LLM-Orchestrated Diagnose–Plan–Treat for Mixed-Degradation CT Reconstruction

Yongqiang Huang[1] , **Yingyu Chen**[2] , **Fengzhi Xu**[1] , **Tao Wang**[2] ,
Wenjun Xia[3] , **Hongming Shan**[4] and **Yi Zhang**[1,*]
[1]School of Cyber Science and Engineering, Sichuan University
[2]College of Computer Science, Sichuan University
[3]School of Computer Science and Engineering, Southeast University
[4]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University

Abstract

Clinical Computed Tomography (CT) reconstruction often faces mixed degradations, where quantum noise, streak artifacts, and geometric distortions co-occur with various compositions and severities. Recently, all-in-one frameworks have outperformed traditional single-task models through degradation-specialized modulation mechanisms. However, attempting to resolve these mixed degradations simultaneously forces the model into suboptimal trade-offs, failing to balance conflicting reconstruction objectives. Our empirical study shows that decomposing CT reconstruction into a sequence of ordered steps effectively mitigates this conflict, with the execution order being a critical performance factor. Leveraging this insight, we propose **AgenticCT**, an LLM-orchestrated multi-agent framework designed to autonomously plan the optimal reconstruction trajectory. Specifically, AgenticCT operates through a *diagnose-plan-treat* workflow: a *Supervisor Agent* explicitly estimates a structured degradation state; a *Planner Agent* then selects a topology-optimized execution order conditioned on this state; and a library of *Processor Agents* sequentially executes the trajectory to achieve high-fidelity reconstruction. Extensive experiments demonstrate that AgenticCT consistently improves reconstruction fidelity across single and mixed degradations, and generalizes better to out-of-distribution datasets compared with strong specialized and all-in-one baselines. Code is available at: https://github.com/yqhuang2912/AgenticCT.

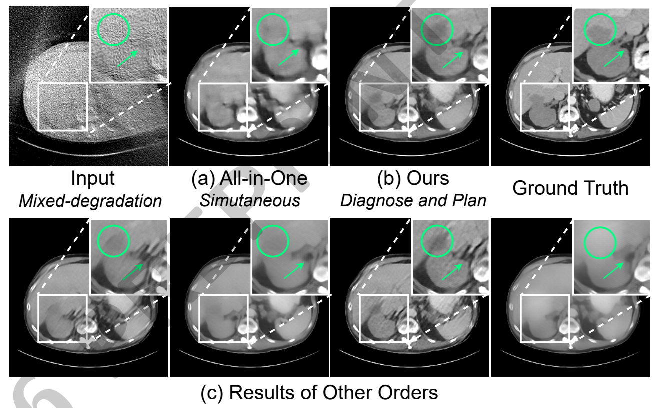

Figure 1: The reconstruction results of empirical study for order sensitivity in mixed-degradation CT reconstruction.

1 Introduction

Computed Tomography (CT) is a cornerstone of modern diagnostic radiology, yet its acquisition is persistently constrained by the balance among radiation safety, scanning speed, and image fidelity [Wang *et al.*, 2020; Zhang *et al.*, 2024; Zhang and Wu, 2019]. Following the As Low As Reasonably Achievable (ALARA) principle [ICRP and others, 1977; Hendee and Edwards, 1986], clinical CT protocols strive to minimize radiation exposure to patients mainly by lowering photon flux or reducing projection angles. The former yields Low-Dose CT (LDCT) [Chen *et al.*, 2024] with amplified quantum noise, while the latter produces Sparse-View CT (SVCT) [Jin *et al.*, 2017] via view down-sampling, which introduces streak artifacts; additionally, mechanical or anatomical constraints can enforce Limited-Angle CT (LACT) [Würfl *et al.*, 2018], causing geometric distortions. In practice, these factors rarely appear in isolation: heterogeneous degradations often co-occur with various compositions and severities, producing mixed artifacts that vary across scanners, protocols, and patient conditions [Hsieh, 2003; Barrett and Keat, 2004]. Therefore, the central challenge is to reconstruct high-fidelity anatomy under such mixtures.

