FROM FORWARD PREDICTION TO SELF-CORRECTIVE WORLD MODELING FOR MRI CONTRAST ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models offer strong visual quality for video generation and are promising for MRI enhancement world modeling. However, existing diffusion-based MRI enhancement world models remain passive forward simulators: they can predict future states, but cannot recognize and correct emerging kinetic errors. For the first time, we propose ReKinetics, a self-corrective world model that extends diffusion-based world modeling from forward prediction to explicit error correction. For self-correction *action supervision*, we introduce Corrective Credit Assignment Learning (CCAL), which shifts supervision from trajectory-level error fitting to action-level corrective credit assignment, turning latent kinetic failures into executable *where–when–how* action targets. For *action execution*, we introduce Factorized Corrective Transition Learning (FCTL), which replaces the shared denoising transition with a factorized corrective transition parameterized by state-dependent activation, spatial support, and action-specific residual transformation, decoupling heterogeneous corrective effects within diffusion. For *action feedback*, we introduce Reliability-Guided Corrective Optimization (RGCO), which shifts optimization from prediction-error-driven learning to correction-outcome-driven learning by converting reliability changes into trajectory-specific optimization weights. Together, these mechanisms form a closed self-corrective learning loop. Across Hand DCE-MRI and Duke Breast, ReKinetics achieves strong frame and kinetic performance. On the Hand benchmark, it reduces late-vs.-early ROI kinetic error by 42.1%, whereas competing methods accumulate 23.7%–44.7% error, confirming that explicit kinetic correction substantially benefits dynamic contrast-enhancement modeling.

1 INTRODUCTION

World models LeCun et al. (2022); Kang et al. (2024); Mendonca et al. (2023) learn to simulate how the state of an underlying system evolves under different conditions. Among them, diffusion-based world models Gao et al. (2024); Kong et al. (2026) have emerged as a powerful paradigm for visual dynamics modeling, leveraging the strong distribution-modeling capability of diffusion models Ho et al. (2020); Rombach et al. (2022); Blattmann et al. (2023) to generate high-fidelity and temporally coherent future observations. This paradigm is particularly promising for dynamic contrast-enhanced MRI (DCE-MRI). The model directly infers the kinetic evolution of contrast agents from initial non-contrast MRI, enabling estimation of contrast distribution at arbitrary time points and synthesis of corresponding contrast-enhanced MRI images. Its

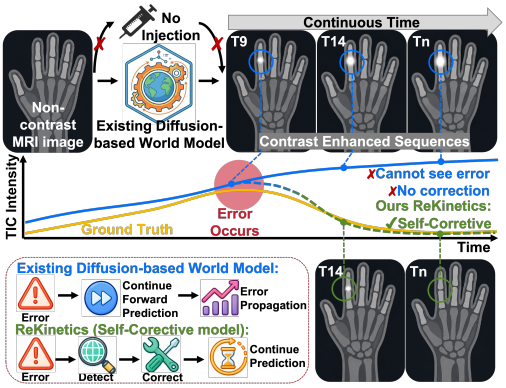

Figure 1: Motivation of ReKinetics. Conventional diffusion-based world models rely on forward prediction, where kinetic errors can propagate despite visually plausible frames. ReKinetics instead introduces a self-corrective loop that detects and revises kinetic deviations during trajectory generation.

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.

goal is not merely visual plausibility, but also capturing region-specific enhancement trends over time through time-intensity curves (TIC) Kong et al. (2026). Such a formulation offers two major advantages (Fig. 1). First, it enables contrast-free MRI by reconstructing enhancement states without additional contrast administration, reducing injection-related risks and acquisition burden Müller-Franzes et al. (2023); Shankar et al. (2018). Second, it supports high-temporal-resolution enhancement modeling by generating dense enhancement trajectories beyond sparsely acquired clinical phases Ingrisch & Sourbron (2013); Strijkers et al. (2007); Kong et al. (2026). Together, these capabilities make diffusion-based world models a compelling framework for simulating contrast-enhancement evolution and synthesizing dynamic contrast-enhanced MRI from non-contrast observations Kong et al. (2026).

In diffusion based world modeling, future visual states are typically predicted from the current observation together with conditioning signals such as steering, navigation commands, or robotic actions Gao et al. (2024); Rigter et al. (2024); Bar et al. (2025); Bruce et al. (2024). When applied to MRI contrast enhancement, the conditioning signal changes from external control to acquisition time, while the model similarly predicts how an initial non contrast MRI evolves into subsequent enhancement states (Fig. 1). These predicted enhancement states jointly define the evolving kinetic context for trajectory modeling. Once early predictions deviate from the underlying dynamics, the estimated kinetic context becomes biased, allowing the deviation to persist across later phases. This limitation affects not only generic video diffusion models Gao et al. (2024); Blattmann et al. (2023) but also specialized DCE-MRI predictors Kong et al. (2026), where the same forward-only objective is applied without an explicit correction mechanism. More importantly, forward prediction methods provide no explicit mechanism to recognize that such a deviation has occurred or to correct it before further prediction. This limitation is particularly consequential in DCE-MRI, where small errors in early enhancement can propagate across temporally coupled phases and progressively distort regional TIC and the overall kinetic trajectory (Fig. 1).

This raises a more fundamental question: *Can a world model go beyond forward prediction and close the loop by detecting errors and correcting its own trajectory?*

To answer this question, *for the first time,* we propose ReKinetics, a self-corrective kinetic world model for DCE-MRI contrast-enhancement modeling. Inspired by how a learner improves through mistakes, we reformulate kinetic trajectory generation as a closed supervision–execution–feedback learning cycle. During training, trajectory failures are converted into targeted where–when–how corrective supervision, executed within the diffusion transition, and verified through their subsequent outcomes. The learned corrective policy is then internalized into the model, enabling reference-free corrective execution at inference. This requires three fundamental learning changes: 1) corrective supervision is intrinsically underdetermined because the desired intervention is latent and would otherwise require counterfactual evaluation over alternative trajectory rollouts; we address this by reformulating trajectory failure learning as structured corrective credit assignment, making the latent responsibility of executable interventions learnable from observed failures. 2) heterogeneous corrective effects are entangled within a shared denoising transition; we reparameterize correction into independently controllable when–where–how degrees of freedom, allowing different failures to induce distinct corrective state transitions. 3) conventional prediction-error optimization cannot determine whether an executed correction is actually effective; we therefore elevate correction outcomes to optimization signals, allowing the model to adapt subsequent learning according to what has or has not been successfully corrected. Together, these learning reformulations extend diffusion-based world modeling from passive forward prediction to explicit self-correction.

Our contributions are summarized as follows: 1) For the first time, we propose ReKinetics, the self-corrective world model for DCE-MRI contrast-enhancement modeling, which extends forward prediction with internal corrective actions to revise kinetic errors during trajectory generation. 2) Our Corrective Credit Assignment Learning (CCAL) establishes a new action-level supervision paradigm for DCE-MRI world modeling by using a structured Failure Card to assign latent corrective credit from trajectory failures to executable *where–when–how* action targets without native action annotations. 3) Our Factorized Corrective Transition Learning (FCTL) replaces the shared corrective pathway with a factorized action space parameterized by state-dependent activation, spatial support, and action-specific residual transformation, enabling executable corrective transitions within diffusion. 4) Our Reliability-Guided Corrective Optimization (RGCO) converts correction-induced reliability changes into trajectory-specific optimization weights, making correction outcomes directly guide subsequent learning and close the self-corrective loop.

2 RELATED WORK

DCE-MRI world models increasingly capture temporal enhancement dynamics, but remain primarily forward predictive and lack an explicit mechanism for correcting kinetic failures that emerge within the generated trajectory; ReKinetics addresses this gap by detecting such failures and executing corrective actions during trajectory modeling. Early MRI synthesis methods learned mappings between predefined contrasts using GANs and multi-task learning Dar et al. (2019); Wang et al. (2020); Yurt et al. (2021), while virtual contrast-enhancement methods introduced volumetric architectures, tumor-aware attention, and region-specific objectives Chen et al. (2022); Kim et al. (2022). Recent work extends this setting to dynamic DCE-MRI through early/late phase prediction and time-conditioned enhancement synthesis Ramanarayanan et al. (2024); Osuala et al. (2024; 2025). Medical world models further capture temporal state evolution Yue et al. (2025); Shin et al. (2025); Kong et al. (2026), including continuous contrast-enhancement kinetics in MRI CEKWorld Kong et al. (2026). However, kinetic deviations are still handled through generation objectives rather than explicit corrective actions.

Action-conditioned world models can guide generation, but their control signals are externally provided rather than derived from failures emerging within the model's own trajectory. Consequently, when the predicted trajectory deviates, the model lacks an internal mechanism to identify the failure and translate it into a corrective action, allowing the error to persist in subsequent prediction. ReKinetics addresses this limitation by converting kinetic failures into corrective supervision and distilling the resulting intervention decisions into internal control. Action-conditioned world models use navigation commands, driving controls, or learned interaction signals to guide future-state prediction Gao et al. (2024); Bruce et al. (2024); Bar et al. (2025); Rigter et al. (2024); Zhang et al. (2025), but these actions are not generated from prediction failures. Medical VLMs can provide diagnostic reasoning and structured feedback Li et al. (2023); Tu et al. (2024); Bannur et al. (2024), but this feedback is not normally internalized as executable corrective control. ReKinetics uses a diagnostic VLM only during training to convert kinetic failures into failure cards and distills their corrective decisions into the internal router and KAN controller, enabling correction without an external VLM or reference trajectory at inference.

3 SELF-CORRECTIVE KINETIC WORLD MODEL

As shown in Fig. 2, ReKinetics implements self-corrective kinetic world modeling by converting action-unsupervised trajectory failures into executable corrective supervision through a structured Failure Card (CCAL, Sec. 3.1), factorizing the corrective effect into state-dependent activation, spatial support, and residual transformation within the diffusion transition (FCTL, Sec. 3.2), and feeding correction-induced reliability changes back into subsequent optimization (RGCO, Sec. 3.3).

3.1 CORRECTIVE CREDIT ASSIGNMENT LEARNING FOR *Action Supervision*

CCAL is the first formulation in DCE-MRI world modeling that converts action-unsupervised trajectory-level kinetic failures into executable corrective-action supervision. It resolves the supervision-level mismatch between observable trajectory errors and unavailable action labels by making latent corrective credit explicitly learnable through a structured Failure Card. The formulation proceeds from underdetermined action supervision, through Failure Card construction and failure-to-action credit assignment, to Failure Card distillation into the internal router.

Underdetermined Corrective Action Supervision. DCE-MRI provides the target trajectory Y_p but no corrective-action label A_p^* for a generated failure. Thus, $\mathcal{D}(x_p, \hat{Y}_p^{(r)}, Y_p)$ reveals what is wrong, but not which action should correct it. Ideally,

$$A_p^* \in \arg\min_{A \in \mathcal{R}_T} \mathcal{L}_{\text{kin}}(G_\theta(x_p, \mathcal{T}_p; A), Y_p)\,, \tag{1}$$

but obtaining such supervision requires impractical action-conditioned counterfactual trajectories. The Failure Card instead represents this latent corrective responsibility as explicit action-selection and spatial-support targets.

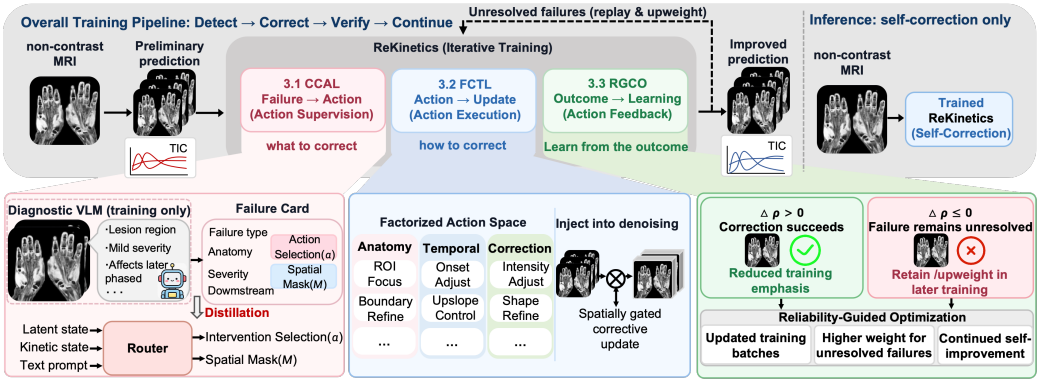

Figure 2: Overview of ReKinetics. ReKinetics realizes self-correction through three coupled learning mechanisms: CCAL converts action-unsupervised trajectory failures into executable corrective-action supervision via the Failure Card; FCTL reparameterizes shared denoising as a factorized corrective transition for action execution; and RGCO converts correction-induced reliability changes into trajectory-specific optimization weights, closing the self-corrective learning loop.

Failure Card Construction. A training-only diagnostic VLM constructs the Failure Card from the current trajectory discrepancy and executable action set:

$$O_p^{(r)} = \{\mathcal{T}_p, K_p, \mathcal{R}_T, \mathcal{D}(x_p, \hat{Y}_p^{(r)}, Y_p)\}, \tag{2}$$

where $\mathcal{D}$ extracts spatial and kinetic evidence. The diagnostic protocol and executable action registry are detailed in Appendix A. These observations form

$$\mathcal{C}_p^{(r)} = (\mathcal{F}_p^{(r)}, \mathcal{A}_p^{(r)}), \mathcal{F}_p^{(r)} = \{q_{p,i}, h_{p,i}, s^D_{p,i,m}, \gamma^D_{p,i\to j}\}_{i,m,j}, \mathcal{A}_p^{(r)} = \{\alpha^D_{p,i,m}, M^D_{p,i,m}\}_{i,m}. \tag{3}$$

Here, q, h, i, s^D, and γ^D encode failure type, anatomy, phase, severity, and downstream influence, respectively, providing the *how–where–when–strength–downstream* attribution required for corrective assignment.

Failure-to-Action Credit Assignment. $\mathcal{A}_p^{(r)}$ maps this attribution to executable actions in $\mathcal{R}_T$, where $\alpha^D_{p,i,m}$ assigns corrective credit and $M^D_{p,i,m}$ specifies its spatial support. Restricting assignment to $\mathcal{R}_T$ yields executable *where–when–how* action targets. Reconstructing $\mathcal{C}_p^{(r)}$ from $\hat{Y}_p^{(r)}$ further updates the assigned credit as failures evolve across rounds.

Failure Card Distillation. The Failure Card is distilled into the internal router by supervising its routing logits:

$$\mathcal{L}_{route} = \frac{1}{BT|\mathcal{G}|} \sum_{p,i} \sum_{g\in\mathcal{G}} \frac{1}{|\mathcal{T}_g|} \sum_{m\in\mathcal{T}_g} \mathrm{BCELogit}(r_{p,i,m}, \alpha^D_{p,i,m}), \tag{4}$$

where $\mathcal{G}$ contains the anatomy, temporal, and correction groups. Independent binary targets permit multiple simultaneous actions, while group-wise averaging prevents group-size bias. Spatial, severity, and downstream supervision are detailed in Appendix B.

Overall, CCAL replaces trajectory-level error supervision with structured action-level corrective credit assignment by coupling failure attribution with executable action targets in the Failure Card. This makes otherwise latent corrective responsibility explicit and learnable, allowing corrective-action supervision to be distilled into the internal router without native action annotations.

3.2 Factorized Corrective Transition Learning for *Action Execution*

Building on the structured corrective actions learned by CCAL, FCTL makes action-level supervision directly executable by reparameterizing the corrective component of the diffusion transition itself. Instead of forcing heterogeneous corrections through a shared denoising pathway, FCTL exposes independently learnable *when–where–how* degrees of freedom through state-dependent activation, spatial support, and action-specific residual transformation. Specifically, we first formulate

the execution problem of a shared denoiser, then construct a factorized executable action space, learn state-dependent action selection, and finally perform corrective-action injection directly within denoising.

Entangled Corrective Action Execution. A conventional shared denoiser couples heterogeneous corrective effects into the same transition, whereas DCE correction must independently determine *when*, *where*, and *how* a transition should change. We therefore factorize the corrective effect as

$$\Delta\epsilon^{\mathrm{corr}}_{\ell,p,i} = \sum_{m=1}^{M} \alpha_{p,i,m} M_{p,i,m} \odot T_{\phi_m}(\epsilon_{\ell,p,i}, x_{\ell,p,i}), \tag{5}$$

where $\alpha_{p,i,m}$, $M_{p,i,m}$, and T_{ϕ_m} respectively parameterize state-dependent activation, spatial support, and residual transformation. This gives CCAL's *where–when–how* supervision explicit degrees of freedom inside the world-model transition.

Factorized Action Space Construction. The executable action space consists of learnable residual operators

$$\mathcal{R}_T = \{(n_m, g_m, c_m, T_{\phi_m})\}_{m=1}^{M}, \tag{6}$$

where n_m, g_m, c_m, and T_{ϕ_m} denote the action, semantic group, scope, and residual transformation, with $g_m \in \{\mathrm{anatomy}, \mathrm{temporal}, \mathrm{correction}\}$. This separates heterogeneous corrections into independently selectable and localizable transformations.

State-Dependent Action Selection. Action activation is predicted jointly from the current diffusion and kinetic states, while spatial support is predicted from the latent representation. We represent the DCE state as

$$u_{p,i} = [\tau_{p,i}, \Delta\tau_{p,i}, \mathbf{I}_{p,i}, \dot{\mathbf{I}}_{p,i}, \ddot{\mathbf{I}}_{p,i}, \hat{e}_{p,i}, \hat{s}_{p,i}], \tag{7}$$

containing acquisition timing, regional intensities and their temporal derivatives, kinetic deviation, and failure severity. The router then combines kinetic and latent states:

$$\begin{aligned} r_{p,i,m} &= R^{lat}_{\psi,m}(x_{\ell,p,i}, \ell) + \lambda_K F^{route}_{\eta,m}(u_{p,i}), \\ \alpha_{p,i,m} &= \sigma(r_{p,i,m}), \\ P_{p,i,m} &= \sigma(R^{map}_{\psi,m}(x_{\ell,p,i}, \ell)), \end{aligned} \tag{8}$$

where $\alpha_{p,i,m}$ and $P_{p,i,m}$ predict action activation and spatial support.

Corrective Action Injection. During training, CCAL supervision is mixed with the internal routing prediction,

$$\bar{\alpha}_{p,i,m} = \lambda_D \alpha^{D}_{p,i,m} + (1 - \lambda_D)\alpha_{p,i,m}, \tag{9}$$

while $\bar{M}_{p,i,m}$ uses the spatial target during training and $P_{p,i,m}$ at inference. The corrective transition is injected into denoising as

$$\epsilon'_{\ell,p,i} = \epsilon_{\ell,p,i} + \lambda_T \sum_{m=1}^{M} \bar{\alpha}_{p,i,m} \bar{M}_{p,i,m} \odot T_{\phi_m}(\epsilon_{\ell,p,i}, x_{\ell,p,i}), \tag{10}$$

where λ_T controls the correction strength. At inference, $\hat{e}_{p,i}$ is estimated from the deviation of the preliminary trajectory from the non-enhanced baseline, while $\hat{s}_{p,i} = 0$ because no diagnostic Failure Card is available. Together with the phase-wise ROI intensities and their temporal derivatives, these reference-free quantities form $u_{p,i}$ for the refined rollout.

Overall, FCTL replaces an action-agnostic shared denoising pathway with an explicitly factorized corrective transition whose activation, spatial support, and residual transformation are separately learnable. This makes the structured action supervision produced by CCAL executable within the world-model transition itself, allowing different DCE kinetic failures to induce different corrective state updates rather than sharing the same denoising response.

3.3 Reliability-Guided Corrective Optimization for *Action Feedback*

Building on the action supervision of CCAL and action execution of FCTL, RGCO closes the training-time self-corrective learning loop by using the outcome of each corrective round to determine subsequent optimization. Rather than continuing to optimize corrected and persistently

unreliable trajectories similarly, RGCO converts the reliability change caused by each corrective round into trajectory-specific optimization emphasis, so the model learns not only from *what remains wrong* but from *whether its previous correction actually worked*. Specifically, we first encode corrective effectiveness through correction-induced reliability change, and then feed this outcome back through reliability-guided reweighting of subsequent optimization.

Corrective Outcome Encoding. At round r, ReKinetics regenerates the trajectory and computes its mean normalized reconstruction error and reliability as

$$\bar{e}_p^{(r)} = \frac{1}{T_p CHW} \sum_{i=1}^{T_p} \|\hat{y}_{p,i}^{(r)} - y_{p,i}\|_1, \qquad \rho_p^{(r)} = \mathrm{clip}(0.35 - \bar{e}_p^{(r)}, 0, 1). \tag{11}$$

The correction outcome is measured against the best reliability previously recorded for that trajectory:

$$\Delta\rho_p^{(r)} = \rho_p^{(r)} - \max_{q<r} \rho_p^{(q)}, \tag{12}$$

where the first observation initializes the trajectory memory. Positive $\Delta\rho_p^{(r)}$ indicates improvement over all preceding rounds, while non-positive values identify an unresolved trajectory.

Reliability-Guided Reweighting. RGCO reduces the contribution of improved trajectories while retaining or upweighting unresolved ones, progressively concentrating optimization on remaining failures. Each selected sample follows

$$\mathcal{L} = \mathcal{L}_{diff} + \lambda_1 \mathcal{L}_{diag\text{-}rec} + \lambda_2 \mathcal{L}_{route} + \mathcal{L}_{temp} + \mathcal{L}_{KAN} + \lambda_3 \mathcal{L}_{balance} + \lambda_4 \mathcal{L}_{specialize}, \tag{13}$$

where the terms respectively preserve denoising, diagnosed regions and downstream phases, corrective routing, temporal kinetics, KAN-based kinetic supervision, routing balance, and localized action specialization. Detailed reliability weighting is provided in Appendix B.

Overall, RGCO changes self-corrective learning from error-driven optimization to correction-outcome-guided optimization: the effect of an executed action determines whether its trajectory is suppressed or emphasized in subsequent learning. Together with CCAL and FCTL, this completes the failure–action–outcome loop, making corrective actions not only learnable and executable, but also improvable according to their actual outcomes.

4 Experiments

4.1 Experimental Setup

We evaluate ReKinetics on a hand DCE-MRI benchmark with long and nonuniform enhancement trajectories and on Duke Breast DCE-MRI Saha et al. (2021). We use case-level splits with no subject overlap, and all methods receive the same non enhanced input and acquisition time condition. We report PSNR, SSIM, LPIPS, and rMSE for frame fidelity. On hand DCE-MRI, crSSIM measures agreement between predicted and target contrast responses, and ROI intensity MAE measures the absolute error between predicted and target mean enhancement intensities within each anatomical ROI across phases. On Duke, cSSIM measures adjacent-phase temporal consistency under the published protocol and does not compare generated phases with their references. The same definitions, split, and DDIM-20 sampler are used for all locally evaluated methods. The Duke comparison follows the corresponding published protocols. Additional implementation and metric definitions are provided in Appendix C.

4.2 ReKinetics Corrects Predictions through Executed Actions at Inference

Trajectory errors remain controlled toward later acquisitions. Figure 3 shows the trajectory-level consequence of self-corrective learning: ReKinetics suppresses ROI-error accumulation toward later acquisitions. Figure 3(a) tracks test-set ROI intensity error and shows lower error toward later acquisitions, whereas competing methods exhibit persistent or increasing deviations. Figure 3(b) summarizes this behavior: ReKinetics reduces late-stage ROI error by 42.1% relative to early error, while all competing methods accumulate 23.7%–44.7% error. Figure 5 provides the controlled mechanistic evidence that matched corrective actions produce localized and downstream error reductions with the latent state, sampler noise, and non-tested actions fixed. Additional within-trajectory visualizations are provided in Appendix D.2.

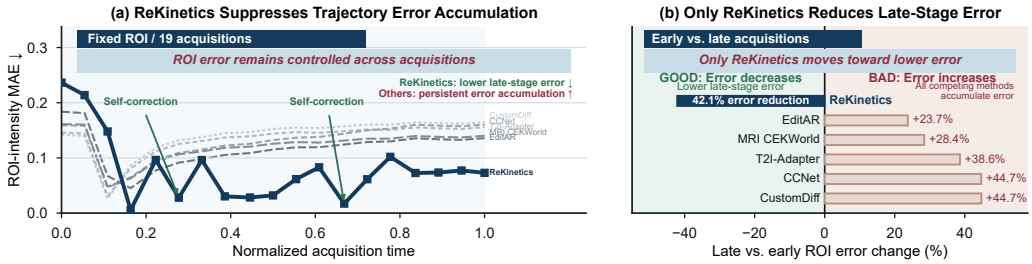

Figure 3: ReKinetics suppresses trajectory-level ROI-error accumulation toward later DCE acquisitions. (a) Test-set mean ROI intensity MAE across acquisition time. ReKinetics maintains lower late-stage error, whereas competing methods retain larger deviations. (b) Relative change between late- and early-stage ROI errors, $100(E_{\text{late}} - E_{\text{early}})/E_{\text{early}}$. ReKinetics is the only compared method with a negative change, reducing error by 42.1%.

Table 1: Results on Hand and Duke Breast DCE-MRI. Hand crSSIM is target-referenced, while Duke cSSIM measures adjacent-phase temporal consistency. Best and second-best results within each dataset are shown in red and blue.

	Hand DCE-MRI					Duke Breast DCE-MRI				
Method	PSNR ↑	SSIM ↑	LPIPS ↓	rMSE ↓	crSSIM ↑	PSNR ↑	SSIM ↑	LPIPS ↓	rMSE ↓	cSSIM ↑
CustomDiff Kumari et al. (2023)	26.348	0.8962	**0.0904**	83.8615	0.8034	11.19	0.2463	0.4706	1094.0	0.3835
T2I Adapter Mou et al. (2024)	**26.535**	**0.9010**	0.1041	**80.9761**	**0.8060**	17.73	0.4130	0.2905	297.4	0.1347
CCNet Osuala et al. (2024)	26.118	0.8936	**0.0953**	84.5692	0.7813	**21.47**	0.4043	0.3128	289.3	0.3155
EditAR Mu et al. (2025)	26.421	**0.8991**	0.1034	82.6697	0.8057	19.85	0.4170	0.3119	288.9	0.3886
ControlNet Zhang et al. (2023)	20.753	0.5241	0.1238	127.0699	0.4325	19.79	**0.5196**	0.2625	308.0	0.3370
MRI CEKWorld Kong et al. (2026)	25.678	0.8882	0.0990	90.8840	0.7564	21.09	**0.5599**	**0.2620**	**243.5**	**0.3900**
ReKinetics (Ours)	**27.3594**	0.8801	0.1034	**77.8730**	**0.8161**	**27.21**	0.4663	**0.1548**	**195.5**	**0.7673**

4.3 ReKinetics Improves Frame and Kinetic Fidelity Across Anatomies

Table 1 shows that the self-corrective formulation improves generated enhancement states and their temporal behavior across anatomies. ReKinetics achieves the strongest PSNR and rMSE on both datasets, the strongest target-referenced crSSIM on hand, and the strongest adjacent-phase cSSIM under the published Duke protocol. Several methods retain higher SSIM or lower LPIPS on hand, and MRI CEKWorld retains higher SSIM on breast, showing that the gains are concentrated in enhancement magnitude and temporal behavior rather than static texture similarity.

Figure 4 shows that ReKinetics corrects the trajectory without sacrificing regional appearance, supporting self-correction of both spatial structure and enhancement kinetics. Across early, middle, and late phases, ReKinetics remains closest to the ground truth in the R1 region, while its TIC follows the target enhancement trend more closely than competing methods. The agreement between the phase-wise images and TIC shows that ReKinetics reproduces both plausible individual phases and the correct phase-dependent enhancement kinetics, consistent with the state-dependent activation and spatial support learned by FCTL rather than generic image refinement. More cases (including Duke) are provided in Appendix D.

4.4 The Proposed Mechanisms Explain the Self-Corrective Gains

FCTL executes targeted where–when–how corrective actions. Figure 5 shows that the self-corrective behavior is produced by targeted FCTL corrective actions rather than generic refinement. The corrections are anatomy-, temporal-, and correction-specific, spatially localized, and temporally persistent, supporting the factorized corrective transitions and their direct composition into the denoising prediction. All controlled comparisons keep the latent state, sampler noise, and non-tested corrective actions fixed, so the measured changes isolate the tested corrective action. In (a), matched HOW and WHEN reduce target-region error by 18.7%, while a wrong HOW or WHEN reduces the gain to 4.1% and 2.8%, showing that correction depends on both correction type and enhancement phase. In (b), the inference-time predicted spatial support remains close to the training-time diagnostic spatial support (0.68 vs. 0.72 overlap; 0.17 vs. 0.14 leakage), showing that spatially localized

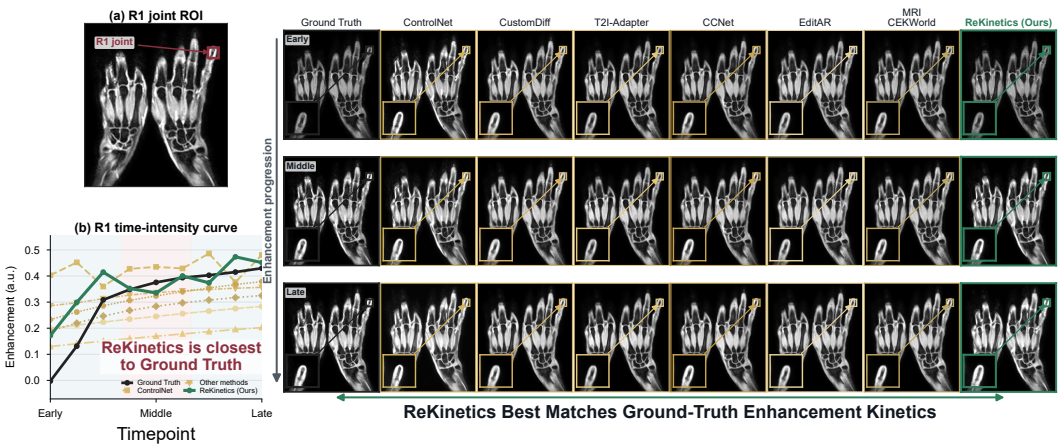

Figure 4: ReKinetics best matches the ground truth in both TIC-based quantitative kinetics (left) and visual appearance (right). The left panels identify the R1 joint and show its TIC across early, middle, and late enhancement stages. The right panel visualizes the corresponding R1 appearance at these three stages.

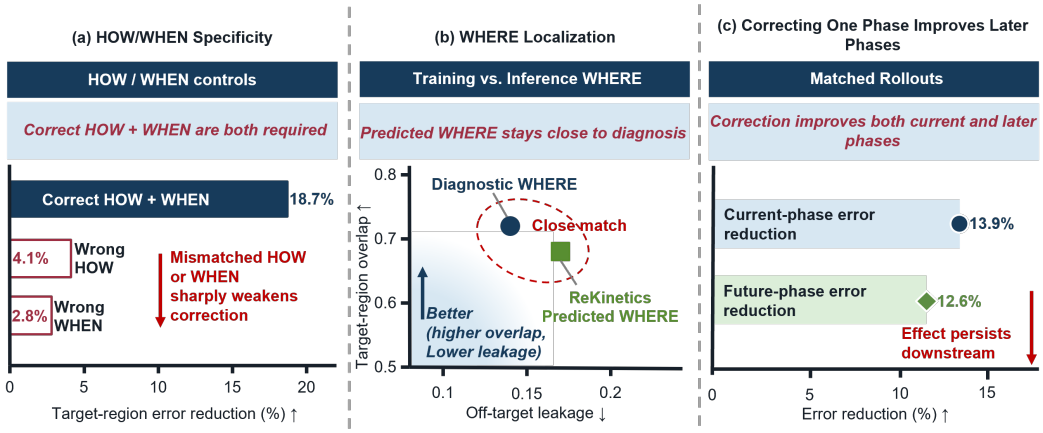

Figure 5: Factorized Corrective Transition Learning in ReKinetics provides targeted where–when–how action execution. (a) Matching HOW and WHEN produces substantially stronger correction than mismatching either factor, confirming their functional specificity. (b) The inference-time predicted spatial support closely matches the training-time diagnostic spatial support, confirming that spatial support is learned and retained at inference. (c) Correcting the current phase also reduces future-phase error, showing that the corrective action affects subsequent kinetic evolution rather than only the current frame.

action execution is internalized without external diagnostic localization at inference. In (c), correcting the current phase reduces current- and future-phase error by 13.9% and 12.6%, respectively, showing that the corrective action revises the kinetic transition rather than only the current image. The full-test, same-time action-on/off control in Appendix C.5 further shows that corrective execution improves ROI MAE in 71.4% of trajectories and 64.5% of acquisitions, while improving crSSIM in 70.2% of trajectories.

Round-wise CCAL updates and KAN-based kinetic-state routing improve corrective-action learning. Table 2 shows that ReKinetics improves corrective-action learning through two mechanisms: round-wise updated failure cards in CCAL and KAN-based kinetic-state routing in FCTL, rather than through additional optimization or controller capacity alone. Panel A isolates the effect of updating failure cards from additional optimization. Moving from one-shot diagnosis to fixed failure cards raises PSNR from 26.1030 to 26.6420 dB. With the number of rounds fixed, updating

Table 2: Round-wise CCAL updates and KAN-based kinetic-state routing improve corrective-action learning beyond additional optimization or controller capacity. Panel A shows that updating failure cards across rounds outperforms both one-shot diagnosis and fixed failure cards, demonstrating the benefit of adapting action supervision to evolving model failures. Panel B shows that KAN outperforms a capacity-matched MLP under identical kinetic states and rollouts, supporting its effectiveness for kinetic-state routing.

Variant	PSNR ↑	SSIM ↑	LPIPS ↓	rMSE ↓	crSSIM ↑	ROI intensity MAE ↓
A. Round-Wise Failure-Card Updates in CCAL						
One-shot failure diagnosis	26.1030	0.8124	0.1092	87.4253	0.7490	0.063
Fixed failure cards across rounds	26.6420	0.8348	0.1070	83.6034	0.7840	0.055
Round-wise updated failure cards	26.9038	0.8429	0.1059	81.2846	0.8007	0.052
B. KAN-Based Kinetic-State Routing						
Matched MLP kinetic controller	25.9870	0.8017	0.1098	89.2945	0.7310	0.066
KAN kinetic controller	26.7066	0.8452	0.1053	82.6703	0.7785	0.057

Table 3: Progressively completing the self-corrective mechanisms consistently improves ReKinetics, showing that its gains arise from the corrective design. Best results are shown in red.

Variant	PSNR ↑	SSIM ↑	LPIPS ↓	rMSE ↓	crSSIM ↑	ROI intensity MAE ↓
ControlNet backbone	20.7530	0.5241	0.1238	127.0699	0.4325	0.132
+ Monolithic residual (capacity control)	20.8319	0.5313	0.1228	125.8464	0.4401	0.129
+ Factorized Corrective Transition Learning	22.1374	0.6011	0.1189	116.2517	0.5153	0.113
+ Corrective Credit Assignment Learning	24.4220	0.7231	0.1121	99.4610	0.6469	0.085
+ KAN-Based Kinetic-State Routing	26.7066	0.8452	0.1053	82.6703	0.7785	0.057
+ Reliability-Guided Corrective Optimization	**27.3594**	**0.8801**	**0.1034**	**77.8730**	**0.8161**	**0.049**

failure cards from newly generated trajectories further raises PSNR to 26.9038 dB and crSSIM to 0.8007, while reducing ROI intensity MAE to 0.052, supporting the round-wise action supervision in CCAL. Panel B isolates the KAN controller under identical kinetic state $u_{p,i}$, preliminary and refined rollouts, and matched parameter budgets. KAN raises PSNR from 25.9870 to 26.7066 dB and crSSIM from 0.7310 to 0.7785, while reducing ROI intensity MAE from 0.066 to 0.057, supporting the kinetic-state routing in FCTL.

Progressive gains confirm the full self-corrective loop. Table 3 shows that ReKinetics becomes progressively more effective as the self-corrective loop is completed, rather than simply gaining from additional capacity. It progressively constructs ReKinetics from the same backbone to separate these effects. The capacity-matched monolithic residual improves PSNR by only 0.0789 dB, showing that capacity alone cannot explain the gains. Adding FCTL raises PSNR from 20.8319 to 22.1374 dB and crSSIM from 0.4401 to 0.5153; CCAL further raises them to 24.4220 dB and 0.6469; and KAN-Based Kinetic-State Routing reaches 26.7066 dB and 0.7785. Finally, RGCO reaches 27.3594 dB PSNR and 0.8161 crSSIM by converting correction-induced reliability changes into trajectory-specific optimization weights for subsequent learning.

5 Conclusion

We introduced ReKinetics, the first self-corrective world model for DCE-MRI contrast-enhancement modeling. Unlike forward-only diffusion models that can retain and propagate kinetic errors, ReKinetics assigns corrective credit from trajectory failures to executable actions, executes factorized corrective transitions, and uses correction-induced reliability changes to guide subsequent learning. Experiments on hand and Duke Breast DCE MRI show that this self-corrective formulation improves enhancement kinetics and suppresses error accumulation, while controlled actions and ablations verify the roles of corrective credit assignment, factorized corrective transitions, kinetic-state routing, and reliability-guided corrective optimization. Despite its strong performance, future work will extend ReKinetics to broader anatomies and multi-center acquisitions, and investigate more general self-corrective world models that can learn corrective actions with reduced external supervision.

AI USE STATEMENT

GPT-5.5 was used within the proposed method as a training-only assistant VLM to transform locally computed spatial and kinetic diagnostic summaries into structured corrective supervision. The authors reviewed and verified all AI-assisted content.

REFERENCES

Shruthi Bannur, Kenza Bouzid, Daniel C Castro, Anton Schwaighofer, Anja Thieme, Sam Bond-Taylor, Maximilian Ilse, Fernando Pérez-García, Valentina Salvatelli, Harshita Sharma, et al. Maira-2: Grounded radiology report generation. *arXiv preprint arXiv:2406.04449*, 2024.

Amir Bar, Gaoyue Zhou, Danny Tran, Trevor Darrell, and Yann LeCun. Navigation world models. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15791–15801. IEEE, 2025.

Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, et al. Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127*, 2023.

Jake Bruce, Michael D Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, et al. Genie: Generative interactive environments. In *Forty-first international conference on machine learning*, 2024.

Chao Chen, Catalina Raymond, William Speier, Xinyu Jin, Timothy F Cloughesy, Dieter Enzmann, Benjamin M Ellingson, and Corey W Arnold. Synthesizing mr image contrast enhancement using 3d high-resolution convnets. *IEEE Transactions on Biomedical Engineering*, 70(2):401–412, 2022.

Salman UH Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Cukur. Image synthesis in multi-contrast mri with conditional generative adversarial networks. *IEEE transactions on medical imaging*, 38(10):2375–2388, 2019.

Shenyuan Gao, Jiazhi Yang, Li Chen, Kashyap Chitta, Yihang Qiu, Andreas Geiger, Jun Zhang, and Hongyang Li. Vista: A generalizable driving world model with high fidelity and versatile controllability. *Advances in Neural Information Processing Systems*, 37:91560–91596, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Michael Ingrisch and Steven Sourbron. Tracer-kinetic modeling of dynamic contrast-enhanced mri and ct: a primer. *Journal of pharmacokinetics and pharmacodynamics*, 40(3):281–300, 2013.

Bingyi Kang, Yang Yue, Rui Lu, Zhijie Lin, Yang Zhao, Kaixin Wang, Gao Huang, and Jiashi Feng. How far is video generation from world model: A physical law perspective. *arXiv preprint arXiv:2411.02385*, 2024.

Eunjin Kim, Hwan-Ho Cho, Junmo Kwon, Young-Tack Oh, Eun Sook Ko, and Hyunjin Park. Tumor-attentive segmentation-guided gan for synthesizing breast contrast-enhanced mri without contrast agents. *IEEE journal of translational engineering in health and medicine*, 11:32–43, 2022.

Jindi Kong, Yuting He, Cong Xia, Rongjun Ge, and Shuo Li. MRI contrast enhancement kinetics world model. *arXiv preprint arXiv:2602.19285*, 2026.

Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-concept customization of text-to-image diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1931–1941, 2023.

Yann LeCun et al. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *Open Review*, 62(1):1–62, 2022.

Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in neural information processing systems*, 36:28541–28564, 2023.

Russell Mendonca, Shikhar Bahl, and Deepak Pathak. Structured world models from human videos. *arXiv preprint arXiv:2308.10901*, 2023.

Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2I-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 4296–4304, 2024.

Jiteng Mu, Nuno Vasconcelos, and Xiaolong Wang. EditAR: Unified conditional generation with autoregressive models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 7899–7909, 2025.

Gustav Müller-Franzes, Luisa Huck, Soroosh Tayebi Arasteh, Firas Khader, Tianyu Han, Volkmar Schulz, Ebba Dethlefsen, Jakob Nikolas Kather, Sven Nebelung, Teresa Nolte, et al. Using machine learning to reduce the need for contrast agents in breast mri through synthetic images. *Radiology*, 307(3):e222211, 2023.

Richard Osuala, Daniel M Lang, Preeti Verma, Smriti Joshi, Apostolia Tsirikoglou, Grzegorz Skorupko, Kaisar Kushibar, Lidia Garrucho, Walter HL Pinaya, Oliver Diaz, et al. Towards learning contrast kinetics with multi-condition latent diffusion models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 713–723. Springer, 2024.

Richard Osuala, Smriti Joshi, Apostolia Tsirikoglou, Lidia Garrucho, Walter HL Pinaya, Daniel M Lang, Julia A Schnabel, Oliver Diaz, and Karim Lekadir. Simulating dynamic tumor contrast enhancement in breast mri using conditional generative adversarial networks. *Journal of Medical Imaging*, 12(S2):S22014–S22014, 2025.

Sriprabha Ramanarayanan, Arunima Sarkar, Matcha Naga Gayathri, Keerthi Ram, Mohanasankar Sivaprakasam, et al. Dce-diff: Diffusion model for synthesis of early and late dynamic contrast-enhanced mr images from non-contrast multimodal inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5174–5183, 2024.

Marc Rigter, Tarun Gupta, Agrin Hilmkil, and Chao Ma. Avid: Adapting video diffusion models to world models. *arXiv preprint arXiv:2410.12822*, 2024.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 10674–10685. ieee, 2022.

Ashirbani Saha, Michael R. Harowicz, Lars J. Grimm, Jingxi Weng, Elizabeth H. Cain, Connie E. Kim, Sujata V. Ghate, Ryan Walsh, and Maciej A. Mazurowski. Dynamic contrast-enhanced magnetic resonance images of breast cancer patients with tumor locations. The Cancer Imaging Archive, 2021.

Prasad R Shankar, Kushal Parikh, and Matthew S Davenport. Financial implications of revised acr guidelines for estimated glomerular filtration rate testing before contrast-enhanced mri. *Journal of the American College of Radiology*, 15(2):250–257, 2018.

Yejee Shin, Yeeun Lee, Hanbyol Jang, Geonhui Son, Hyeongyu Kim, and Dosik Hwang. Anatomical consistency and adaptive prior-informed transformation for multi-contrast mr image synthesis via diffusion model. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 30918–30927, 2025.

Gustav J Strijkers, Willem J M. Mulder, Geralda A F. van Tilborg, and Klaas Nicolay. Mri contrast agents: current status and future perspectives. *Anti-Cancer Agents in Medicinal Chemistry-Anti-Cancer Agents)*, 7(3):291–305, 2007.

Tao Tu, Shekoofeh Azizi, Danny Driess, Mike Schaekermann, Mohamed Amin, Pi-Chuan Chang, Andrew Carroll, Charles Lau, Ryutaro Tanno, Ira Ktena, et al. Towards generalist biomedical ai. *Nejm Ai*, 1(3):AIoa2300138, 2024.

Guanhua Wang, Enhao Gong, Suchandrima Banerjee, Dann Martin, Elizabeth Tong, Jay Choi, Huijun Chen, Max Wintermark, John M Pauly, and Greg Zaharchuk. Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model. *IEEE transactions on medical imaging*, 39(10):3089–3099, 2020.

Yang Yue, Yulin Wang, Haojun Jiang, Pan Liu, Shiji Song, and Gao Huang. Echoworld: Learning motion-aware world models for echocardiography probe guidance. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 25993–26003, 2025.

Mahmut Yurt, Salman UH Dar, Aykut Erdem, Erkut Erdem, Kader K Oguz, and Tolga Çukur. mustgan: multi-stream generative adversarial networks for mr image synthesis. *Medical image analysis*, 70:101944, 2021.

Hongxin Zhang, Zeyuan Wang, Qiushi Lyu, Zheyuan Zhang, Sunli Chen, Tianmin Shu, Behzad Dariush, Kwonjoon Lee, Yilun Du, and Chuang Gan. Combo: Compositional world models for embodied multi-agent cooperation. In *International Conference on Learning Representations*, volume 2025, pp. 49996–50019, 2025.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3836–3847, 2023.

Diagnostic VLM prompt template

```
You are the medical kinetic VLM controlling a contrast-free DCE-MRI
world model.  Your environment is a generated multi-phase enhancement
trajectory.  Actively inspect it by calling local diagnostic
functions, reason about regional temporal kinetic failures, and
output a routing action for the registered corrective operators.
Context ID: [opaque context identifier]
Time labels:  [ordered acquisition-time labels]
Reference available:  [true/false]
Call list_kinetic_tools before final routing and use only returned
operator names.  Call at least one diagnostic function.  Return JSON
only with:  trajectory_reliability; global_tools; up to six
local_actions, each containing frame_index, phase, anatomy, failure,
normalized bbox, severity, tools, and causal_targets; and
memory_update.  Prefer a factorized route containing an anatomy,
temporal, and correction operator.  All weights and severities must
lie in [0, 1].
```

Figure 6: Prompt template for the training-only diagnostic VLM. The runtime fills only the opaque context identifier, acquisition times, and reference-availability flag; diagnostic evidence is obtained through the local functions in Table 4.

A Corrective Supervision and Action Registry

This appendix records the implementation details that are necessary to reproduce the corrective supervision used by CCAL. The external diagnostic VLM is used only to construct training targets; it is absent at inference.

A.1 Training-Only Diagnostic VLM

Figure 6 reports the diagnostic prompt template used to obtain structured corrective supervision. Bracketed fields are populated at runtime; the callable function definitions are those in Table 4.

The first response requires a function call (Table 4), parallel calls are disabled, and at most four sequential function-call rounds are allowed. The API calls do not override temperature, top-p, or output-length parameters and therefore use the provider defaults; JSON is requested in the prompt and parsed locally rather than enforced through a structured-output mode. A final response contains at most six local failure records and is parsed into the Failure Card in Eq. 3. Invalid responses are retried once after a two-second delay. These callable functions (Table 4) are lightweight, parameter-free auxiliary modules. The VLM diagnoses the failure and may select additional registered actions; a deterministic schema normalizer then converts the diagnosed anatomy, phase, failure type, and severity into a complete executable routing target.

For completeness, the soft maps use $d^{+} = \mathrm{Norm}_{+}(\hat{y} - x + 0.15)$, where Norm_{+} applies ReLU and divides each map by its maximum. A 7×7 average filter gives the local mean and an 11×11 filter defines local sharpness. The vessel, lesion, and parenchyma scores are respectively normalized versions of $0.65 d_{\mathrm{sharp}} + 0.35 d_{\mathrm{grad}}$, $0.70 d_{\mathrm{mean}} + 0.30 d^{+} - 0.25 d_{\mathrm{vessel}}$, and $d_{\mathrm{mean}} - 0.20 d_{\mathrm{lesion}}$. The locator centers a box whose height and width are 28% of the image dimensions on the maximum-error pixel. The rule checker assigns under- and over-enhancement when the signed mean error is below -0.035 or above 0.035, respectively; otherwise it assigns blur. Temporal inconsistency overrides that label when the first-difference error exceeds $\max(0.035, 0.75 e_i)$, where e_i is the current frame error. Severity is the frame error divided by the maximum frame error, clipped to $[0.15, 1]$. These fixed rules generate diagnostic evidence and semantic fields.

The complete normalized output contract is given in Table 5. It makes every selectable category and numerical constraint explicit. The implementation key `causal_targets` is a legacy field name and stores only supervised downstream influence γ^{D}; no causal identification is claimed. The local fields correspond to $(q, h, i, M^{D}, s^{D}, \alpha^{D}, \gamma^{D})$ in Eq. 3. This is an output-schema definition: "Card"

Table 4: Parameter-free local diagnostic modules exposed to the training-only VLM. Each module performs a deterministic calculation in the local environment .

Module	Deterministic computation	Evidence returned to the VLM
`list_kinetic_tools`	Reads the static corrective-action registry defined in `tool_schema.py`; it performs no image analysis.	Valid operator names, WHERE/WHEN/HOW groups, scopes, intended targets, and compatible failure categories.
`roi_curve_analyzer`	Builds soft vessel, lesion, and parenchyma maps from normalized enhancement, local averages, and spatial gradients, then averages intensities inside each map over acquisition time.	Candidate ROI time-intensity curves and, when a reference is available, the corresponding reference curves.
`temporal_difference_analyzer`	Computes frame-wise means and first differences of the candidate trajectory; with a reference, it also computes absolute errors between candidate and reference temporal differences.	Time labels, frame means, temporal-difference summaries, and optional reference-based temporal errors.
`anatomy_locator`	Ranks acquisitions by mean absolute candidate–reference error, or by candidate–baseline deviation when no reference is available. A fixed-size box is centered at the maximum-error pixel and assigned the dominant soft-map label.	Candidate frame indices, normalized high-error boxes, regional labels, and error scores.
`kinetic_rule_checker`	Applies fixed thresholds to signed candidate–reference intensity differences and first-difference errors. It distinguishes under-enhancement, over-enhancement, blur, and temporal inconsistency in the current code.	Ranked local findings containing phase, regional label, normalized box, failure type, and error-normalized severity.

fields occur once per trajectory, whereas "Local record" fields are repeated for each diagnosed failure.

A.2 Representative Failure Cards and Parsing

Table 6 shows how the output contract expresses an early uptake discontinuity, a middle-phase over-enhancement, and a late lesion rebound. In each case, the record combines independent anatomy, temporal, and correction operators rather than selecting a pre-enumerated Cartesian action. The three records are examples of the normalized training cache.

The corresponding intended corrections are to recover a smooth early uptake ramp, damp a middle-phase overshoot while preserving redistribution, and suppress a late lesion rebound while retaining delayed-phase continuity. These textual summaries are retained for audit and analysis; trainable supervision is provided by the structured phase, box, severity, and operator weights.

During parsing, frame indices and scalar values are clipped to their valid ranges and unknown operator names are discarded. A malformed or non-four-value box falls back to $[0, 0, 1, 1]$; numeric coordinates are clipped to $[0, 1]$, and a non-positive width or height is expanded by 0.1 where the image boundary permits. Anatomy aliases such as `background` and `uncertain` are normalized to the non-enhancing and artifact/uncertain operators, respectively.

The routing target is made schema-complete by a deterministic normalization step. For a local record with severity s, the anatomy, phase, and failure fields instantiate their canonical WHERE, WHEN, and HOW operators with weights $\max(0.40, s)$, $\max(0.35, 0.85s)$, and $\max(0.45, s)$, respectively. Temporal inconsistency additionally activates temporal smoothness with weight s. Any explicit

Table 5: Failure Card field definitions. Direct diagnostic responses and offline-expanded records share one normalized JSON card per trajectory. Card-level fields occur once; local-record fields repeat for each diagnosed failure, up to six records. Tool-valued fields may select any subset of the 14 operators in Table 7.

JSON field	Allowed output	Role in corrective supervision
Trajectory-level fields: emitted once per Failure Card		
`trajectory_reliability`	Real value in $[0, 1]$	Stores the schema-level VLM quality score for audit. Formal GPT-replay runs recompute the ρ used by RGCO as specified in Sec. B.3. Lower values indicate less reliable kinetics.
`global_tools`	Sparse map from any registered operator name to a weight in $[0, 1]$	Specifies trajectory-wide or shared corrective activation.
`local_actions`	Array of 0–6 local records	Collects the phase- and region-specific Failure Card entries below.
`memory_update`	`unresolved_failure_score`$\in [0, 1]$ and a fixed diagnostic-mode tag	Stores cached training metadata used to construct subsequent-round sampling and weighting.
Local-record fields: repeated for each diagnosed failure (at most six)		
`frame_index`	Integer in $\{0, \ldots, T-1\}$	Selects the acquisition on which the diagnosed failure is localized.
`phase`	`early`, `middle`, or `late`	Provides the discrete WHEN factor associated with the selected frame.
`anatomy`	`vessel`, `lesion`, `parenchyma`, `non_enhancing_structure`, `artifact`, or `unknown`	Provides the categorical WHERE factor.
`failure`	`under_enhancement`, `over_enhancement`, `shape_drift`, `blur`, `fake_structure`, or `temporal_inconsistency`	Defines the diagnosed failure category that requires correction.
`bbox`	$[x_0, y_0, x_1, y_1] \in [0, 1]^4$, with $x_1 > x_0$ and $y_1 > y_0$	Rasterizes the normalized spatial support M^D.
`severity`	Real value in $[0, 1]$	Scales regional loss weighting and the default action strength.
`tools`	Sparse multi-label map from registered operator names to weights in $[0, 1]$	Supplies the executable routing target α^D; multiple WHERE, WHEN, and HOW operators may be active together.
`causal_targets`	Possibly empty list of downstream phase/influence descriptors	Stores supervised downstream influence γ^D for propagation weighting and auditing.

operator weights returned by GPT-5.5 are merged by taking the larger weight. Consequently, the semantic triplet is a method-defined action schema rather than an unconstrained prompt discovery; the action parameters, spatial predictor, and inference-time router remain learned from these structured targets.

A.3 Factorized Corrective Action Registry

The registry contains 14 independently parameterized actions. Each action has the same executable architecture: it receives the base denoising residual and current noisy latent, applies 1×1, 3×3,

Table 6: Representative local records sampled from the GPT-5.5 training cache. Each row instantiates the schema in Table 5. Boxes are shown as normalized $x : [x_0, x_1]$ and $y : [y_0, y_1]$ intervals; sample identifiers are omitted and weights are rounded for presentation.

ID	Failure and WHEN	WHERE	Selected factorized action	Severity
A	Temporal inconsistency Early, frame 0	Parenchyma $x : [.22, .50]$, $y : [.00, .28]$	`parenchyma_tool` (0.93) `early_uptake_tool` (0.80) `temporal_smoothness_tool` (0.95)	0.931
B	Over-enhancement Middle, frame 13	Parenchyma $x : [.33, .61]$, $y : [.00, .28]$	`parenchyma_tool` (0.85) `middle_redistribution_tool` (0.75) `over_enhancement_tool` (0.90) `temporal_smoothness_tool` (0.80)	0.845
C	Temporal inconsistency Late, frame 12	Lesion $x : [.63, .91]$, $y : [.11, .39]$	`lesion_tool` (0.85) `late_washout_delayed_tool` (0.85) `temporal_smoothness_tool` (0.90) `over_enhancement_tool` (0.75)	0.721

and 1×1 convolutions with SiLU activations, and returns a latent residual correction. The final layer is zero-initialized. Operator semantics are attached by the Failure Card routing and spatial targets, while the actual residual transformations are learned. At execution, the router multiplies each residual by its activation weight and spatial map before summing the selected corrections as in Eq. 10. The three groups form a semantic action taxonomy: a corrective action jointly activates one or more anatomy, temporal, and correction operators. Table 7 lists the complete registry exposed to the diagnostic VLM and internal router.

Table 7: Complete FCTL action registry, including operator scope, compatible diagnostic categories, and the intended specialization of each learned residual action.

Operator	Target and scope	Compatible failures	Learned corrective role
WHERE: anatomical support of the correction			
`vessel_tool`	Vascular enhancement; spatial	Under-/over-enhancement; temporal inconsistency	Correct rapid uptake, continuity, and vessel-specific enhancement amplitude.
`lesion_tool`	Focal lesion-like enhancement; spatial	Under-/over-enhancement; shape drift; fake structure	Correct focal uptake, heterogeneity, unstable boundaries, and lesion-like delayed behavior.
`parenchyma_tool`	Diffuse tissue perfusion; spatial	Under-/over-enhancement; blur	Correct diffuse tissue enhancement level and smooth parenchymal perfusion changes.
`non_enhancing_structure_tool`	Stable non-enhancing anatomy; spatial	Over-enhancement; fake structure	Suppress inappropriate enhancement in anatomy expected to remain stable.
`artifact_uncertain_tool`	Ambiguous or artifact-dominated regions; spatial	Blur; fake structure; temporal inconsistency	Downweight unreliable local feedback and handle uncertain high-error regions.
WHEN: acquisition stage or trajectory behavior			
`early_uptake_tool`	Early wash-in; global or phase-level	Under-/over-enhancement	Control enhancement onset and rapid early uptake.
`middle_redistribution_tool`	Intermediate redistribution; global or phase-level	Temporal inconsistency; shape drift	Control the transition from early uptake to middle-phase redistribution.
`late_washout_delayed_tool`	Late washout or delayed enhancement; global or phase-level	Under-/over-enhancement; temporal inconsistency	Control residual enhancement, washout, and delayed behavior.
`temporal_smoothness_tool`	Full trajectory; global or phase-level	Temporal inconsistency	Suppress flicker and inconsistent phase-to-phase changes.
HOW: corrective residual applied to the selected support			
`under_enhancement_tool`	Low enhancement amplitude; spatial	Under-enhancement	Increase insufficient enhancement in the responsible region or phase.
`over_enhancement_tool`	Excessive enhancement amplitude; spatial	Over-enhancement	Suppress unrealistically strong enhancement in the responsible region or phase.
`shape_drift_tool`	Anatomical shape consistency; spatial	Shape drift	Reduce boundary drift and local anatomical deformation across phases.
`blur_detail_loss_tool`	Local detail preservation; spatial	Blur; detail loss	Restore local detail and counter overly smooth enhancement predictions.
`fake_structure_suppression_tool`	Unsupported enhancing structures; spatial	Fake structure	Suppress hallucinated structures and enhancement unsupported by the input anatomy.

B Detailed Training Objectives

This section supplies only the implementation details omitted from the compact objective in Eq. 13. Superscript D denotes a target extracted from a diagnostic Failure Card.

B.1 Spatial, Severity, and Downstream Weighting

For frame i, local boxes are rasterized into spatial supports and scaled by their diagnosed severity. The resulting loss mask and downstream frame weight are

$$L^D_{p,i} = \max_{a \in \mathcal{C}_{p,i}} s^D_a M^D_a, \tag{14}$$

$$\omega^D_{p,i} = \max\left(1,\ \max_{a:\, k \to i} 1 + \gamma^D_{a,k\to i} s^D_a\right), \tag{15}$$

where maxima are element-wise for spatial masks. In the implementation, $L^D_{p,i}$ is clipped to $[0,1]$ and $\omega^D_{p,i}$ to $[1,3]$. The diagnosed-region reconstruction term is

$$\mathcal{L}_{\text{diag-rec}} = \frac{1}{BT} \sum_{p,i} w_{p,i} \left\| (1 + L^D_{p,i}) \odot (\hat{z}^0_{p,i} - z_{p,i}) \right\|_1, \tag{16}$$

where $\hat{z}^0$ and z are the predicted and reference clean latents, and $w_{p,i}$ includes the downstream and RGCO weights defined below.

The route loss is given in Eq. 4. To prevent route collapse and train spatial support, we additionally use

$$\mathcal{L}_{\text{balance}} = \frac{1}{M} \sum_m \left[0.75 \max(0, \alpha_{\min} - \bar{\alpha}_m)^2 + 0.25(\bar{\alpha}_m - \pi_m)^2\right], \tag{17}$$

$$\mathcal{L}_{\text{specialize}} = \frac{\sum_{p,i} \sum_{m \in \mathcal{S}} \alpha^D_{p,i,m} \| P_{p,i,m} - \bar{M}^D_{p,i,m} \|_2^2}{\sum_{p,i} \sum_{m \in \mathcal{S}} \alpha^D_{p,i,m} + \varepsilon}, \tag{18}$$

where $\alpha_{\min} = 0.03$ and $\mathcal{S}$ contains spatial anatomy and correction operators. In the registry order of Table 7, the action-frequency prior is

$$\boldsymbol{\pi} = (.22, .18, .22, .10, .08,\ .18, .20, .18, .20,\ .18, .14, .10, .18, .08).$$

B.2 Temporal and Kinetic-State Supervision

For nonuniform acquisition times, velocity is normalized by the observed interval, with $\varepsilon_t = 10^{-3}$:

$$v_{p,i} = \frac{y_{p,i} - y_{p,i-1}}{\max(|t_{p,i} - t_{p,i-1}|, \varepsilon_t)}, \qquad \hat{v}_{p,i} = \frac{\hat{y}_{p,i} - \hat{y}_{p,i-1}}{\max(|t_{p,i} - t_{p,i-1}|, \varepsilon_t)}, \tag{19}$$

$$\mathcal{L}_{\text{vel}} = \sum_{p,i} \|\hat{v}_{p,i} - v_{p,i}\|_1, \tag{20}$$

$$\mathcal{L}_{\text{acc}} = \sum_{p,i} \|(\hat{v}_{p,i+1} - \hat{v}_{p,i}) - (v_{p,i+1} - v_{p,i})\|_1. \tag{21}$$

The same diagnostic spatial weights are averaged over adjacent frames when computing these terms.

The KAN heads are supervised by

$$\mathcal{L}_{\text{curve-fit}} = \sum_{p,i,a} |\tilde{I}^a_{p,i} - I^a_{p,i}(Y)|, \tag{22}$$

$$\mathcal{L}_{\text{curve-prior}} = \sum_{p,i,a} |I^a_{p,i}(\hat{Y}) - \text{sg}(\tilde{I}^a_{p,i})|, \tag{23}$$

$$\mathcal{L}_{\text{down}} = \sum_p \sum_{i<j} \|\gamma^{KAN}_{p,i\to j} - \gamma^D_{p,i\to j}\|_2^2. \tag{24}$$

The curve-prior target is detached before supervising the generated ROI intensity. $\mathcal{L}_{\text{down}}$ distills a downstream-influence score.

Table 8: Training and execution coefficients for the reported 256×256 configuration.

Term	Weight	Term	Weight
$\mathcal{L}_{\text{diag-rec}}$	2.6	$\mathcal{L}_{\text{route}}$	0.03
$\mathcal{L}_{\text{vel}}$	0.15	$\mathcal{L}_{\text{acc}}$	0.08
$\mathcal{L}_{\text{balance}}$	0.003	$\mathcal{L}_{\text{specialize}}$	0.008
$\mathcal{L}_{\text{curve-fit}}$	0.012	$\mathcal{L}_{\text{curve-prior}}$	0.015
$\mathcal{L}_{\text{down}}$	0.012	λ_D	0.30
λ_K	0.22	λ_T	0.04

B.3 Reliability-Guided Corrective Optimization

For the formal GPT-replay runs, the schema-level reliability returned in the initial VLM response is retained for audit, while the routing cache recomputes the value used by RGCO deterministically according to Eq. 11. Thus, $\rho_p^{(r)} \in [0, 1]$ is reference-informed during training and larger values indicate lower trajectory reconstruction error. The memory bank stores the best value observed before round r and evaluates $\Delta\rho_p^{(r)} = \rho_p^{(r)} - \max_{q<r} \rho_p^{(q)}$ as defined in Eq. 12. The first observation initializes this history.

A trajectory is marked unresolved when $\Delta\rho_p^{(r)} \leq 0$. The weight applied in the subsequent round is

$$w_p^{\text{RGCO}} = 1 + \eta_{\mathcal{M}}(1 - \rho_p^{(r)}) + \mu_{\mathcal{M}} \mathbb{1}[\Delta\rho_p^{(r)} \leq 0], \tag{25}$$

with $\eta_{\mathcal{M}} = 1.0$ and $\mu_{\mathcal{M}} = 0.5$. The sign of the reliability change identifies whether the current trajectory exceeds its best previous reliability. The per-frame optimization weight is the product of this trajectory weight, the downstream frame weight $\omega_{p,i}^D$, and the cached score $\text{clip}((1 - \rho_p^{(r)}) + 0.5\bar{s}, 0.1, 2.0)$, where $\bar{s}$ is the mean diagnosed severity. Thus, low-reliability or non-improving trajectories receive greater emphasis in the next round, whereas corrected trajectories are relatively suppressed.

For the reported 256×256 configuration, the compact terms in Eq. 13 correspond to

$$\mathcal{L}_{\text{temp}} = 0.15\mathcal{L}_{\text{vel}} + 0.08\mathcal{L}_{\text{acc}}, \tag{26}$$

$$\mathcal{L}_{\text{KAN}} = 0.012\mathcal{L}_{\text{curve-fit}} + 0.015\mathcal{L}_{\text{curve-prior}} + 0.012\mathcal{L}_{\text{down}}. \tag{27}$$

Rather than repeating Eq. 13, Table 8 reports its implementation weights.

C Experimental and Evaluation Protocol

C.1 Hand DCE-MRI Dataset

The institutional hand benchmark contains DCE-MRI examinations from 34 subjects with rheumatoid arthritis. The examinations were acquired at AIC Yaesu Clinic, Tokyo, Japan, on a Philips 3T scanner. The cohort comprises 26 female and 8 male subjects, with an age of 53.6 ± 18.5 years (range: 17–82 years). Each examination contains one non-enhanced baseline series and 12–26 temporally ordered post-contrast acquisitions (20.1 ± 4.9 per subject). All readable slices are resized to 256×256 and normalized to $[-1, 1]$. A slice trajectory denotes one spatially matched slice followed across all available acquisitions; a frame pair denotes one baseline slice paired with one post-contrast target from that trajectory.

We use the fixed patient-level split in Table 9. Subjects are disjoint across training, validation, and test sets. Failure Cards and all other corrective supervision are constructed exclusively from the 20 training subjects. The validation and test sets are excluded from parameter optimization and diagnostic-cache construction. The dataset was approved by the Hokkaido University ethics review board (24-104) and the Institute of Science Tokyo (A24672); access is restricted by the corresponding ethics and institutional data-use requirements.

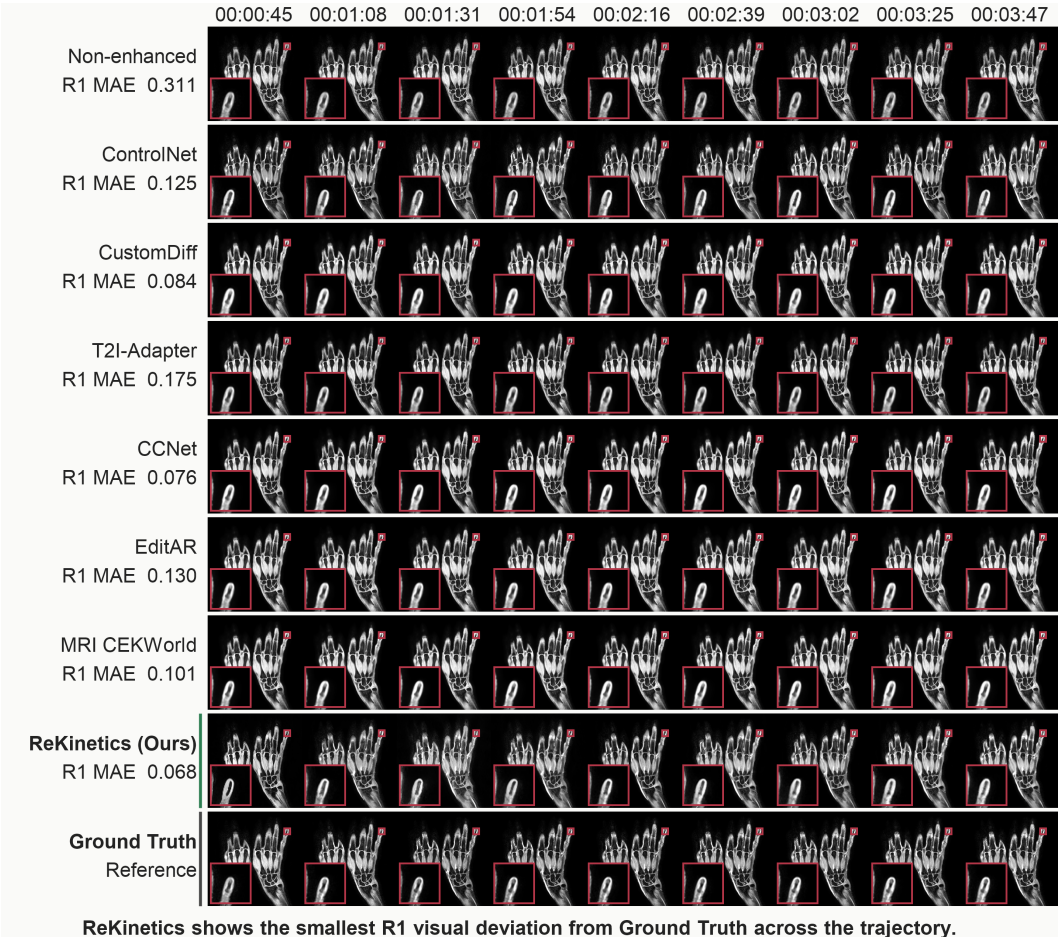

Figure 7: Nine-acquisition comparison at the little-finger R1 joint. Columns follow acquisition time; rows compare the non-enhanced input, competing methods, ReKinetics, and ground truth. The identical R1 box and bottom-left magnification in every image expose local differences that are difficult to inspect at full-hand scale. The row-wise R1 trajectory MAE quantifies the difference over all nine acquisitions (lower is better); ReKinetics achieves the lowest error among the compared methods.

Table 9: Patient-level split of the hand DCE-MRI benchmark. A frame pair is one non-enhanced input slice paired with one post-contrast target.

Split	Subjects	Trajectories	Frame pairs
Training	20	1,154	24,317
Validation	7	406	7,344
Test	7	406	7,582
Total	34	1,966	39,243

C.2 Training and Compute Details

ReKinetics is initialized from the reported ControlNet-conditioned latent diffusion checkpoint. The latent-diffusion backbone, first-stage VAE, and text encoder remain frozen; the full ControlNet branch, internal router, KAN modules, and factorized corrective operators are optimized. This gives 361.45M trainable parameters in the reported configuration. We use AdamW with $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 0.01, and a fixed learning rate of 2×10^{-7} for all trainable parameter groups. No learning-rate scheduler is used.

Table 10: Training and compute configuration for the reported hand model.

Item	Setting
Input resolution	256×256
Optimizer	AdamW, $\beta = (0.9, 0.999)$, weight decay 0.01
Learning rate	Fixed 2×10^{-7}
Rounds / epochs	2 rounds $\times$ 8 epochs (16 epochs total)
Optimizer updates	Approximately 16,211 per round; 32,423 total
Batch size	3 physical; 12 effective with 4-step accumulation
Numerical precision	16-bit mixed precision
Stability settings	Gradient clipping 1.0; CUDA cache cleared every 20 steps
Data loading	2 workers; sequence-aware shuffled mini-batches
Validation / test split	7 subjects / 7 subjects; no per-epoch validation sampling
Random seed	20260618
Hardware	One NVIDIA GeForce RTX 4090 (24 GB)
Wall-clock time	$>$ 72 h end-to-end, including data I/O and diagnostic VLM requests; 7.8 h for the final 8-epoch optimization round alone
Sampling	DDIM-10 for diagnostic review; DDIM-20 for evaluation

Table 10 reports the complete optimization and compute setup. Training comprises two corrective rounds of eight epochs each. The final checkpoint of the first round initializes the second round. Based on 24,317 training pairs and an effective batch size of 12, each round contains approximately 16,211 optimizer updates, giving approximately 32,423 updates over the full 16-epoch run. These are exposure-based estimates; sequence-aware batching forms mini-batches within trajectories and drops incomplete groups, so the exact update count is obtained from the training log. The physical mini-batch contains three slice–acquisition pairs; four-step gradient accumulation gives an effective batch size of 12. The complete end-to-end run required more than 72 hours on the reported workstation. This wall-clock time includes on-demand reading and preprocessing of the DCE-MRI series, local diagnostic computation, diagnostic VLM requests, Failure Card and routing-cache construction, and both optimization rounds. The retained optimization log for the second eight-epoch round records 7.8 hours; this shorter value excludes the preceding data-access and corrective-supervision construction stages and therefore is not the total training time.

C.3 Common Evaluation Setup

All hand experiments use a fixed case-level split with no patient overlap. Locally evaluated methods receive the same non-enhanced image, acquisition time, image resolution, and test samples. Results in the main paper use DDIM-20 sampling. Training-time Failure Cards are constructed only from training cases. Duke Breast is evaluated under the protocol stated in the corresponding comparison.

C.4 Metric Definitions

Let x be the non-enhanced image, y_i the reference at acquisition i, and $\hat{y}_i$ its prediction. PSNR and SSIM are computed after mapping images to $[0, 1]$; LPIPS uses the standard $[-1, 1]$ input. rMSE is computed after restoring each image to its recorded original intensity range.

crSSIM evaluates the target-referenced contrast response:

$$\mathrm{crSSIM} = \frac{1}{T} \sum_{i=1}^{T} \mathrm{SSIM}(\hat{y}_i - x,\, y_i - x), \tag{28}$$

with data range four for differences of normalized $[-1, 1]$ images. In contrast, cSSIM measures only adjacent-phase structural consistency:

$$\mathrm{cSSIM} = \frac{1}{T-1} \sum_{i=1}^{T-1} \mathrm{SSIM}(\hat{y}_i, \hat{y}_{i+1}), \tag{29}$$

after mapping predictions to $[0, 1]$. It does not compare predictions with the reference and is therefore interpreted together with frame fidelity.

For a fixed evaluation ROI R, the regional kinetic error is

$$\text{ROI-MAE} = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{1}{|R|}\sum_{v\in R}\hat{y}_i(v) - \frac{1}{|R|}\sum_{v\in R}y_i(v)\right|. \tag{30}$$

The ROI is fixed across methods and acquisitions. The TIC is the ordered sequence of mean intensities extracted from that fixed region.

C.5 Full-Test Matched Action-Execution Control

This experiment directly tests whether executing the learned corrective actions improves a prediction at the same target acquisition. We evaluate all 406 complete test trajectories (7582 frames) twice with the same input, checkpoint, acquisition time, DDIM-20 sampler, initial sampling noise, and random seed. In *Action ON*, the internal router executes the learned WHERE–WHEN–HOW actions normally. In *Action OFF*, all routed action weights are set to zero at every denoising step, making the summed corrective residual in Eq. 10 zero. All routing, state estimation, and sampling operations are otherwise retained, so the two conditions differ only in corrective residual execution.

For trajectory p, define the paired improvement indicator

$$I_p = \mathbb{1}\left[\text{ROI-MAE}_p^{\text{ON}} < \text{ROI-MAE}_p^{\text{OFF}}\right], \tag{31}$$

where $I_p = 1$ indicates lower regional kinetic error when corrective actions are executed. We report the proportion of comparisons with $I_p = 1$ over complete trajectories, individual acquisitions, and pixels inside the fixed evaluation ROI. The ROI-pixel improvement proportion is

$$P_{\text{pix}} = \frac{1}{|R|}\sum_{v\in R}\mathbb{1}[|\hat{y}_{\text{ON}}(v) - y(v)| < |\hat{y}_{\text{OFF}}(v) - y(v)|]. \tag{32}$$

Table 11 reports the resulting improvement rates.

Table 11: Improvement consistency in the full-test matched action-execution control. ON and OFF use identical inputs, target times, sampling noise, DDIM-20 steps, and computation; OFF zeros only the routed corrective residual. The primary result is the proportion of comparisons improved by Action ON.

Criterion	Evaluation unit	ON better / total	Improvement rate ↑
ROI MAE	Trajectory ($N = 406$)	290 / 406	71.4%
ROI MAE	Acquisition ($N = 7582$)	4890 / 7582	64.5%
Absolute ROI error	Pixel within ROI	$N_{\text{win}}/N_{\text{pix}}$	56.8%
PSNR	Trajectory ($N = 406$)	255 / 406	62.8%
crSSIM	Trajectory ($N = 406$)	285 / 406	70.2%

Corrective execution improves a clear majority of matched comparisons at every reported level. Action ON lowers trajectory-level ROI MAE in 290 of 406 trajectories (71.4%) and lowers acquisition-level ROI error in 4890 of 7582 acquisitions (64.5%). The effect is spatially distributed: 56.8% of evaluated ROI pixels move closer to their references under Action ON. The same majority-win pattern holds for complementary image and kinetic criteria, with PSNR improving in 62.8% and crSSIM in 70.2% of trajectories. Here, N_{win} denotes the number of ROI pixels with lower absolute error under Action ON and N_{pix} the total number of evaluated ROI pixels.

D Qualitative Visualization Analysis

D.1 Expanded Hand Enhancement Trajectory

Figure 7 expands the three-stage comparison in Fig. 4 to nine acquisitions from the same hand sequence. Every panel marks and magnifies the identical little-finger joint (R1), directing the comparison to the local enhancement emphasized in the main text. Across the trajectory, ReKinetics

most closely follows the ground-truth R1 appearance and obtains the lowest nine-phase R1 trajectory MAE (0.0676 versus 0.0760 for the next-best method). Competing methods retain weaker, over-enhanced, or less phase-sensitive responses at multiple acquisitions.

D.2 Within-Trajectory Local Error Evolution

Figure 8 shows that ReKinetics maintains lower localized discrepancies toward later acquisitions within the generated DCE trajectories. All nine spatially matched comparisons have lower local MAE at the later acquisition, with relative reductions of 37.3–75.8% and a mean reduction of 59.1%. Every comparison is produced by the same model with fixed weights; "earlier" and "later" identify acquisition states within one generated trajectory.

The three examples exhibit this behavior across both joint and soft-tissue regions. In Example 1, the early joint error decreases from 14.5% to 3.7% (74.2%), the middle joint error from 9.5% to 6.0% (37.3%), and the late soft-tissue error from 12.0% to 3.8% (68.6%). Example 2 shows corresponding reductions from 14.8% to 4.2% (71.4%), 12.5% to 5.4% (57.1%), and 11.5% to 6.2% (46.4%). In Example 3, the two early joint errors decrease from 18.8% to 7.1% (62.4%) and from 10.9% to 6.7% (38.4%), while the late soft-tissue error decreases from 8.2% to 2.0% (75.8%).

The later-acquisition decrease is consistent across acquisition stages and tissue types. Six of the nine regions achieve reductions above 50%. The largest decrease occurs in the late soft-tissue region of Example 3 (75.8%), followed by the early joint regions of Examples 1 and 2 (74.2% and 71.4%). These cases show that localized discrepancies do not persist uniformly through the generated trajectory. The smaller reductions in the middle joint of Example 1 (37.3%) and the second early joint of Example 3 (38.4%) retain residual MAEs of 6.0% and 6.7%, identifying localized joint detail as the more difficult setting. Nevertheless, every evaluated region moves closer to its time-matched ground truth. The shared trend across early, middle, and late acquisitions complements the aggregate error-accumulation result in Fig. 3; controlled action effects are isolated separately in Fig. 5.

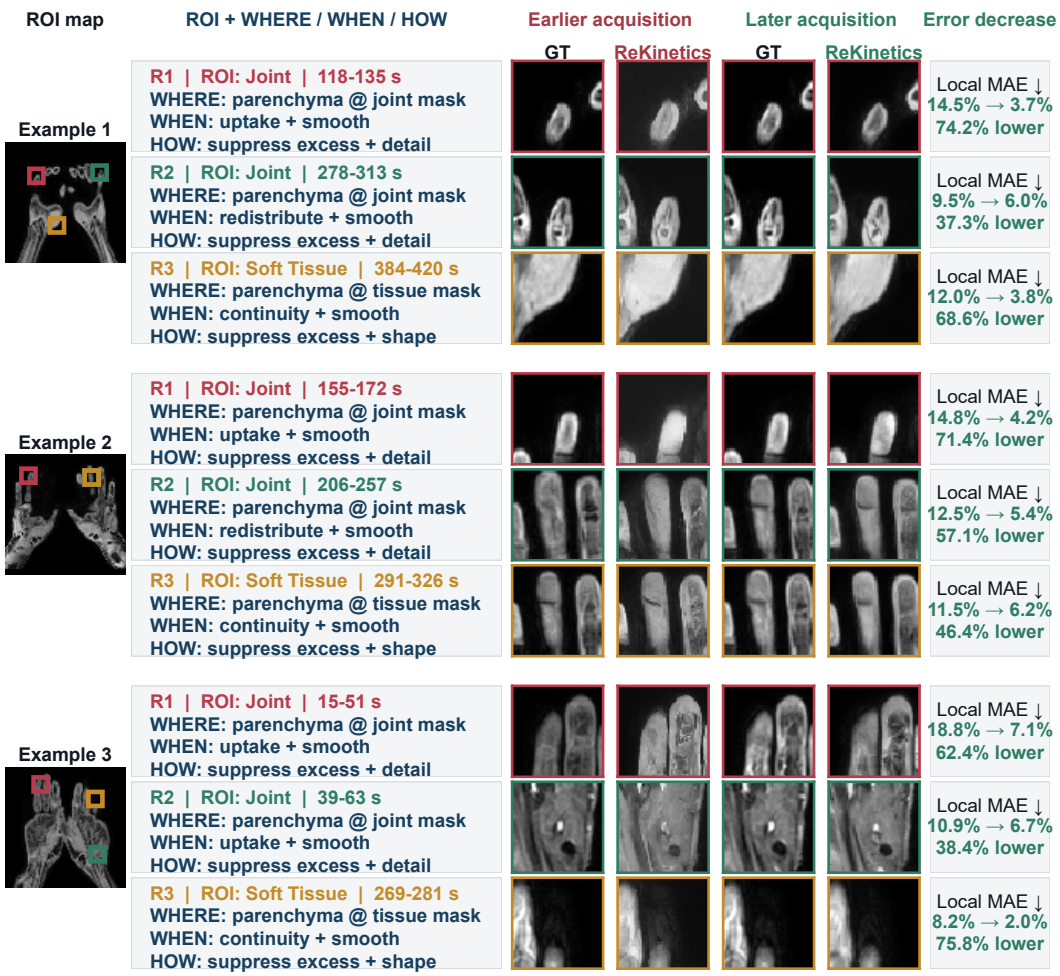

Figure 8: ReKinetics exhibits lower local error at later acquisitions in all nine spatially matched comparisons from the same three DCE trajectories. Read each example from left to right: (1) the ROI map localizes two joint regions (R1–R2) and one soft-tissue region (R3), with color linking each box to its table row; (2) the factorized route lists the registered anatomy-group WHERE action and its spatial support, temporal-group WHEN actions, and the ROI-matched HOW actions. The notation $a@M$ means that action a is executed within spatial mask M. The aliases map to existing registry entries: `parenchyma` to `parenchyma_tool`; `uptake`, `redistribute`, `continuity`, and `smooth` to the four temporal operators; and `suppress excess`, `detail`, and `shape` to `over_enhancement_tool`, `blur_detail_loss_tool`, and `shape_drift_tool`, respectively; (3) the *Earlier acquisition* columns compare ReKinetics with the corresponding time-matched ground truth; (4) the *Later acquisition* columns repeat this comparison for the same ROI; and (5) the final column reports the earlier and later local MAE together with their relative decrease (lower MAE and a larger decrease are better). All states are generated with the same fixed model weights; "earlier" and "later" denote acquisitions within one trajectory. Local MAE is computed inside the same colored box against the time-matched ground truth and is reported as a percentage of the normalized $[0, 1]$ intensity range.

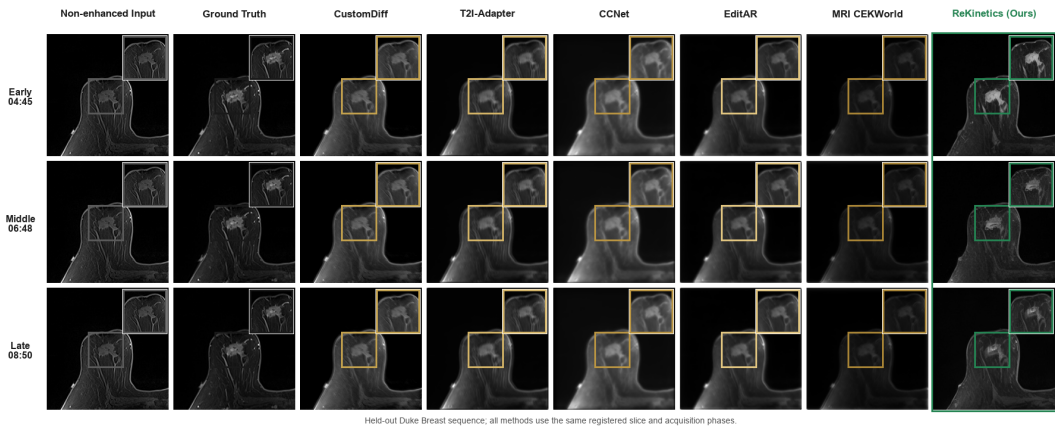

Figure 9: Held-out Duke Breast qualitative comparison. Rows show early, middle, and late acquisitions; columns show the shared non-enhanced input, ground truth, competing methods, and ReKinetics.

D.3 Duke Breast Visualization Analysis

Figure 9 compares the non-enhanced input, five competing methods, ReKinetics, and ground truth on a complete held-out Duke Breast sequence. ReKinetics preserves the surrounding breast structure while recovering the location, shape, and phase-dependent evolution of the central enhancement. Competing outputs are visibly smoother and exhibit weaker phase-dependent changes in the magnified region.