

**Maker-less Lung Tumor Tracking from Real-time  
Color X-ray Fluoroscopic Images for Radiotherapy**

放射線治療のためのリアルタイムカラーX線透視  
画像からのマーカーレス肺腫瘍追跡

March, 2026

Yamaguchi University Graduate School of Sciences and Technology for  
Innovation

Yan Yongxuan

# Contents

|       |                                                                                   |    |
|-------|-----------------------------------------------------------------------------------|----|
| 1     | Introduction . . . . .                                                            | 3  |
| 2     | Modeling of Degradation in SyncTraX Fluoroscopic Imaging . . . . .                | 7  |
| 2.1   | Style Difference . . . . .                                                        | 7  |
| 2.2   | Degradation Modeling . . . . .                                                    | 12 |
| 2.3   | Results . . . . .                                                                 | 17 |
| 2.4   | Conclusion . . . . .                                                              | 17 |
| 3     | Workflow for Tumor Tracking . . . . .                                             | 19 |
| 3.1   | Equipment and Data . . . . .                                                      | 19 |
| 3.2   | Overview of Workflow . . . . .                                                    | 20 |
| 3.3   | Algorithms and Neural Network Models . . . . .                                    | 21 |
| 3.3.1 | Grayscale Conversion Algorithm . . . . .                                          | 21 |
| 3.3.2 | Transfer Learning of Restormer . . . . .                                          | 22 |
| 3.4   | Tumor Segmentation Network . . . . .                                              | 23 |
| 3.4.1 | Network Input and Configuration . . . . .                                         | 23 |
| 3.4.2 | Training Settings . . . . .                                                       | 24 |
| 3.5   | Trajectory Extraction and Post-processing Analysis . . . . .                      | 24 |
| 3.5.1 | Low-pass Filtering . . . . .                                                      | 25 |
| 3.5.2 | 3D Motion Trajectory Comparison . . . . .                                         | 27 |
| 3.5.3 | 2D Motion Trajectory Comparison . . . . .                                         | 29 |
| 3.6   | Results . . . . .                                                                 | 29 |
| 3.6.1 | Image Grayscale and DRR Style Transfer Results . . . . .                          | 29 |
| 3.6.2 | DUCK-Net Segmentation Results and Motion Trajectory Accuracy Assessment . . . . . | 30 |
| 3.7   | Discussion . . . . .                                                              | 36 |
| 3.8   | Conclusion . . . . .                                                              | 42 |
| 4     | Development of An Integrated Software Platform . . . . .                          | 44 |
| 4.1   | Introduction . . . . .                                                            | 44 |
| 4.2   | System Design and Architecture . . . . .                                          | 45 |

|       |                                                             |    |
|-------|-------------------------------------------------------------|----|
| 4.2.1 | Requirements Analysis . . . . .                             | 45 |
| 4.2.2 | Technology Stack and Architecture . . . . .                 | 46 |
| 4.3   | Implementation of Key Functional Modules . . . . .          | 46 |
| 4.3.1 | GUI Layer and Interface Implementation . . . . .            | 47 |
| 4.3.2 | Data Management and Cross-interface Communication . . . . . | 48 |
| 4.3.3 | Storage Management and Extensibility . . . . .              | 48 |
| 4.3.4 | Virtual Imaging Geometry and DRR Generation . . . . .       | 48 |
| 4.3.5 | AI Model Batch Processing and Evaluation . . . . .          | 50 |
| 4.4   | Discussion and Summary . . . . .                            | 52 |
|       | Acknowledgment . . . . .                                    | 60 |

# Chapter 1

## Introduction

According to data from the World Health Organization, lung cancer has the highest incidence and mortality rates among all cancers [1]. Radiotherapy remains one of the primary treatment modalities for lung cancer [2–4]. Lung tumors exhibit individualized periodic displacement due to respiratory motion, which can easily lead to discrepancies between the planned and actual irradiation areas, thereby affecting the accuracy of radiation dose delivery and increasing the risk of complications. Image-Guided Radiation Therapy (IGRT), which relies on continuous or quasi-continuous imaging during treatment to monitor the positional changes of tumors and organs at risk in real-time, is a key technology for ensuring precise dose delivery and reducing irradiation to normal tissues [5–7]. In recent years, extensive research has been conducted in both clinical and engineering fields on respiratory motion management, respiratory gating, and real-time tumor tracking, promoting the in-depth development of accelerators and imaging systems [8–13].

Kilovoltage (kV) X-ray fluoroscopic image-guided radiotherapy is a common solution, yet the low soft-tissue contrast and organ occlusion often result in poor tumor visibility and blurred edges in fluoroscopic images. Clinically, it is necessary to implant fiducial markers around the tumor to assist in computer-aided tumor localization [14–16]. However, the implantation procedure of these markers is invasive and may lead to complications such as pneumothorax, intra-alveolar hemorrhage, and acute pneumonia [17–20]. Marker-less methods, which eliminate the need for marker implantation surgery, can avoid the associated surgical risks and complications, offering higher patient acceptance and greater potential for widespread adoption, and have thus gradually become a research hotspot. These methods often rely on temporal information and anatomical priors from imaging modalities, achieving tumor localization and trajectory estimation through strategies such as image segmentation, registration, optical flow estimation, or deep learning regression. Segmentation-based approaches using kV X-ray images face significant challenges due to low soft-tissue contrast and the style discrepancy between real fluoroscopic



images and digitally reconstructed radiograph (DRR) images used for training neural network models. These challenges severely limit segmentation performance and generalization across patients, posing a major bottleneck for clinical translation.

Benefiting from recent advances in machine learning technology and its broad application and research in the field of medical image processing [21–24], several deep learning-based marker-less tumor tracking methods have been proposed using kV X-ray fluoroscopic images. For instance, Roggen et al. [25] utilized Mask R-CNN for vertebral bone detection in intraoperatively acquired projection images, achieving high accuracy in spinal motion detection. He et al. [26] proposed a neural network method based on ResNetGAN, achieving sub-millimeter accuracy for marker-less motion tracking in paravertebral stereotactic body radiation therapy. However, neither method is applicable to tracking soft-tissue tumors such as those in the liver or lungs.

Among the reports on marker-less tracking suitable for lung cancer, it is common practice to train a model specifically for each individual patient. For example, Terunuma et al. [27] proposed an attention enhancement strategy combining random overlay and noise injection, which endowed the deep learning model with stronger interpretability and controllability, improving the real-time segmentation accuracy of lung tumors in kV fluoroscopic images. Their ground truth generation method was relatively complex, while the validation was primarily conducted on large tumors in the right lung, lacking coverage for cases significantly affected by heartbeats (e.g., in the left lung) or cases with smaller tumors. Zhou et al. [7] proposed using detectron2 [28], pre-trained on the COCO dataset [29], and performed transfer learning using Gaussian-filtered DRR images. The prediction results from this patient-specific model showed a total median 3D deviation of 2.27 mm, but the tumor detection rate was not high enough. Hirai et al. [30] proposed using a deep neural network (DNN) to develop regression models for motion detection based on the image patterns of one or more regions of interest (ROIs). This method does not directly detect the tumor position and contour, thus potentially overcoming situations where the tumor is severely obscured, such as by the liver. They achieved a positional accuracy with a Euclidean distance of  $1.64 \pm 0.73$  mm in 10 cases involving the liver and lung. However, certain limitations remain: besides requiring training a separate DNN for each patient, manual setting of one or more ROIs in the region of maximum displacement on DRR images is needed during the training phase. Furthermore, changes in the shape and size of patient organs and tumors as treatment progresses may lead to substantial differences between the ROIs defined during training and the corresponding regions in subsequently acquired images, potentially degrading model performance. Yan et al. [31] optimized the tumor label generation and grayscale conversion algorithms to produce labels whose boundaries better fit the actual visible tumor edges and to preserve more contrast information during grayscale conversion. They

used a cascaded patient-specific U-Net for tumor segmentation, localization, and tracking. This method could not handle cases with severe liver occlusion, and the overall accuracy in tumor localization and tracking was relatively low.

These studies collectively highlight the persistent challenges in utilizing kV X-ray fluoroscopy for tumor tracking, namely the poor generalizability of segmentation models across patients and the inherent difficulty of segmenting tumors from noisy images. To address these issues, we propose a novel workflow for real-time, marker-less tumor tracking designed for cross-patient application. While the development of more advanced neural network architectures is a viable path forward, the performance of any deep learning model is fundamentally capped by the quality of the data it is trained on. A significant and widely recognized challenge in this field is the "domain gap" between the clean, synthetic Digitally Reconstructed Radiographs (DRRs) used for training and the noisy, lower-contrast fluoroscopic images acquired during treatment. This discrepancy can cause a substantial degradation in model performance when transitioning from simulation to clinical application. Recent studies have demonstrated that explicitly addressing this domain gap by generating more realistic training data can lead to significant improvements in model generalization and accuracy.

For instance, Unberath et al. [32] addressed the domain gap in fluoroscopy by introducing DeepDRR, a framework that generates highly realistic X-ray images from CT scans. It integrates learning-based material decomposition and scatter estimation with physics-aware forward projection and noise injection. Models trained on their synthetic data successfully generalized to real clinical fluoroscopy without any domain adaptation, whereas models trained on conventional DRRs failed entirely. Kataria et al. [33] proposed to bridge the domain gap for medical image segmentation by directly matching the feature-space distributions of the source and target domains using kernel density estimation and Jensen-Shannon divergence. Their approach demonstrated significant performance improvements on target datasets and, importantly, enhanced generalization to held-out domains, outperforming adversarial and moment-matching methods, especially when the target data is limited. Guan et al. [34] comprehensively survey the pivotal role of domain adaptation in mitigating performance degradation caused by domain shift in medical image analysis.

Furthermore, the research and development workflow for marker-less image-guided therapy systems is often fragmented and inefficient. While a number of powerful open-source software platforms exist, none offer an integrated, end-to-end solution tailored for the specific cycle of marker-less tumor tracking research. For example, platforms like 3D Slicer are designed as general-purpose, extensible platforms for quantitative imaging, requiring significant customization and scripting to connect disparate modules for a specialized task [35]. Specialized tools



like ITK-SNAP excel at specific functions such as interactive 3D segmentation and navigation [36], but they do not natively support the downstream steps of deep learning model training and inference. Similarly, command-line toolkits such as Plastimatch, while robust for tasks like image registration and DRR computation, are fundamentally a collection of discrete processing functions that lack a unified graphical interface, forcing researchers to manually script the entire pipeline [37]. This necessity to manually export data, process it across multiple tools, and write custom code for model application creates a significant bottleneck. This lack of a cohesive environment has long been recognized in the field as an impediment to progress. On the other hand, commercial treatment planning systems (e.g., RayStation, Eclipse) are closed-source and designed for clinical use, lacking the flexibility required for rapid prototyping and validation of novel research algorithms. This landscape underscores the critical need for a user-friendly, research-oriented platform that consolidates these essential functionalities into a streamlined workflow, thereby accelerating the research cycle from hypothesis to validation.

This dissertation aims to tackle these challenges by presenting a cohesive framework for marker-less lung tumor tracking. The main contributions are threefold:

1. We introduce a novel methodology for generating high-quality training data. This includes an improved tumor label generation technique and a physics-informed stylization pipeline that transforms standard DRRs into realistic, fluoroscopy-like images, effectively narrowing the domain gap.
2. We propose an integrated workflow for tumor tracking, centered around an enhanced deep learning model that leverages the high-fidelity training data to achieve robust and accurate segmentation performance on actual clinical fluoroscopic images.
3. We describe the design and implementation of a dedicated desktop application that provides an end-to-end solution for medical image processing in this domain. The software integrates modules for CT/DRR handling, image stylization, AI-driven batch processing, and automated evaluation, streamlining the entire research workflow.

By addressing the fundamental issues of data quality and workflow efficiency, this work seeks to advance the state of the art in marker-less tumor tracking and pave the way for its broader clinical adoption.

## Chapter 2

# Modeling of Degradation in SyncTraX Fluoroscopic Imaging

### 2.1 Style Difference

Bridging the gap between simulated DRR images and real X-ray fluoroscopic images is a critical prerequisite for achieving robust tumor segmentation and tracking performance. A straightforward solution would be to directly transfer the visual style of real fluoroscopic images to DRRs using image denoising or style transfer techniques. In particular, recent advances in deep learning-based denoising methods, including both supervised models such as DnCNN and self-supervised approaches such as Noise2Void [38], appear promising, especially under scenarios where clean ground-truth images are unavailable.

However, extensive preliminary experiments revealed that these methods fail to effectively remove noise or reproduce the visual characteristics of real SyncTraX fluoroscopic images. As illustrated in Fig. 2.1, the outputs of both DnCNN and Noise2Void exhibit minimal visual differences from the original noisy inputs, indicating that the underlying noise characteristics cannot be adequately captured or suppressed by these models. This observation suggests that the noise present in real fluoroscopic images violates the fundamental assumptions commonly adopted by conventional denoising frameworks, such as pixel-wise independence and zero-mean noise distribution.

In particular, X-ray fluoroscopic noise acquired during radiotherapy is strongly influenced by high-energy photon scattering and detector-related effects, resulting in spatially correlated noise patterns and intensity-dependent degradation [39,40]. These characteristics render direct denoising or style transfer from real fluoroscopic images to DRRs ineffective. This empirical failure motivated a paradigm shift from direct image translation toward a Simulation-to-Real strategy, in which the degradation process itself is explicitly modeled and analyzed. Consequently, I believe that understanding and characterizing the statistical properties of noise and



blur in real fluoroscopic images becomes a necessary step for constructing realistic simulated training data.

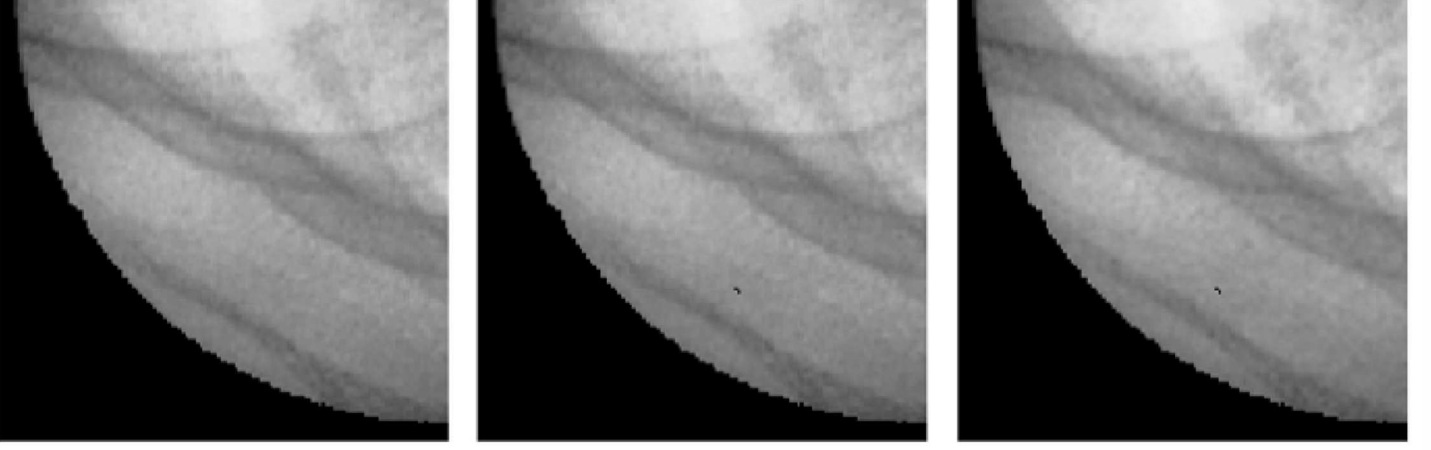


Figure 2.1: Comparison of deep learning-based denoising methods applied to real SyncTraX fluoroscopic images. From left to right: original noisy fluoroscopic image, DnCNN-denoised result, and Noise2Void-denoised result. Both supervised (DnCNN) and self-supervised (Noise2Void) approaches exhibit limited noise suppression, with minimal visual differences compared to the original image. This suggests that the noise characteristics in real fluoroscopic images cannot be effectively modeled by conventional denoising assumptions.

For the tumor segmentation model, minimizing the style discrepancy between the training images and the real images is expected to enhance tumor segmentation performance of the model, thereby benefiting tumor localization and tracking accuracy. Compared to real projection images, DRR images exhibit poorer sharpness and different contrast [7]. Furthermore, real fluoroscopic images contain significant noise and blur. This noise and blur primarily stem from two factors: firstly, the scattering of high-energy photons during propagation through tissue, causing the grayscale value of a pixel to be influenced not only by the line integral at that location but also by surrounding pixels, especially with the additional scatter from megavoltage X-rays during treatment; secondly, quantum noise and electronic noise inherent to the X-ray detector are unavoidable.

To better understand the style difference between real fluoroscopic images and DRRs, we captured multiple images of the same target, computed their average image as a clean reference, and derived noise in the images by subtracting individual frames from this average. Specifically, given a sequence of  $N$  fluoroscopic images  $\{I_1, I_2, \dots, I_N\}$ , the clean reference image  $I_{\text{avg}}$  is first computed by pixel-wise averaging:

$$I_{\text{avg}}(x, y) = \frac{1}{N} \sum_{k=1}^N I_k(x, y)$$



where  $(x, y)$  denotes the pixel coordinates. Subsequently, a set of noise-only images  $\{N_1, N_2, \dots, N_N\}$  is derived by subtracting this average from each individual frame:

$$N_k(x, y) = I_k(x, y) - I_{\text{avg}}(x, y)$$

The statistical heatmaps are then generated from this set of noise images. The noise probability,  $P_N(x, y)$ , is defined as the fraction of frames in which the absolute noise value  $|N_k(x, y)|$  exceeds a specific threshold  $\tau$ . The calculation is performed using the indicator function  $\mathbb{I}(\cdot)$ :

$$P_N(x, y) = \frac{1}{N} \sum_{k=1}^N \mathbb{I}(|N_k(x, y)| > \tau), \quad \text{where } \tau = 2$$

The selection of the threshold  $\tau$  is a critical step that directly influences the sensitivity of noise detection. The optimal value for this threshold is contingent upon the quality of the 'clean' reference image,  $I_{\text{avg}}$ . An ideal reference image would be entirely free of noise, allowing for a very low threshold. However, in practice, an image averaged from a finite number of frames always retains some residual structural or random noise. If this residual noise in  $I_{\text{avg}}$  is significant, i.e., the reference image itself appears grainy, a low threshold  $\tau$  would erroneously classify these baseline fluctuations as significant temporal noise events, leading to an overestimation of noise across the entire image. Therefore, a noisier reference image necessitates a higher threshold to avoid false positives. Here, the reference image was generated by averaging a large set of 1800 frames. This large sample size ensures a very high signal-to-noise ratio and effective suppression of random fluctuations, resulting in a very clean and stable baseline. Consequently, a relatively low threshold of  $\tau = 2$  was sufficient to capture genuine noise events without misinterpreting the baseline of the reference image.

The noise amplitude,  $A_N(x, y)$ , is quantified by the mean absolute deviation of the noise values across all frames:

$$A_N(x, y) = \frac{1}{N} \sum_{k=1}^N |N_k(x, y)|$$

This approach allows for a direct, empirical characterization of the noise's frequency and typical magnitude at each pixel location. To ensure that the target remained motionless between frames, we fabricated a gelatin cube and acquired 1800 X-ray fluoroscopic images under 100 kV, 63 mA, and 4 ms conditions using the SyncTraX system as illustrated in Figure 2.2 and Figure 2.3. The results of the statistical analysis of the subtracted noise images are shown in Figures 2.4 and 2.5. Our analysis revealed that the noise probability and amplitude in real X-ray fluoroscopic images are not globally random but exhibit systematic correlations with both pixel location and intensity values.