In modern CT reconstruction, deep learning has emerged as the dominant approach. Current methods generally fall into two categories: (1) Specialized models [Chen *et al.*, 2017; Yang *et al.*, 2018], which are typically designed to address a single specific degradation type. While effective in controlled settings, their ability to generalize under mixed degradations is limited. (2) All-in-one models [Li *et al.*, 2022; Potlapalli *et al.*, 2023], which aim to learn a unified mapping using implicit conditioning techniques like prompts or latent

*Corresponding author.

modulation. Although flexible, treating all degradations as a unified task often induces optimization interference [Mehri *et al.*, 2021; Standley *et al.*, 2020], resulting in suboptimal trade-offs. For instance, as shown in Figure 1(a), simultaneous correction in all-in-one models risks prematurely smoothing critical structures, causing artifacts to become entangled with anatomy and leading to irreversible fidelity loss.

A key observation of this work is that mixed-degradation reconstruction is inherently *order-sensitive*. Our empirical study in Figure 1(b) shows that decomposing reconstruction into a sequence of specialized steps by explicitly diagnosis and planning can mitigate objective interference and progressively recover anatomy. However, the effectiveness of this decomposition is strictly contingent on the *execution order*. This stems from the intrinsic order-sensitivity of mixed degradations, given that reconstruction operators are mathematically non-commutative. As illustrated in Figure 1(c), deviating from the optimal order yields drastically different failure modes, where some suboptimal sequences lead to over-smoothing while others fail to remove streak artifacts.

Anchored by the aforementioned critical observation, we propose **AgenticCT**, an LLM-orchestrated multi-agent framework. To the best of our knowledge, this is the first work that reformulates mixed-degradation CT reconstruction for a unified model into perception-driven sequential decision-making paradigm. Specifically, AgenticCT operates under a transparent *Diagnose–Plan–Treat* workflow, comprising three distinct roles: A *Supervisor Agent* performs disentangled perception and outputs a structured degradation state that indicates which artifacts are present and how severe they are. Conditioned on this state, a *Planner Agent* predicts a causality-aware reconstruction trajectory that respects physical dependencies and avoids premature smoothing. A library of *Processor Agents* then executes the trajectory step-by-step to achieve high-fidelity reconstruction. The LLM serves as a system-level orchestrator that renders the process executable and auditable. It converts predicted trajectories into structured, human-readable plans, binds abstract operations to concrete tool invocations, and produces an auditable decision trace for clinical inspection.

Overall, our contributions are summarized as follows.

- We introduce the first LLM-orchestrated multi-agent framework for mixed-degradation CT reconstruction that replaces implicit typical one-step correction typical one-step correction with an explicit sequential *Diagnose–Plan–Treat* workflow.
- We model mixed-degradation reconstruction as order-sensitive trajectory selection and develop a causality-aware planner that determines physically consistent execution orders based on explicit degradation states.
- We provide extensive empirical evidence that mixed-degradation CT reconstruction is better treated as explicit, order-aware sequential decision-making rather than a single-shot correction approach, establishing the practicality and benefits of the proposed paradigm.

2 Related Work

2.1 Traditional CT Reconstruction

From a mathematical standpoint, CT reconstruction from incomplete (e.g. SVCT, LACT) and noisy (e.g. LDCT) measurements constitutes an ill-posed inverse problem [Pedersen *et al.*, 2025]. The central challenge lies in effectively suppressing complex reconstruction artifacts while simultaneously preserving fine anatomical structures and texture details. Mainstream approaches prioritize image-domain reconstruction for its efficiency and simplicity, typically learning a direct mapping from degraded images to clean ground-truth targets. As a representative method, RED-CNN [Chen *et al.*, 2017] preserves spatial resolution by removing pooling layers and introducing skip connections for detail recovery. Similarly, FBPConvNet [Jin *et al.*, 2017] improves performance by incorporating multi-resolution decomposition and residual learning to progressively reduce artifacts. Later, with the introduction of attention mechanisms, TransCT [Zhang *et al.*, 2021] uses a dual-path transformer to split images into high- and low-frequency components, restoring textures through latent low-frequency features. Although these methods show significant performance improvements, they are designed only for single degradation types. This narrow focus overlooks real-world CT imaging scenarios, where multiple degradation factors often occur together, resulting in complex artifact patterns and limiting the ability to generalize.

