FLOWSEED: CONTINUOUS-TIME CONTRAST-ENHANCED CT SYNTHESIS VIA ORGAN-AWARE GAUSSIAN-WEIGHTED RESIDUAL SPLATTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multiphase contrast-enhanced computed tomography (CECT) sparsely samples a continuous, organ-dependent enhancement process. However, existing synthesis methods treat arterial, venous, and delayed scans as discrete targets and reconstruct anatomy already available in non-contrast CT (NCCT). We introduce FlowSeed, a continuous-time residual contrast-rendering framework that retains NCCT as the anatomical base and renders only its time-dependent enhancement. FlowSeed uses soft organ maps to route time-conditioned affine modulation at each voxel. A slice-aware Gaussian residual decoder then predicts intensity, opacity, axis-aligned scale, and center offset to render enhancement onto queried axial slices. Physiology-inspired ordinal constraints regularize liver and aortic enhancement, while a conditional PatchGAN provides adversarial supervision for local appearance fidelity. On held-out MCT-LTDiag studies, FlowSeed reduces mean absolute error by 6.7–18.8% relative to the strongest comparator across three anatomical groups. On external WAW-TACE data, it achieves the lowest mean absolute error and highest SSIM in all groups. Ablations support the contributions of organ-aware temporal routing and Gaussian-weighted residual decoding. Dense-query visualizations further show that FlowSeed can synthesize continuous enhancement sequences with distinct organ-dependent responses across time.

1 INTRODUCTION

Contrast-enhanced computed tomography reveals vascular structures, organ perfusion, and phase-dependent lesion appearance. A multiphase study samples this process at a few conventional points, commonly described as arterial, portal venous, and delayed phases (Singal et al., 2023). These acquisitions increase radiation exposure and require iodinated contrast (Smith-Bindman et al., 2015; van der Molen et al., 2025). Their timing also follows protocol-level conventions, although contrast transport varies with patient physiology and acquisition conditions (Alyani Nezhad et al., 2026). Synthesizing enhancement from a single NCCT could reduce dependence on additional acquisitions and permit retrospective exploration of virtual timepoints that were not captured clinically.

Current NCCT-to-CECT synthesis methods largely inherit the formulation of image-to-image translation (Zhong et al., 2025). Conditional adversarial models generate a target domain from a source image (Isola et al., 2017), while recent CT-specific methods improve structural preservation (Chen et al., 2024) or use diffusion models to capture cross-phase relationships (Zheng et al., 2025; Zhu et al., 2025; Liu et al., 2025). These approaches have advanced phase-specific synthesis, but two limitations remain. First, a phase label collapses a broad and continuous range of acquisition times into one category. This is especially restrictive for delayed scans, whose timing can vary by several minutes. Second, full-volume generation asks the model to reconstruct anatomy already present in the NCCT. This entangles anatomical preservation with contrast prediction and can expose supervision to residual mismatch between scans.

We instead formulate CECT synthesis as continuous-time residual contrast rendering (Figure 1). The NCCT supplies the anatomical base, and the model predicts only its non-negative contrast

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

increment. A soft anatomical map routes organ-specific temporal modulation, allowing the same time to produce different responses in liver parenchyma, kidneys, and vessels. The resulting time-modulated features define the intensity, opacity, axis-aligned scales, and center offsets of dense residual splats. Gaussian splatting was originally developed as an explicit representation for efficient scene rendering (Kerbl et al., 2023). However, unlike conventional Gaussian splatting, which optimizes an explicit primitive set for each scene, we predicts its residual-splat parameters directly from the input NCCT and query time in a single feed-forward pass. Viewed through this rendering geometry, each axial coordinate defines a parallel query plane through the 3D anatomical field. Whereas scene Gaussians carry RGB appearance for camera-view rendering, our residual splats carry view-independent scalar contrast increments that are accumulated on the queried CT slice. The name FlowSeed reflects this intuition: the Gaussian kernels act as local contrast seeds whose time-conditioned contributions evoke enhancement flowing through vessels and organs. Adding the accumulated contrast field to the NCCT preserves the input anatomy by construction.

We make three contributions. First, to the best of our knowledge, FlowSeed is the first work to explore continuous-time, whole-abdomen CECT synthesis from NCCT, treating post-injection time as a directly queryable input rather than reducing enhancement to three categorical phases. Our organ-aware temporal modulation transforms a shared anatomical representation using spatially routed, organ-specific dynamics. Second, we propose an end-to-end, feed-forward Gaussian-weighted residual splatting decoder that directly predicts the complete splat parameter field from time-modulated volumetric features, giving time control over both contrast magnitude and spatial distribution without per-scene optimization. Third, we introduce a physiology-inspired temporal objective that constrains liver and aortic enhancement through reference-time renderings. On the registered MCT-LTDiag test set, the complete framework outperforms other methods across all reported anatomical groups and metrics. Dense queries show distinct kidney, liver, and vascular responses to the time input, a capability unavailable to existing methods.

Anonymous code is available at `https://anonymous.4open.science/r/iclr2027-anonymous-code-1B2F/`.

2 Related Work

2.1 Contrast-enhanced CT synthesis

General paired translation methods such as Pix2Pix (Isola et al., 2017) learn a conditional mapping with reconstruction and adversarial objectives, but assume that source and target pixels are well aligned. Abdominal multiphase CT violates this assumption because respiration, bowel motion, and patient repositioning cause residual differences even after preprocessing. Structure-aware CT translation methods introduce anatomical masks or invariant objectives (Chen et al., 2024). SMILE uses anatomy-aware diffusion and supervision designed to avoid unnecessary changes outside clinically relevant regions (Liu et al., 2025). These methods emphasize structural fidelity, but generate a finite set of target phases.

Multiphase synthesis requires temporal coordination. CSGen uses phase-causal and spatiotemporal modeling to jointly synthesize the conventional enhanced phases (Zhu et al., 2025), while CFPS-Diff models ordered phase transitions and cross-phase specificity within a conditional diffusion framework (Zheng et al., 2025). Both use categorical phase representations. We condition directly on elapsed post-injection time and learn one continuous function whose organ-specific response can be queried between conventional acquisitions.

2.2 Gaussian representations and conditional modulation

3D Gaussian Splatting represents a scene with explicit primitives and a differentiable renderer (Kerbl et al., 2023). GaSpCT adapts this representation to synthesize CT projection views from limited projection data (Nikolakakis et al., 2024). Conventional 3D Gaussian splatting optimizes its primitive parameters separately for each scene. In contrast, our method directly predicts a patient-space contrast increment from NCCT and accumulates dense, Gaussian-weighted residual contributions on parallel axial query planes. The decoder anchors candidate splats to an anatomical lattice and gates them by predicted organ probability and opacity.

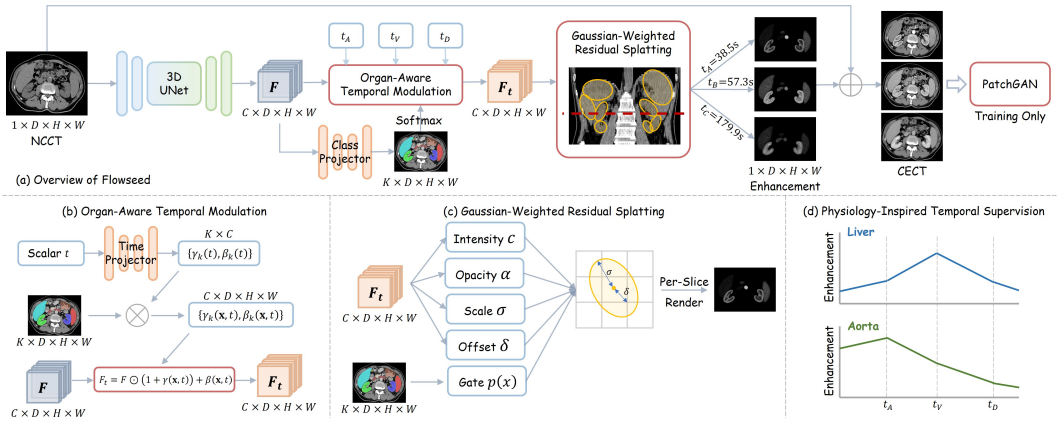


Figure 1: Overview of FlowSeed. (a) A shared 3D U-Net encodes NCCT X into the anatomical feature field F and predicts soft organ probabilities. Query times use the same organ-aware modulation and Gaussian-weighted residual-splatting decoder to produce enhancement volumes, which are added to X. The conditional PatchGAN is used only during training. (b) A scalar time produces class-specific affine parameters that are spatially blended by the soft organ probabilities to obtain F_t. (c) Four dense heads predict intensity c_i, opacity α_i, axis-aligned scale $\boldsymbol{\sigma}_i$, and center offset $\boldsymbol{\delta}_i$. The target-organ gate p_i and slice-aware renderer assemble the residual enhancement. (d) Reference-time renderings impose physiology-inspired ordinal relations for liver and aorta.

Our temporal conditioning is related to feature-wise affine modulation (Perez et al., 2018) and spatially adaptive conditioning (Park et al., 2019). The distinction is that a scalar time produces separate parameters for every anatomical class, while a predicted soft organ map blends them at each voxel. Modulation precedes splat-parameter prediction, so time changes intensity, opacity, scale, and center position throughout the rendered field.

3 Method

3.1 Problem Formulation

Let $Y(t) \in \mathbb{R}^{1\times D\times H\times W}$ denote a patient-specific CECT volume at post-injection time t, and let X denote the corresponding NCCT volume. After registration and resampling, X and each observed target $Y(t_m)$ share the same patient coordinate system. A conventional multiphase study provides only a few samples $\{Y(t_m)\}$, typically from the arterial, venous, and delayed phases, although contrast enhancement evolves continuously between these acquisitions.

Our goal is to learn one conditional mapping

$$\mathcal{G}_\theta : (X, t) \mapsto \widehat{Y}_\theta(t), \tag{1}$$

whose parameters are shared across all observed phases and query times. During training, the same model is supervised at the recorded times t_m available for each study. At inference, it can be evaluated at any query time within the modeled interval. This shared scalar-time parameterization captures within-phase timing variation and provides one directly queryable function across the conventional phases.

