinement.