Project RAPTOR+ Proposal 16:

AI-Driven Anatomical and Response Adapted Proton Therapy

PhD Project Proposal

1 Academic State-of-the-Art: The Convergence of Generative AI and Deformable Registration (2020-2025)

The landscape of adaptive proton therapy (APT) has undergone a seismic shift in the first half of the 2020s, driven by the intersecting trajectories of generative artificial intelligence and quantitative image guidance. While photon-based radiotherapy has long benefited from the "robustness" of dose delivery—where the integral dose washes out minor geometric uncertainties—proton therapy remains uniquely unforgiving. The Bragg peak's finite range means that millimeter-scale anatomical deviations or minor fluctuations in tissue density can result in catastrophic under-dosing of the target or, conversely, the deposition of maximum dose into distal organs at risk (OARs).[1, 2] Consequently, the academic focus has narrowed on two critical challenges: the restoration of photometric integrity to daily Cone-Beam Computed Tomography (CBCT) images to enable accurate stopping power calculations, and the rigorous decoupling of geometric deformation from biological response to facilitate true dose accumulation.

1.1 The Evolution of Synthetic CT (sCT) Generation: From GANs to Diffusion Models

The period from 2020 to 2025 has been defined by the transition from adversarial learning to probabilistic diffusion as the gold standard for synthetic CT (sCT) generation. The clinical imperative is clear: CBCT images, while geometrically sufficient for positioning, suffer from scatter artifacts, beam hardening, and limited field-of-view (FOV) disparities that render their Hounsfield Units (HU) unreliable for proton stopping power ratio (SPR) calculation.[3, 4, 5]

1.1.1 The Era of CycleGAN and Unpaired Translation (2020-2023)

In the early part of this decade, Generative Adversarial Networks (GANs), particularly the CycleGAN architecture, dominated the literature.[6, 7] The primary advantage of CycleGAN lay in its ability to perform unpaired image-to-image translation. In a clinical setting, obtaining perfectly registered pairs of Planning CT (pCT) and daily CBCT is physically impossible due to inevitable non-rigid patient motion between scans. CycleGAN circumvented this by enforcing a cycle-consistency loss—learning a mapping $G : CBCT \rightarrow sCT$ and an inverse mapping $F : sCT \rightarrow CBCT$ such that $F(G(x)) \approx x$.[8]

Research involving Head and Neck (H&N) cohorts demonstrated that CycleGAN-generated sCTs could reduce the Mean Absolute Error (MAE) of HU values significantly compared to raw CBCTs. For instance, studies showed MAE reductions from over 100 HU (in raw images) to approximately 30–50 HU in sCTs.[6] However, applying these models to proton therapy revealed critical safety flaws. GANs are prone to "hallucination"—the generation of realistic-looking but anatomically nonexistent structures. In proton therapy, a hallucinated pocket of air or a misrepresented bone interface changes the water-equivalent path length (WEPL), potentially shifting the Bragg peak by centimeters.[5] Furthermore, standard CycleGANs often struggled with high-frequency texture preservation, leading to "waxy" or blurred bone trabeculae, which are essential for accurate SPR estimation in osteophilic tumors.[9]

1.1.2 The Paradigm Shift to Diffusion Probabilistic Models (2024-2025)

Recognizing the limitations of adversarial training (mode collapse and instability), the field shifted toward Denoising Diffusion Probabilistic Models (DDPMs) and Score-Based Generative Models (SGMs).[4, 10] Unlike GANs, which learn an implicit distribution via a minimax game, diffusion models learn the data distribution explicitly by reversing a gradual noise addition process (Forward SDE vs. Reverse SDE).

Recent literature from 2024 and 2025 indicates that DDPMs offer superior fidelity in medical image synthesis. By conditioning the reverse denoising process on the daily CBCT anatomy (Conditional DDPM), these models can generate sCTs that are not only geometrically faithful to the daily anatomy but also texturally indistinguishable from diagnostic CTs.[11, 12]

Comparative Performance Metrics: A synthesis of recent validation studies highlights the superiority of Diffusion models over CycleGANs in the context of proton dosimetry:

Table 1: Comparison of Generative Models for sCT Generation

Metric	Raw CBCT	CycleGAN sCT	Diffusion (DDPM) sCT	Impact on Proton Therapy
MAE (Soft Tissue)	¿ 100 HU	∼29.9 ± 4.9 HU [6]	∼**24.5 ± 3.3 HU** [11]	Improved range accuracy in homogenous tissue.
MAE (Bone)	¿ 300 HU	∼75 HU [13]	∼**61 HU** [13]	Critical for avoiding distal undershoot in bone.
SSIM	0.65 - 0.70	0.86 ± 0.05 [14]	¿ **0.96 ± 0.01** [12]	High structural fidelity reduces geometric distortion.
SPR Error (Bone)	∼6.8%	∼1.5 - 2.0%	¡ **0.20%** [15, 16]	**Major Milestone:** Negligible range uncertainty.
Gamma (1%/1mm)	¡ 80%	∼95% (Photon)	¿ **97.9% (Proton)** [17]	Enables clinical confidence for online adaptation.

The data implies that while GANs effectively solved the "photon problem" (where electron density accuracy is less critical), only Diffusion models have achieved the precision required for the "proton problem," specifically in reducing SPR error in bone to below 0.5%.[15, 16]

1.2 The Deformable Registration Dilemma: Mass Conservation vs. Biological Reality

While sCT generation addresses the photometric consistency of the image, Adaptive Radiotherapy (ART) also requires addressing the geometric consistency. This is achieved through Deformable Image Registration (DIR). However, a fundamental theoretical conflict exists in current DIR implementations applied to oncology: the assumption of mass conservation versus the reality of tumor response.[18]

1.2.1 The Limits of Diffeomorphic Registration

Standard clinical DIR algorithms (e.g., B-spline, Demons, Diffeomorphic Log-Demons) optimize a cost function that balances image similarity with a regularization term enforcing smoothness.[19, 20] These algorithms typically assume a diffeomorphism—a transformation that is smooth, invertible, and topology-preserving. Mathematically, this implies that the mass within the volume is conserved; tissue can stretch, compress, or shear, but it cannot vanish or appear.[21]

This assumption holds for physiological motion (e.g., rectal filling, respiration) but fails catastrophically for therapeutic response. In H&N cancer, tumors often undergo significant regression (mass loss) and necrosis (density change) during the 6-7 weeks of fractionated proton therapy.[2, 22] When a diffeomorphic DIR is applied to a shrinking tumor, the algorithm is mathematically forced to "fill the void" by stretching the surrounding healthy tissue into the space previously occupied by the tumor. This results in physically implausible Deformation Vector Fields (DVFs) and incorrect dose accumulation.[23, 24, 25]

1.2.2 Residuals and Jacobian Maps as Biomarkers

To address this, research has turned to analyzing the "residuals" of registration—the information that the transformation fails to explain.

- **Jacobian Analysis:** The Jacobian determinant of the DVF (J) serves as a proxy for local volume change. $J < 1$ indicates compression, while $J > 1$ indicates expansion. Studies have successfully used Jacobian maps to quantify parotid gland shrinkage, correlating specific J thresholds (e.g., $J < 0.5$) with xerostomia outcomes.[23, 26]
- **Intensity Residuals:** The residual map, defined as $|I_{fixed} - I_{warped}|$, highlights regions where intensity matching failed. In the context of tumor regression, high residuals often cluster around the tumor boundary.[27, 28]
- **Unbalanced Optimal Mass Transport (OMT):** Emerging theoretical frameworks (2023-2025) propose "unbalanced" OMT, which relaxes the mass conservation constraint. These methods introduce a source term in the continuity equation, explicitly modeling the creation or destruction of mass.[29, 30]

1.3 Deep Learning for Response Prediction: The Two-Stream Approach

The convergence of residuals and deep learning has given rise to "Response Prediction" models. Rather than relying on static pre-treatment radiomics, these models exploit longitudinal changes.

- **Difference Images as Input:** Several studies have demonstrated that feeding the element-wise difference between longitudinal scans ($I_t - I_{t-1}$) into a CNN yields higher predictive accuracy for Pathological Complete Response (pCR) than using static images alone.[31, 32, 33]
- **Multi-Stream Architectures:** The state-of-the-art architectures utilize "Two-Stream" or "Dual-Path" networks.[34, 35, 36] One stream processes the **Geometry** (DVF, Jacobian), learning patterns of elastic deformation. The second stream processes the **Texture/Biology** (sCT values, Residuals), learning patterns of necrosis and regression.

1.4 Dosimetric Implications for Proton Therapy

The inability to distinguish anatomy from biology has severe dosimetric consequences in proton therapy.