Because the NCCT already contains the patient anatomy, we factor both the target and its prediction into an anatomical base and a time-dependent contrast increment:

$$Y(t) = X + \Delta C(X, t), \qquad \widehat{Y}_\theta(t) = X + \widehat{\Delta C}_\theta(X, t). \tag{2}$$

The predicted contrast increment is constrained to be non-negative. The model learns the enhancement field rather than reconstructing the complete CT appearance at every time. For an axial coordinate z_q, the renderer produces the residual slice $\widehat{\Delta C}_{\theta,q}(X, t)$, which is added to the corresponding NCCT slice X_q. Stacking the queried slices forms $\widehat{Y}_\theta(t)$. The following sections describe how anatomy routes the temporal conditioning and how the resulting residual field is rendered.

3.2 Anatomy encoding and organ-aware temporal modulation

A 3D residual U-Net (Isensee et al., 2021) encodes X and returns a full-resolution feature field $F \in \mathbb{R}^{C\times D\times H\times W}$. An auxiliary $1\times 1\times 1$ convolution predicts logits for $K = 43$ segmentation classes: background and 42 anatomical labels from masks generated with TotalSegmentator v2 (Wasserthal et al., 2023). Their softmax probabilities are denoted by $S_k(\mathbf{x})$ at voxel $\mathbf{x}$. Appendix Table 5 lists the complete label map.

An MLP maps t to one scale and shift vector for every class:

$$\{\gamma_k(t), \beta_k(t)\}_{k=0}^{K-1} = \mathrm{MLP}(t), \qquad \gamma_k(t), \beta_k(t) \in \mathbb{R}^C. \tag{3}$$

The soft anatomical probabilities spatially route these parameters:

$$\begin{aligned} \big(\gamma(\mathbf{x}, t), \beta(\mathbf{x}, t)\big) &= \sum_{k=0}^{K-1} S_k(\mathbf{x})\big(\gamma_k(t), \beta_k(t)\big), \\ F_t(\mathbf{x}) &= F(\mathbf{x}) \odot (1 + \gamma(\mathbf{x}, t)) + \beta(\mathbf{x}, t). \end{aligned} \tag{4}$$

This operation preserves one shared anatomical encoding while allowing different structures to follow different time-dependent transformations.

3.3 Time-conditioned Gaussian-weighted residual splatting

We interpret each normalized axial coordinate z_q as a parallel query plane through the 3D residual-enhancement field. In place of the RGB appearance used in scene rendering, each candidate residual splat carries a view-independent scalar contrast intensity c_i. Its spatial Gaussian weights determine its contribution to the queried CT slice. Four convolutional heads transform F_t into a dense set of these candidate splats. At lattice location i, the model predicts contrast intensity c_i, opacity α_i, axis-aligned scale $\boldsymbol{\sigma}_i$, and center offset $\boldsymbol{\delta}_i$:

$$\begin{aligned} c_i &= \mathrm{softplus}(h_c(F_t)_i), & \alpha_i &= \mathrm{sigmoid}(h_\alpha(F_t)_i), \\ \boldsymbol{\sigma}_i &= 0.05 + 0.5\,\mathrm{sigmoid}(h_\sigma(F_t)_i), & \boldsymbol{\delta}_i &= 0.5\tanh(h_\delta(F_t)_i). \end{aligned} \tag{5}$$

The four prediction heads are shared across patients and produce the complete parameter field directly from F_t, so inference requires no patient-specific fitting.

Let $\mathcal{O}$ be the contrast-relevant classes, with label indices $\{1, 2, 3, 5, 7, 8, 9, 25, \ldots, 42\}$, and let $p_i = \sum_{k\in\mathcal{O}} S_k(i)$. We gate intensity and opacity separately as $\widetilde{c}_i = p_i c_i$ and $\widetilde{\alpha}_i = p_i\alpha_i$. Intensity gating limits residual magnitude outside the target anatomy, while opacity gating controls candidate retention and accumulated contribution. The splat center is $\boldsymbol{\mu}_i = \mathbf{g}_i + \boldsymbol{\delta}_i$, where $\mathbf{g}_i$ is a normalized voxel-grid coordinate. All four parameter fields depend on F_t and time.

For an axial slice at normalized coordinate z_q, let $\mathcal{A}_q$ be the candidates satisfying $|\mu_{i,z} - z_q| < 3\sigma_{i,z}$ and $\widetilde{\alpha}_i > 0.05$. Their through-plane weight is

$$w_{iq}^z = \exp\left(-\frac{(\mu_{i,z} - z_q)^2}{2(\sigma_{i,z}^2 + \epsilon)}\right). \tag{6}$$

For in-plane accumulation, the normalized center is mapped to raster coordinates as $r_{i,y} = (\mu_{i,y} + 1)(H-1)/2$ and $r_{i,x} = (\mu_{i,x} + 1)(W-1)/2$. Let $\mathcal{N}_i$ denote the 3×3 integer stencil centered at the rounded raster location. Define the in-plane weight

$$w_i^{xy}(\mathbf{u}) = \exp\left[-\frac{1}{2}\sum_{a\in\{x,y\}} \frac{(u_a - r_{i,a})^2}{\sigma_{i,a}^2 + \epsilon}\right] \mathbb{I}[\mathbf{u} \in \mathcal{N}_i]. \tag{7}$$

The rendered residual is the weighted sum of retained candidates:

$$\widehat{\Delta C}_{\theta,q}(X, t; \mathbf{u}) = \sum_{i\in\mathcal{A}_q} \widetilde{c}_i\widetilde{\alpha}_i w_{iq}^z w_i^{xy}(\mathbf{u}). \tag{8}$$

The fixed stencil bounds each candidate's in-plane support to a local (3×3) neighborhood around its predicted raster center. The scales $\sigma_{i,y}$ and $\sigma_{i,x}$ modulate the Gaussian weights within this neighborhood, while the center offset can move the contribution away from its lattice anchor. The resulting

operation is a local Gaussian-weighted scatter decoder: each dense lattice candidate predicts its destination, intensity, opacity, and axis-specific scales, and its contribution is accumulated by scatter addition. In contrast to a stationary convolution kernel, these destinations and weights vary spatially and with query time. As a result, camera projection, visibility ordering, and view-dependent appearance used in scene-rendering 3DGS are outside this slice-query formulation.

3.4 Learning objectives

Training samples axial slices containing target organs. Let Ω denote the sampled pixel domain and $\Omega_T \subset \Omega$ the pixels whose labels belong to $\mathcal{O}$. Reconstruction is evaluated on the synthesized CECT, $\widehat{Y}_q = X_q + \widehat{\Delta C}_{\theta,q}$:

$$\mathcal{L}_1 = \frac{1}{|\Omega_T|} \sum_{\mathbf{u}\in\Omega_T} \left| \widehat{Y}_q(\mathbf{u}) - Y_q(\mathbf{u}) \right|. \tag{9}$$

The image objective combines this reconstruction term with a gradient-difference loss used in medical image synthesis (Nie et al., 2018), weak residual shrinkage inside Ω_T, stronger suppression outside Ω_T, and cross-entropy supervision for the organ map:

$$\mathcal{L}_{\text{img}} = \lambda_{\text{rec}}\mathcal{L}_1 + \lambda_{\text{grad}}\mathcal{L}_{\text{grad}} + \lambda_{\text{in}}\mathcal{L}_{\text{sparse,in}} + \lambda_{\text{out}}\mathcal{L}_{\text{sparse,out}} + \lambda_{\text{seg}}\mathcal{L}_{\text{seg}}. \tag{10}$$

Appendix A.3 gives the exact definitions of these regularizers.

Let (t_A, t_V, t_D) denote the corresponding model inputs, and $E_o(t)$ be the mean rendered increment in organ o. We impose

$$\begin{aligned} \mathcal{L}_{\text{liver}} &= [E_{\text{liver}}(t_A) - E_{\text{liver}}(t_V) + m]_+ + [E_{\text{liver}}(t_D) - E_{\text{liver}}(t_V) + m]_+, \\ \mathcal{L}_{\text{aorta}} &= [E_{\text{aorta}}(t_V) - E_{\text{aorta}}(t_A) + m]_+ + [E_{\text{aorta}}(t_D) - E_{\text{aorta}}(t_V) + m]_+, \end{aligned} \tag{11}$$

where $[\cdot]_+$ is ReLU and $m = 0.03$ in normalized intensity units (approximately 9.6 HU). The complete model also uses a conditional 2D PatchGAN (Isola et al., 2017) to encourage local appearance fidelity. The discriminator receives the NCCT slice together with either the organ-masked real or synthesized CECT slice. The resulting least-squares generator loss is $\mathcal{L}_{\text{adv}}$. The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{img}} + \lambda_{\text{temp}}(\mathcal{L}_{\text{liver}} + \mathcal{L}_{\text{aorta}}) + \lambda_{\text{adv}}\mathcal{L}_{\text{adv}}. \tag{12}$$

4 Experiments

4.1 Datasets and Experimental Setup

We use three public multiphase CT datasets. MCT-LTDiag contains 517 four-phase liver-tumor studies and provides standardized registered volumes (Wu et al., 2026). After preprocessing and registration quality control, the patient-disjoint split contains 364 training studies and 147 held-out test studies, each with NCCT and arterial, venous, and delayed targets. UCSF RMaC contains 831 renal-mass exams with heterogeneous phase availability and per-volume registration status (Larson et al., 2025). Of these, 707 exams provide NCCT and at least one enhanced target. All 1,276 usable pairs (333 arterial, 523 venous, and 420 delayed) contribute additional training data. UCSF serves as an additional training cohort. Because official UCSF metadata do not include post-injection times, we estimate them with a gradient-boosted regressor using organ-level HU summary features. Appendix A gives further details. Finally, we use WAW-TACE as a generalization test set, a public multiphase abdominal CT dataset of 233 patients with hepatocellular carcinoma (Bartnik et al., 2024). Following a frozen registration-QC protocol, 548 phase pairs from 199 patients (196 arterial, 193 venous, and 159 delayed) are included in total. Cases with poor registration quality are excluded.