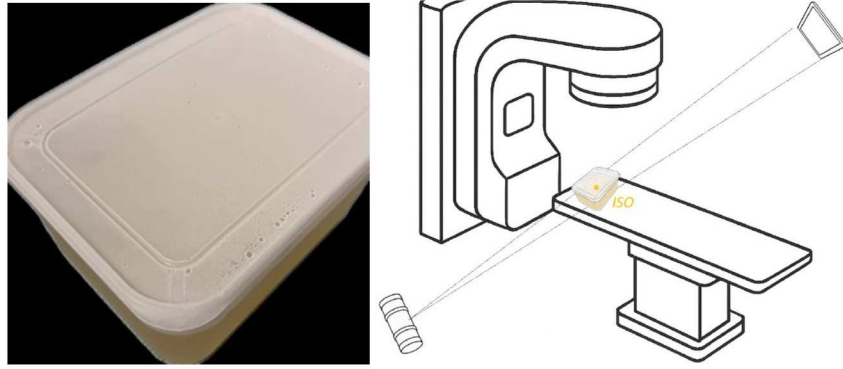


Figure 2.2: Gelatin cube preparation and image acquisition: (left) Container with gelatin inside; (right) Schematic diagram of image acquisition site, the gelatin container was placed at the ISO center.

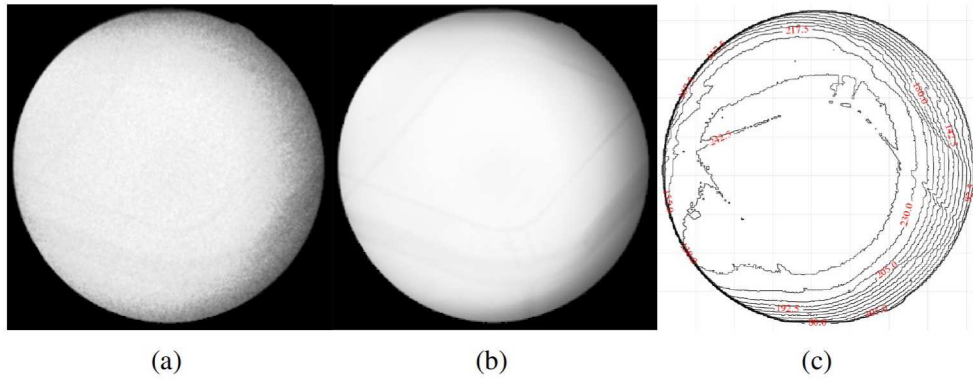


Figure 2.3: Captured X-ray images of gelatin cube. **(a)** An X-ray fluoroscopic image captured by the SyncTraX system, grayscaled using CorrC2G; **(b)** Clean image calculated by averaging 1800 images; **(c)** Brightness contour map of the average image. The brightness contour map **(c)** shows that the brightness center of the image is shifted slightly to the left and does not overlap with the center of the image, which indicates that the central axis of the X-ray beam is offset. Since the gelatin container was placed at the ISO center and its shape is not circular, it is unlikely that this brightness deviation is caused by the container.

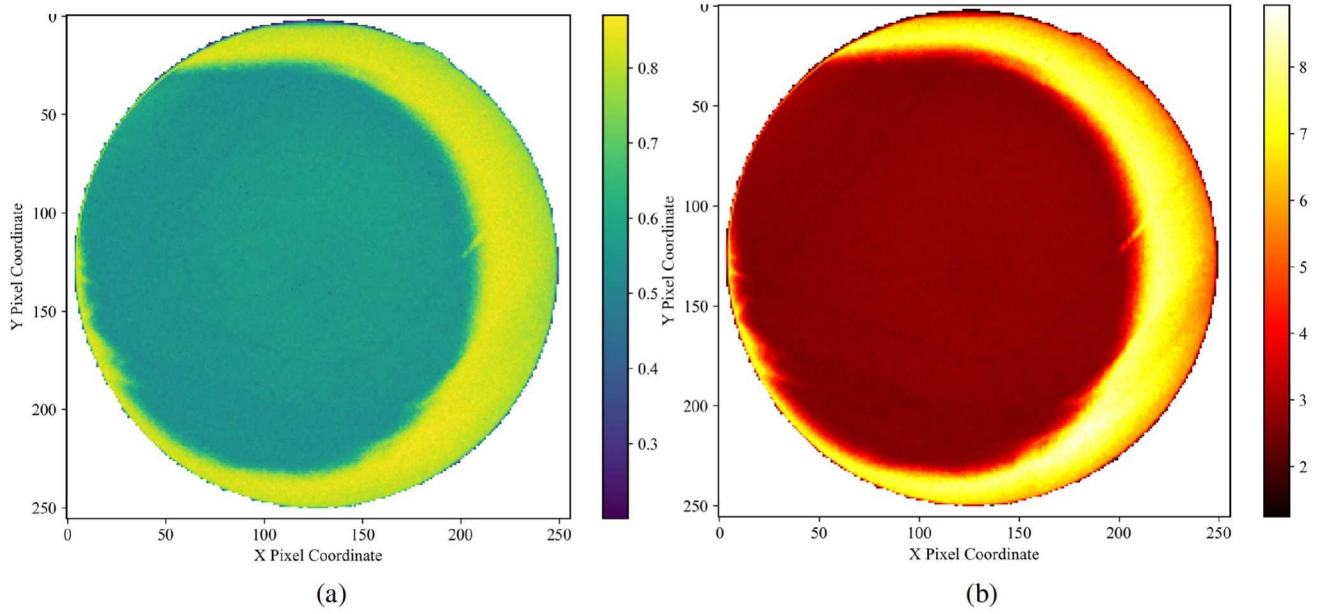


Figure 2.4: Noise characteristics in a gelatin cube fluoroscopic image (background pixels excluded). **(a)** Heatmap of degradation-induced noise probability and **(b)** amplitude, derived by computing the pixel-wise variance across 1800 subtracted images (individual frames minus the average of all frames) and normalizing it to a  $[0, 1]$  range. In **(a)**, the mean noise probability within the central circular region is approximately 0.56, while that in the right peripheral arc region reaches about 0.83. In **(b)**, the mean noise amplitude in the central circular region is approximately 2.7, increasing to around 7.2 in the right peripheral arc. Regions with lower grayscale values in the image, such as the cube's edges, exhibit lower noise probability and amplitude, and the farther the area is from the center of the X-ray beam (note: the beam center, not the image center), the higher the noise probability and amplitude. Importantly, the empirical profile is not strictly monotonic: noise probability and amplitude increase with radial distance from the beam center until approximately one radius, after which a decline is observed. This decline becomes noticeable primarily on the right-hand side of the image, due to a slight leftward shift of the X-ray beam center, which exposes more of the far-right region beyond the effective beam axis. Here, the radius refers specifically to the radius of the circular foreground region in the image.



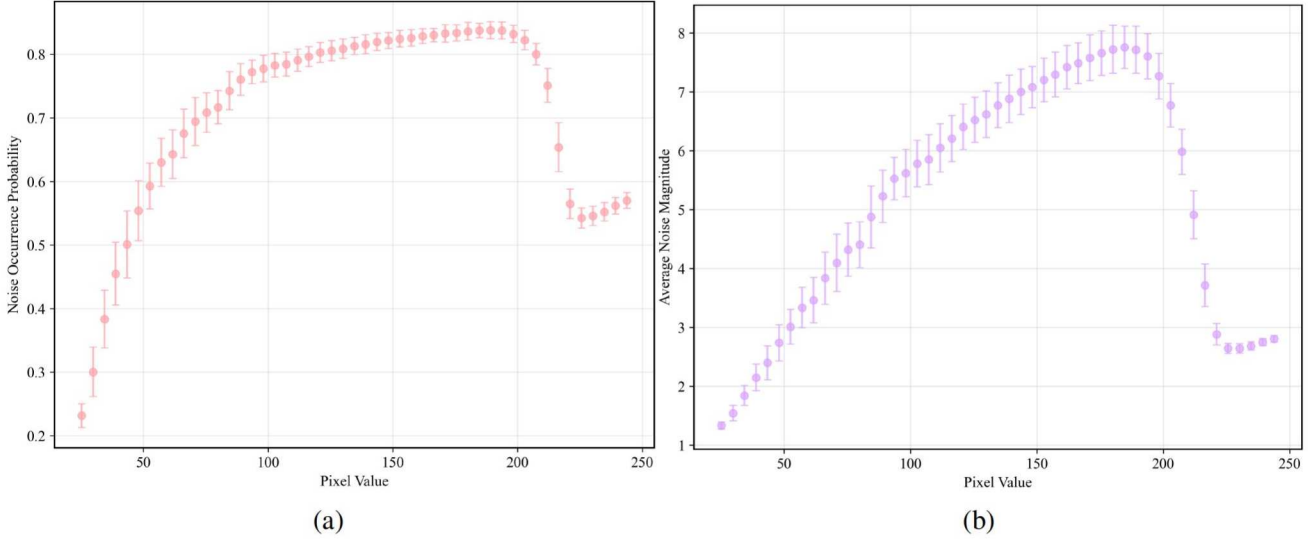


Figure 2.5: **(a)** Relationship between pixel grayscale value and noise probability. **(b)** Relationship between pixel grayscale value and noise amplitude. From a holistic perspective, the relationship can be broadly divided into four phases: rapid ascent phase (pixel values between 0 and 85), slow ascent phase (pixel values between 85 and 190, the average noise probability and amplitude are around 0.81 and 6.5, respectively), sharp decline phase (pixel values between 190 and 225), and recovery phase (pixel values between 225 and 255, the average noise probability and amplitude are around 0.56 and 2.75, respectively). The sharp decline phase is likely attributable to the clipping of high-amplitude noise fluctuations that exceed the upper grayscale limit of 255, while the recovery phase corresponds to the re-elevation of the overall noise level by low-amplitude noise. Given the presence of various types of noise [41, 42] in X-ray imaging—such as quantum noise, electronic noise and structural noise from the imaging device—which exhibit different frequencies and amplitudes, this may account for the observed four-phase relationship.

Simulating these physical characteristics enables us to generate simulated fluoroscopic images from DRRs that possess noise and blur similar to real fluoroscopic images. This allows us to use paired image data of DRRs and simulated images to train a neural network model, equipping it with the ability to transform real X-ray fluoroscopic images into the DRR image style. To this end, we designed a DRR degradation modeling method, primarily including a Gaussian mist scattering model and a content- and spatially-dependent noise model.

## 2.2 Degradation Modeling

To accurately simulate the complex fluoroscopic image degradation style observed in clinical settings, we propose a stochastic degradation model. This model explicitly accounts for both the non-uniform, signal-dependent noise and the inherent diffusion caused by physical and electronic effects. The degradation process is composed of several carefully designed stages to generate highly realistic simulated images. The key tunable parameters that control the behav-

ior of the proposed degradation model and potentially affect tumor segmentation and tracking performance are summarized in Table 2.1.

**Noise Characteristic Functions** The noise characteristics at any pixel  $\mathbf{p} = (x, y)$  are defined by the noise probability ( $\mathcal{N}_p$ ) and the noise amplitude ( $\mathcal{N}_a$ ) maps. These maps are calculated as follows:

$$\mathcal{N}_p(V, d) = \text{scale}(\mathcal{P}_{\max}, \mathcal{B}_p, f_{vp}(V) \cdot f_{sp}(d, X_c, \sigma_d)) \quad (2.1)$$

$$\mathcal{N}_a(V, d) = \text{scale}(\mathcal{M}_{\max}, \mathcal{B}_a, f_{va}(V) \cdot f_{sa}(d, X_c, \sigma_d)) \quad (2.2)$$

where  $V$  is the pixel intensity value.  $\mathcal{P}_{\max}$  and  $\mathcal{M}_{\max}$  are the upper limits for noise probability and amplitude, respectively, which are set to 0.92 and 16 based on experimental observations.  $f_{vp}(\cdot)$  and  $f_{va}(\cdot)$  are functions describing the content dependency of noise.  $f_{sp}(\cdot)$  and  $f_{sa}(\cdot)$  are functions describing the spatial dependency of noise.  $\sigma_d$  is the central flatness radius parameter.  $X_c = (x_c, y_c)$  denotes the random offset of the X-ray beam center  $\mathbf{d}_{\text{beam}}$  from the image geometric center  $\mathbf{d}_{\text{img}}$ :

$$X_c = \mathbf{d}_{\text{beam}} - \mathbf{d}_{\text{img}} \quad (2.3)$$

and  $d$  is the radial distance from the current pixel  $\mathbf{p}$  to the actual beam center:

$$d = \|\mathbf{p} - \mathbf{d}_{\text{img}} - X_c\|_2 \quad (2.4)$$

The function  $\text{scale}(\max, \min, \cdot)$  represents a linear mapping of the raw calculated noise characteristic to a target range defined by the base value ( $\mathcal{B}_p$  or  $\mathcal{B}_a$ ) and the maximum value ( $\mathcal{P}_{\max}$  or  $\mathcal{M}_{\max}$ ).

**Image Content Dependency** The content dependency functions  $f_{vp}(V)$  and  $f_{va}(V)$  capture the complex, multi-phase relationship between pixel intensity and noise level. To accurately represent the observed non-linear trends (e.g., rapid ascent, slow ascent, sharp decline, and recovery), we fit a 10th-order polynomial to curves smoothed from experimental data, as illustrated in Figure 2.6. In our implementation, the output of  $f_{va}(V)$  is multiplied by a factor of 2 to match empirical data.

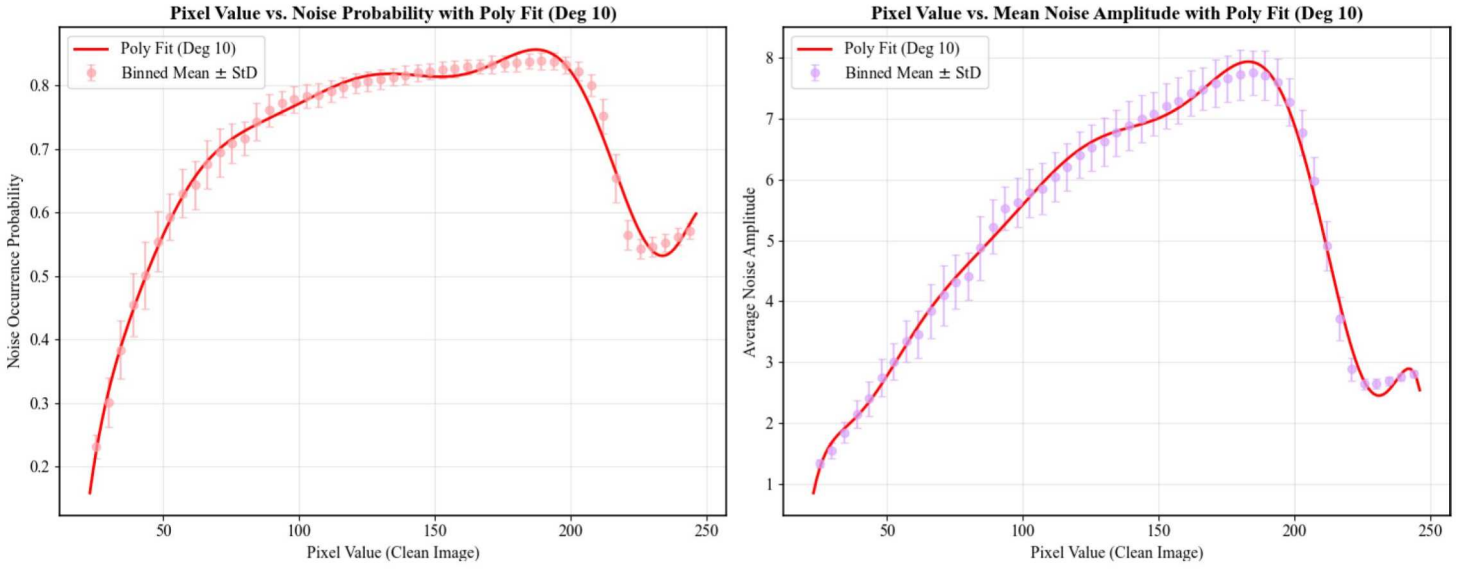


Figure 2.6: Fitting results using a 10th-order polynomial.

**Spatial Dependency** The spatial dependency term models the phenomenon of minimum noise at the beam center, which gradually increases towards the periphery. The spatial functions  $f_{sp}$  and  $f_{sa}$  are defined as:

$$f_{s\zeta}(d, X_c, \sigma_d) = \mathcal{B}_\zeta + \mathcal{F}_d(d, \sigma_d) \quad (2.5)$$

where  $\zeta \in \{p, a\}$ , and  $\mathcal{B}_\zeta$  is the base noise level (either probability  $\mathcal{B}_p$  or amplitude  $\mathcal{B}_a$ ), ensuring a baseline level of noise even at the center.  $\mathcal{F}_d(\cdot)$  is the core radial modification function:

$$\mathcal{F}_d(d, \sigma_d) = 1 - \exp\left(-\frac{d^2}{2\sigma_d^2}\right) \quad (2.6)$$

**Noise Generation and Diffusion** Instead of direct noise addition, our model employs a multi-step process to simulate the physical diffusion properties of noise, as detailed in Algorithm 1.



---

**Algorithm 1** Stochastic Noise Generation and Diffusion Process

---

Input: Original DRR image  $I_{DRR}$ . Output: Degraded Image  $I_{deg}$ .

---

Generate a uniform random matrix  $R \sim U(0, 1)$  of the same size as  $I_{DRR}$ .

Create a binary noise mask  $M$  where  $M(\mathbf{p}) = 1$  if  $R(\mathbf{p}) < \mathcal{N}_p(\mathbf{p})$ .

Initialize an empty noise array  $I_{noise\_init}$ .

Loop for each pixel  $\mathbf{p}$  where  $M(\mathbf{p}) = 1$ .

Sample value  $v \sim \mathcal{N}(\mu, \sigma = \mathcal{N}_a(\mathbf{p}))$  and assign to  $I_{noise\_init}(\mathbf{p})$ , where  $\mu$  is the global noise mean.

End loop.

Apply a Gaussian filter  $G(\cdot)$  with  $\sigma = 2.0$  to the sparse noise array:

$$I_{noise\_blur} = G(I_{noise\_init}).$$

Blend the original image with the blurred noise:

$$I_{blend} = I_{DRR} + I_{noise\_blur}.$$

Generate a secondary noise array  $I_{noise\_sec}$  by sampling from  $\mathcal{N}(\mu, \sigma = \mathcal{N}_a(\mathbf{p}))$  at locations defined by the mask  $M$ .

Add the secondary noise to the blended image:

$$I_{deg} = I_{blend} + I_{noise\_sec}.$$

Clip pixel values of  $I_{deg}$  to the range  $[0, 255]$  and convert to 8-bit integer.

Return  $I_{deg}$ .

---

**Stochastic Strategy for Generalization** Stochastic variables are introduced into the key parameters of the model as shown in Table 2.2. By resampling these stochastic variables during the generation of each degraded image, the simulated data achieves high diversity and is expected to simulate from different devices and acquisition settings, thereby enhancing the generalization performance of the Restormer model.

Table 2.2: Stochastic Strategy for Synthesizing Degraded Images

| Parameter                      | Random Strategy / Distribution                                                         | Physical Significance   |
|--------------------------------|----------------------------------------------------------------------------------------|-------------------------|
| $X_c$                          | $X_c \sim \mathcal{N}(\mathbf{0}, 30)$                                                 | Beam center offset      |
| $\sigma_d$                     | $\sigma_d \sim \mathcal{U}(103, 113)$                                                  | Central flatness radius |
| $\mathcal{B}_p, \mathcal{B}_m$ | $\mathcal{B}_p \sim \mathcal{U}(0.35, 0.37), \mathcal{B}_m \sim \mathcal{U}(1.8, 2.0)$ | Base noise levels       |

Table 2.1: Tunable parameters in the proposed degradation model and their impact on tumor tracking performance.

| Parameter                 | Symbol               | Typical Value | Influence on Tracking Performance                                                                                                      |
|---------------------------|----------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------|
| Maximum noise probability | $\mathcal{P}_{\max}$ | 0.92          | Controls the overall frequency of noise occurrence. Higher values increase image graininess and may degrade tumor boundary visibility. |
| Maximum noise amplitude   | $\mathcal{M}_{\max}$ | 16            | Determines the strength of noise fluctuations. Excessive amplitude can obscure tumor contours and reduce segmentation accuracy.        |
| Base noise probability    | $\mathcal{B}_p$      | Empirical     | Ensures baseline noise presence. Affects robustness to low-contrast regions near the beam center.                                      |
| Base noise amplitude      | $\mathcal{B}_a$      | Empirical     | Controls minimum noise level. Influences background texture realism and model generalization.                                          |
| Beam flatness radius      | $\sigma_d$           | Empirical     | Regulates spatial noise distribution. Smaller values emphasize peripheral degradation, affecting tumors located near image edges.      |
| Gaussian blur kernel      | $\sigma$             | 2.0           | Models spatial diffusion of noise. Larger kernels increase blur and may reduce fine tumor structure visibility.                        |
| Noise mean                | $\mu$                | 0             | Controls noise bias. Non-zero values may introduce systematic intensity shifts.                                                        |
| Polynomial order          | $N$                  | 10            | Determines flexibility of intensity-dependent noise modeling. Higher orders improve fit but risk overfitting noise statistics.         |

The noise probability and amplitude generated using our modeling method are illustrated in Figure 2.7.

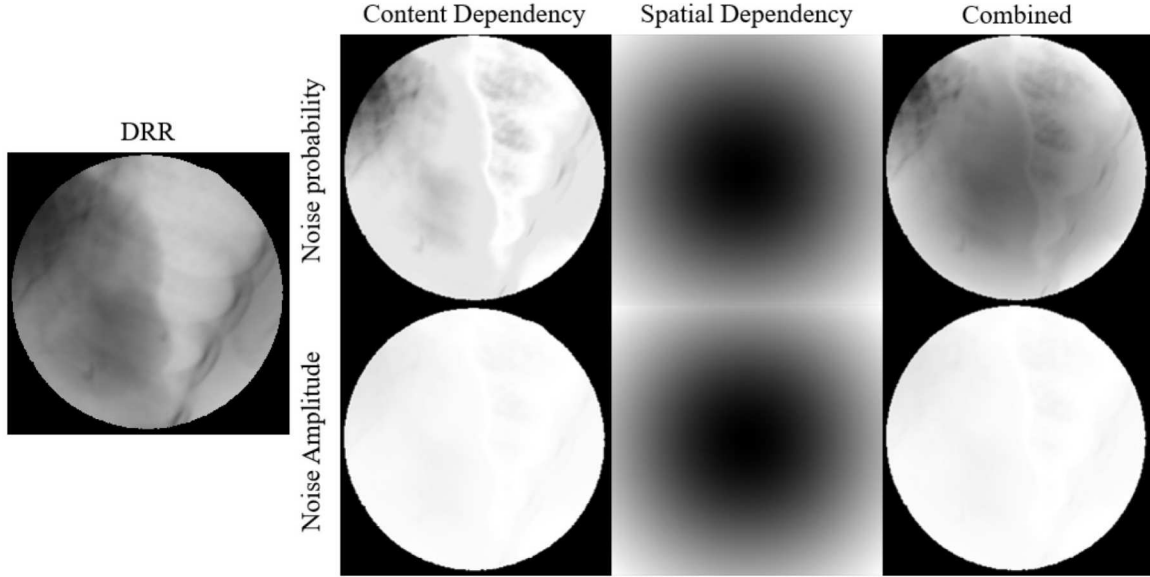


Figure 2.7: Visual representation of how pixel values and spatial position relationship affects noise probability and amplitude.

## 2.3 Results

Figure 2.8 shows the visual comparison using proposed image degradation model and commonly used image augmentation methods.

## 2.4 Conclusion

This chapter presented a comprehensive analysis of the style differences between real fluoroscopic images and synthetic DRRs, leading to the development of a sophisticated degradation model for simulating realistic fluoroscopic image characteristics. Through empirical analysis using gelatin phantoms, we characterized the non-uniform, content-dependent nature of noise in clinical imaging systems, revealing systematic correlations between noise characteristics and both pixel intensity values and spatial location relative to the X-ray beam center. The proposed stochastic degradation model successfully captures these complex relationships through content-dependent and spatially-aware noise functions, incorporating beam center offset and central flatness parameters to enhance realism.



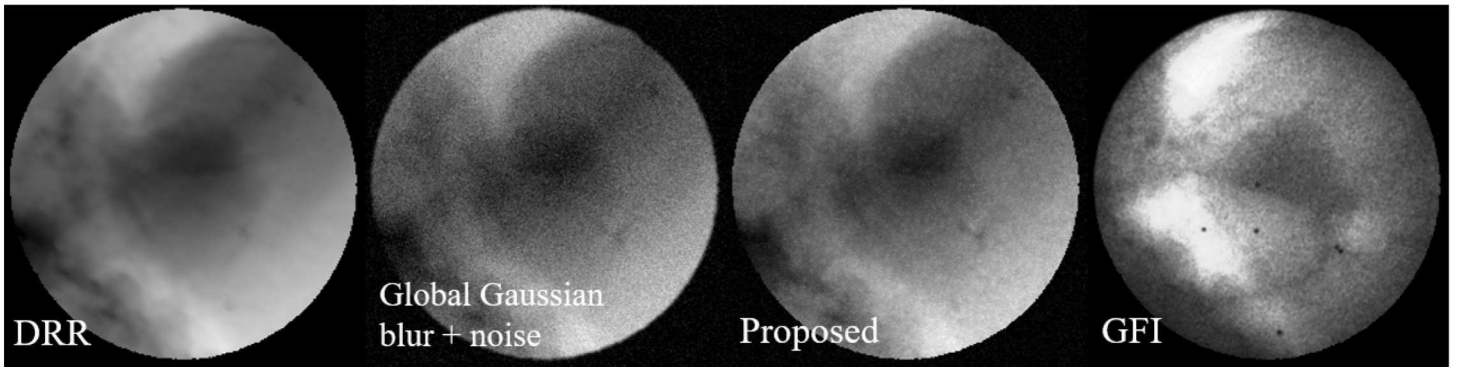


Figure 2.8: Visualization comparison of the image degradation model. From left to right: DRR, degraded image with Gaussian noise, degraded image using proposed method, and GFI. It can be seen that the degraded image generated by the proposed method can better simulate the noise pattern in GFI. It is notable that while the noise density is visibly reduced in the two highlighted areas of the GFI, the corresponding areas in our generated image, although also showing reduced noise, appear comparatively noisier. This discrepancy is because the corresponding areas in the source DRR are not as bright as those in the GFI, leading to a different noise expression after degradation.

## Chapter 3

### Workflow for Tumor Tracking

#### 3.1 Equipment and Data

The computer configuration used for algorithm development and model training is detailed in Table 3.1.

The real-time motion management platform for radiation therapy utilized in this chapter integrates a medical linear accelerator (TrueBeam, Varian Medical Systems, Palo Alto, CA, USA) with a real-time tumor tracking system (SyncTraX, Shimadzu Co., Japan). The SyncTraX system employs two pairs of kilovoltage (kV) X-ray tubes and color image intensifiers, mounted on ceiling and floor rails, respectively, enabling stereoscopic fluoroscopic imaging from various preset positions during treatment [43].

The dataset includes 4D planning CT data and CFIs from 7 lung cancer patients undergoing real-time tumor tracking radiotherapy. The tumors were located in the right lower lobe (1 case), right middle lobe (4 cases), and left upper lobe (2 cases). All patients had fiducial markers implanted; these markers served solely as a reference standard for evaluation and were not used as input for the tracking algorithm. The dataset comprises 50 pairs of consecutively acquired CFI sequences from seven different patients. Each pair consists of X-ray fluoroscopic images captured from two stereoscopic imaging viewpoints at a frequency of 30 frames per second,

Table 3.1: Computer Configuration

| Component  | Specification       |
|------------|---------------------|
| CPU        | Intel Core i9-9940X |
| Memory     | 64 GB               |
| GPU        | GeForce RTX 3090    |
| GPU Memory | 24 GB GDDR6X        |

resulting in a total of 100 CFI image sequences.

The use of patient data was approved by the Institutional Review Board of Yamaguchi University Hospital. Furthermore, written informed consent was waived due to the retrospective nature of this study. This retrospective study was conducted in accordance with the relevant guidelines and regulations outlined by the institution.

The 4D planning CT for each patient had a slice thickness of 2 mm and covered 10 respiratory phases. For all 4D-CT datasets in this study, inter-slice mean interpolation was performed to reduce the slice thickness from 2 mm to 1 mm, thereby enhancing the fidelity of subsequent DRR generation. The tumor contours delineated by doctors across all respiratory phases were also interpolated to obtain smooth volumetric labels.

Using the interpolated 4D-CT data, DRRs and corresponding tumor label maps were generated at couch angles ranging from  $0^\circ$  to  $355^\circ$  in  $5^\circ$  increments, resulting in 72 couch positions per respiratory phase. The resolution of each DRR-label pair was set to  $256 \times 256$  pixels. Tumor labels were generated using a threshold projection algorithm with a threshold of 7 voxels [31]. Subsequently, morphological opening and closing operations were applied using a  $10 \times 10$  elliptical structuring element to eliminate small isolated white regions and fill minor black voids. In this study, DRRs and tumor labels were used to train the segmentation network. DRR rendering adhered to the same imaging geometry as the SyncTraX system to ensure spatial consistency with CFI acquisition.

To ensure spatial consistency and facilitate training, background standardization was applied to all DRRs and CFIs [31]. Specifically, a Standard Background (SB) mask was extracted, where pixel values within a circular region were set to 1 and pixels outside this region were set to 0. The same SB mask was applied to CFIs, DRRs, tumor labels, and all neural network model outputs, ensuring identical background regions to improve network convergence speed.

### 3.2 Overview of Workflow

Our workflow sequentially processes the images as follows: First, intraoperative color fluoroscopic images are converted to grayscale using the CorrC2G algorithm [44]. Next, a transfer-learned Restormer model transforms the image into a clean, DRR-like style. Subsequently, a DUCK-Net model performs tumor segmentation, from which the tumor's central trajectory is extracted and smoothed via a low-pass filter (LPF). The complete pipeline is illustrated in Figure 3.1.



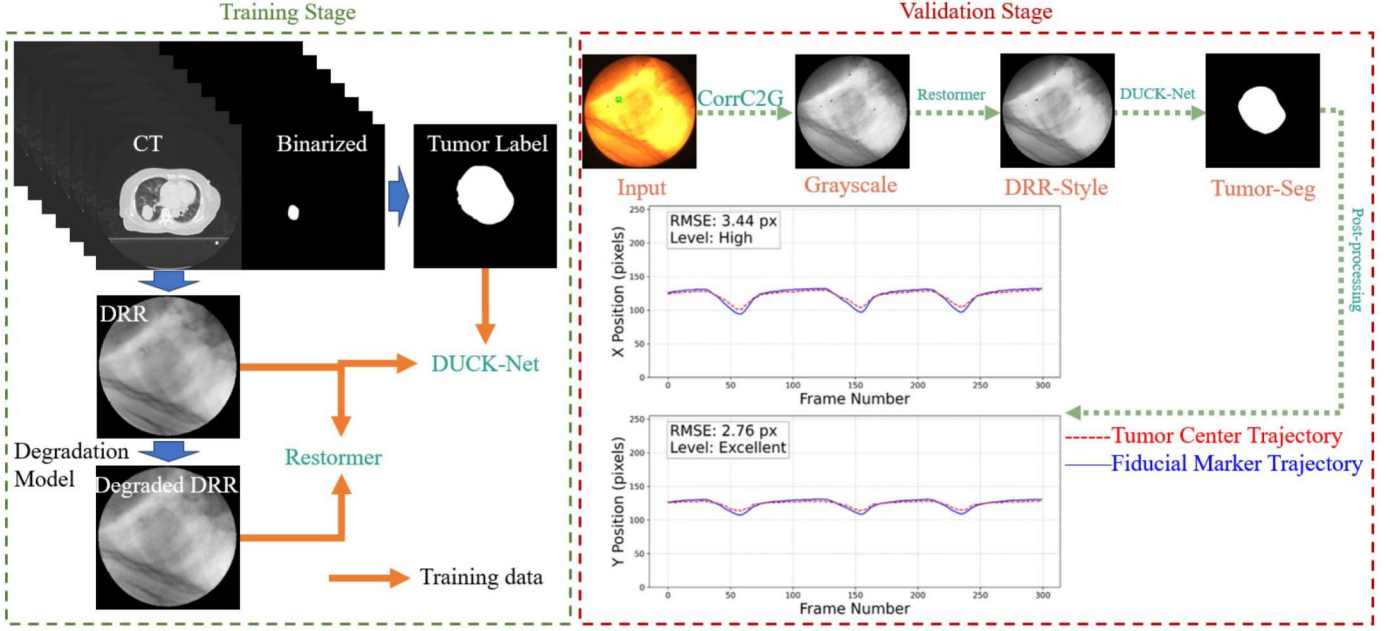


Figure 3.1: Workflow of the proposed pipeline. The green square from the input CFI represents the selected fiducial marker, which was implanted around the tumor to assist in determining the tumor location. In this study, we used the optical flow method to obtain the motion trajectory of the selected fiducial marker, which was used for comparison against the predicted tumor trajectory.

### 3.3 Algorithms and Neural Network Models

#### 3.3.1 Grayscale Conversion Algorithm

CorrC2G [44] is a correlation-based global color-to-grayscale conversion method. This method estimates three linear weighting parameters directly by calculating the Pearson correlation coefficients between each RGB channel and a contrast image. We selected CorrC2G for several reasons: It supports real-time processing without iterative optimization or training; Its computational complexity is controllable via a parameter  $r$ , which balances speed and quality. The average processing time was 3.62 ms per image in this study; CorrC2G provides two contrast mappings and automatically selects the best result via histogram analysis. It offers performance superior to most existing methods and produces natural-looking output images with preserved contrast, which aligns with human visual perception.

### 3.3.2 Transfer Learning of Restormer

To mitigate the domain gap between CFI and DRR, this study employs Restormer [45] for style transfer. Restormer is an efficient Transformer model specifically designed for high-resolution image restoration tasks. Its core innovation lies in the redesign of the traditional Transformer modules, namely the multi-head self-attention and feed-forward network, which significantly reduce computational complexity while retaining global modeling capability. The choice of Restormer is motivated by its several key advantages: it enables efficient high-resolution image processing through its multi-dconv head transposed attention module, which reduces computational cost by shifting attention computation from spatial to channel dimensions; it effectively integrates global and local contexts by combining the global modeling capability of the Transformer with local details captured through depthwise convolution, thereby enhancing restoration quality while preserving fine details; it exhibits superior cross-task generalization, having achieved state-of-the-art performance across multiple restoration tasks such as image deraining, deblurring, and denoising; it supports a progressive training strategy that improves the modeling of global image statistics; and its lightweight design, with fewer parameters and lower computational cost compared to other Transformer models like IPT and SwinIR, offers faster inference speed and higher practicality.

In order to fully leverage its prior knowledge, we initialize the Restormer model with publicly available Real Image Denoising (RealDenoising) pre-trained weights released by the authors.

The input consists of simulated fluoroscopic-style images generated using the degradation model proposed in Section 2.2, while the target output is the corresponding clean DRR. This paired design enforces the network to learn the mapping from degraded inputs to DRR-style outputs, thereby narrowing the gap between the two domains. We trained a baseline version by applying globally random Gaussian noise. The total training epochs are 345, except for the configurations mentioned in Table 3.2, all other training details for the Restormer model remain the same as the author's.

After fine-tuning, Restormer is integrated as part of the preprocessing pipeline, transforming CorrC2G converted grayscale fluoroscopic images into DRR-style images, which are then fed into DUCK-Net for tumor segmentation.



Table 3.2: Training Configurations

| Parameter                      | Original (Restormer)      | Ours                      |
|--------------------------------|---------------------------|---------------------------|
| Number of GPUs                 | 8                         | 1                         |
| Number of Workers per GPU      | 8                         | 4                         |
| Batch Size per GPU             | 8                         | 4                         |
| Mini Batch Sizes (Progressive) | [8, 5, 4, 2, 1, 1]        | [4, 3, 2, 2, 1, 1]        |
| Max Patch Size (GT Size)       | 384                       | 320                       |
| Patch Sizes (Progressive)      | [160, 192, 256, 320, 384] | [128, 160, 192, 256, 320] |
| Max Minibatch (Validation)     | 8                         | 4                         |

### 3.4 Tumor Segmentation Network

To segment the tumor region from the output of the restorer, we employed DUCK-Net [46], an encoder-decoder convolutional neural network specifically designed for medical image segmentation, particularly for colon polyp segmentation. DUCK-Net was chosen as the polyp segmentation model due to its combination of powerful multi-scale feature fusion, exceptional generalization capability, and practical efficiency. The model’s DUCK block captures both local details and global context through a parallel multi-path architecture, enabling it to effectively handle polyps of diverse sizes, shapes, and textures. Notably, DUCK-Net achieves state-of-the-art performance on multiple public datasets without relying on external pre-training, unlike many Transformer-based approaches. Its robustness is further enhanced by a strong data augmentation strategy implemented via the Albumentations library, which improves adaptability to real-world variations in lighting, angle, and scale. In addition, the model is computationally efficient, featuring fewer parameters and faster inference compared to Transformer-based alternatives, while its multi-path convolutional design offers strong interpretability, yielding well-defined segmentation masks with reduced false positives. DUCK-Net also demonstrates excellent generalization in cross-dataset tests, maintaining high accuracy even when trained and tested on data from different sources. This reliability under varying clinical conditions underscores its strong practical applicability, making it particularly suitable for lung tumor segmentation where tumor appearance and scale can vary significantly.

#### 3.4.1 Network Input and Configuration

Since the output from the previous Restormer stage is a grayscale image, the input to DUCK-Net is a single-channel image with a uniform size of  $256 \times 256$  pixels. The network outputs a binary tumor mask with the same spatial resolution as the input. The number of convolutional filters in the initial stage was set to 17.



### 3.4.2 Training Settings

The training was conducted for 700 epochs. The initial learning rate  $\eta_0$  was set to  $2 \times 10^{-3}$ , decaying by  $1 \times 10^{-5}$  after each epoch until reaching a minimum value of  $1 \times 10^{-5}$ :

$$\eta_{t+1} = \max(10^{-5}, \eta_t - 10^{-5}), \quad (t \geq 0). \quad (3.1)$$

Most parameter settings were consistent with those in the original DUCK-Net paper. Specifically, optimization was performed using the Adam optimizer with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and weight decay of  $10^{-4}$ . A fixed random seed was used when splitting training, validation and test datasets to ensure reproducibility. Data augmentation included horizontal and vertical flipping, random affine transformations (scaling factor 0.5–1.5, translation range  $\pm 12.5\%$ , rotation range  $\pm 180^\circ$ , shear range  $\pm 22.5^\circ$ ), and color jitter (brightness range 0.6–1.6, contrast range  $\pm 0.2$ , saturation range  $\pm 0.1$ , hue range  $\pm 0.01$ ). We additionally incorporated Gaussian blur and Gaussian noise to handle cases where Restormer failed to achieve complete DRR stylization. All augmentation operations were applied with a probability of 1.0, except for Gaussian blur and Gaussian noise, which were applied with a probability of 0.5. The loss function remained the same as in the original paper, using dice loss.

### 3.5 Trajectory Extraction and Post-processing Analysis

Due to the lack of ground truth tumor labels for evaluation of the output, we extracted the temporal trajectory of the tumor via centroid calculation and compared it with the motion trajectory of the fiducial markers to assess the tumor segmentation performance. For the segmentation result  $\Omega_t \subset \mathbb{R}^2$  at time  $t$ , the tumor center position  $(x_t, y_t)$  is defined as:

$$x_t = \frac{1}{|\Omega_t|} \sum_{(i,j) \in \Omega_t} i, \quad y_t = \frac{1}{|\Omega_t|} \sum_{(i,j) \in \Omega_t} j, \quad (3.2)$$

where  $|\Omega_t|$  denotes the number of tumor pixels at time  $t$ , and  $(i, j)$  represents the pixel coordinates. Using this method, we obtain the complete trajectory sequence:

$$\mathcal{T} = (x_t, y_t)_{t=1}^N, \quad (3.3)$$

where  $N$  is the number of frames in the sequence, in this study, images were collected at a rate of 30 frames, so the relationship between time  $t$  and  $N$  is:  $t = N/30$  s.

### 3.5.1 Low-pass Filtering

As the megavoltage (MV) level X-ray scatter noise during radiation therapy can affect tumor segmentation results in some frames, the extracted motion trajectory sequence often contains high-frequency jitter components. This study employs a fourth-order Butterworth low-pass filter to smooth the trajectory. This is very important for systems that determine whether the tumor has reached the beam output position based on detected coordinates. The digital filter was designed based on the following parameters: cutoff frequency  $f_c = 8$  Hz, sampling frequency  $f_s = 30$  Hz. Its transfer function in the continuous-time domain is expressed as:

$$H(\omega) = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}}, \quad (3.4)$$

where  $\omega_c = 2\pi f_c$  is the angular cutoff frequency and  $n = 4$  is the filter order. In the practical digital implementation, it was discretized using the bilinear transform, with the corresponding normalized cutoff frequency calculated as  $\omega_{\text{norm}} = f_c/(0.5 \times f_s)$ .

A global frequency analysis was performed on all reference marker trajectories to accurately characterize the frequency components of the motion signals. By applying the fast fourier transform, the frequency spectra and corresponding amplitudes were computed. The frequency and amplitude data from all trajectories were then aggregated to generate a frequency–amplitude scatter plot, as illustrated in Figure 3.2. The motion of the fiducial markers or the tumor center point is primarily influenced by the patient’s respiration and heart rate, with the frequency energy concentrated mainly below 2 Hz. This study selected a cutoff frequency of 8 Hz to filter the motion trajectory of the tumor center point.

The filtering operation was implemented using a zero-phase filtering method to avoid introducing phase distortion, ensuring that the filtered trajectory remains precisely aligned in time with the original data.

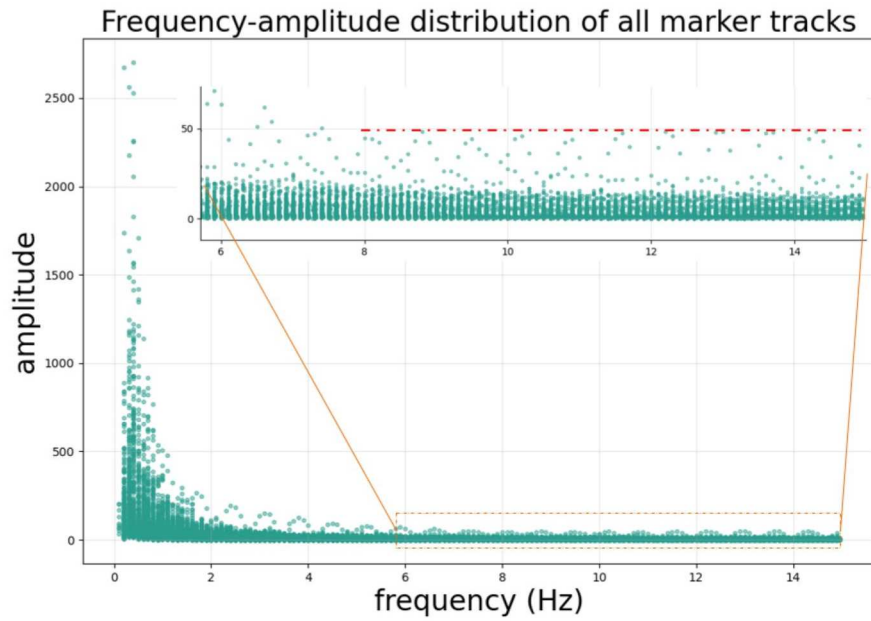


Figure 3.2: Frequency-amplitude distribution of the fiducial marker motion trajectory. Frequency analysis reveals the typical harmonic structure of respiratory motion, with a fundamental frequency of approximately 0.33 Hz (20 breaths/min) and multiple harmonics extending up to 15 Hz. Notably, the amplitude of harmonic components above 8 Hz remains stable at a low level (amplitude below 50 as indicated by the red dashed line). Based on this observation, we selected 8 Hz as the cutoff frequency to retain physiologically relevant motion components while filtering out trajectory jitter introduced by segmentation errors in individual frames, which corresponds to the high-frequency components.



### 3.5.2 3D Motion Trajectory Comparison

As introduced in the equipment section, the stereoscopic imaging system consists of two pairs of imaging devices acquiring two image sequences simultaneously. Since identifying and pairing the same fiducial marker across images from these two different imaging systems is challenging by visual inspection, we developed a method to determine the position of selected fiducial marker from images captured under different imaging positions. This paper proposes a fiducial marker dilation method for marker matching, illustrated in Figure 3.3. The method involves three steps: 1) A suitable fiducial marker is selected from the CT data. The selection principle is to choose one as close to the tumor as possible and sufficiently distant from other markers to avoid overlap or proximity issues in certain angles. 2) Dilating selected marker: set all voxels within a 5mm radius sphere centered on the marker's center voxel to 3000 Hounsfield Unit (HU). This process is akin to inflating the marker into a sphere with a diameter of 1 cm. 3) DRR Generation: DRRs using geometry parameters identical to those of the acquired X-ray fluoroscopic images are then generated.

In the generated DRR images at any angle, the selected marker region appears as a circular area, with its center indicating the marker's position. By comparing the DRR with the actual acquired fluoroscopic image, and using nearby bone structures, tissues, and the relative positions to other markers as references, the position of the selected fiducial marker in the real X-ray fluoroscopic image can be determined.

For markers that remain difficult to identify, one can either select a different marker or reduce the dilation volume for confirmation. Subsequently, optical flow is used to track the target fiducial marker. The 3D spatial motion trajectory is then calculated using the imaging geometry information and the marker's coordinates in the images.

To accurately compare the motion trajectories of the tumor center point and the fiducial marker, it is essential to analyze their relative movement while eliminating the constant spatial offset arising from their non-overlapping initial positions in the image plane. A direct comparison of the raw coordinates would introduce a large baseline error, even if their motion patterns were identical. Therefore, we aligned the trajectories by normalizing the coordinate system to the average position of the fiducial marker. This alignment removes the global displacement component caused by imaging geometry, ensuring that the comparison focuses solely on the dynamic motion relative to a common baseline.

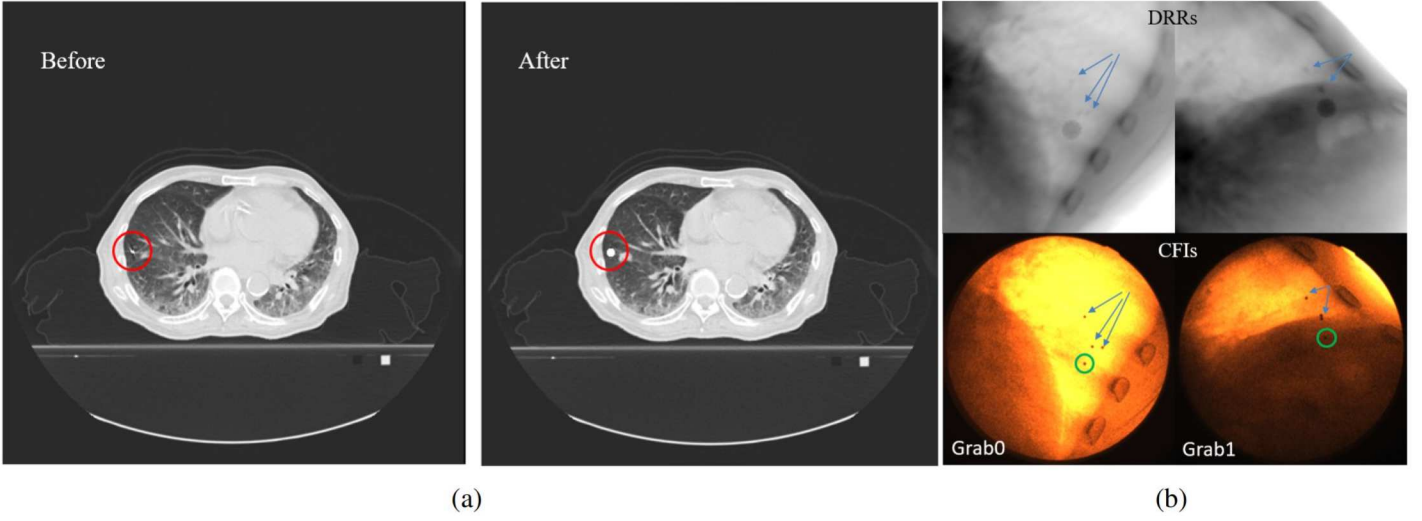


Figure 3.3: Fiducial marker expansion and matching method. (a): Illustration of the fiducial marker expansion process in CT data: the left panel shows the original CT slice with a red circle highlighting the target fiducial marker; the right panel shows the same slice after expanding the marker region into a spherical volume (5mm radius, 3000 HU). (b): Fiducial marker matching between DRR and clinical fluoroscopic image: the upper row shows DRR images generated with identical geometric parameters to the actual imaging setup; the lower row shows the corresponding clinically acquired CFIs. The green circles indicate the identified target fiducial marker in the CFIs. Arrows in both DRR and CFIs point to surrounding other fiducial markers, demonstrating successful spatial correspondence between the expanded marker region in DRRs and the actual marker position in CFIs.

The alignment is performed using the following equations:

$$x'_t = x_t - \left( \frac{1}{N} \sum_{t=1}^N x_t - \frac{1}{N} \sum_{t=1}^N x_t^m \right), \quad y'_t = y_t - \left( \frac{1}{N} \sum_{t=1}^N y_t - \frac{1}{N} \sum_{t=1}^N y_t^m \right). \quad (3.5)$$

In these equations,  $x_t$  and  $y_t$  represent the raw 2D image coordinates of the tumor center point at frame  $t$ , while  $x_t^m$  and  $y_t^m$  represent the corresponding coordinates of the fiducial marker. The total number of frames in the sequence is denoted by  $N$ . The terms  $\frac{1}{N} \sum_{t=1}^N x_t$  and  $\frac{1}{N} \sum_{t=1}^N y_t$  calculate the mean position of the tumor center point over the entire sequence, and similarly,  $\frac{1}{N} \sum_{t=1}^N x_t^m$  and  $\frac{1}{N} \sum_{t=1}^N y_t^m$  calculate the mean position of the fiducial marker, which serves as the reference baseline. The resulting coordinates  $x'_t$  and  $y'_t$  are the aligned coordinates of the tumor center point. The key operation involves subtracting the constant offset—the difference between the average positions of the tumor and the marker—which effectively centers the tumor's trajectory around the marker's average position.

The registered trajectory  $\mathcal{T}' = \{(x'_t, y'_t)\}$  is then filtered using a low-pass filter. The 3D motion trajectory of the tumor center point is calculated from this aligned 2D trajectory and compared frame-by-frame with the reference trajectory derived from the fiducial marker.



Table 3.3: Grading Standard Based on 2D Trajectory error

| Grade     | $E$ (pixels)    | Description                                 |
|-----------|-----------------|---------------------------------------------|
| Excellent | $< 3$           | Excellent performance in almost all frames. |
| High      | $3 \leq E < 8$  | Few frames are not accurate enough.         |
| Moderate  | $8 \leq E < 13$ | Some frames had poor results.               |
| Low       | $\geq 13$       | Poor accuracy in many frames.               |

### 3.5.3 2D Motion Trajectory Comparison

In addition to comparing 3D motion trajectories, we also performed 2D motion trajectory comparisons. This analysis focused on the in-plane motion of both the tumor center point and the fiducial marker, which was conducted to facilitate a direct comparison with a previous study [31]. Trajectory accuracy was assessed using the Root Mean Square Error (RMSE):

$$\text{RMSE}_x = \sqrt{\frac{1}{N} \sum_{t=1}^N (x'_t - x_t^m)^2}, \quad \text{RMSE}_y = \sqrt{\frac{1}{N} \sum_{t=1}^N (y'_t - y_t^m)^2}. \quad (3.6)$$

The final error  $E$  is taken as the maximum of the errors in the two directions:

$$E = \max(\text{RMSE}_x, \text{RMSE}_y). \quad (3.7)$$

A grading standard based on the final error  $E$  was established, as shown in Table 3.3.

## 3.6 Results

### 3.6.1 Image Grayscale and DRR Style Transfer Results

Figure 3.4 shows a comparison between the input CFI, the gray fluoroscopic image (GFI) processed by the CorrC2G algorithm, the output of Restormer (GFI\_DRR\_Style), and the DRR image generated under the same imaging geometry configuration as the CFI. It can be observed that the CorrC2G algorithm preserves the contrast information in the image. Visually, compared to the original grayscale fluoroscopic image, Restormer suppresses noise and blur in the GFI and brings the image style closer to that of the DRR image.

Figure 3.5 presents a visual comparison of denoising results, showing that the output im-



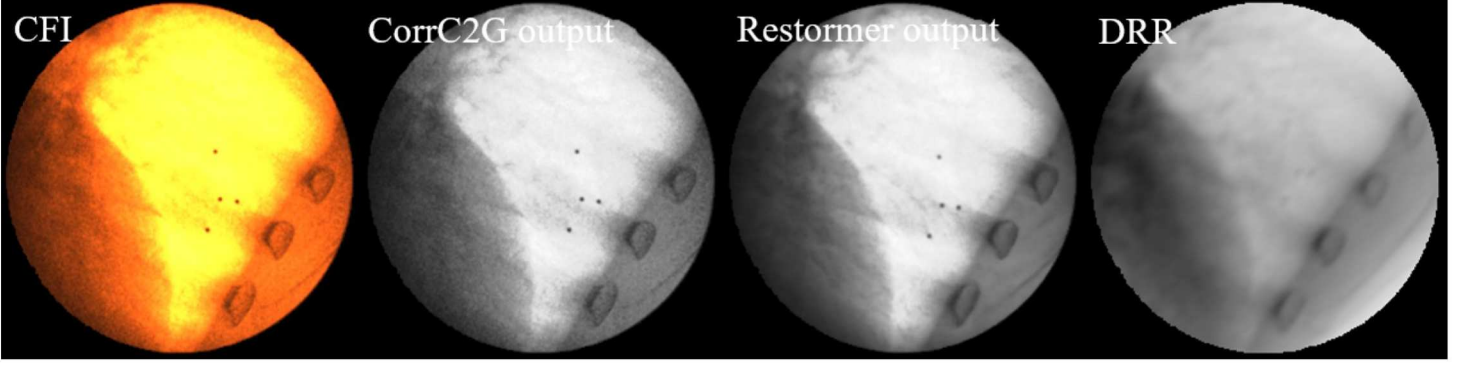


Figure 3.4: Comparison of image processing results at different stages. From left to right: Original CFI; CorrC2G output; Restormer output; DRR: generated with identical imaging geometry. The CorrC2G algorithm preserves contrast information while Restormer effectively suppresses noise and blur, producing image characteristics closer to those of DRR images.

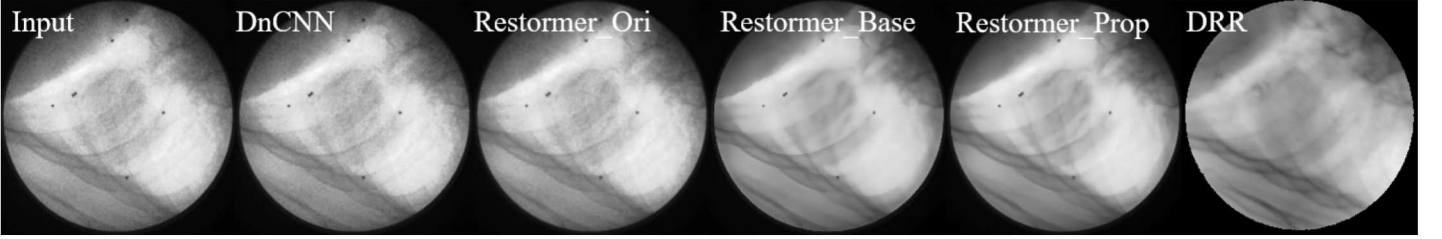


Figure 3.5: Visual comparison of denoising results using different methods on a sample X-ray fluoroscopic image. From left to right: Grayscale fluoroscopic image (Input); Output from pre-trained DnCNN; Output from pre-trained Restormer (Restormer\_Ori); Output from Restormer fine-tuned on baseline dataset (Restormer\_Base); Output from Restormer fine-tuned on our proposed dataset (Restormer\_Prop); Reference DRR. The results demonstrate that output from Restormer fine-tuned on our proposed dataset has better visual quality and closer resemblance to DRR-style images.

ages from the Restormer trained on our proposed dataset exhibit better visual quality and closer resemblance to real X-ray fluoroscopic images.

As highlighted in Figure 3.6, we conducted a focused comparison between the baseline method and our proposed approach. While both methods demonstrate excellent noise suppression capabilities, the Restormer fine-tuned on our proposed dataset preserves significantly more image details, leading to superior visual fidelity and more accurate representation of anatomical structures.

### 3.6.2 DUCK-Net Segmentation Results and Motion Trajectory Accuracy Assessment

For the 2D motion trajectory comparison, representative cases before trajectory post-processing from each grade are presented in Figure 3.7. In each figure, the left part displays the workflow

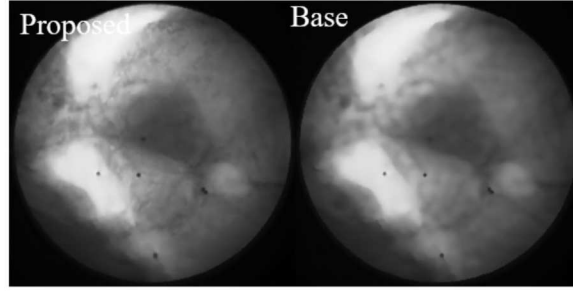


Figure 3.6: Focused comparison between baseline method and our proposed approach. While both methods demonstrate effective noise suppression, the Restormer fine-tuned on our proposed dataset (left) preserves significantly more image details compared to the baseline method (right). This enhanced detail preservation contributes to better visual quality and more accurate representation of anatomical structures.

input CFI with a green square showing the selected fiducial marker on the top and the corresponding tumor segmentation result at the bottom, and the right part shows the aligned x and y coordinate values over time for the tumor center and the fiducial marker within the image plane.

With LPF post-processing, the performance is further improved (average final error  $E$  dropped by 0.11) as illustrated in Figure 3.8.

Compared to the results of previous studies, the proportion of cases achieving *Excellent* and *High* grades with the workflow used in this study increased significantly, while the number of *Moderate* and *Low* grade cases substantially decreased as shown in Table 3.4. Table 3.5 shows the comparison of 3D tumor tracking accuracy between different methods.

Table 3.4: 2D trajectory evaluation comparison. All compared methods were evaluated on the exact same dataset.

| Grade     | $E$             | Method 1 <sup>1</sup> | Method 2 <sup>2</sup> | Method 3 <sup>3</sup> | Method 4 <sup>4</sup> | Method 5 <sup>5</sup> |
|-----------|-----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Excellent | (0, 3]          | 22.73%                | 50%                   | 48%                   | 50%                   | 51%                   |
| High      | (3, 8]          | 39.9%                 | 45%                   | 47%                   | 46%                   | 45%                   |
| Moderate  | (8, 13]         | 22.73%                | 4%                    | 4%                    | 3%                    | 3%                    |
| Low       | (13, $\infty$ ] | 14.64%                | 1%                    | 1%                    | 1%                    | 1%                    |

<sup>1</sup> Existing method [31].

<sup>2</sup> Omitting style conversion of proposed method.

<sup>3</sup> Restormer trained with baseline dataset of proposed method.

<sup>4</sup> Proposed method without LPF.

<sup>5</sup> Proposed method.

Partial visual results of the 3D motion trajectory using proposed whole framework are shown in Figure 3.9.

To analyze the effect of training data scale at the patient level, multiple DUCK-Net mod-



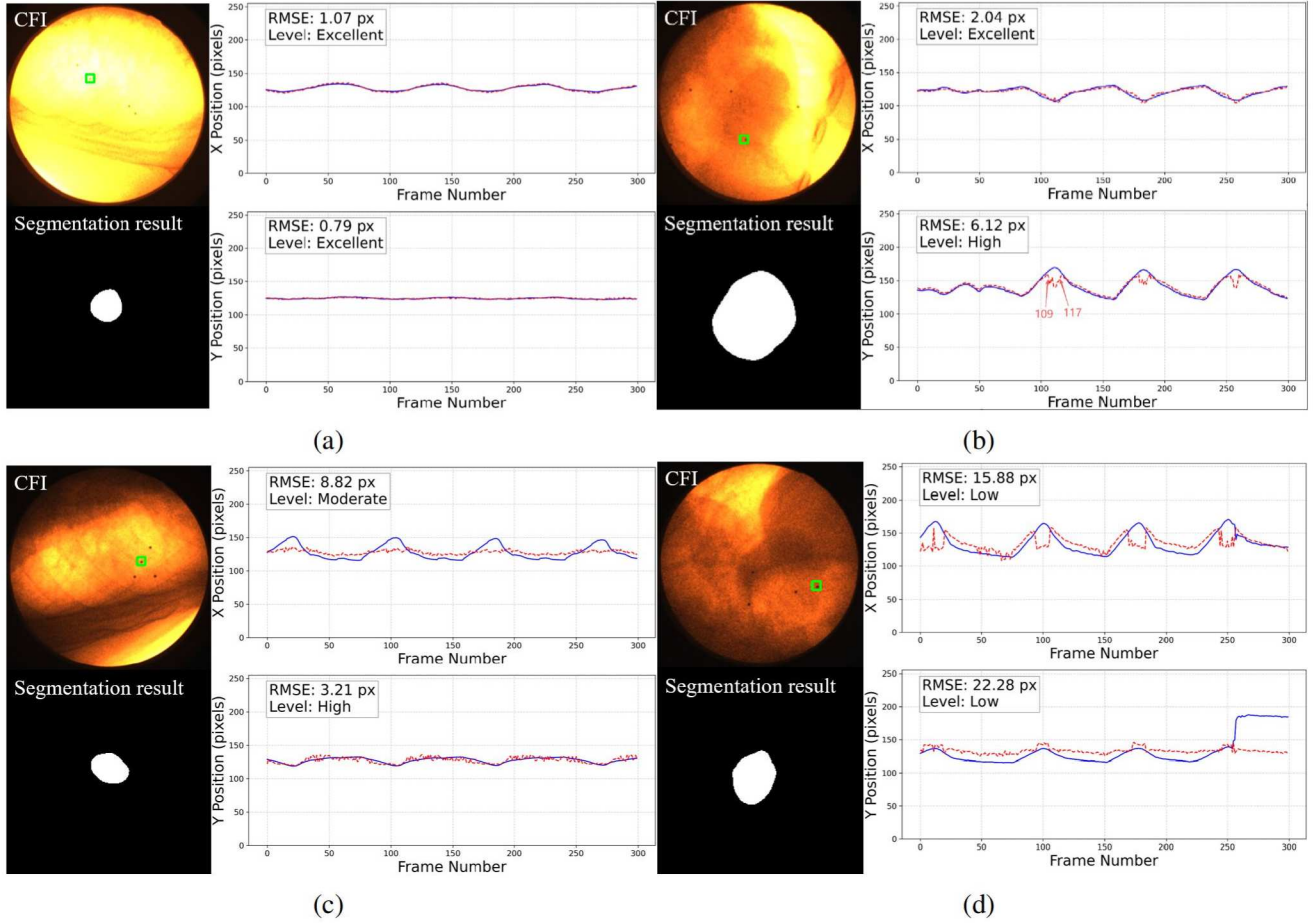


Figure 3.7: Representative cases of 2D motion trajectory comparison across different accuracy grades (LPF not applied). For each case: Left panel shows the workflow input CFI with green square indicating the selected fiducial marker (top) and corresponding tumor segmentation result (bottom); Right panel displays the aligned temporal trajectories of tumor center (red dotted line) and fiducial marker (blue line) in image coordinates (x and y directions). **(a)** Excellent cases (50%) have near-perfect trajectory alignment. **(b)** High accuracy cases (46%) show certain deviations in some frames (such as frame 109 to 117, we will discuss about it later) but the trajectory remains roughly aligned with the reference marker. This is common in cases where the image contrast is too poor, with severe noise and large tumor motion amplitude; **(c)** Moderate cases (3%) exhibit larger errors; **(d)** The only Low-graded case (1%) is caused by two factors: significant trajectory divergence in inspiratory frames and marker tracking error in the last 44 frames. The overall tumor detection rate is 100%.



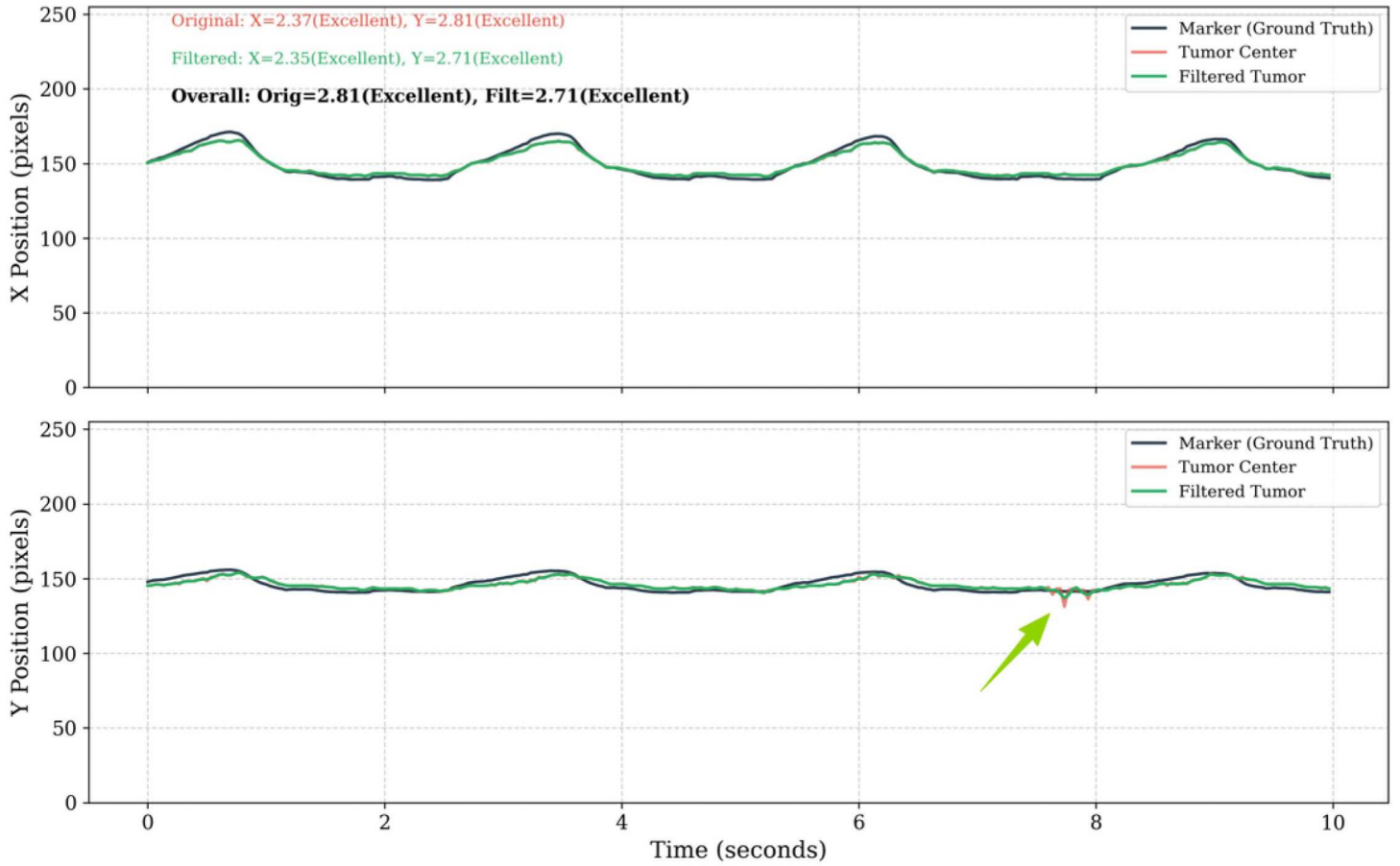


Figure 3.8: Comparison of filtered and unfiltered tumor trajectories. Unfiltered trajectory showing high-frequency jitter from segmentation error (red); Low-pass filtered trajectory (8 Hz cutoff) preserving respiratory motion pattern while suppressing the jitter. The trajectory using the low-pass filter deviates less at the points indicated by the arrow.

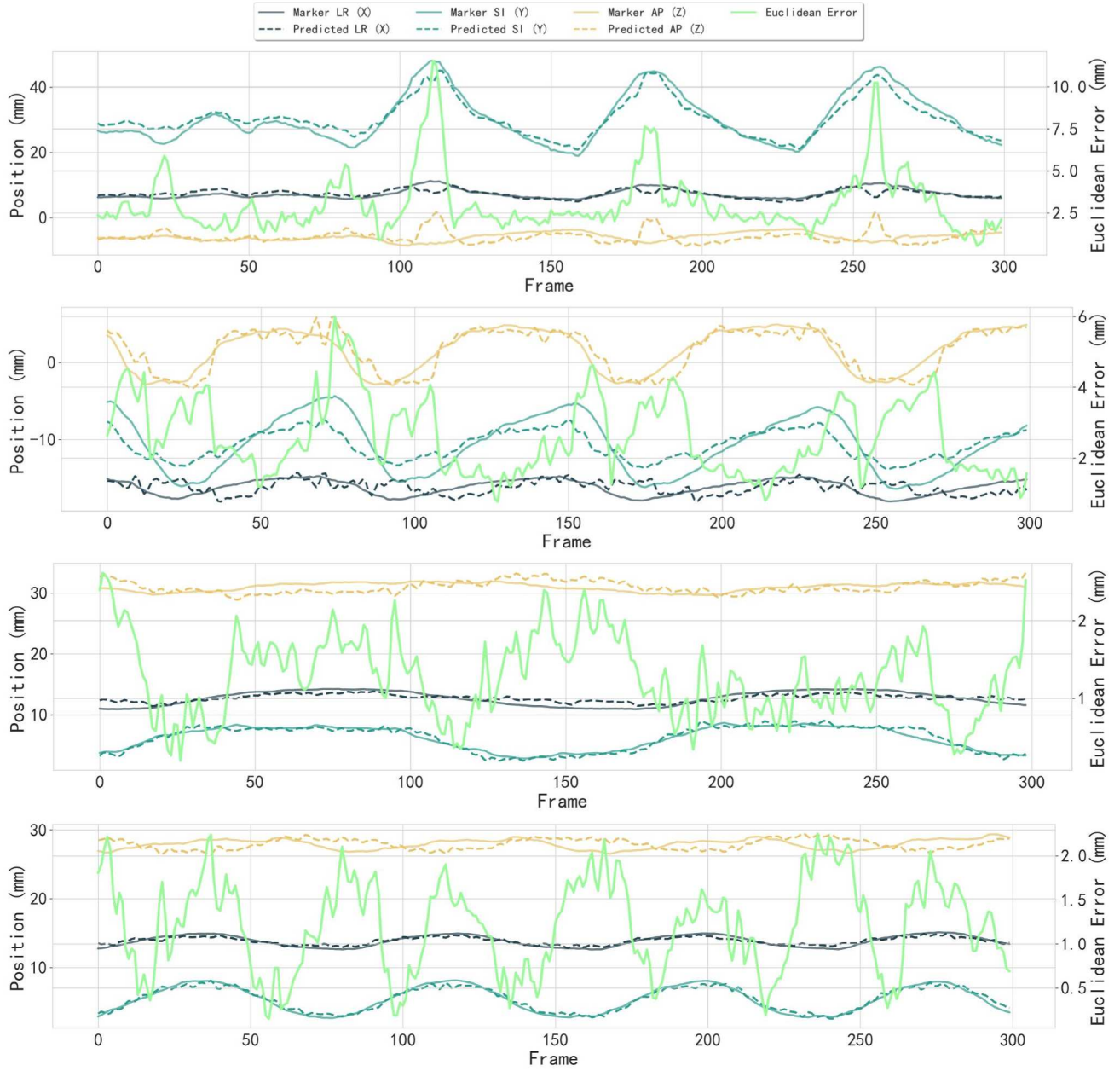


Figure 3.9: Partial results of 3D motion trajectory comparison between fiducial marker (reference) and tracked tumor center. Each subfigure shows: Individual trajectory components in LR (X), SI (Y), and AP (Z) directions with fiducial marker and detected tumor center; Euclidean distance error (bright green, right y-axis) between fiducial marker and tumor center across frames (x-axis). The median Euclidean error in all cases was 1.53 mm, with directional errors of  $0.98 \pm 0.70$  mm (LR),  $1.09 \pm 0.74$  mm (SI), and  $1.34 \pm 0.94$  mm (AP).

Table 3.5: 3D trajectory evaluation comparison.

|      | Method 1 <sup>1</sup> | Method 2 <sup>2</sup> | Method 3 <sup>3</sup> |
|------|-----------------------|-----------------------|-----------------------|
| LR   | $0.98 \pm 0.67mm$     | $1.00 \pm 0.69mm$     | $0.98 \pm 0.70mm$     |
| SI   | $1.14 \pm 0.76mm$     | $1.15 \pm 0.77mm$     | $1.09 \pm 0.74mm$     |
| AP   | $1.34 \pm 0.93mm$     | $1.35 \pm 0.91mm$     | $1.34 \pm 0.94mm$     |
| MEE* | 1.62 mm               | 1.64 mm               | 1.53 mm               |
| IQR* | 1.04 mm               | 0.99 mm               | 1.04 mm               |

\*MEE: Median euclidean error.

\*IQR: Interquartile Range.

<sup>1</sup> Omitting style conversion of proposed method.

<sup>2</sup> Restormer trained with baseline dataset of proposed method.

<sup>3</sup> Proposed method.

els were trained from scratch using different subsets of the available patient data. The number of training patients was gradually increased from one to seven. To ensure fair comparison, all trained models were evaluated on the same test set consisting of all seven patients. Figure 3.10 illustrates the relationship between the number of patients used for training and the resulting segmentation performance. A clear positive correlation can be observed, with Dice scores improving substantially as more patient data are included. When trained with only one or two patients, the model shows limited generalization capability, indicating high sensitivity to inter-patient variability. As the number of training patients increases, segmentation performance improves steadily, suggesting that additional patient diversity enables the model to learn more robust and transferable tumor representations. Notably, the performance gain becomes marginal beyond six training patients, indicating a saturation effect under the current dataset size and network capacity.



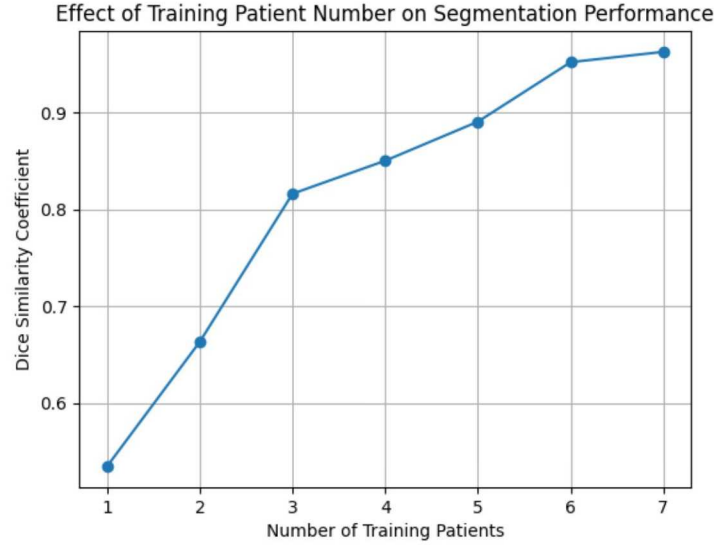


Figure 3.10: Relationship between the number of training patients and tumor segmentation performance. DUCK-Net was trained using data from an increasing number of patients (from 1 to 7), and each trained model was evaluated on the full dataset consisting of all seven patients. Segmentation accuracy was quantified using the Dice similarity coefficient.

### 3.7 Discussion

In the analysis of frames with discrepancies between the tracking trajectory and the reference markers, we observed that these frames are typically accompanied by segmentation errors. The main contributing factors can be summarized as follows:

**1. Abrupt variations in brightness and contrast:** As shown in Figure 3.11, the left panel presents a trajectory rated as *High*, where the error significantly increases around frame 40. The right panel displays the original images and segmentation results from frames 28–45, where noticeable fluctuations in brightness and contrast occurred between frames 40–43, leading to reduced segmentation accuracy. Such variations are likely caused by the interference of MV treatment beams and scattered radiation.

**2. Tumor motion-induced occlusion and contrast fluctuation:**

- When the tumor moves close to adjacent structures such as the liver, the boundary becomes severely blurred, resulting in decreased segmentation accuracy. As illustrated in Figure 3.12, the trajectory of this sequence is shown in Figure 3.7 (b). During frames 109–117, corresponding to the end of inhalation, the tumor enters the liver region and becomes partially occluded.

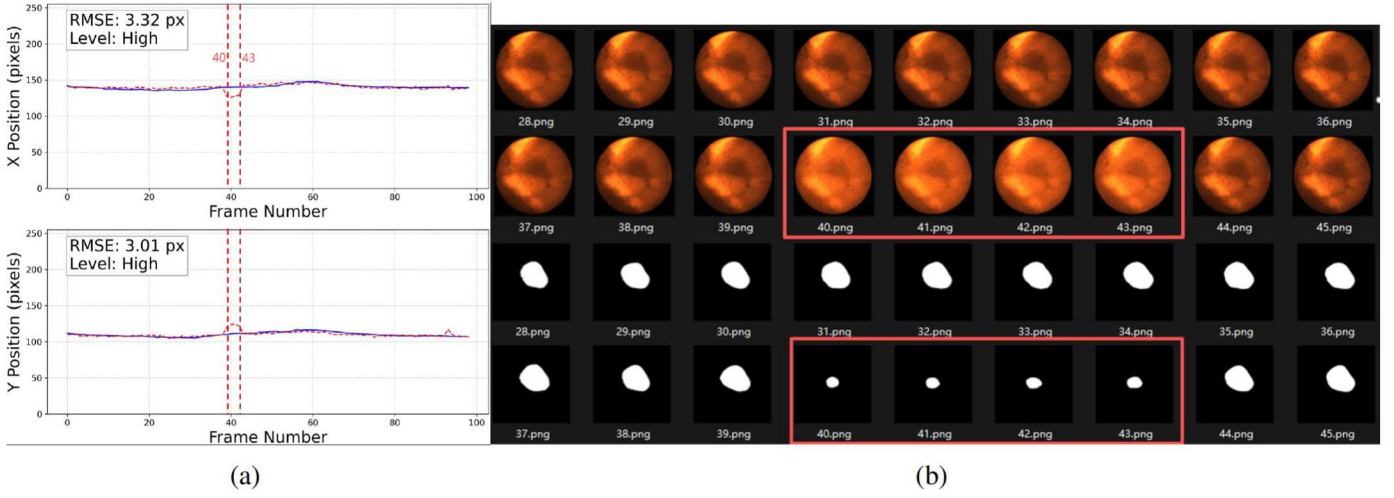


Figure 3.11: Abrupt variations in brightness and contrast. The left figure (a) shows a trajectory rated as *High*, the solid blue line represents the baseline trajectory, the dashed red line represents the predicted trajectory of the tumor center point. The tracking error markedly increases from frame 40 to 43 (as illustrated by the vertical dashed red lines). The right figure (b) presents the original CFIs and segmentation results, highlighting frames 40–43 (inside the red boxes) with strong brightness and contrast fluctuations. Note that the trajectory plot in (a) starts from frame 0, whereas the image filenames start from frame 1; therefore, frame 40 corresponds to position 39 in the trajectory plot.

- At the end of inhalation, the lung volume expands, and because lung tissue has weaker absorption of X-rays, the overall image contrast decreases. This effect is further amplified during grayscale conversion and style transfer, leading to degraded segmentation performance (Figure 3.13).

Since both conditions often occur simultaneously at the end phases of inhalation, their combined influence further increases segmentation errors in some cases. In this study, six image sequences were affected by this effect, resulting in significant deviations in the maximum inspiratory stages, which in turn increased the predicted trajectory error to an  $E > 5$  pixels.

**3. Severe noise compounded with low contrast:** When the image suffers simultaneously from strong noise and insufficient contrast, the Restormer fails to restore sufficient detail, resulting in significantly degraded segmentation accuracy. As shown in Figure 3.7 (d), this sequence is affected not only by tumor motion but also by severe noise and low contrast, which exacerbate the overall error.

Our initial cohort included 10 patients, 3 of these patients were excluded from this study because our method is not suitable for tumors severely obscured by the liver, when the tumor is located in the lower lobe of the lung and obscured by tissues such as the liver, making its position difficult to discern even for the human eye. The accuracy of marker-less tracking based on tumor segmentation is insufficient for such cases. For these scenarios, alternative approaches such as

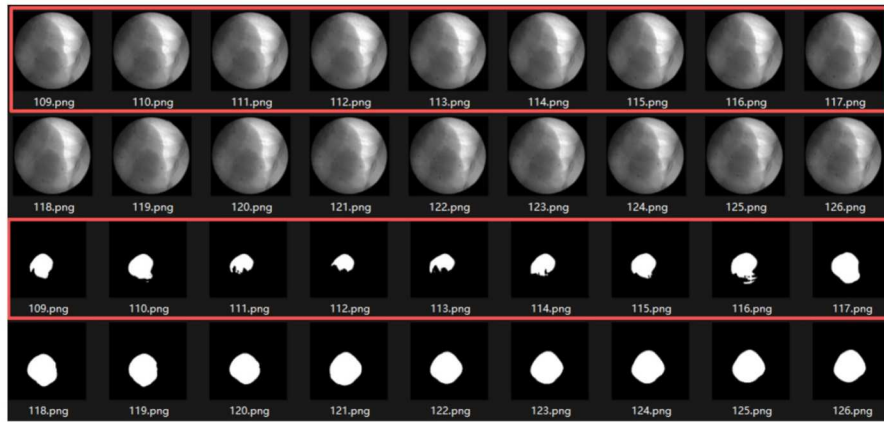


Figure 3.12: Motion-induced occlusion. During frames 109–117 (end of inhalation, inside the red boxes)), the tumor enters the liver region and becomes partially occluded, resulting in blurred boundaries and decreased segmentation accuracy.

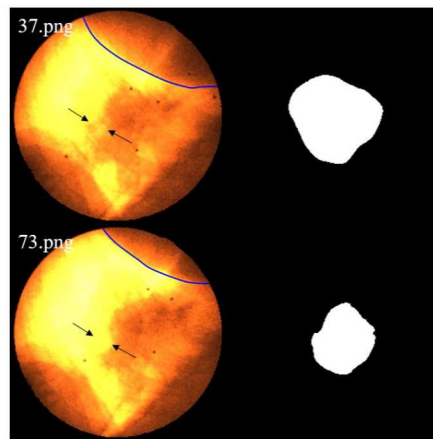


Figure 3.13: As lung expansion increases the proportion of lung regions in the image, the relatively low X-ray absorption of lung tissue leads to blurring tumor-background boundaries and consequently degrades the segmentation performance. Image 37.png corresponds to the end-expiration phase, while 73.png shows the end-inspiration phase. The blue contour delineates the liver boundary. A marked reduction in liver area and a concurrent expansion of the pulmonary region are observed in 73.png. Additionally, the black arrows indicate regions where edges become blurred due to contrast variation.



regression-based motion detection using ROIs (e.g., as proposed by Hirai et al. [30]) or keypoint detection methods are recommended, as they do not rely on direct tumor contour segmentation. For tumor segmentation-driven approaches, enhancing tracking robustness and resolution in cases without severe obstructions like the liver could be a key pathway to improving clinical applicability.

The average processing time of our method is 179.8 ms per image, although it is smaller than the total system latency tolerance of 500 ms for real-time tumor tracking recommended by AAPM TG-76. However, it remains insufficient for the 30 fps requirement of the SyncTraX system. This is primarily due to the relatively high computational complexity of the Restormer and DUCK-Net models.

Although visually our method does reduce the style differences between X-ray and DRR images, in this study, image style transfer improved tumor tracking accuracy less than expected. To quantitatively analyze this trade-off, we compared the end-to-end performance of the workflow with and without the style transfer module. As summarized in Table 3.6, the results clearly show that while including style transfer improves tracking accuracy by approximately 5.6% in median Euclidean error, it increases the processing time by a substantial 76.6%.

Table 3.6: Performance trade-off analysis of the style transfer module.

| Workflow Configuration         | MEE (mm)                            | Processing Time (ms/frame)     |
|--------------------------------|-------------------------------------|--------------------------------|
| Without Style Transfer         | 1.62                                | 101.8                          |
| With Style Transfer (Proposed) | 1.53                                | 179.8                          |
| <b>Change</b>                  | <b>5.6% performance improvement</b> | <b>Time increased by 76.6%</b> |

This limited accuracy gain is primarily due to the inclusion of random noise and blur during DUCK-Net model training, which improves the model’s resilience to style differences. Our analysis indicates that for practical implementation, the substantial increase in processing time introduced by the style transfer step may outweigh its modest contribution to accuracy.

From a clinical and system-design perspective, the target accuracy and computational speed of a tumor tracking system must be jointly considered under explicit safety constraints. In radiotherapy, a commonly accepted geometric error tolerance is 2.0 mm, beyond which tracking errors may compromise treatment safety. Therefore, when the baseline tracking accuracy exceeds this threshold, improving accuracy must be prioritized, even at the cost of increased computational latency. In such cases, the inclusion of computationally expensive modules, such as image style transfer, can be justified.

However, once the tracking accuracy is consistently within the clinically acceptable safety margin, further improvements in segmentation or tracking accuracy yield diminishing clinical

benefit. In this regime, processing speed becomes the dominant factor, as excessive system latency can itself introduce additional geometric errors due to continuous tumor motion during inference. Our results indicate that although the style transfer module improves tracking accuracy by approximately 5.6%, it increases the processing time by 76.6%, which may lead to latency-induced errors that offset its accuracy gain. Consequently, when operating within the clinical safety zone, a streamlined workflow that omits the style transfer step and directly applies DUCK-Net represents a more practical solution. This trade-off highlights the importance of balancing accuracy and computational efficiency based on clinically meaningful thresholds rather than pursuing marginal accuracy improvements in isolation.

Therefore, a streamlined workflow using DUCK-Net directly is a viable alternative. In future optimizations, adopting a more diverse data augmentation strategy for DUCK-Net could potentially further narrow the performance gap with the Restormer-based workflow, or even eliminate the need for a separate style transfer step. However, aggressive or poorly designed augmentations might introduce excessive stochasticity and noise, potentially causing the model to learn non-semantic features and thereby diminishing its ability to learn key anatomical features [47]. We therefore believe a targeted augmentation strategy could be more effective. For instance, introducing augmentations during training that are specifically designed to simulate the failure cases identified in our discussion—such as abrupt brightness changes from MV scatter, temporary tumor-liver occlusion, and low contrast at full inhalation—could be highly beneficial. This remains a critical and promising avenue for future optimization.

A clarification on the evaluation methodology is warranted. Our approach of aligning the mean positions of the predicted and ground-truth trajectories, while standard for comparing motion patterns, effectively cancels out systematic offsets. This means our reported accuracy primarily reflects the fidelity of tracking relative motion, rather than absolute spatial localization, and any constant positional bias would not be captured by this metric, which could be critical for beam gating. The scale of the dataset, which comprises nearly 30,000 images, renders manual tumor segmentation infeasible. Therefore, our study adopts the use of fiducial markers as a reference, which is a standard and practical methodology for validation in marker-less tumor tracking field.

In the absence of implanted fiducial markers, direct point-wise ground-truth error measurement is not feasible for markerless tumor tracking. Nevertheless, the applicability of the proposed system can be evaluated through clinically relevant indirect validation. Specifically, the temporal consistency and physical plausibility of the tracked tumor trajectory can be assessed, as respiratory-induced tumor motion is known to be smooth and quasi-periodic. Metrics such as trajectory smoothness, velocity continuity, and consistency with respiratory surrogates



(e.g., diaphragm motion or external breathing signals) provide objective indicators of tracking reliability. In addition, robustness across different patients, imaging conditions, and tumor locations can be analyzed to ensure generalizability.

Similar indirect validation strategies have been adopted in clinically deployed commercial markerless tracking systems. For example, in the Elekta Unity 1.5T MR-Linac system, real-time cine MRI is used to directly visualize the tumor or surrounding soft-tissue surrogates without implanted markers. Instead of point-wise ground truth comparison, tracking reliability is evaluated through the temporal stability of the tracked contour, the consistency between predicted and observed tumor positions during respiratory motion [48].

Similarly, the internal tumor motion is inferred from multi-angle kV imaging combined with a continuously updated respiratory motion model in Accuray Radixact Synchrony system without fiducial markers [49]. Its system performance is evaluated using a combination of model stability analysis, motion-compensated dosimetric quality assurance, and gamma analysis under dynamic conditions, rather than direct point-wise tumor localization error. Both phantom studies and early clinical outcomes have demonstrated that physically consistent motion modeling and stable tracking behavior are sufficient to ensure clinically acceptable dose delivery.

These established commercial systems demonstrate that, in the absence of implanted fiducial markers, markerless tumor tracking can be justified through indirect yet clinically meaningful validation criteria, including motion regularity, trajectory smoothness, surrogate consistency, and robustness across patients and imaging conditions. Following the same rationale, the proposed system can be reasonably evaluated and justified for future lung cancer patients without fiducial implantation. To further position the proposed approach within the landscape of existing markerless tracking solutions, Table 3.7 summarizes and compares representative clinically deployed systems with the proposed method, highlighting their respective strengths and practical constraints.

Future research could be expanded in several meaningful directions: 1. Incorporate of multi-center validation to assess robustness against variations in imaging protocols and equipment. This would enhance confidence in its broader clinical applicability. 2. Real-time optimization through model lightweighting or the use of more powerful GPUs to accelerate inference speed. 3. Multi-modal integration by incorporating signals from respiratory monitoring systems or other tumor tracking sources to improve robustness via multi-source information fusion, as well as clinical workflow integration that embeds the method into radiotherapy treatment planning systems and online image navigation platforms to develop a readily usable auxiliary localization tool.



Table 3.7: Comparison of representative markerless tumor tracking approaches and the proposed method

| Method                     | Imaging modality              | Strengths                                                                                                | Limitations                                                                                                             | Research accessibility |
|----------------------------|-------------------------------|----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|------------------------|
| Elekta Unity MR-Linac      | Real-time cine MRI (1.5T)     | Excellent soft-tissue contrast; direct visualization of tumor or surrogates; clinically validated gating | Limited tracking frame rate ( $\sim 5$ fps); high system cost; magnetic field may affect dose distribution              | Low                    |
| Accuray Radixact Synchrony | kV images + respiratory model | Clinically successful markerless lung tracking; real-time beam adaptation                                | Proprietary tracking and modeling algorithms; limited transparency; difficult for independent reproduction or extension | Low                    |
| Proposed method            | kV images only                | Explicit modeling of temporal consistency and motion plausibility                                        | Currently evaluated as a research system; not a turnkey clinical product                                                | High                   |

### 3.8 Conclusion

This chapter proposed and validated a novel deep learning framework for marker-less lung tumor tracking using a cross-patient model. By applying the realistic degradation model developed in Chapter 2 to train the Restormer network, we successfully bridged the style gap between DRRs and clinical fluoroscopic images. The subsequent DUCK-Net segmentation network achieved high tracking accuracy across a dataset from 7 patients (100 fields), with a median 3D Euclidean error (MEE) of 1.53 mm.

This result is clinically promising, falling well within the 2 mm clinical acceptance threshold. The primary contribution of this study is the demonstration of the feasibility of a general, "train-once, apply-to-all" cross-patient model, successfully eliminating the need for patient-specific retraining.

However, while the framework shows high accuracy, the analysis of processing time (179.8 ms total) revealed that the style transfer step (Restormer at 78 ms) is the main performance bottleneck. Ablation studies (as shown in Table 3.6) indicated that omitting this step (Method 1) results in only a minor decrease in accuracy (1.62 mm MEE). This presents a practical trade-off for real-time clinical application: sacrificing a marginal degree of accuracy for crucial gains in speed.

Therefore, future work must focus on optimizing the entire pipeline. Lightweighting the

Restormer and DUCK-Net models and employing hardware acceleration are essential to reduce the processing time to a clinically viable real-time target.

## Chapter 4

# Development of An Integrated Software Platform

### 4.1 Introduction

In the preceding chapters, we explored methods to enhance training data quality (Chapter 2) and developed a deep learning workflow for tumor tracking (Chapter 3). A significant practical challenge encountered during this research was the fragmented and inefficient nature of the existing software landscape. As discussed in the Introduction, the typical workflow for developing and validating marker-less tracking algorithms is cumbersome. It requires researchers to sequentially operate multiple, disparate software tools: one for DICOM visualization and segmentation (e.g., 3D Slicer, ITK-SNAP) [35,36], another for simulating imaging physics and generating DRRs (e.g., Plastimatch, or custom scripts) [37], and separate Python environments for training, executing, and evaluating the AI models.

This disjointed process introduces several bottlenecks:

- **Time Inefficiency:** Significant time is lost in data conversion, manual file handling, and context-switching between different applications.
- **High Error-Proneness:** Manual data transfer between steps increases the risk of human error, such as mismatched patient data, incorrect coordinate systems, or inconsistent parameter settings.
- **Steep Learning Curve:** Researchers must gain proficiency in multiple complex tools, each with its own interface and data conventions.
- **Lack of Reproducibility:** A workflow reliant on manual steps and disparate scripts is inherently difficult to standardize and reproduce.



To address these critical workflow inefficiencies and accelerate research in this domain, we identified the need for a unified, end-to-end software platform. This chapter details the design, development, and evaluation of our custom-built application, provisionally named **YULAB-B504**. The primary contribution of this software is not a novel algorithm, but rather the novel *integration* of all essential functions into a single, cohesive graphical user interface (GUI). It is designed to provide a "one-stop" solution, managing the entire research cycle from initial CT data loading to the final quantitative evaluation of AI-driven tumor tracking.

This chapter is organized as follows: Section 4.2 details the requirements analysis and system architecture. Section 4.3 describes the implementation of key functional modules, and Section 4.4 concludes with a discussion and summary.

## 4.2 System Design and Architecture

### 4.2.1 Requirements Analysis

Based on the identified challenges, we defined a set of core functional requirements for the integrated platform:

1. **Data Management:** Load, display, and manage CT datasets in DICOM format.
2. **Virtual Geometry Setup:** Allow users to interactively set the radiotherapy isocenter and define the complete geometry of a virtual X-ray imaging system using a few simple parameters.
3. **DRR Generation:** Integrate a physics-based DRR generation engine capable of producing single or batch DRR images automatically without user supervision, which is particularly useful for generating large volumes of training data.
4. **AI Model Integration:** Provide an interface to load pre-trained deep learning models (currently supporting PyTorch or TensorFlow/Keras models).
5. **Batch Processing:** Automate the application of the AI model to entire sequences of DRRs or clinical fluoroscopic images.
6. **3D Tracking Visualization:** Automatically extract tumor center coordinates from the AI model's results and visualize them.

Table 4.1: Core technologies and libraries used in the development of the YULAB-B504 platform.

| Component        | Technology/Library        | Purpose                         |
|------------------|---------------------------|---------------------------------|
| GUI              | PyQt5                     | User interface and interaction  |
| Data I/O         | Pydicom                   | Reading DICOM files             |
| Image Processing | PIL, OpenCV, NumPy        | Image processing                |
| Data Management  | json                      | Managing data and state         |
| DRR Generation   | Custom Implementation     | Ray-projection DRR Generation   |
| AI Inference     | PyTorch, TensorFlow/Keras | Loading and executing DL models |

7. **Automated Evaluation (Under Development):** Implement a "one-click" evaluation module that compares model predictions against ground-truth labels and automatically computes standard performance metrics (e.g., Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), Target Registration Error (TRE)).

#### 4.2.2 Technology Stack and Architecture

To meet the defined requirements, we designed a modular three-layer architecture that ensures scalability and maintainability. The application was developed primarily in Python 3.7, leveraging its rich ecosystem of libraries for scientific computing and medical imaging. The technology stack is summarized in Table 4.1.

The system is designed as a desktop application with a clear separation between the front-end GUI and the back-end processing core. The GUI was built using **PyQt5**, providing a responsive and native user experience. The back-end relies on libraries such as **Pydicom** for DICOM I/O, **PIL (Python Imaging Library)** and **OpenCV** for image processing and visualization, and **PyTorch/TensorFlow** for AI model inference. This modular architecture, illustrated in Figure 4.1, ensures that individual components can be updated or replaced without affecting the rest of the application. More importantly, the design explicitly supports extensibility - new interfaces can be seamlessly integrated into the existing framework by adhering to the standardized data access protocols, enabling continuous platform evolution and feature expansion.

### 4.3 Implementation of Key Functional Modules

This section describes the implementation of the core modules corresponding to the layered architecture and interface organization.

## YULAB-B504 System Architecture Diagram

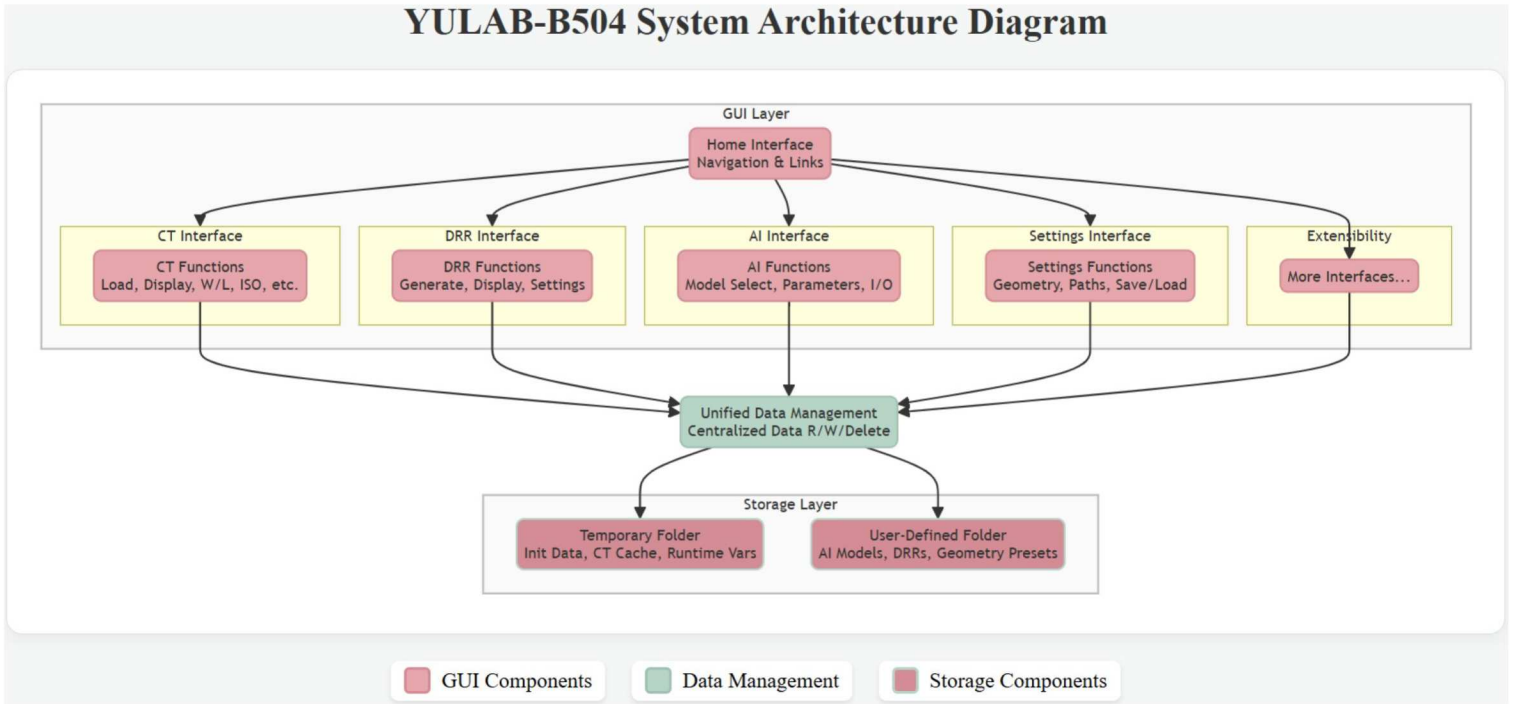


Figure 4.1: The modular system architecture of the YULAB-B504 application.

### 4.3.1 GUI Layer and Interface Implementation

The GUI layer is implemented using PyQt5, with a main window managing the Home interface and multiple specialized interfaces. Each interface operates as an independent module while maintaining seamless data exchange through the unified data management layer:

- **CT Interface:** Utilizes PyQt5's graphics view framework for interactive CT slice navigation. The interface incorporates mouse wheel support for slice browsing, interactive window width/level adjustment. Tumor contour processing is implemented using adaptive morphological operations with OpenCV. Another critical function is the ability to define the treatment iso center, this coordinate becomes the center of rotation for the virtual imaging system.
- **DRR Interface:** Provides a comprehensive parameter panel for DRR generation configuration, including X-ray source parameters, DRR type, couch angle, batch generation mode, etc. It is able to efficiently produce large DRR datasets, with GPU acceleration.
- **AI Interface:** Implements a dynamic model loading system that automatically detects available AI architectures from the user directory. The interface includes real-time visualization of input images and model predictions using PyQt5's graphics pipeline.



- **Settings Interface:** Set up geometry configurations. It is able to save to/load/delete local saved JSON format geometry configuration files.

### 4.3.2 Data Management and Cross-interface Communication

The data management layer employs a custom function that provides essential functions including `set_var`, `get_var`, `cleanup_temp_dir`, `initialize_global_vars`, and `del_var`.

The implementation follows a publisher-subscriber pattern, where data modifications in one interface automatically trigger updates in dependent interfaces, ensuring data consistency throughout the workflow.

### 4.3.3 Storage Management and Extensibility

The storage layer implements a dual-directory strategy managed through the data management module:

- **Temporary Storage:** Created in the system's temp directory during application initialization, this stores runtime data including parsed DICOM volumes, intermediate processing results, and current session variables. Automatic cleanup routines when process ends ensure efficient resource management.
- **User Directory:** Configurable persistent storage organized into subdirectories for AI models, DRR outputs, and geometry presets. This separation allows users to maintain project-specific configurations and reuse assets across multiple sessions.

The architecture's extensibility is demonstrated through standardized interface contracts and data access patterns. New sub-interfaces can be added by implementing the base interface class and registering with the main application controller, automatically gaining access to the shared data ecosystem and storage management capabilities.

### 4.3.4 Virtual Imaging Geometry and DRR Generation

Once the iso center and geometry information are registered (Figure 4.4), the user accesses the DRR generation module (Figure 4.3). This panel allows for the complete specification of

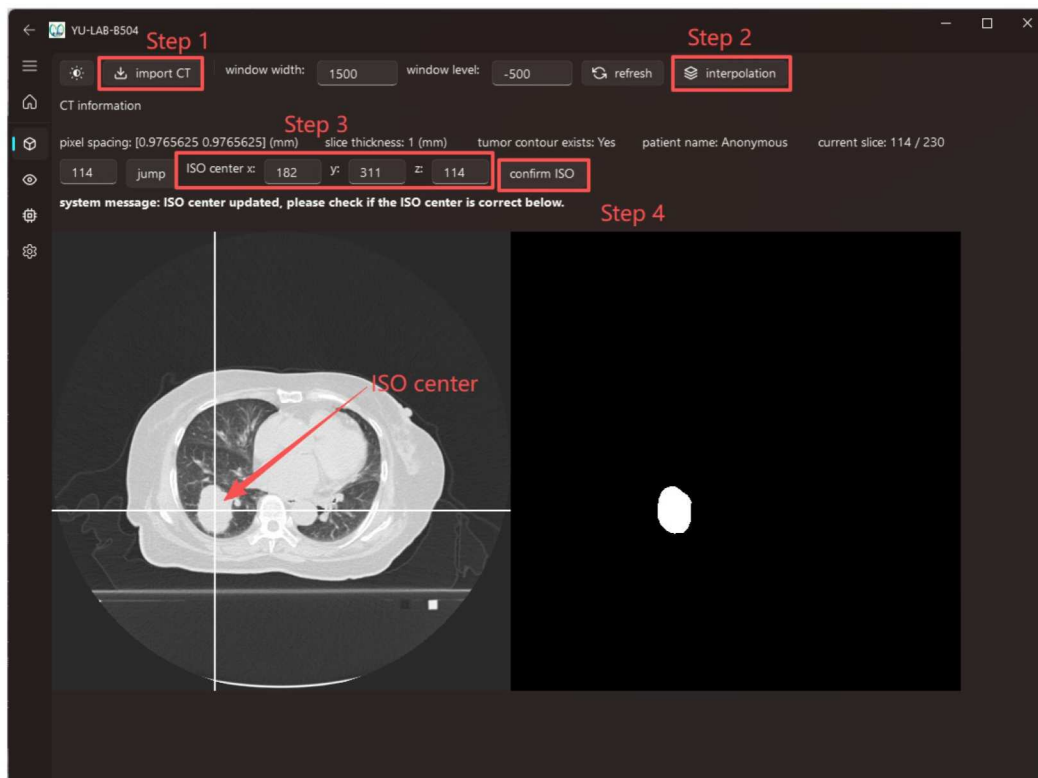


Figure 4.2: The CT interface. CT dataset is loaded, displaying CT and binary CT, with ISO center (white crosshairs) registered.

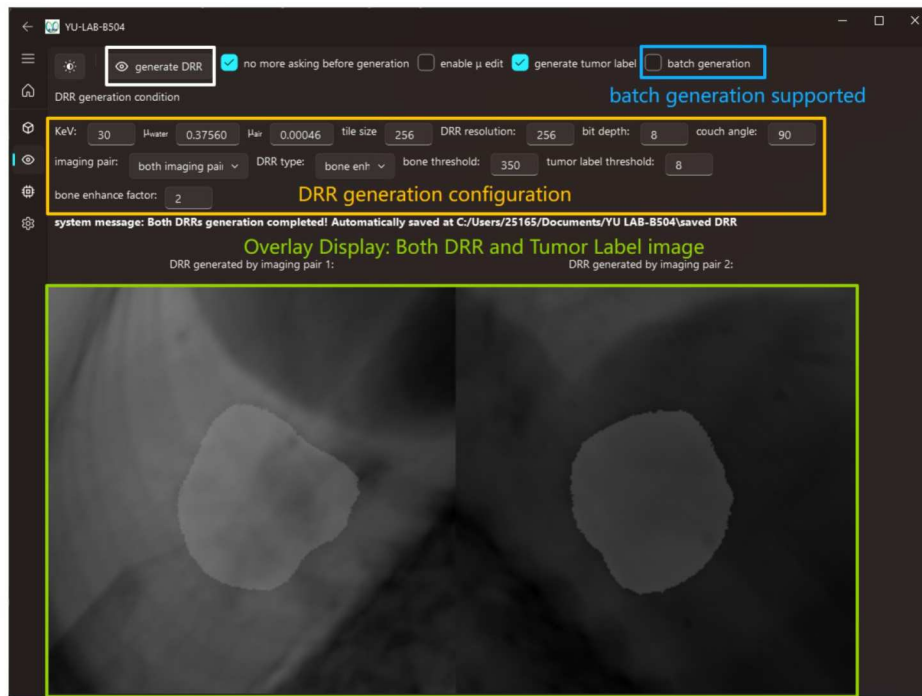


Figure 4.3: The DRR interface. The user defines DRR parameters and output settings. The generated DRR and tumor label are displayed in the preview window. We can use the mouse wheel to scroll through the pages to view tumor labels.

the DRR generation, including X-ray source configuration, DRR resolution, DRR type, DRR grayscale bit depth, and couch angle.

User can define a single projection angle or a range of angles to generate DRR sequence. The back-end DRR engine then utilizes ray-casting algorithm to compute the projection images. This module also integrates the stylization pipeline developed in Chapter 2, allowing users to generate "realistic" fluoroscopy-style images directly from the CT data.

#### 4.3.5 AI Model Batch Processing and Evaluation

The user can use integrated AI module as shown in Figure 4.5) to:

1. Load a pre-trained model file (e.g., a '\*.pth') from Chapter 3.
2. Select an input directory containing a sequence of DRRs or clinical fluoroscopic images.
3. Select a corresponding directory for ground-truth segmentation masks (if available).
4. Click "start" button to predict and "evaluate" button to evaluate.



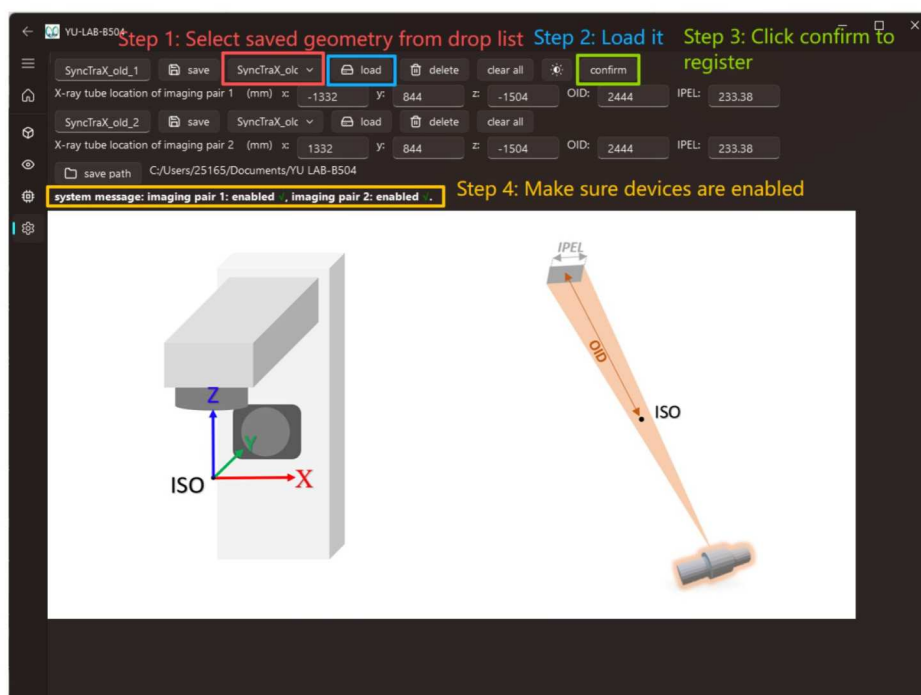


Figure 4.4: The Setting interface. The user defines imaging device geometric configurations here. It is allowed to name and save, load/delete saved geometric configurations. Please note that it is necessary to click the confirm button to register the geometric info, user can register up to two imaging devices, once the info is registered, system message will give hints. The parameters include: coordinates of X-ray source, object-to-image distance (OID) and edge length of imaging plane (IPEL).

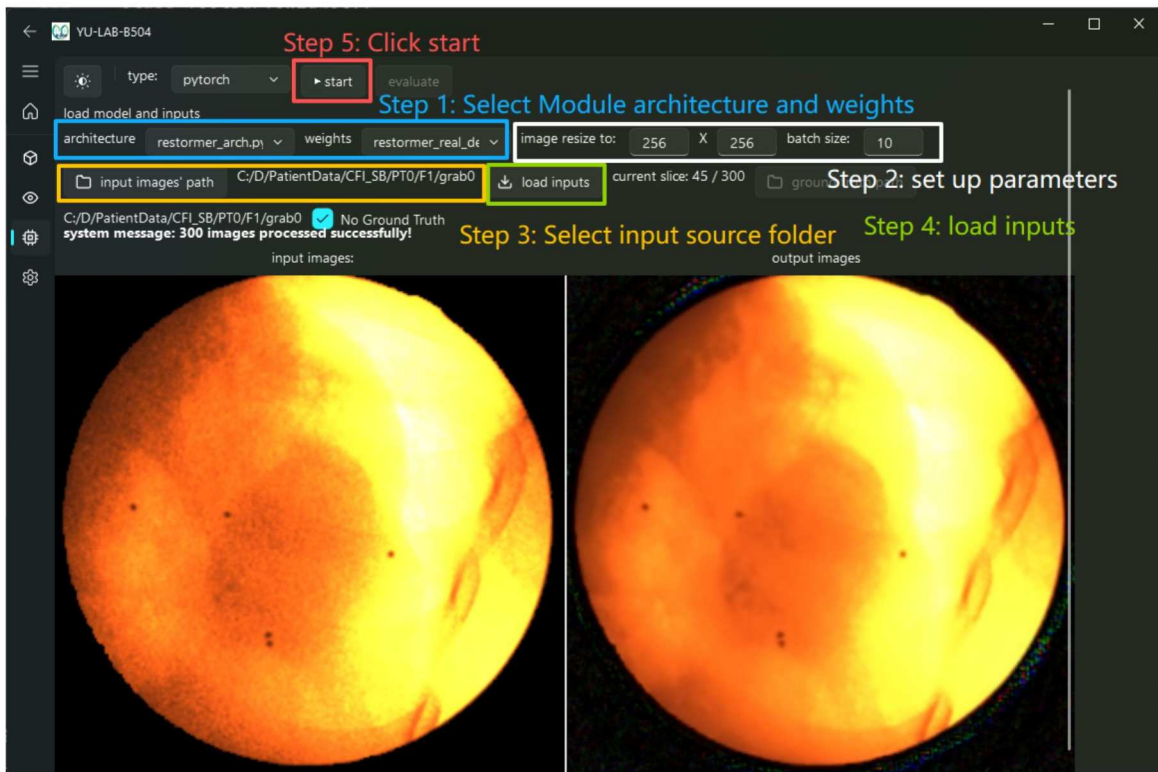


Figure 4.5: The integrated AI and Evaluation module. The user loaded a model and specifies input directories. The input and output results are shown below.

The application then iterates through all images, performs inference, and compares the predicted results with the ground truth (if available).

The results are presented in two forms:

- **Visual Overlay:** The system displays the input image with the predicted mask and ground-truth mask stacked in different layers, scroll the mouse wheel to turn pages, allowing for immediate visual inspection of the model's performance.
- **Quantitative Report:** A table is populated with patient/image identifiers and the computed metrics (SSIM, IoU, PSNR, Dice, etc) for each frame. This entire report can be exported as a CSV file for further statistical analysis.

## 4.4 Discussion and Summary

This chapter presented the design and implementation of YULAB-B504, an integrated software platform for marker-less tumor tracking research. We have demonstrated that by analyzing

the specific needs of the research workflow and consolidating essential tools into a single application, we can overcome significant bottlenecks related to efficiency, error-proneness, and reproducibility.

The key contribution of this work is the provision of a "one-stop" solution that bridges the gap between CT data, imaging simulation, AI model execution, and quantitative evaluation. While the individual components (e.g., DRR generation, AI models) are based on existing methods, their novel integration provides a substantial practical and methodological advantage for researchers in this field. This tool was instrumental in enabling the large-scale experiments and validations presented in Chapters 2 and 3.

The limitation of this software is that it currently only supports retrospective research analysis and cannot be used for real-time processing during surgery.. Future work will focus on expanding the platform's capabilities, such as integrating deformable image registration modules, supporting a wider range of AI model architectures, and further optimizing the computational backend for near real-time performance.



## Bibliography

- [1] J. Ferlay, M. Ervik, F. Lam, M. Laversanne, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray, “Global cancer observatory: Cancer today,” <https://gco.iarc.who.int/today>, Lyon, France, 2024, accessed: 2025-09-02.
- [2] L. Wang, C. Lynch, S. P. Pitroda, A. Piffkó, K. Yang, A. K. Huser, H. L. Liang, and R. R. Weichselbaum, “Radiotherapy and immunology,” *Journal of Experimental Medicine*, vol. 221, no. 7, p. e20232101, 2024.
- [3] O. M. Dona Lemus, M. Cao, B. Cai, M. Cummings, and D. Zheng, “Adaptive radiotherapy: next-generation radiotherapy,” *Cancers*, vol. 16, no. 6, p. 1206, 2024.
- [4] H. Zhu, M. L. K. Chua, I. Chitapanarux, O. Kaidar-Person, C. Mwaba, M. Alghamdi, A. R. Mignola, N. Amrogowicz, G. Yazici, Z. Bourhaleb *et al.*, “Global radiotherapy demands and corresponding radiotherapy-professional workforce requirements in 2022 and predicted to 2050: a population-based study,” *The Lancet Global Health*, vol. 12, no. 12, pp. e1945–e1953, 2024.
- [5] J. Bertholet, A. Knopf, B. Eiben, J. McClelland, A. Grimwood, E. Harris, M. Menten, P. Poulsen, D. T. Nguyen, P. Keall *et al.*, “Real-time intrafraction motion monitoring in external beam radiotherapy,” *Physics in medicine & biology*, vol. 64, no. 15, p. 15TR01, 2019.
- [6] Y. Wang, E. Lombardo, A. Thummerer, T. Blöcker, Y. Fan, Y. Zhao, C. I. Papadopoulou, C. Hurkmans, R. H. Tijssen, P. A. Görts *et al.*, “Trackrad2025 challenge dataset: real-time tumor tracking for mri-guided radiotherapy,” *Medical physics*, vol. 52, no. 7, p. e17964, 2025.
- [7] D. Zhou, M. Nakamura, N. Mukumoto, Y. Matsuo, and T. Mizowaki, “Feasibility study of deep learning-based markerless real-time lung tumor tracking with orthogonal x-ray projection images,” *Journal of applied clinical medical physics*, vol. 24, no. 4, p. e13894, 2023.

- [8] G. Li, “Advances and potential of optical surface imaging in radiotherapy,” *Physics in Medicine & Biology*, vol. 67, no. 16, p. 16TR02, 2022.
- [9] R. De Crevoisier, C. Lafond, A. Mervoyer, C. Hulot, N. Jaksic, I. Bessi res, and G. Delpon, “Image-guided radiotherapy,” *Cancer/Radioth rapie*, vol. 26, no. 1-2, pp. 34–49, 2022.
- [10] N. Rammohan, J. W. Randall, and P. Yadav, “History of technological advancements towards mr-linac: The future of image-guided radiotherapy,” *Journal of Clinical Medicine*, vol. 11, no. 16, p. 4730, 2022.
- [11] D. Chrystall, A. Mylonas, E. Hewson, J. Martin, P. Keall, J. Booth, and D. T. Nguyen, “Deep learning enables mv-based real-time image guided radiation therapy for prostate cancer patients,” *Physics in Medicine & Biology*, vol. 68, no. 9, p. 095016, 2023.
- [12] Q. Lu, D. Guo, Y. Wei, J. Zheng, and J. Li, “Role of gastrointestinal ultrasound in image-guided radiation therapy: A review,” *Journal of Radiation Research and Applied Sciences*, vol. 16, no. 1, p. 100520, 2023.
- [13] X. Liu, L.-S. Geng, D. Huang, J. Cai, and R. Yang, “Deep learning-based target tracking with x-ray images for radiotherapy: a narrative review,” *Quantitative imaging in medicine and surgery*, vol. 14, no. 3, p. 2671, 2024.
- [14] W. G. Campbell, M. Miften, and B. L. Jones, “Automated target tracking in kilovoltage images using dynamic templates of fiducial marker clusters,” *Medical physics*, vol. 44, no. 2, pp. 364–374, 2017.
- [15] R. Li, E. Mok, D. T. Chang, M. Daly, B. W. Loo Jr, M. Diehn, Q.-T. Le, A. Koong, and L. Xing, “Intrafraction verification of gated rapidarc by using beam-level kilovoltage x-ray images,” *International Journal of Radiation Oncology\* Biology\* Physics*, vol. 83, no. 5, pp. e709–e715, 2012.
- [16] A. Van der Horst, S. Wognum, R. D. Fajardo, R. De Jong, J. E. Van Hooft, P. Fockens, G. Van Tienhoven, and A. Bel, “Interfractional position variation of pancreatic tumors quantified using intratumoral fiducial markers and daily cone beam computed tomography,” *International Journal of Radiation Oncology\* Biology\* Physics*, vol. 87, no. 1, pp. 202–208, 2013.
- [17] A. Casutt, R. Kinj, E.-M. Ozsahin, C. von Garnier, and A. Lovis, “Fiducial markers for stereotactic lung radiation therapy: review of the transthoracic, endovascular and endobronchial approaches,” *European Respiratory Review*, vol. 31, no. 163, 2022.

- [18] V. Oudin, J. Salleron, V. Marchesi, D. Peiffert, M. Khadige, and J.-C. Faivre, “Cyberknife® stereotactic radiation therapy for stage i lung cancer and pulmonary oligometastases: is fiducial implantation still relevant?—a cohort study,” *Journal of thoracic disease*, vol. 15, no. 9, p. 4636, 2023.
- [19] A. Casutt, L. Noirez, M. Bernasconi, A. Koutsokera, C. Beigelman-Aubry, R. Kinj, E.-M. Ozsahin, A.-D. Durham, C. von Garnier, and A. Lovis, “Endobronchial coil spring fiducial markers for cyberknife® stereotactic body radiation therapy,” *Respirology*, vol. 26, no. 5, pp. 469–476, 2021.
- [20] K. Khullar, S. T. Dhawan, J. Nosher, and S. K. Jabbour, “Fiducial marker migration following computed tomography-guided placement in the liver: a case report,” *AME Case Reports*, vol. 5, p. 15, 2021.
- [21] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [24] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [25] T. Roggen, M. Bobic, N. Givchchi, and S. G. Scheib, “Deep learning model for markerless tracking in spinal sbt,” *Physica Medica*, vol. 74, pp. 66–73, 2020.
- [26] X. He, W. Cai, F. Li, Q. Fan, P. Zhang, J. J. Cuaron, L. I. Cerviño, X. Li, and T. Li, “Decompose kv projection using neural network for improved motion tracking in paraspinal sbt,” *Medical physics*, vol. 48, no. 12, pp. 7590–7601, 2021.
- [27] T. Terunuma, T. Sakae, Y. Hu, H. Takei, S. Moriya, T. Okumura, and H. Sakurai, “Explainability and controllability of patient-specific deep learning with attention-based augmentation for markerless image-guided radiotherapy,” *Medical Physics*, vol. 50, no. 1, pp. 480–494, 2023.



- [28] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, “Detectron2,” 2019.
- [29] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *European conference on computer vision*. Springer, 2014, pp. 740–755.
- [30] R. Hirai, Y. Sakata, A. Tanizawa, and S. Mori, “Real-time tumor tracking using fluoroscopic imaging with deep neural network analysis,” *Physica Medica*, vol. 59, pp. 22–29, 2019.
- [31] Y. Yan, F. Fujii, T. Shiinoki, and S. Liu, “Markerless lung tumor localization from intra-operative stereo color fluoroscopic images for radiotherapy,” *IEEE Access*, vol. 12, pp. 40 809–40 826, 2024.
- [32] M. Unberath, J.-N. Zaech, S. C. Lee, B. Bier, J. Fotouhi, M. Armand, and N. Navab, “Deepdrr—a catalyst for machine learning in fluoroscopy-guided procedures,” in *International conference on medical image computing and computer-assisted intervention*. Springer, 2018, pp. 98–106.
- [33] W. Feng, L. Ju, L. Wang, K. Song, X. Zhao, and Z. Ge, “Unsupervised domain adaptation for medical image segmentation by selective entropy constraints and adaptive semantic alignment,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 1, 2023, pp. 623–631.
- [34] H. Guan and M. Liu, “Domain adaptation for medical image analysis: a survey,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1173–1185, 2021.
- [35] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka *et al.*, “3d slicer as an image computing platform for the quantitative imaging network,” *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.
- [36] P. A. Yushkevich, Y. Gao, and G. Gerig, “Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images,” in *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2016, pp. 3342–3345.
- [37] G. C. Sharp, R. Li, J. Wolfgang, G. Chen, M. Peroni, M. F. Spadea, S. Mori, J. Zhang, J. Shackleford, and N. Kandasamy, “Plastimatch: an open source software suite for radiotherapy image processing,” in *Proceedings of the XVI<sup>th</sup> International Conference on the use of Computers in Radiotherapy (ICCR)*, Amsterdam, Netherlands, vol. 3, 2010.

- [38] A. Krull, T.-O. Buchholz, and F. Jug, “Noise2void-learning denoising from single noisy images,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2129–2137.
- [39] P. Hammersberg, M. Stenström, H. Hedtjärn, and M. Mångård, “Image noise in x-ray imaging caused by radiation scattering and source leakage, a qualitative and quantitative analysis,” *Journal of X-Ray Science and Technology*, vol. 8, no. 1, pp. 19–29, 1998.
- [40] M. Cesarelli, P. Bifulco, T. Cerciello, M. Romano, and L. Paura, “X-ray fluoroscopy noise modeling for filter design,” *International journal of computer assisted radiology and surgery*, vol. 8, no. 2, pp. 269–278, 2013.
- [41] A. Sarno, E. Andreozzi, D. De Caro, G. Di Meo, A. Strollo, M. Cesarelli, and P. Bifulco, “Real-time algorithm for poissonian noise reduction in low-dose fluoroscopy: performance evaluation,” *Biomedical engineering online*, vol. 18, no. 1, p. 94, 2019.
- [42] W. Huda and R. B. Abrahams, “Radiographic techniques, contrast, and noise in x-ray imaging,” *American Journal of Roentgenology*, vol. 204, no. 2, pp. W126–W131, 2015.
- [43] T. Shiinoki, F. Fujii, K. Fujimoto, Y. Yuasa, and T. Sera, “A novel dynamic robotic moving phantom system for patient-specific quality assurance in real-time tumor-tracking radiotherapy,” *Journal of Applied Clinical Medical Physics*, vol. 21, no. 7, pp. 16–28, 2020.
- [44] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, “Corrc2g: Color to gray conversion by correlation,” *IEEE Signal Processing Letters*, vol. 24, no. 11, pp. 1651–1655, 2017.
- [45] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, “Restormer: Efficient transformer for high-resolution image restoration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.
- [46] R.-G. Dumitru, D. Peteleaza, and C. Craciun, “Using duck-net for polyp image segmentation,” *Scientific reports*, vol. 13, no. 1, p. 9803, 2023.
- [47] S. A. Koohpayegani, A. Tejankar, and H. Pirsiavash, “Mean shift for self-supervised learning,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 326–10 335.
- [48] B. Smith, B. Allen, S. Rusu, J. St-Aubin, and D. Hyer, “Real-time tracking of lung tumors using a 1.5 t elekta unity mr-linac: first clinical experiences,” *Frontiers in oncology*, vol. 15, p. 1569428, 2025.

- [49] E. O. Goksel, Z. Ozen, S. Gunduz, A. Ture, H. Kucucuk, A. Arifoglu, E. Tezcanli, U. Abacioglu, and M. Sengoz, “Markerless tumor tracking for lung sbt with the radixact synchrony system: Initial experience and short-term clinical outcomes,” *Journal of Applied Clinical Medical Physics*, vol. 26, no. 12, p. e70362, 2025.



## Acknowledgment

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Fumitake Fujii, for his invaluable guidance, patience, and meticulous instruction throughout my doctoral journey. His insightful advice, rigorous academic attitude, and continuous encouragement have profoundly shaped both my research and my personal growth. I am equally grateful to Associate Professor Takehiro Shiinoki for his generous support, which greatly helped me overcome many research challenges and refine my ideas.

I would also like to extend my heartfelt thanks to Yamaguchi University for providing such a rich and inspiring academic environment, and to Japan, as well as the Japan Science and Technology Agency (JST), for their financial and logistical support that made both my life and research possible during these years.

My deepest appreciation also goes to my family, especially my wife, for her unwavering love, understanding, and support. Her encouragement gave me the strength to persevere through difficulties and to keep moving forward even when the path seemed uncertain.

Finally, I would like to thank myself—for not giving up, for continuing to strive to become a better version of who I am, and for learning from every experience, whether joyful or painful. I am also thankful to all the people I have met along the way—those who have supported me and even those who disliked or challenged me. Each of them, in their own way, has taught me valuable lessons and helped shape the person I have become today.

Looking back, I have spent twenty-five years in total—over 83% of my life—being a student. Now, I think I am finally ready to step into the next stage of life. Also, I sincerely hope to become a good father and eagerly await the arrival of my son. I have imagined countless times how I will give him a childhood full of love, and how I will teach him the knowledge, skills, strength, and resilience needed to live as a kind and capable adult in this world. Though my son has not yet come into this world, I already feel a deep love and responsibility for him. I hope one day he will understand how much he has inspired me to become a better person.

## Appendix

The non-uniform noise modeling approach proposed in this study is designed to maximize its generalizability across different imaging devices and acquisition settings. Although the primary parameterization in this research relies on data from one specific device, preliminary noise analysis conducted on a newer generation SyncTraX system provides important corroboration for the adaptability of the model.

We fabricated a gelatin cylinder and acquired 300 X-ray fluoroscopic images under 100 kV, 80 mA, and 4 ms conditions using the newer SyncTraX system. We embedded four aluminum blocks of different thicknesses inside the gelatin to simulate bones as shown in Figure 4.6. Statistical analysis results of the subtracted noise images are shown in Figures 4.7 and 4.8.

In the newer device, the characteristics of noise probability ( $\mathcal{N}_p$ ) were found to be highly consistent with our primary findings:  $\mathcal{N}_p$  remains lower in regions of low material density (high pixel values) and gradually increases with distance from the X-ray beam center. However, a notable difference was observed in the noise mean amplitude ( $\mathcal{N}_m$ ). This contrast is evident when comparing the noise amplitude characteristics of the older device shown in Figure 2.5 (b), where the amplitude is clearly influenced by pixel grayscale values, with those of the newer device shown in Figure 4.8 (b). In the newer device, the noise amplitude exhibited uniformity across the entire image area; no significant variation in amplitude was observed between the regions corresponding to the aluminum blocks and the surrounding gelatin area. This indicates that the noise amplitude in the newer system has little sensitivity to changes in the pixel value  $V$  or the radial distance  $d$ .

This uniform amplitude characteristic is likely a result of the newer device employing improved X-ray collimating technology, which ensures highly uniform X-ray intensity, thereby eliminating the spatial intensity variations typical of older systems. For devices exhibiting such high amplitude uniformity, our proposed noise model demonstrates adaptability: the  $\sigma_d$  and  $\beta$  within the amplitude functions  $f_{sm}(\cdot)$  in Equation 2.6 can be adjusted to be larger until the  $f_{sm}(\cdot)$  become insensitive to changes in  $d$ . This principle is analogous to the concept that "the



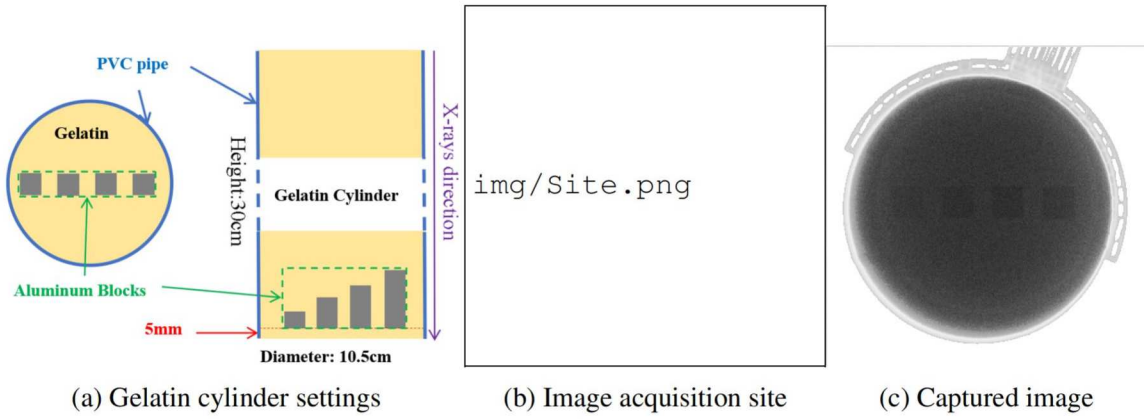


Figure 4.6: Gelatin preparation and image acquisition. The aluminum blocks are labeled 'aluminum\_1' to 'aluminum\_4' from left to right, with heights of 1, 2, 3, and 4 mm, respectively.

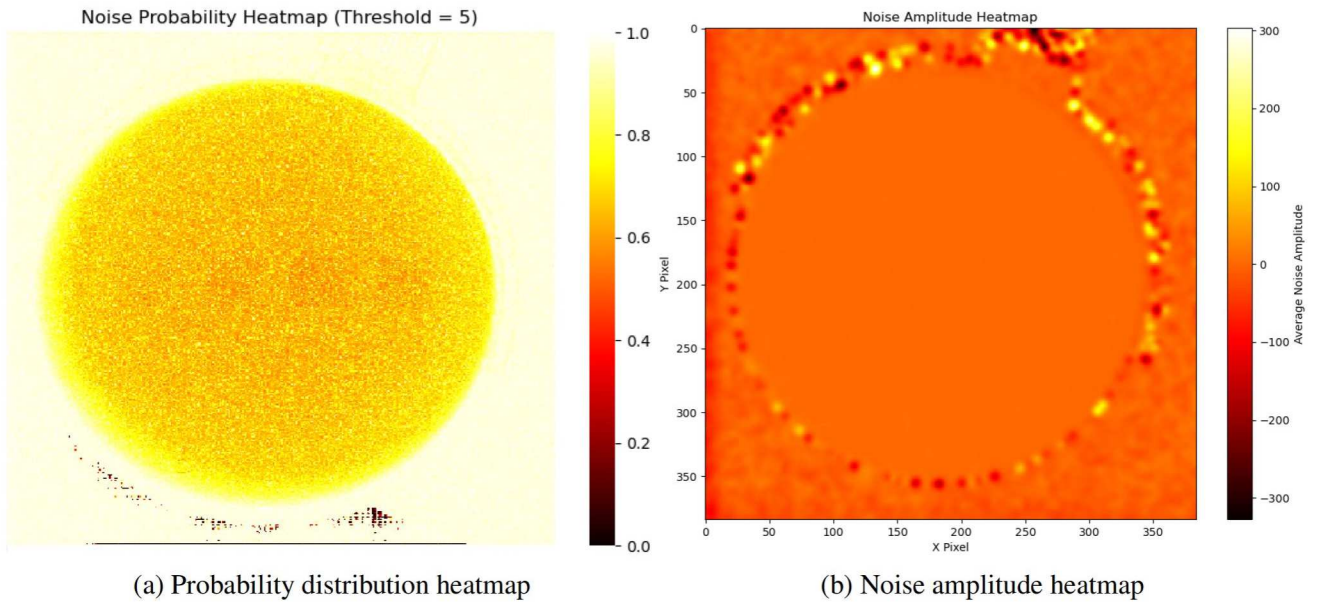


Figure 4.7: Noise characteristics in a gelatin cylinder fluoroscopic image. **(a)** Probability distribution map of degradation-induced noise, derived by computing the pixel-wise variance across 300 subtracted images (individual frames minus the average of all frames) and normalizing it to a  $[0, 1]$  range, where higher values indicate greater degradation likelihood. To account for the fact that the average image is not perfectly clean, a threshold of 5 grayscale values was applied, such that only differences exceeding this threshold were considered as noise. Square regions with embedded aluminum blocks show lower probabilities than gelatin areas, decreasing toward the image center within the circular field. **(b)** Heatmap of average noise amplitude across the image, calculated as the mean absolute difference between individual frames and the average image. The noise amplitude remains stable inside the gelatin area.



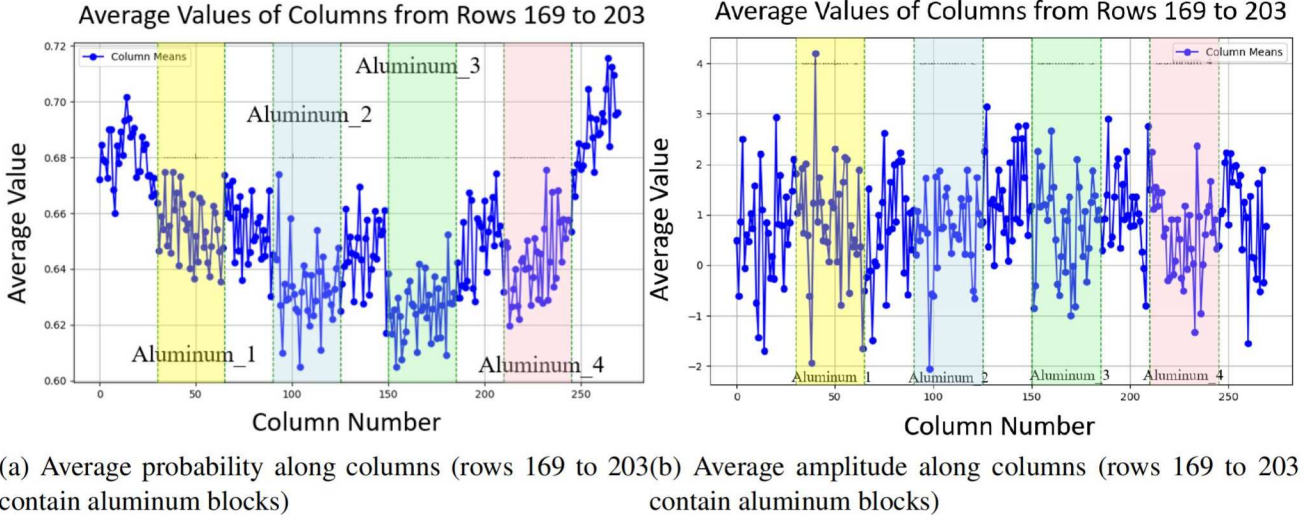


Figure 4.8: Line plots of noise characteristics along image columns, with edge pixels excluded to focus on internal variations. **(a)** Average probability along columns (rows 169 to 203 contain aluminum blocks); aluminum block regions exhibit lower probability (around 0.015 to 0.02 lower, depending on how close the pixel is to the center) compared to gelatin areas. **(b)** Average amplitude along columns (rows 169 to 203 contain aluminum blocks), showing stable amplitude in aluminum block regions compared to gelatin areas.

Earth is large enough that the local terrain appears flat," demonstrating the model's flexible capacity. Although noise characteristics in fluoroscopic images were observed on two sets of equipment in this study, but their generalizability to other devices remains unclear. Furthermore, this finding also confirms that a limitation of our proposed modeling method is that it requires a device-specific calibration prior to its application.

To guide future adoption, this calibration procedure can be formalized into the following steps: (1) Acquire a series of images of a static phantom (e.g., a gelatin or stationary phantom) under the desired imaging conditions. (2) Calculate the mean of these images to obtain a "clean" average image, which is then subtracted from the individual frames to isolate the noise characteristics. (3) Determine the functional relationship between grayscale intensity and noise by selecting an appropriate fitting method. In this study, a 10th-order polynomial was chosen based on empirical observation, as it was simple and provided a good fit for our specific device's non-linear noise characteristics. (4) Finally, determine the optimal values for spatial parameters  $\sigma_d$  and  $\beta$ , through a traversal or search algorithm to match the observed spatial distribution.

Validation across different SyncTraX system generations demonstrated the model's adaptability to varying hardware characteristics, though it requires device-specific calibration for optimal performance. This degradation modeling approach provides a crucial foundation for generating realistic training data, effectively bridging the domain gap between synthetic DRRs

and clinical fluoroscopic images - a fundamental requirement for developing robust marker-less tumor tracking algorithms that can generalize to real clinical scenarios.