2.2 All-in-one CT Reconstruction

To mitigate the intrinsic rigidity and limited transferability of specialized models, recent studies have shifted toward all-in-one reconstruction frameworks, aiming to cultivate a unified representation capable of adapting to diverse degradation conditions. Central to this paradigm is conditional modulation, which leverages physical scanning parameters as priors or prompts to dynamically steer network behavior, facilitating degradation-aware and adaptive reconstruction. ProCT [Ma *et al.*, 2023] is a representative example, incorporating projection view–aware prompting to embed acquisition geometry cues for degradation-aware guidance. AMIR [Yang *et al.*, 2024] introduces a task-adaptive routing mechanism, allowing different degradation types to activate distinct computational paths to effectively disentangle conflicting reconstruction objectives. Furthermore, taking a parameter-generation approach, TAT [Yang *et al.*, 2025] employs hypernetworks to dynamically synthesize task-specific weights, thereby customizing filtering behaviors. Despite their improved flexibility, these frameworks typically rely on a monolithic shared network to handle diverse degradation simultaneously. However, different degradations often impose conflicting reconstruction demands, affecting image quality at disparate scales and stages. Consequently, a unified treatment forces the network into suboptimal trade-offs, often obscuring fine details and leading to excessive smoothing.

3 Method

3.1 Problem Formulation

We study mixed-degradation CT reconstruction, where multiple physical factors co-occur with varying compositions and

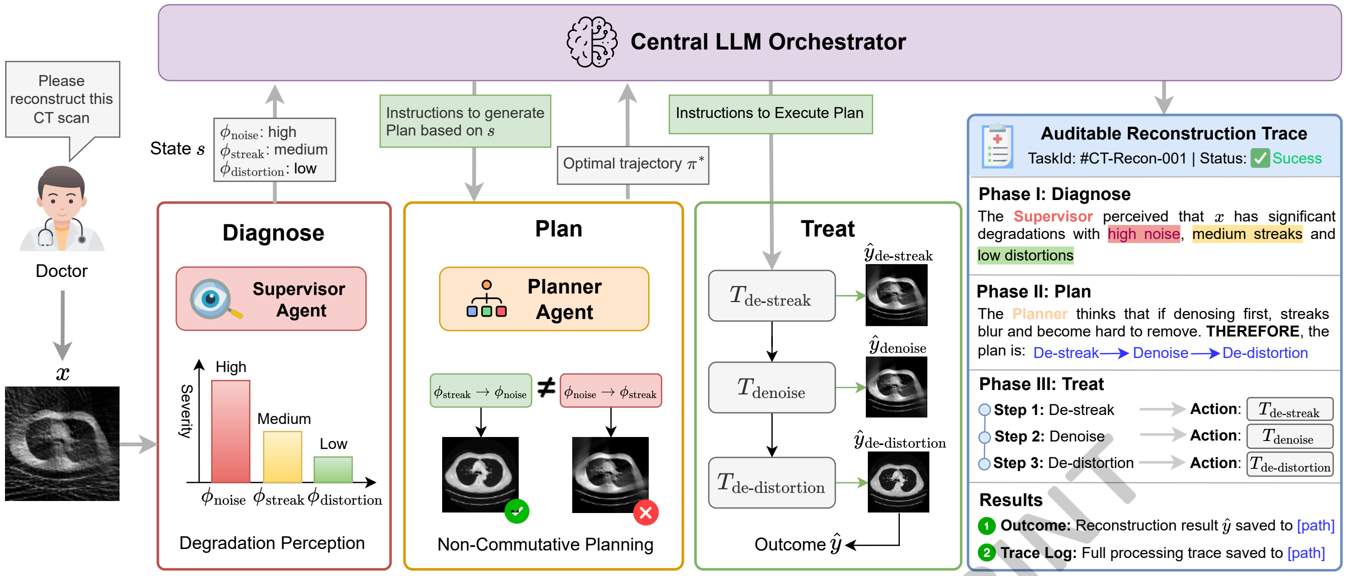

Figure 2: Overview of the AgenticCT framework.

severities. Let $y \in \mathbb{R}^{H \times W}$ represent the clean CT image and x the observed degraded scan. We model the degradation process as a sequence of operators $\Phi = \{\phi\}_{i=1}^{K}$:

$$x = \mathcal{D}(y) = (\phi_k \circ \cdots \circ \phi_1)(y), \tag{1}$$

where each ϕ_i corresponds to a physical factor such as low-dose noise, sparse-view streaks, or limited-angle distortions.

A key property is order sensitivity. Degradation and reconstruction operators are generally non-commutative, and the effectiveness of one correction may depend on whether structural cues have been preserved by previous steps. Therefore, we formulate reconstruction as selecting an ordered trajectory over a processor library. Let Π denote a finite set of candidate trajectories composed from the available processors for efficiency. Our objective is to find

$$\pi^* = \arg\min_{\pi \in \Pi} \mathbb{E}_{(x,y)} \Big[\mathcal{L}_{\text{rec}} \big(\text{Execute}(\pi, x),\ y \big) \Big], \tag{2}$$

where $\text{Execute}(\pi, x)$ applies processors according to π.

3.2 AgenticCT Framework

AgenticCT implements an explicit *Diagnose–Plan–Treat* loop for mixed-degradation reconstruction. Given a degraded input scan x, it produces three explicit outputs: a structured degradation state s for diagnosis, an ordered optimal reconstruction trajectory π^* for planning, and an auditable execution trace that records intermediate states and verification signals.

AgenticCT contains three types of agents. A *Supervisor Agent* estimates the degradation state from the current image. Conditioned on this state, a *Planner Agent* predicts a causality-aware trajectory over the reconstruction tool library. A set of *Processor Agents* then executes the trajectory sequentially. An LLM serves as the orchestrator that converts predicted trajectories into structured, human-readable plans, binds abstract operations to concrete tool calls, and records the execution trace.

Supervisor Agent. Explicit planning requires an explicit state to plan on. In mixed-degradation CT, degradations are coupled in the observation x, so we introduce a Supervisor Agent that performs explicit perception by converting the current image into a structured degradation state s as:

$$s = \mathcal{S}(x) = \{(\kappa_\phi, \rho_\phi)\}_{\phi \in \Phi}, \tag{3}$$

where ϕ indexes degradation type, such as noise, streaks, distortions, $\kappa_\phi \in \{0, 1\}$ indicates whether degradation ϕ is present, and $\rho_\phi \in [0, 1]$ quantifies its normalized severity. This state s provides interpretable evidence for downstream decision-making by exposing what degradations are present and how severe they are.

To make the diagnosis actionable and auditable at inference time, an LLM orchestrator serializes s into a structured diagnostic instruction that can be directly consumed by the Planner, while recording the diagnosis in the execution trace.

Planner Agent. Given explicit perception, the next step is to *decide how to act*. Mixed-degradation reconstruction is order-sensitive, and choosing an execution order that respects physical dependencies requires selecting a trajectory from the candidate set Π. Since $|\Pi|$ grows rapidly with the number of processors, online combinatorial search is impractical under clinical latency constraints. We therefore learn an Planner that maps the current evidence to a distribution over candidate trajectories.

Specifically, the Planner $\mathcal{M}_\theta$ conditions on both the current image x and the explicit degradation state s, and predicts

$$P(\pi|x, s) = \text{Softmax}\big(\mathcal{M}_\theta(\psi_{\text{vis}}(x) \oplus \psi_{\text{state}}(s))\big), \tag{4}$$

where ψ_{vis} and ψ_{state} are lightweight encoders and $\oplus$ denotes concatenation. The selected trajectory is

$$\pi^* = \arg\max_{\pi \in \Pi} P(\pi|x, s). \tag{5}$$

This single forward pass yields an explicit, traceable execution order that replaces fixed-ordering correction with causality-aware sequential decisions.

To operationalize the predicted trajectory at inference time, an LLM orchestrator converts π^* into a structured plan, binds each abstract operation to the corresponding Processor tool call, and records the planning and execution trace for auditing.

Processor Agents. Given the trajectory plan, the next step is to *execute actions*. We introduce a library of Processor Agents $\mathcal{T} = \{T_k\}_{k=1}^{K}$, where each processor represents a specialized reconstruction operation, such as denoising, de-streaking, and de-distortion. Each Processor Agent takes the current image as input and outputs an updated image after applying its targeted correction.

Let the Planner output a trajectory $\pi^* = [T_1, \ldots, T_m]$, where T_k indexes a processor in $\mathcal{T}$ and $m \leq K$. Starting from the current state $\hat{y}_0 = x$, the Treat stage composes these actions sequentially

$$\hat{y}_k = T_k\left(\hat{y}_{k-1}\right), \quad k = 1, \ldots, m. \tag{6}$$

The final result is denoted as $\hat{y} = \hat{y}_m$. At inference time, the LLM orchestrator binds each T_k to the corresponding processor call and records the executed processor information and intermediate states as part of the auditable trace.

3.3 Optimization

We adopt a decoupled protocol that optimizes explicit perception (Supervisor), trajectory selection (Planner), and specialist execution (Processors) with separate objectives. This separation mirrors the *Diagnose–Plan–Treat* workflow and prevents heterogeneous supervision signals from being entangled into a single monolithic training objective.

Supervisor Optimization. The Supervisor predicts a structured degradation state $s = \{(\kappa_\phi, \rho_\phi)\}_{\phi \in \Phi}$. We optimize it with a multi-task objective that combines presence classification and severity regression:

$$\mathcal{L}_{\text{sup}} = \sum_{\phi \in \Phi} \lambda_\kappa \, \mathrm{BCE}(\hat{\kappa}_\phi, \kappa_\phi) + \lambda_\rho \, \|\hat{\rho}_\phi - \rho_\phi\|_1, \tag{7}$$

where Φ indexes degradation types and (κ_ϕ, ρ_ϕ) denote the presence and severity labels.

Planner Optimization. Training the Planner requires trajectory supervision. We obtain this supervision by offline oracle mining on a held-out set. Further details of the held-out set construction and offline oracle mining procedure are provided in the Appendix. For each mixed-degradation pair (x_i, y_i), we evaluate candidate trajectories $\pi \in \Pi$ and select the one that minimizes the reconstruction error under the same execution mechanism used at inference time:

$$\pi_i^* = \arg\min_{\pi \in \Pi} \left\|\mathrm{Execute}(\pi, x_i) - y_i\right\|_1. \tag{8}$$

This yields a planning dataset $\mathcal{D}_{\text{plan}} = \{(x_i, s_i, \pi_i^*)\}_{i=1}^{M}$, where $s_i = \mathcal{S}(x_i)$. Here $\mathrm{Execute}(\pi, x_i)$ applies the candidate trajectory using Processor library, so the mined oracle reflects the actual execution behavior. Given $\mathcal{D}_{\text{plan}}$, we train the Planner by:

$$\mathcal{L}_{\text{plan}}(\theta) = -\frac{1}{M} \sum_{i=1}^{M} \log P(\pi_i^* | x_i, s_i; \theta). \tag{9}$$

After training, prediction requires only a single forward pass.

Processor Optimization. We take all reconstruction tools as Processor Agents indexed by $T_k \in \mathcal{T}$. Each Processor T_k is trained independently on its corresponding training set $\mathcal{D}_k$, constructed to match sequential execution via intermediate targets. We optimize each Processor with an ℓ_1 reconstruction objective:

$$\theta_k = \arg\min_{\theta_k} \, \mathbb{E}_{(x_{\text{in}}, x_{\text{tgt}}) \sim \mathcal{D}_k} \left[\left\|T_k(x_{\text{in}}; \theta_k) - x_{\text{tgt}}\right\|_1\right]. \tag{10}$$

In particular, $\mathcal{D}_k$ includes denoising, de-streaking, and de-distortion pairs with intermediate targets.

4 Experiments

4.1 Settings

We use DeepSeek-Chat [Liu *et al.*, 2025] as the LLM orchestrator for structured plan generation and tool-call binding. In this section, we summarize the evaluation datasets, baseline methods, and metrics. More details about data construction, model architectures, hyperparameters, optimization strategies and implementation details are provided in the Appendix.

Datasets. We benchmark AgenticCT on three clinical CT datasets to assess both in-distribution fidelity and out-of-distribution robustness. The large-scale DeepLesion dataset [Yan *et al.*, 2017] serves as our primary source for both training and in-domain evaluation. For zero-shot generalization, we use the AAPM-Mayo Clinic dataset [McCollough *et al.*, 2017] and the COVID-19 Diagnosis dataset [Afshar *et al.*, 2022] exclusively for testing, which allows us to measure robustness against domain shifts for all baselines.

Baselines and Metrics. We evaluate AgenticCT against a comprehensive set of baselines, representing both *specialized* and *all-in-one* reconstruction methods. Specialized baselines include RED-CNN [Chen *et al.*, 2017], FBPConvNet [Jin *et al.*, 2017], and TransCT [Zhang *et al.*, 2021], targeting noise, streak artifacts, and geometric distortions, respectively. All-in-one baselines include AirNet [Li *et al.*, 2022], PromptIR [Potlapalli *et al.*, 2023], ProCT [Ma *et al.*, 2023], AMIR [Yang *et al.*, 2024], TAT [Yang *et al.*, 2025], and Focal-IR [Cui *et al.*, 2026]. We report PSNR, SSIM, RMSE, and LPIPS under a unified evaluation protocol for all methods.