Volumes are resampled to 1.5 mm isotropic spacing. Intensity is standardized using a training-set mean of -74.48 HU and standard deviation of 319.06 HU. All experiments are conducted on 8 NVIDIA GeForce RTX 3090 GPUs. Training uses 96^3 patches, random affine transforms and flips, and 12 target-containing axial slices per volume. The model uses a six-stage residual U-Net as the encoder and a two-layer time MLP with a 64-dimensional hidden representation. We train for 1000 epochs with a per-GPU batch size of 4 using AdamW, an initial learning rate of 1×10^{-4}, weight decay 10^{-5}, and cosine annealing to one tenth of the initial rate. Each volume contributes 12 target-containing axial slices per iteration. At inference, overlapping 96^3 windows are combined with

Table 1: Comparison on the MCT-LTDiag test set. Values are macro-averaged over organs within each anatomical group. Best results are bold.

Group	Method	MAE (HU) ↓	RMSE (HU) ↓	PSNR-ST (dB) ↑	SSIM ↑	Mean-HU CCC ↑
Abdominal parenchyma	Pix2Pix	163.41	179.15	8.45	0.176	0.012
	SMILE	99.60	131.36	12.98	0.408	0.072
	MAFormer	48.48	60.43	17.77	0.451	0.165
	CFPS-Diff	49.87	62.79	17.29	0.522	0.213
	FlowSeed	**42.50**	**54.19**	**19.46**	**0.569**	**0.270**
Abdominal vessels	Pix2Pix	151.57	168.26	8.83	0.247	0.002
	SMILE	93.13	122.29	13.65	0.448	0.193
	CFPS-Diff	60.82	72.71	16.43	0.572	0.060
	MAFormer	61.71	74.84	16.24	0.468	0.120
	FlowSeed	**49.41**	**59.52**	**18.91**	**0.632**	**0.320**
Cardiothoracic vessels	Pix2Pix	235.12	252.06	6.76	0.144	0.070
	SMILE	124.95	156.30	11.56	0.407	−0.129
	CFPS-Diff	98.18	119.92	12.93	0.473	0.069
	MAFormer	83.42	98.69	14.18	0.419	0.118
	FlowSeed	**77.83**	**94.39**	**15.84**	**0.529**	**0.282**

Gaussian blending at 0.5 overlap. We use $(\lambda_{\text{rec}}, \lambda_{\text{grad}}, \lambda_{\text{in}}, \lambda_{\text{out}}, \lambda_{\text{seg}}, \lambda_{\text{temp}}) = (1, 1, 0.05, 5, 2, 2)$ and $\lambda_{\text{adv}} = 0.1$.

4.2 Baselines and metrics

We compare with Pix2Pix (Isola et al., 2017), MAFormer (Chen et al., 2024), CFPS-Diff (Zheng et al., 2025), and SMILE (Liu et al., 2025). All methods use the same training cohort and held-out MCT-LTDiag test split, with common source–target pairs and spatial preprocessing. Each baseline retains its intended architecture and training objective while its data interface is adapted to this common cohort. The categorical baselines generate the requested arterial, venous, or delayed phase. MAFormer uses three separately trained phase-specific models. FlowSeed instead receives the recorded acquisition time. We report MAE and RMSE in HU, soft-tissue-window PSNR (PSNR-ST) and SSIM, and CCC between predicted and real organ-level mean HU.

Abdominal parenchyma contains liver, spleen, bilateral kidneys, pancreas, and bilateral adrenal glands. Abdominal vessels contain aorta, inferior vena cava, portal/splenic vein, bilateral iliac arteries, and bilateral iliac veins. Cardiothoracic vessels contain heart, pulmonary vein, left atrial appendage, and superior vena cava.

5 Results

5.1 Comparison with prior methods

Table 1 shows that FlowSeed is best on every reported metric and group. Relative to the strongest comparator in each group, MAE decreases by 12.3% for abdominal parenchyma, 18.8% for abdominal vessels, and 6.7% for cardiothoracic vessels. The corresponding PSNR-ST gains are 1.69, 2.48, and 1.66 dB. The improvements extend beyond large parenchymal organs to vessels, where enhancement changes rapidly with acquisition time. Cardiothoracic structures remain the most difficult group in absolute terms.

Relative to the strongest comparator for this metric, Mean-HU CCC improves from 0.213 to 0.270 in abdominal parenchyma, from 0.193 to 0.320 in abdominal vessels, and from 0.118 to 0.282 in cardiothoracic vessels. These gains indicate better agreement of organ-level enhancement magnitude in addition to lower voxel error. Voxelwise metrics penalize HU discrepancies even when they are visually subtle after soft-tissue windowing, so qualitative comparison provides complementary evidence of appearance and anatomical fidelity.

Figure 2 shows held-out arterial-liver, venous-liver, and delayed-kidney examples. FlowSeed preserves the organ contour and surrounding context while concentrating lower error within the evaluated organs. The adjacent model without adversarial supervision provides a visual counterpart to the discriminator ablation, while the delayed example extends the comparison to late enhancement.

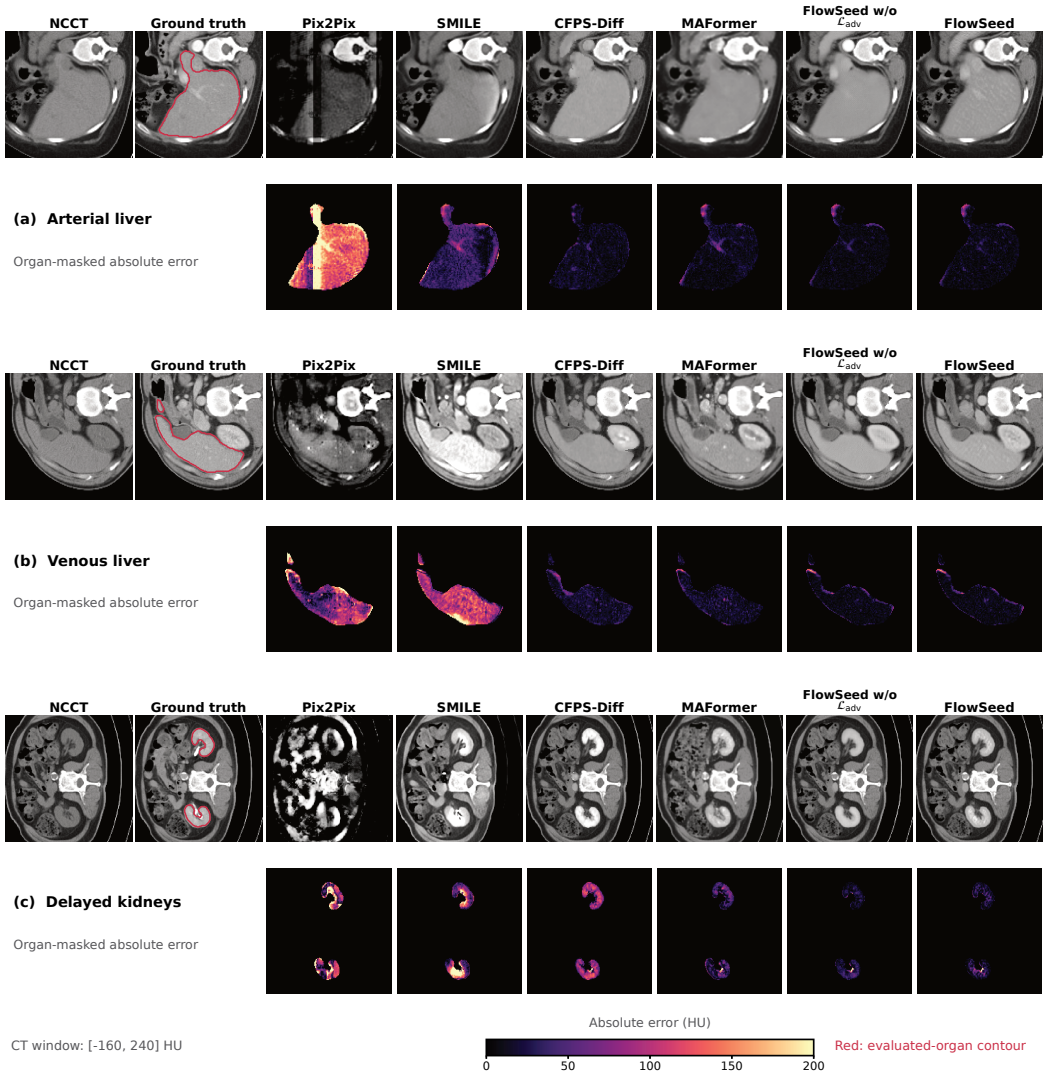


Figure 2: Qualitative comparison on MCT-LTDiag. Rows show arterial-liver, venous-liver, and delayed-kidney examples. Red contours identify the evaluated organ. The two rightmost columns compare FlowSeed without $\mathcal{L}_{\text{adv}}$ and the complete model. Both receive the recorded time, whereas categorical baselines receive the target phase.

These images illustrate the cohort-wide results in Table 1. Wider views and an additional venous-kidney example appear in Appendix C.

5.2 External evaluation on WAW-TACE

We test phase synthesis with the held out WAW-TACE dataset. The comparison results are shown in Table 2. Measured post-injection times are unavailable, so FlowSeed is conditioned with the preregistered reference-time surrogates of 35, 75, and 150 s. FlowSeed has the lowest MAE, highest SSIM, and highest Mean-HU CCC in all three groups on this complete external cohort. Relative to the strongest available baseline for each group, MAE decreases by 25.1% in parenchyma, 13.3% in abdominal vessels, and 22.2% in cardiothoracic vessels. SMILE is the strongest vascular baseline and approaches FlowSeed in cardiothoracic Mean-HU CCC (0.283 versus 0.287). The abdominal-vessel CCC remains only 0.058, so the external results support transfer of reconstruction fidelity more strongly than quantitative vascular enhancement agreement.

Table 2: External comparison on the full frozen WAW-TACE cohort. Each entry is MAE (HU) / SSIM / Mean-HU CCC.

Method	Parenchyma	Abd. vessels	Cardiothoracic
Pix2Pix	66.89 / 0.396 / 0.048	264.22 / 0.521 / −0.030	182.51 / 0.567 / 0.054
SMILE	68.78 / 0.537 / 0.316	187.59 / 0.690 / 0.045	112.63 / 0.740 / 0.283
MAFormer	40.20 / 0.519 / 0.434	239.01 / 0.614 / −0.013	153.44 / 0.674 / 0.083
CFPS-Diff	47.13 / 0.556 / 0.193	305.85 / 0.406 / −0.010	227.93 / 0.411 / 0.007
FlowSeed	**30.09 / 0.657 / 0.561**	**162.61 / 0.767 / 0.058**	**87.64 / 0.847 / 0.287**

Table 3: Ablation results. Each entry is MAE (HU) / SSIM / Mean-HU CCC.

Variant	Parenchyma	Abd. vessels	Cardiothoracic
w/o temporal loss	44.86 / 0.547 / 0.243	50.63 / 0.616 / 0.297	79.37 / 0.513 / 0.226
w/o discriminator	42.88 / **0.582** / 0.264	51.48 / 0.627 / 0.315	79.73 / **0.530** / 0.238
Global time modulation	47.88 / 0.539 / 0.229	58.21 / 0.594 / 0.236	82.92 / 0.506 / 0.186
Direct residual decoder	48.71 / 0.538 / 0.248	54.88 / 0.603 / 0.242	82.07 / 0.516 / 0.198
FlowSeed	**42.50** / 0.569 / **0.270**	**49.41** / **0.632** / **0.320**	**77.83** / 0.529 / **0.282**

5.3 Ablation studies

Removing the physiological temporal loss degrades all three metrics in every group (Table 3). The largest CCC change occurs in cardiothoracic vessels, from 0.282 to 0.226. The discriminator has a mixed effect. The complete model lowers MAE and improves CCC in all groups, but its parenchymal SSIM is lower than the model without the discriminator. The adjacent qualitative columns in Figure 2 show that both variants retain the same overall anatomy while differing locally in synthesized appearance. Together, these results are consistent with adversarial supervision changing local texture or intensity transitions. Replacing organ-routed modulation with one global time transform consistently degrades performance. In paired patient-level analysis, organ-aware routing improves MAE by 4.15–8.74 HU and SSIM by 0.024–0.038 across the three groups. All corresponding 95% bootstrap intervals exclude zero. Replacing Gaussian-weighted splatting with the direct residual decoder degrades MAE by 3.85–6.22 HU and SSIM by 0.015–0.031, while reducing Mean-HU CCC in every group. These results support both soft anatomical routing and structured residual splatting. Appendix B gives paired bootstrap intervals and phase-level results for the variants.

5.4 Temporal sensitivity

Table 4 summarizes a local ± 10-s oracle search on MCT-LTDiag. The oracle selects the prediction with the lowest target-image error and is therefore an upper bound. Improvements are modest: MAE decreases by 0.33, 1.23, and 0.65 HU in parenchyma, abdominal vessels, and cardiothoracic vessels, respectively. The largest change occurs for abdominal vessels, where Mean-HU CCC increases by 0.022. This is physiologically plausible because vascular enhancement changes rapidly around the bolus.

A key advance of FlowSeed is that a single model can synthesize CECT at arbitrary intermediate post-injection times, including timepoints without corresponding acquired targets in a patient's multiphase study. Figure 3 visualizes this capability through dense queries between the sparsely observed phases. Each case provides one observed A, V, and D acquisition, while the displayed columns show six direct queries across the 10–400 s range together with denser ROI trajectories. Kidney enhancement develops and later decreases, liver parenchyma responds more gradually, and the combined aorta–IVC ROI is strongest at early queries and declines thereafter. These within-case trajectories demonstrate an organ-dependent conditioning response between the sparse observations.

Taken together, the results support acquisition time as a directly queryable conditioning variable. The organ-routing and decoder ablations further show that this temporal control is most effective when it is spatially adapted to anatomy and expressed through a structured residual field. Residual

Table 4: Sensitivity to local oracle time search. "Oracle" selects the best prediction within ± 10 s using the target image.

Group	MAE: recorded / oracle	PSNR-ST: recorded / oracle	CCC: recorded / oracle
Parenchyma	42.50 / 42.17	19.46 / 19.52	0.270 / 0.277
Abd. vessels	49.41 / 48.18	18.91 / 19.08	0.320 / 0.342
Cardiothoracic	77.83 / 77.18	15.84 / 15.94	0.282 / 0.289

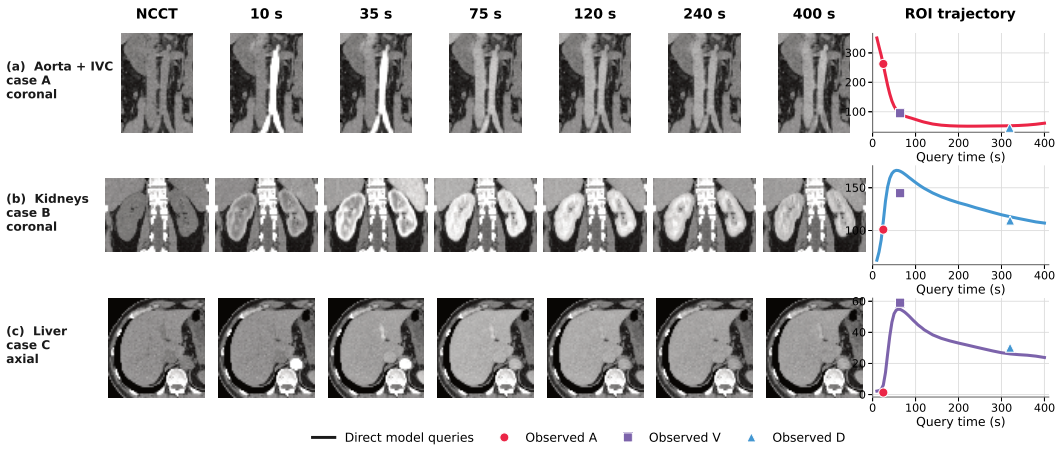


Figure 3: Dense temporal queries for anatomical visibility. Each row uses a fixed slice. Columns show the NCCT and direct model outputs at six query times. Curves report mean predicted enhancement relative to NCCT within the corresponding 3D ROI, and markers show the observed arterial (A), venous (V), and delayed (D) acquisitions at their recorded times.

composition preserves the NCCT as the anatomical base, while the dense queries reveal distinct kidney, liver, and vascular responses between the sparsely observed phases.

6 Conclusion

We presented FlowSeed, a continuous-time NCCT-to-CECT synthesis framework that separates anatomy from contrast enhancement. A predicted soft organ map routes time-dependent feature modulation, and the resulting features directly parameterize Gaussian-weighted splats for the residual contrast field. Reference-time physiological ordering supplies additional temporal supervision from model renderings. FlowSeed achieves the best performance, and produces organ-dependent changes across dense temporal queries. These findings establish continuous acquisition time and structured residual rendering as a promising formulation for whole-abdomen CECT synthesis. Future work can focus on the continuous-time synthesis of abdomen tumors.

AI use statement

Generative AI tools assisted with manuscript drafting and revision. The authors reviewed and verified all AI-assisted interpretations and take responsibility for all submitted content.

References

Zahra Alyani Nezhad, Giuseppe V. Toia, Sean D. Rose, Rachel M. Bladorn, and Timothy P. Szczykutowicz. Factors affecting contrast enhancement on CT images. *RadioGraphics*, 46(7), 2026. doi: 10.1148/rg.250192.

Krzysztof Bartnik, Tomasz Bartczak, Mateusz Krzyziński, Krzysztof Korzeniowski, Krzysztof Lamparski, Piotr Wegrzyn, Eric Lam, Mateusz Bartkowiak, Tadeusz Wróblewski, Katarzyna Mech, Magdalena Januszewicz, and Przemysław Biecek. WAW-TACE: A hepatocellular carcinoma multiphase CT dataset with segmentations, radiomics features, and clinical data. *Radiology: Artificial Intelligence*, 6(6):e240296, 2024. doi: 10.1148/ryai.240296.

Wenting Chen, Wei Zhao, Zhen Chen, Tianming Liu, Li Liu, Jun Liu, and Yixuan Yuan. Mask-aware transformer with structure invariant loss for CT translation. *Medical Image Analysis*, 96: 103205, 2024. doi: 10.1016/j.media.2024.103205.

Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021. doi: 10.1038/s41592-020-01008-z.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, 2017.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:14, 2023. doi: 10.1145/3592433.

Peder E. Z. Larson, Zhen Jane Wang, Sule Sahin, Ernesto Diaz, A. Rajagopal, Mansour Abtahi, Samantha Jones, Qing Dai, and Sage Kramer. UCSF RMaC: UCSF renal mass CT dataset. UCSF Center for Intelligent Imaging Datasets for Medical Imaging, 2025. URL `https://imagingdatasets.ucsf.edu/dataset/3`.

Junqi Liu, Zejun Wu, Pedro R. A. S. Bassi, Xinze Zhou, Wenxuan Li, Ibrahim E. Hamamci, Sezgin Er, Tianyu Lin, Yi Luo, Szymon Płotka, Bjoern Menze, Daguang Xu, Kai Ding, Kang Wang, Yang Yang, Yucheng Tang, Alan L. Yuille, and Zongwei Zhou. See more, change less: Anatomy-aware diffusion for contrast enhancement. *arXiv preprint arXiv:2512.07251*, 2025.

Dong Nie, Roger Trullo, Jun Lian, Li Wang, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, 65(12):2720–2730, 2018. doi: 10.1109/TBME.2018.2814538.

Emmanouil Nikolakakis, Utkarsh Gupta, Jonathan Vengosh, Justin Bui, and Razvan Marinescu. GaSpCT: Gaussian splatting for novel CT projection view synthesis. *arXiv preprint arXiv:2404.03126*, 2024.

Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2337–2346, 2019.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pp. 3942–3951, 2018. doi: 10.1609/aaai.v32i1.11671.

Amit G. Singal, Josep M. Llovet, Mark Yarchoan, Neil Mehta, Julie K. Heimbach, Laura A. Dawson, Janice H. Jou, Laura M. Kulik, Vatche G. Agopian, Jorge A. Marrero, Mishal Mendiratta-Lala, Daniel B. Brown, William S. Rilling, Lipika Goyal, Alice C. Wei, and Tamar H. Taddei. AASLD practice guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. *Hepatology*, 78(6):1922–1965, 2023. doi: 10.1097/HEP.0000000000000466.

Rebecca Smith-Bindman, Michelle Moghadassi, Nicole Wilson, Thomas R. Nelson, John M. Boone, Christopher H. Cagnon, Robert Gould, David J. Hall, Mayil Krishnam, Ramit Lamba, Michael McNitt-Gray, Anthony Seibert, and Diana L. Miglioretti. Radiation doses in consecutive CT examinations from five university of california medical centers. *Radiology*, 277(1):134–141, 2015. doi: 10.1148/radiol.2015142728.

Aart J. van der Molen, Annick A. J. M. van de Ven, Francisco Vega, Ilona A. Dekkers, Remy W. F. Geenen, Marie-France Bellin, Michele Bertolotto, Torkel B. Brismar, Olivier Clément, Jean-Michel Correas, Katerina Deike, Gertraud Heinz, Andreas H. Mahnken, Carlo A. Mallio,

Carlo C. Quattrocchi, Alexander Radbruch, Peter Reimer, Giles Roditi, Laura Romanini, Carmen Sebastià, and Fulvio Stacul. Hypersensitivity reactions to contrast media: Part 1. management of immediate and non-immediate hypersensitivity reactions in adults. updated guidelines by the ESUR contrast media safety committee. *European Radiology*, 35(11):6798–6810, 2025. doi: 10.1007/s00330-025-11675-1.

Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel T. Boll, Joshy Cyriac, Shan Yang, Michael Bach, and Martin Segeroth. Totalsegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5):e230024, 2023. doi: 10.1148/ryai.230024.

Xiang'an Wu, Haoyang Su, Yiwei Hua, Yali Xu, Lilong Wang, Xiaosong Wang, Shitian Wang, Bao Jin, Xiao Liu, Xueshuai Wan, Qiang Sun, Xuan Wang, and Shunda Du. A multi-phase CT dataset for automated differential diagnosis of liver tumors. *Scientific Data*, 13:31, 2026. doi: 10.1038/s41597-025-06343-4.

Kaiyi Zheng, Mu Huang, Xinming Li, Jianhua Ma, Qianjin Feng, Wei Yang, and Liming Zhong. Contrast flow pattern and cross-phase specificity-aware diffusion model for NCCT-to-multiphase CECT synthesis. In *Medical Image Computing and Computer Assisted Intervention*, volume 15963, pp. 89–99. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-032-04965-0_9.

Liming Zhong, Ruolin Xiao, Hai Shu, Kaiyi Zheng, Xinming Li, Yuankui Wu, Jianhua Ma, Qianjin Feng, and Wei Yang. NCCT-to-CECT synthesis with contrast-enhanced knowledge and anatomical perception for multi-organ segmentation in non-contrast CT images. *Medical Image Analysis*, 100:103397, 2025. doi: 10.1016/j.media.2024.103397.

Qikui Zhu, Hao Wu, Yanyan Zhang, and Shuo Li. Causality-driven spatio-temporal generator for multi-phase contrast-enhanced CT synthesis. In *Medical Image Computing and Computer Assisted Intervention*, volume 15962, pp. 110–119. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-032-04947-6_11.

A Training Data and Implementation Details

A.1 Organ Classes

Table 5 gives the exact mapping used in preprocessing, training, and evaluation. All classes participate in soft anatomical routing. The marked subset $\mathcal{O}$ is summed to gate the rendered contrast intensity and opacity.

A.2 Time Prediction

Official UCSF metadata do not provide post-injection acquisition times. We fit an XGBoost regressor on scans with recorded delays and apply it to the UCSF targets. Its inputs are organ-level HU summary statistics and intensity ratios derived from the aorta, portal vein, inferior vena cava, liver, spleen, kidneys, gallbladder, and pancreas.

We evaluated the frozen regressor on the held-out MCT-LTDiag test split, whose 147 patients did not overlap the fitting records. The evaluation comprised 441 enhanced volumes, with 147 observations per phase. Across all phases, the predicted times achieved an MAE of 14.91 s (95% CI: 12.17–17.82) with the recorded times. Table 6 compares the predictor with fixed phase centers of 35, 75, and 150 s. The predictor reduced overall MAE by 4.61 s, with the clearest improvement in the more variable delayed acquisitions. Arterial and venous paired intervals crossed zero.

Figure 4 shows the officially recorded delay metadata for MCT-LTDiag training split by phase.

A.3 Image objective

Section 3 defines the target-organ pixel set Ω_T, the sampled slice domain Ω, and the masked reconstruction loss $\mathcal{L}_1$. For direction $d \in \{x, y\}$, let $\mathcal{V}_d = \{\mathbf{u} : \mathbf{u}, \mathbf{u} + \mathbf{e}_d \in \Omega_T\}$. The gradient-difference

Table 5: Segmentation label map used by the organ-routing head. "Target" indicates membership in the contrast-relevant subset $\mathcal{O}$ used by the renderer gate.

ID	Class	Target	ID	Class	Target
0	Background	–	22	Prostate	–
1	Spleen	yes	23	Left kidney cyst	–
2	Right kidney	yes	24	Right kidney cyst	–
3	Left kidney	yes	25	Heart	yes
4	Gallbladder	–	26	Aorta	yes
5	Liver	yes	27	Pulmonary vein	yes
6	Stomach	–	28	Brachiocephalic trunk	yes
7	Pancreas	yes	29	Right subclavian artery	yes
8	Right adrenal gland	yes	30	Left subclavian artery	yes
9	Left adrenal gland	yes	31	Right common carotid artery	yes
10	Left upper lung lobe	–	32	Left common carotid artery	yes
11	Left lower lung lobe	–	33	Left brachiocephalic vein	yes
12	Right upper lung lobe	–	34	Right brachiocephalic vein	yes
13	Right middle lung lobe	–	35	Left atrial appendage	yes
14	Right lower lung lobe	–	36	Superior vena cava	yes
15	Esophagus	–	37	Inferior vena cava	yes
16	Trachea	–	38	Portal and splenic veins	yes
17	Thyroid gland	–	39	Left iliac artery	yes
18	Small bowel	–	40	Right iliac artery	yes
19	Duodenum	–	41	Left iliac vein	yes
20	Colon	–	42	Right iliac vein	yes
21	Urinary bladder	–			

Table 6: Acquisition-time prediction on held-out MCT-LTDiag. ΔMAE is the predictor MAE minus the fixed-center MAE; negative values favor the predictor.

Group	Predictor MAE (s)	Fixed-center MAE (s)	ΔMAE [95% CI] (s)
Overall	14.91	19.52	-4.61 [-8.02, -1.13]
Arterial	4.35	4.60	-0.26 [-1.25, 0.76]
Venous	9.90	8.90	1.00 [-2.41, 5.03]
Delayed	30.48	45.05	-14.57 [-22.33, -6.69]

term is

$$\mathcal{L}_{\text{grad}} = \sum_{d\in\{x,y\}} \frac{1}{|\mathcal{V}_d|} \sum_{\mathbf{u}\in\mathcal{V}_d} \left| \left|\widehat{Y}_q(\mathbf{u}+\mathbf{e}_d) - \widehat{Y}_q(\mathbf{u})\right| - |Y_q(\mathbf{u}+\mathbf{e}_d) - Y_q(\mathbf{u})| \right|. \tag{13}$$

The residual penalties are evaluated inside and outside the target anatomy:

$$\mathcal{L}_{\text{sparse,in}} = \frac{1}{|\Omega_T|} \sum_{\mathbf{u}\in\Omega_T} \left|\widehat{\Delta C}_{\theta,q}(\mathbf{u})\right|, \tag{14}$$

$$\mathcal{L}_{\text{sparse,out}} = \frac{1}{|\Omega \setminus \Omega_T|} \sum_{\mathbf{u}\in\Omega\setminus\Omega_T} \left|\widehat{\Delta C}_{\theta,q}(\mathbf{u})\right|. \tag{15}$$

For ground-truth class label $m(\mathbf{u})$, the organ-map head uses

$$\mathcal{L}_{\text{seg}} = -\frac{1}{|\Omega|} \sum_{\mathbf{u}\in\Omega} \log S_{m(\mathbf{u})}(\mathbf{u}). \tag{16}$$

For adversarial supervision, the generator minimizes the mean squared error between the PatchGAN output and an all-one patch map.

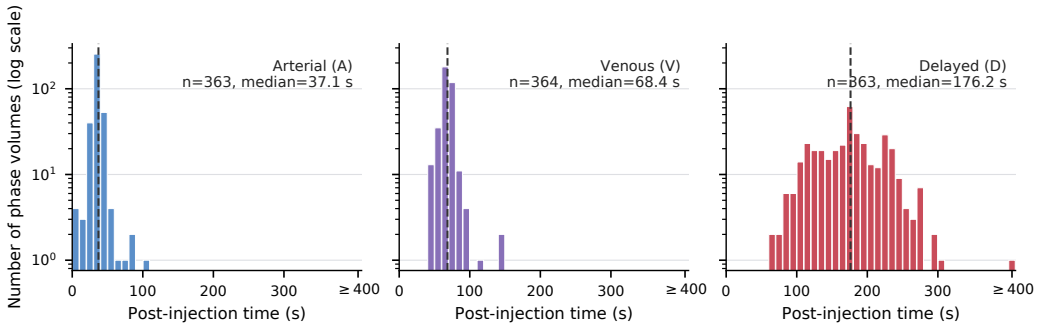


Figure 4: Recorded post-injection times in MCT-LTDiag training split. Dashed lines mark the phase medians. Counts use a logarithmic vertical scale. Values at or above the 400-s conditioning limit are collected in the terminal bin.

B Supplementary Results

Table 7 reports patient-bootstrap intervals for the aggregate FlowSeed results.

For each dataset and anatomical group, Table 8 selects the comparator with the lowest aggregate MAE. Every interval excludes zero for both MAE and SSIM. These paired effects complement the point estimates in the main comparison tables and retain the patient as the sampling unit.

Table 9 separates the main MCT-LTDiag comparison by target phase. FlowSeed achieves the lowest MAE in eight of nine group–phase comparisons and the highest SSIM in all nine. Mean-HU CCC is highest in seven of nine comparisons. MAFormer has the highest arterial parenchymal CCC and the lowest venous cardiothoracic MAE, while CFPS-Diff has the highest venous parenchymal CCC. The phase-level analysis supports consistent structural fidelity while exposing metric-specific exceptions in enhancement concordance.

Table 10 gives paired effect sizes for the complete model relative to each controlled variant. Organ-aware modulation and Gaussian-weighted residual splatting improve both MAE and SSIM in every group, with all corresponding intervals excluding zero. The temporal loss also improves both metrics throughout. The discriminator produces a trade-off between fidelity and structure. It lowers MAE in every group, but the variant without it has higher parenchymal SSIM, and the cardiothoracic SSIM interval includes zero.

Table 11 gives the WAW-TACE comparison by target phase. FlowSeed has the lowest MAE and highest SSIM in all nine group–phase comparisons. Mean-HU CCC is highest in seven of nine comparisons: SMILE has a slightly higher arterial cardiothoracic CCC and delayed abdominal-vessel CCC. The delayed abdominal-vessel CCC is near zero for all methods, reinforcing the distinction between reconstruction fidelity and quantitative vascular enhancement agreement under this external registration protocol.

C Full-View Qualitative Comparisons

Figures 5 and 6 retain the wider anatomical context for the examples for visual inspection. Q1, Q3, and Q4 correspond to the organ-centered comparisons in Figure 2.

Table 7: Aggregate FlowSeed results with patient-cluster bootstrap 95% confidence intervals.

Dataset	Anatomical group	MAE (HU) ↓	Bias (HU) → 0	SSIM ↑	Mean-HU CCC ↑
MCT-LTDiag	Parenchyma	42.50 [38.82, 46.62]	3.94 [−1.70, 9.92]	0.569 [0.538, 0.597]	0.270 [0.184, 0.354]
MCT-LTDiag	Abd. vessels	49.41 [44.70, 53.87]	7.75 [−0.24, 15.71]	0.632 [0.592, 0.675]	0.320 [0.254, 0.396]
MCT-LTDiag	Cardiothoracic	77.83 [72.63, 83.28]	−28.33 [−38.41, −17.92]	0.529 [0.495, 0.561]	0.282 [0.210, 0.345]
WAW-TACE	Parenchyma	30.09 [28.91, 31.40]	8.28 [6.14, 10.47]	0.657 [0.644, 0.669]	0.561 [0.506, 0.622]
WAW-TACE	Abd. vessels	162.61 [116.75, 213.36]	121.41 [74.63, 174.15]	0.767 [0.725, 0.807]	0.058 [0.007, 0.191]
WAW-TACE	Cardiothoracic	87.64 [75.67, 100.96]	50.43 [38.29, 64.05]	0.847 [0.830, 0.862]	0.287 [0.207, 0.452]

Table 8: Paired improvements of FlowSeed in each dataset and anatomical group. Values are mean improvement [95% bootstrap CI].

Dataset	Anatomical group	Comparator	ΔMAE (HU) ↑	ΔSSIM ↑
MCT-LTDiag	Parenchyma	MAFormer	5.99 [4.65, 7.25]	0.118 [0.104, 0.134]
MCT-LTDiag	Abd. vessels	CFPS-Diff	11.30 [8.47, 14.04]	0.060 [0.054, 0.066]
MCT-LTDiag	Cardiothoracic	MAFormer	4.98 [1.43, 8.37]	0.106 [0.090, 0.122]
WAW-TACE	Parenchyma	MAFormer	10.23 [8.79, 11.59]	0.138 [0.127, 0.149]
WAW-TACE	Abd. vessels	SMILE	27.56 [23.07, 32.50]	0.080 [0.070, 0.090]
WAW-TACE	Cardiothoracic	SMILE	25.07 [22.56, 27.39]	0.105 [0.096, 0.114]

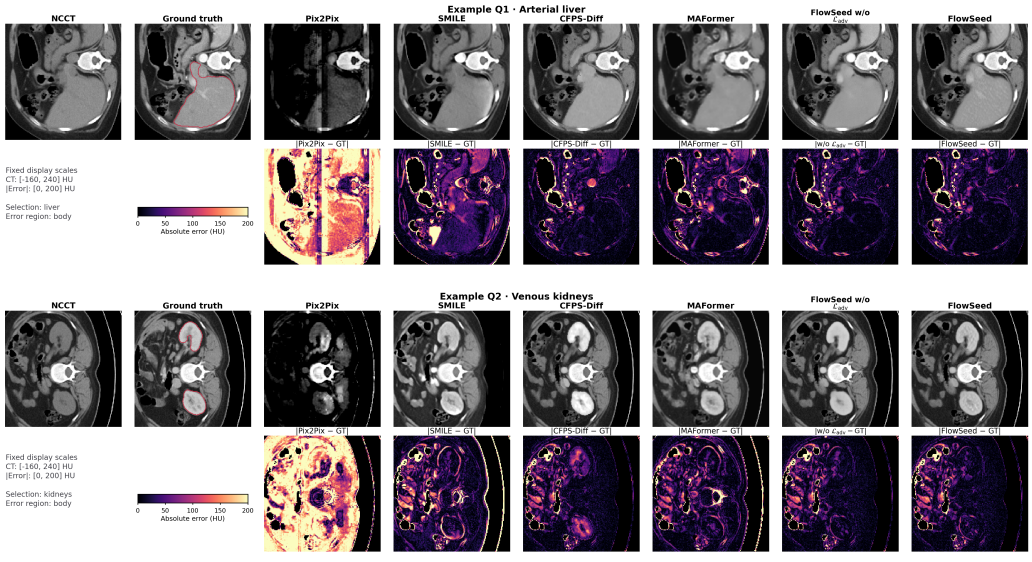


Figure 5: Full-view qualitative comparisons for arterial liver Example Q1 (top) and venous kidneys Example Q2 (bottom).

Table 9: Phase-stratified MCT-LTDiag results. Each entry is MAE (HU) / SSIM / Mean-HU CCC. Metrics are macro-averaged over organs within each anatomical group. The best component in each phase and group is bold.

Phase	Method	Parenchyma	Abd. vessels	Cardiothoracic
A	Pix2Pix	181.26 / 0.147 / 0.007	173.76 / 0.234 / 0.010	237.65 / 0.155 / 0.033
A	SMILE	109.20 / 0.388 / 0.012	123.45 / 0.455 / 0.014	166.63 / 0.357 / 0.070
A	MAFormer	47.52 / 0.461 / **0.244**	78.93 / 0.445 / 0.041	121.14 / 0.363 / 0.072
A	CFPS-Diff	53.93 / 0.494 / 0.214	83.92 / 0.510 / 0.045	138.30 / 0.428 / 0.070
A	**FlowSeed**	**43.27** / **0.562** / 0.234	**60.99** / **0.625** / **0.120**	**108.99** / **0.480** / **0.178**
V	Pix2Pix	145.76 / 0.223 / 0.025	164.47 / 0.254 / 0.012	182.73 / 0.204 / 0.011
V	SMILE	97.56 / 0.423 / 0.051	89.21 / 0.443 / -0.065	118.96 / 0.432 / 0.002
V	MAFormer	50.22 / 0.436 / -0.023	57.80 / 0.478 / -0.039	**68.37** / 0.435 / 0.010
V	CFPS-Diff	47.72 / 0.549 / **0.198**	52.92 / 0.618 / 0.056	85.07 / 0.498 / 0.003
V	**FlowSeed**	**42.13** / **0.576** / 0.174	**46.73** / **0.646** / **0.063**	69.32 / **0.543** / **0.054**
D	Pix2Pix	163.22 / 0.157 / 0.010	116.49 / 0.253 / 0.062	284.99 / 0.073 / 0.022
D	SMILE	92.04 / 0.414 / 0.105	66.71 / 0.444 / -0.023	89.25 / 0.432 / 0.116
D	MAFormer	47.71 / 0.456 / 0.158	48.39 / 0.482 / 0.002	60.75 / 0.460 / 0.117
D	CFPS-Diff	47.98 / 0.524 / 0.235	45.63 / 0.588 / 0.105	71.18 / 0.492 / 0.131
D	**FlowSeed**	**42.09** / **0.569** / **0.245**	**40.51** / **0.626** / **0.132**	**55.18** / **0.563** / **0.284**

Table 10: Paired ablation effects on MCT-LTDiag. Each entry is the mean improvement of complete FlowSeed [95% bootstrap CI].

Comparator	Anatomical group	ΔMAE (HU) $\uparrow$	ΔSSIM $\uparrow$
w/o temporal loss	Parenchyma	2.36 [1.80, 2.94]	0.022 [0.021, 0.024]
w/o temporal loss	Abd. vessels	1.23 [0.70, 1.78]	0.017 [0.015, 0.018]
w/o temporal loss	Cardiothoracic	1.45 [0.77, 2.10]	0.018 [0.015, 0.021]
w/o discriminator	Parenchyma	0.38 [0.10, 0.67]	-0.012 [-0.014, -0.011]
w/o discriminator	Abd. vessels	2.05 [1.44, 2.66]	0.005 [0.004, 0.007]
w/o discriminator	Cardiothoracic	1.35 [0.39, 2.28]	-0.002 [-0.005, 0.001]
Global time modulation	Parenchyma	5.38 [4.65, 6.14]	0.030 [0.028, 0.033]
Global time modulation	Abd. vessels	8.74 [7.95, 9.57]	0.038 [0.036, 0.041]
Global time modulation	Cardiothoracic	4.15 [2.26, 6.01]	0.024 [0.018, 0.029]
Direct residual decoder	Parenchyma	6.22 [5.40, 7.06]	0.031 [0.029, 0.033]
Direct residual decoder	Abd. vessels	5.47 [4.60, 6.27]	0.029 [0.027, 0.031]
Direct residual decoder	Cardiothoracic	3.85 [2.48, 5.18]	0.015 [0.011, 0.019]

Table 11: Phase-stratified WAW-TACE results on the frozen 548-volume cohort. Each entry is MAE (HU) / SSIM / Mean-HU CCC. The best component in each phase and group is bold.