- **Range Uncertainty:** If a tumor shrinks and the DIR algorithm stretches healthy tissue into the void, the sCT used for dose accumulation will have incorrect stopping powers in the path of the beam. This leads to range calculation errors that can overshoot the target.[1, 37]
- **The "Hallucination" Risk:** While Diffusion models are superior, they are not immune to generating realistic-looking features that do not exist. Validation studies utilizing Monte Carlo simulations have shown that while sCTs achieve Gamma Pass Rates ¿97% for photons, they can drop to ∼93% for protons if the model is not explicitly constrained by physical stopping power losses.[38]

2 The Methodology: RAPTOR-Net Architecture

This proposal introduces **RAPTOR-Net** (Response-Aware Proton Therapy Online Re-planning Network), a closed-loop AI pipeline designed to operate on longitudinal CBCT data.

2.1 Module A: High-Fidelity sCT Generation via Physics-Informed Diffusion

We propose a **3D Conditional Denoising Diffusion Probabilistic Model (cDDPM)** specifically optimized for proton dosimetry.

2.1.1 Mathematical Formulation of the Diffusion Process

The diffusion process is modeled as a parameterized Markov chain comprising a forward process q and a reverse process p_θ.

Forward Process (Noise Injection): We define a forward diffusion process that gradually adds Gaussian noise to the ground truth Planning CT (x_0) over T timesteps, transforming it into an isotropic Gaussian distribution $x_T \sim \mathcal{N}(0, I)$.

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

where β_t is a variance schedule fixed a priori.

Reverse Process (Generative Denoising): The reverse process is learned by a neural network ϵ_θ (a 3D U-Net) which predicts the noise component at each step, conditioned on the daily CBCT image c.

$$p_\theta(x_{t-1}|x_t, c) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t, c), \Sigma_\theta(x_t, t, c)) \tag{2}$$

Here, the daily CBCT c is not merely concatenated to the input but is injected via **Cross-Attention layers** at multiple resolutions of the U-Net.[11, 12]

2.1.2 Proton-Specific Loss Function: The SPR Constraint

To prevent the model from prioritizing soft-tissue contrast at the expense of range accuracy, we introduce a **Stopping Power Ratio (SPR) Consistency Loss**. Let $\mathcal{M}(HU)$ be the stoichiometric calibration curve converting HU to proton Stopping Power Ratio. The total loss $\mathcal{L}_{total}$ is:

$$\mathcal{L}_{total} = \|\epsilon - \epsilon_\theta(x_t, t, c)\|^2 + \lambda_{SPR}\|\mathcal{M}(x_0) - \mathcal{M}(\hat{x}_0)\|_1 \tag{3}$$

where $\hat{x}_0$ is the estimated denoised volume at $t = 0$. The weight λ_{SPR} is dynamically adjusted to penalize errors in high-density regions (bone) and low-density regions (air cavities) more heavily.[15, 16]

2.2 Module B: The Dual-Stream Metamorphic Discriminator

Once accurate sCTs are generated for the longitudinal timepoints (e.g., sCT_{t0} and sCT_{tn}), the pipeline must determine if the observed changes are purely anatomical or biological.

2.2.1 Step 1: DeepReg-Based Metamorphic Registration

We utilize a learning-based registration framework (DeepReg) to align the baseline sCT_{t0} to the current sCT_{tn}.

- **Input:** Moving Image (sCT_{t0}), Fixed Image (sCT_{tn}).
- **Outputs:**
  1. **Deformation Vector Field (DVF), ϕ:** Represents geometric transformation.
  2. **Jacobian Determinant Map, J_ϕ:** Calculated as $J_\phi(p) = \det(\nabla\phi(p))$.
  3. **Residual Map, R:** $|sCT_{tn} - sCT_{t0} \circ \phi|$.
  4. **Difference Image, ΔI:** $sCT_{tn} - sCT_{t0}$.[31, 33]

2.2.2 Step 2: Dual-Stream Classification Network

We propose a **Geometry-Texture Disentanglement Network**.

- **Stream 1 (Geometry):** A 3D CNN taking the DVF (ϕ) and Jacobian (J_ϕ) as inputs.
- **Stream 2 (Biology/Texture):** A 3D CNN taking the Residual Map (R) and Difference Image (ΔI) as inputs.

The feature maps are fused via a **Transformer-based Cross-Attention block**.

3 Coding & Implementation

3.1 Technical Stack

- **Core Framework:** Python 3.9+, PyTorch 2.1+ (CUDA 12.x).
- **Generative AI:** MONAI v1.3 (`DiffusionModelUNet`, `DiffusionInferer`).
- **Registration:** DeepReg v0.2.
- **Data I/O:** SimpleITK, Pydicom.

3.2 Implementation Roadmap and Pseudocode

3.2.1 Phase 1: Conditional Diffusion for sCT (Training Loop)

```
import torch
import torch.nn as nn
from monai.networks.nets import DiffusionModelUNet
from monai.inferers import DiffusionInferer
from monai.schedulers import DDPMScheduler

class ProtonDiffusionModel(nn.Module):
    def __init__(self):
        super().__init__()
        # 3D U-Net backbone with cross-attention for conditioning
        self.model = DiffusionModelUNet(
            spatial_dims=3,
            in_channels=1,
            out_channels=1,
            num_channels=(64, 128, 256),
            attention_levels=(False, True, True),
            num_res_blocks=2,
            with_conditioning=True # Enables Cross-Attention for CBCT
        )
        self.scheduler = DDPMScheduler(num_train_timesteps=1000)
        self.inferer = DiffusionInferer(scheduler=self.scheduler)

    def forward(self, x_ct, x_cbct, noise=None, timesteps=None):
        if noise is None:
            noise = torch.randn_like(x_ct)

        # Diffusion Forward Process
        noisy_ct = self.scheduler.add_noise(
            original_samples=x_ct, noise=noise, timesteps=timesteps
        )

        # Reverse Process conditioned on CBCT
        noise_pred = self.model(
            x=noisy_ct, timesteps=timesteps, context=x_cbct
```

```
        )
        return noise_pred

def spr_weighted_loss(noise_pred, noise, true_ct):
    """
    Physics-Informed Loss: Weights gradients based on Proton Stopping Power.
    """
    mse_loss = nn.MSELoss(reduction='none')(noise_pred, noise)

    # Create spatial weight map based on HU of the ground truth
    weights = torch.ones_like(true_ct)
    weights[true_ct > 100] = 2.0  # High penalty for bone errors
    weights[true_ct < -500] = 1.5 # Moderate penalty for air errors

    return (mse_loss * weights).mean()
```

3.2.2 Phase 2: The Dual-Stream Discriminator

```
class DualStreamDiscriminator(nn.Module):
    def __init__(self):
        super().__init__()

        # Stream 1: Geometry (DVF + Jacobian)
        self.geo_stream = nn.Sequential(
            nn.Conv3d(4, 32, kernel_size=3, padding=1),
            nn.BatchNorm3d(32),
            nn.LeakyReLU(0.2),
            nn.MaxPool3d(2)
        )

        # Stream 2: Biology (Residual + Diff Image)
        self.bio_stream = nn.Sequential(
            nn.Conv3d(2, 32, kernel_size=3, padding=1),
            nn.BatchNorm3d(32),
            nn.LeakyReLU(0.2),
            nn.MaxPool3d(2)
        )

        # Attention Fusion Block
        self.attention = nn.MultiheadAttention(embed_dim=64, num_heads=4)

        # Classification Head
        self.classifier = nn.Sequential(
            nn.Linear(64, 32),
            nn.ReLU(),
            nn.Linear(32, 2)
        )

    def forward(self, dvf, jacobian, residual, diff_img):
        # Feature Extraction
        geo_features = self.geo_stream(torch.cat([dvf, jacobian], dim=1))
        bio_features = self.bio_stream(torch.cat([residual, diff_img], dim=1))

        # Cross-Attention: Biology stream queries Geometry stream
        b, c, d, h, w = geo_features.shape
        geo_flat = geo_features.view(b, c, -1).permute(2, 0, 1) # Key/Value
        bio_flat = bio_features.view(b, c, -1).permute(2, 0, 1) # Query

        attn_out, _ = self.attention(query=bio_flat, key=geo_flat, value=
    geo_flat)

        pooled = attn_out.mean(dim=0)
```

```
        logits = self.classifier(pooled)
        return logits
```

4 Data & Validation Strategy

4.1 Datasets

We will leverage high-quality open-source datasets from The Cancer Imaging Archive (TCIA).

1. **HNSCC-3DCT-RT** (Primary Dataset): Contains longitudinal fan-beam CTs at pre-, mid-, and post-treatment.[39, 40]
2. **Head-Neck-PET-CT** (Secondary Dataset): Contains PET/CT data to validate biological target volumes.[41, 42]
3. **QIN-HEADNECK**: Used as an external test set for response prediction.[43, 44]

4.2 Validation Protocols

Protocol A: Geometric & Photometric Accuracy (sCT)

- MAE (Soft Tissue): ¡ 30 HU
- MAE (Bone): ¡ 60 HU
- SSIM: ¿ 0.95

Protocol B: Dosimetric Validation

- **Gamma Index Analysis:** Target pass rate $> 95\%$ at 1%/1mm criteria.[17]
- **Range Difference (ΔR_{90}):** Target accuracy $\Delta R_{90} < 1$mm.