4.2 Main Results

Table 1 summarizes the overall average reconstruction performance under single- and mixed-degradation scenarios. On the in-distribution DeepLesion benchmark, AgenticCT achieves the strongest and most consistent improvements across fidelity, structure, and perceptual metrics. This differs from specialized models, whose single-degradation priors can become misaligned under mixtures, and from all-in-one models that rely on implicit conditioning with simultaneous correction. In mixed settings, simultaneous correction often couples conflicting objectives, where suppressing noise can remove structural cues needed for geometric artifact removal. By diagnosing degradations and following a execution order, AgenticCT helps maintain the key features needed for later steps, reducing the risk of artifacts interfering with anatomy and consistently delivering stronger reconstructions. The advantage holds up in out-of-distribution tests on

Settings	In-Distribution				Out-of-Distribution							
Datasets	DeepLesion				AAPM-Mayo Clinic				COVID-19 Diagnosis			
Metrics	PSNR ↑	SSIM ↑	RMSE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	RMSE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	RMSE ↓	LPIPS ↓
RED-CNN (2017)	34.12	0.8831	0.0136	0.0547	33.18	0.8906	0.0150	0.0661	29.96	0.8292	0.0207	0.1022
FBPConvNet (2017)	36.42	0.9285	0.0101	0.0390	35.91	0.9244	0.0105	0.0460	32.15	0.8606	0.0163	0.0862
TransCT (2021)	32.77	0.8606	0.0160	0.0673	31.75	0.8638	0.0176	0.0793	27.95	0.7867	0.0255	0.1282
AirNet (2022)	36.12	0.9318	0.0106	0.0418	35.37	0.9314	0.0112	0.0510	32.00	0.8656	0.0163	0.0953
PromptIR (2023)	36.48	0.9400	0.0105	0.0388	35.78	0.9367	0.0111	0.0408	32.41	0.8752	0.0161	0.0914
ProCT (2023)	37.45	0.9246	0.0090	0.0429	37.47	0.9403	0.0088	0.0464	33.92	0.8890	0.0124	0.0861
AMIR (2024)	37.70	0.9479	0.0091	0.0317	37.11	0.9488	0.0095	0.0414	33.22	0.8801	0.0145	0.0918
TAT (2025)	38.83	0.9443	0.0084	0.0376	37.59	0.9480	0.0093	0.0363	34.09	0.8930	0.0135	0.0881
Focal-IR (2026)	38.72	0.9555	0.0081	0.0311	38.22	0.9509	0.0083	0.0412	33.92	0.8837	0.0134	0.0926
AgenticCT (Ours)	**40.20**	**0.9564**	**0.0072**	**0.0172**	**39.65**	**0.9554**	**0.0072**	**0.0240**	**35.64**	**0.9026**	**0.0114**	**0.0580**

Table 1: Quantitative comparison of average performance on various combinations of degradations in the DeepLesion, AAPM-Mayo Clinic, and COVID-19 Diagnosis datasets. The best and second-best results are **bolded** and underlined, respectively.

the AAPM-Mayo and COVID-19 datasets, showing that the improvement isn't specific to just one dataset. By grounding decisions in a structured intermediate state and a traceable execution order, AgenticCT is less likely to encounter issues like over-smoothing or leftover structured artifacts, making it more reliable for transfer than simultaneous correction.

4.3 Scenario-wise Performance

Figure 3 displays the scenario-wise results on AAPM-Mayo Clinic test set for both single- and mixed-degradation compositions, providing additional context to the dataset-level averages shown in Table 1. For single-degradation cases, AgenticCT remains consistently competitive and often the strongest across metrics, indicating that our decomposition does not sacrifice performance when only one artifact dominates. This suggests that the Supervisor can reliably identify the active factor and the Planner tends to select a concise, near-specialized trajectory, avoiding unnecessary operations and preventing over-correction. For mixed-degradation cases, the gap between AgenticCT and *all-in-one* baselines becomes smaller, indicating that implicit perception can handle mixed artifacts to some extent. However, most baselines show larger variance and sharper drops, reflecting conflicts under simultaneous correction. Aggressive denoising weakens edge cues for geometric correction, while early artifact suppression propagates errors. In contrast, AgenticCT remains the most stable method across mixtures. This stability is achieved through the *Diagnose–Plan–Treat* design, where explicit diagnosis reveals the degradation mixture and the Planner selects an order-aware trajectory that preserves critical cues and mitigates artifact-anatomy entanglement.

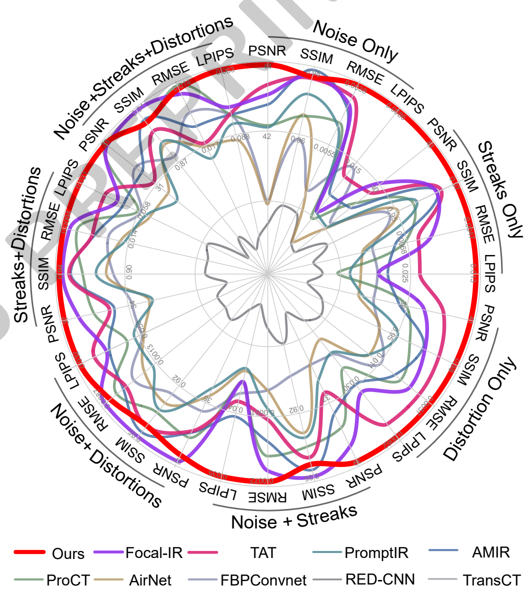

Figure 3: Scenario-wise results on different degradations. Metrics are min–max normalized across all methods; for RMSE and LPIPS, values were inverted as $1 -$ value so that larger values are better.

4.4 Qualitative Results

Figure 4 presents representative visual comparisons for both single-degradation and mixed-degradation cases. For single degradations, specialized baselines can somewhat mitigate dominant artifacts, but they often come with trade-offs, like leaving behind structured patterns or excessively smoothing out fine anatomical details, especially when the degradation is severe. For mixed degradations, these issues are amplified. Simultaneous all-in-one correction frequently produces averaged reconstructions, where streak-like artifacts are partially absorbed into anatomical textures and subtle boundaries become blurred, reflecting artifact entanglement and premature smoothing. In contrast, AgenticCT yields more faithful reconstructions in both regimes. By explicitly diagnosing the degradation mixture and executing an order-aware trajectory, the framework better preserves high-frequency structural cues while suppressing both noise and structured artifacts. This leads to sharper boundaries and more consistent recovery of subtle details in the highlighted regions, providing qualitative evidence that complements the quantitative gains under both single and mixed degradations.