Phase	Method	Parenchyma	Abd. vessels	Cardiothoracic
A	Pix2Pix	51.91 / 0.506 / 0.179	212.39 / 0.668 / 0.159	168.75 / 0.708 / 0.062
A	SMILE	81.97 / 0.522 / 0.164	143.44 / 0.733 / 0.391	107.32 / 0.768 / **0.364**
A	MAFormer	42.66 / 0.527 / 0.393	199.95 / 0.635 / 0.217	154.62 / 0.676 / 0.107
A	CFPS-Diff	49.17 / 0.538 / 0.162	270.18 / 0.410 / 0.026	227.57 / 0.401 / −0.002
A	**FlowSeed**	**33.12 / 0.647 / 0.508**	**121.18 / 0.796 / 0.434**	**86.13 / 0.851** / 0.356
V	Pix2Pix	56.14 / 0.364 / 0.032	258.62 / 0.480 / −0.035	160.93 / 0.548 / 0.061
V	SMILE	59.10 / 0.538 / 0.269	177.12 / 0.703 / 0.083	91.48 / 0.752 / 0.480
V	MAFormer	42.33 / 0.494 / 0.346	234.20 / 0.623 / −0.011	135.03 / 0.690 / 0.111
V	CFPS-Diff	46.92 / 0.563 / 0.198	304.15 / 0.412 / −0.003	214.37 / 0.424 / 0.014
V	**FlowSeed**	**29.80 / 0.665 / 0.471**	**151.45 / 0.777 / 0.091**	**66.55 / 0.864 / 0.489**
D	Pix2Pix	98.42 / 0.297 / −0.103	340.28 / 0.382 / −0.066	226.46 / 0.415 / 0.020
D	SMILE	64.28 / 0.556 / 0.243	259.97 / 0.614 / **0.002**	145.58 / 0.689 / 0.204
D	MAFormer	34.58 / 0.538 / 0.423	297.12 / 0.573 / −0.046	174.95 / 0.653 / 0.059
D	CFPS-Diff	44.88 / 0.570 / 0.232	355.34 / 0.392 / −0.015	245.29 / 0.408 / 0.011
D	**FlowSeed**	**26.70 / 0.658 / 0.508**	**232.31 / 0.717** / 0.001	**115.84 / 0.821 / 0.216**

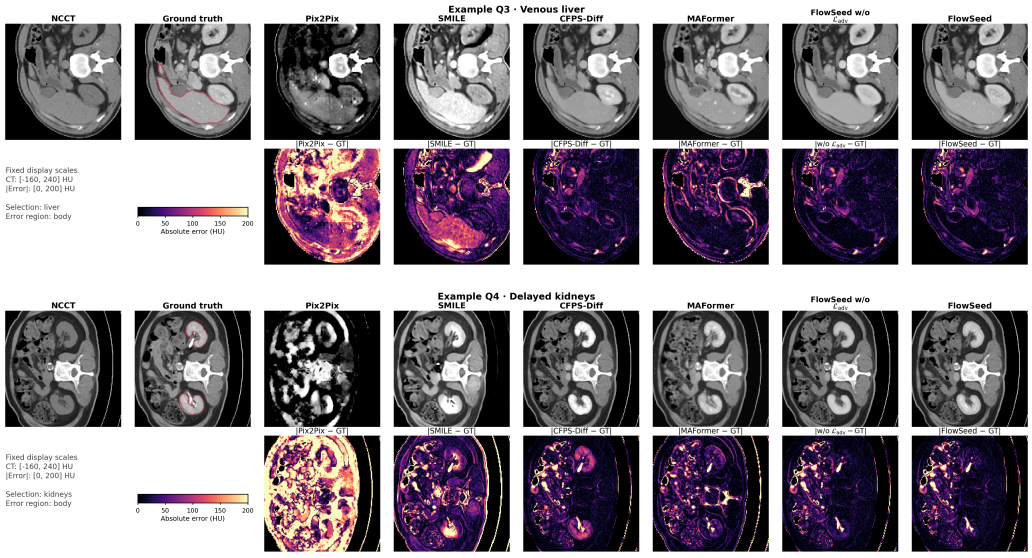


Figure 6: Full-view qualitative comparisons for venous liver Example Q3 (top) and delayed kidneys Example Q4 (bottom).

D ADDITIONAL DENSE TEMPORAL QUERIES

Figures 7–11 show ten representative examples for anatomical visibility. These examples illustrate organ-dependent dense query behavior between the sparsely observed phases.

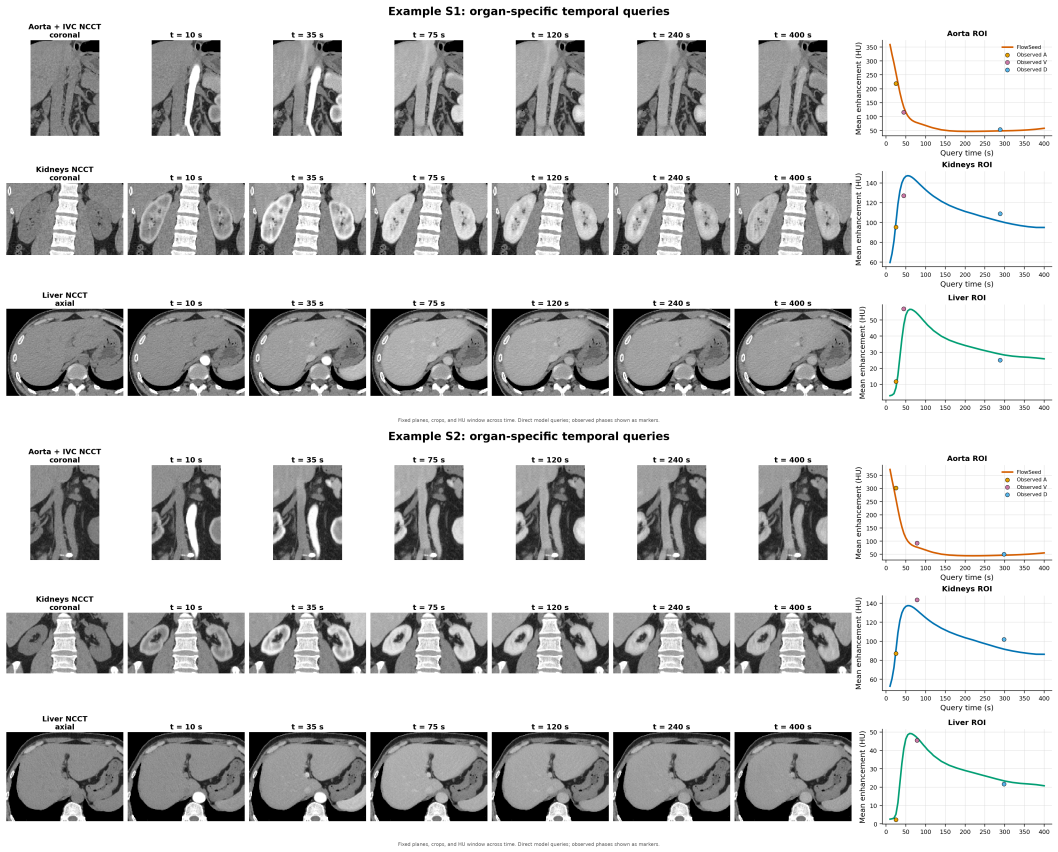


Figure 7: Additional dense temporal queries for Examples S1 (top) and S2 (bottom).

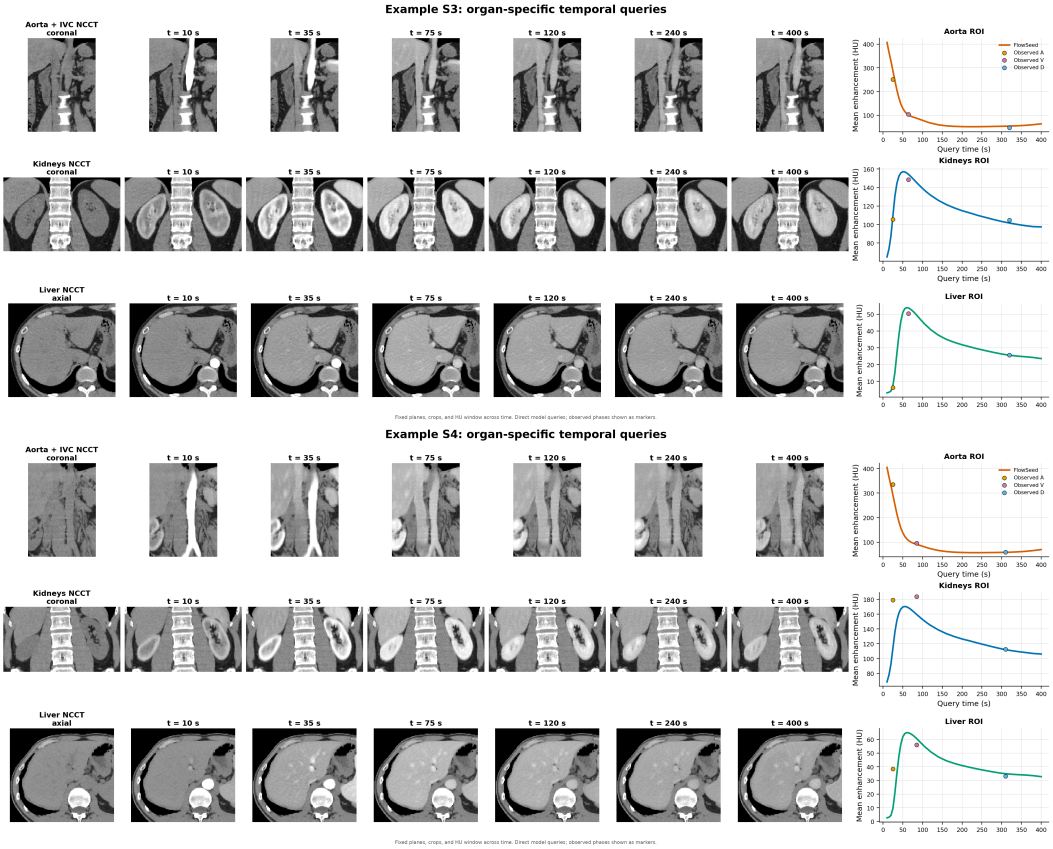


Figure 8: Additional dense temporal queries for Examples S3 (top) and S4 (bottom).

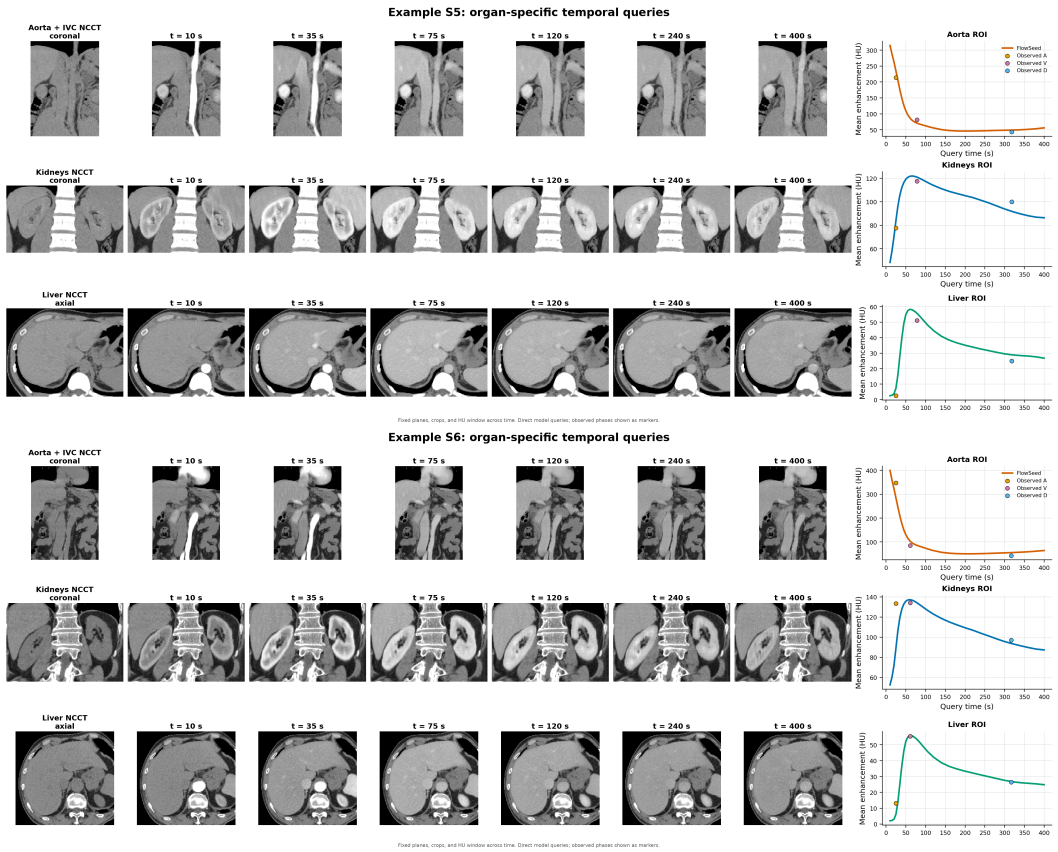


Figure 9: Additional dense temporal queries for Examples S5 (top) and S6 (bottom).

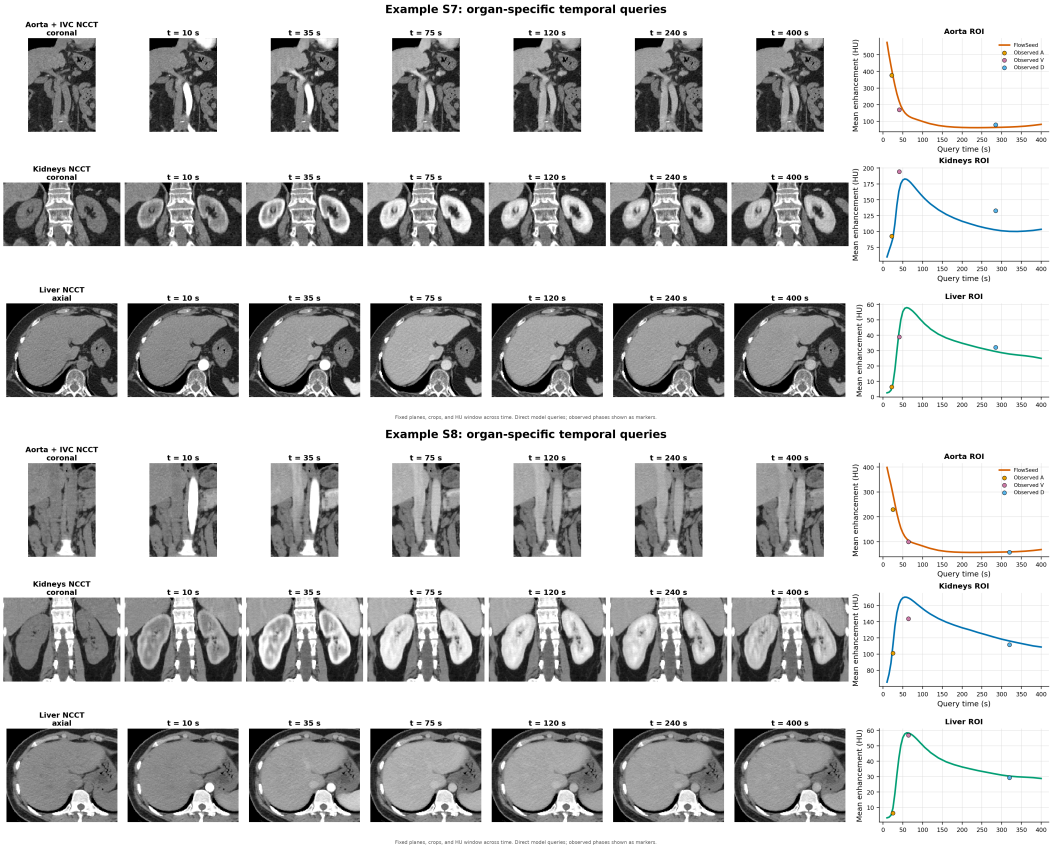


Figure 10: Additional dense temporal queries for Examples S7 (top) and S8 (bottom).

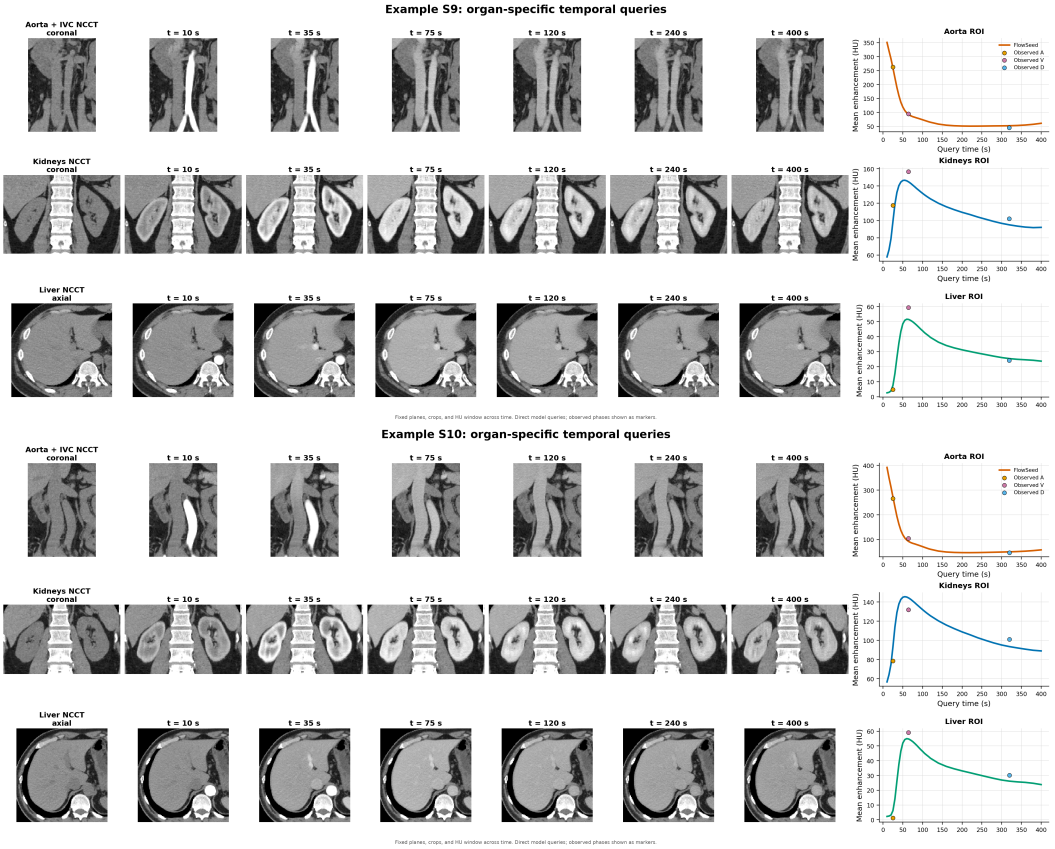


Figure 11: Additional dense temporal queries for Examples S9 (top) and S10 (bottom).