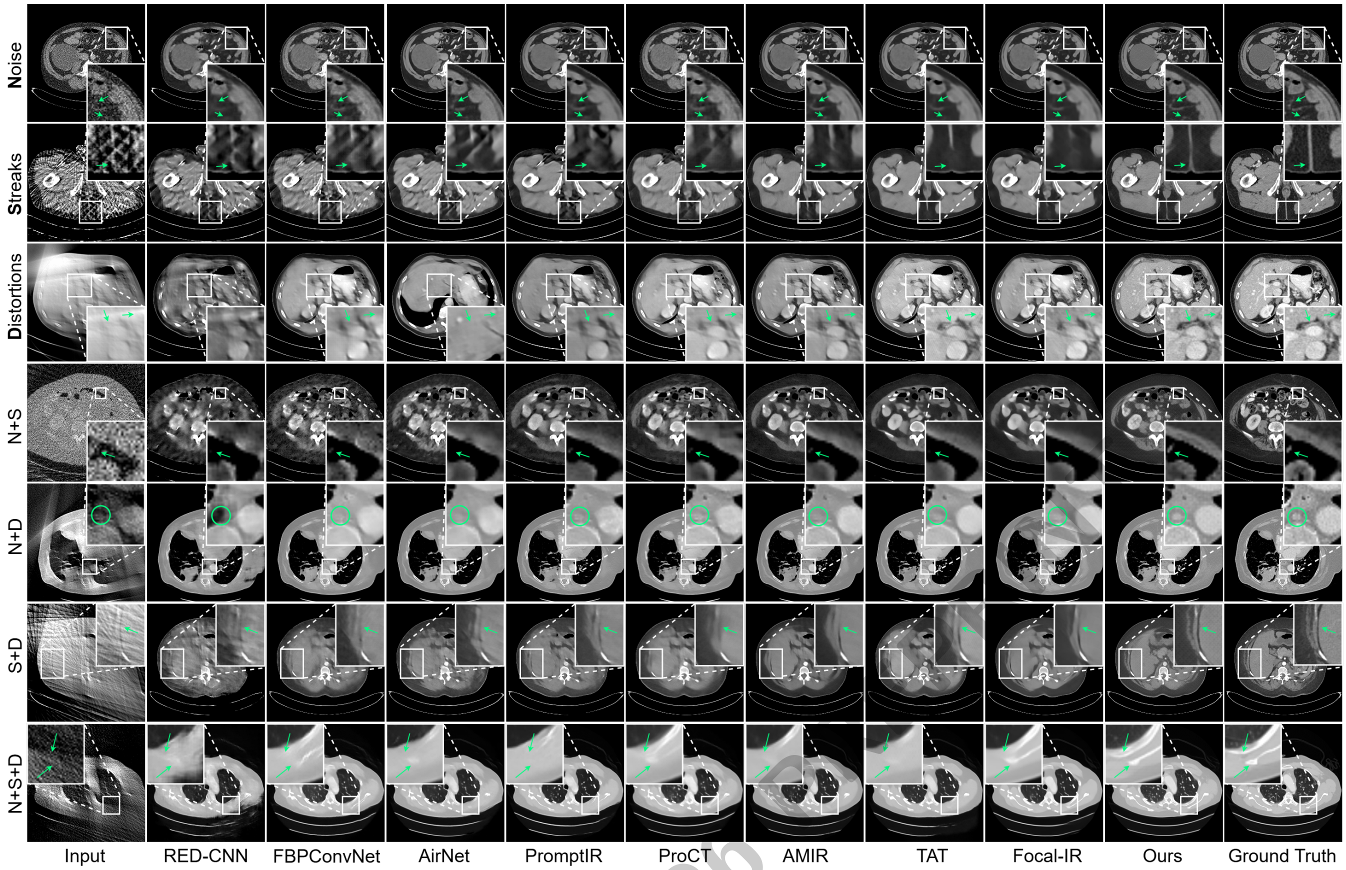

Figure 4: Qualitative comparisons on representative CT slices under single and mixed degradations.

SA	PA	PC	DeepLesion		AAPM-Mayo	
			PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
✗	✗	✗	35.44	0.9204	35.23	0.9166
✗	✗	✓	36.35	0.9312	35.31	0.9225
✗	✓	✓	36.70	0.9352	35.65	0.9253
✓	✓	✓	**37.04**	**0.9378**	**36.03**	**0.9280**

Table 2: Ablation of AgenticCT components.

4.5 Ablation Study

Table 2 evaluates the contributions of explicit diagnosis (SA), trajectory planning (PA), and processor collaboration (PC).

When all components are removed, the system defaults to implicit simultaneous correction. This setting is mismatched to mixed degradations because reconstruction is order-sensitive and objectives can conflict. For example, strong denoising may suppress high-frequency cues needed for distortion correction, while aggressive artifact suppression can blur subtle anatomical boundaries. Without an explicit execution structure, the model is more likely to produce averaged corrections that trade off competing goals, leading to weaker structural preservation and more residual artifacts.

Enabling PC alone already improves results, showing that enforcing sequential execution is beneficial even with a fixed order. Sequential processing reduces direct objective interference by letting each processor focus on a narrower correction at a time. However, a universal order cannot adapt to instance-wise mixture composition and severity, so it remains suboptimal when the dominant degradation changes, and can still induce premature smoothing or error propagation.

Adding PA on top of PC yields further gains, indicating that instance-wise trajectory selection is preferable to a fixed order. In this variant, because SA is absent, the Planner lacks structured degradation state s and must predict the execution order based solely on the degraded image x, learning $P(\pi|x)$ rather than $P(\pi|x, s)$. While this image-conditioned planning can partially adapt the execution order, it is inherently more ambiguous under degradation mixtures and more sensitive to domain shift, which limits the reliability of ordering decisions in challenging cases.

Finally, when all components are enabled, the system achieves the best performance across both datasets. Explicit diagnosis through SA provides a structured and interpretable degradation state, guiding the Planner to select a more informed and precise trajectory. This structured planning enhances the stability of sequential execution, allowing the system to better handle complex degradation mixtures. This supports our central claim that the combination of explicit state estimation, order-aware planning, and sequential execution is crucial for effective mixed-degradation reconstruction.

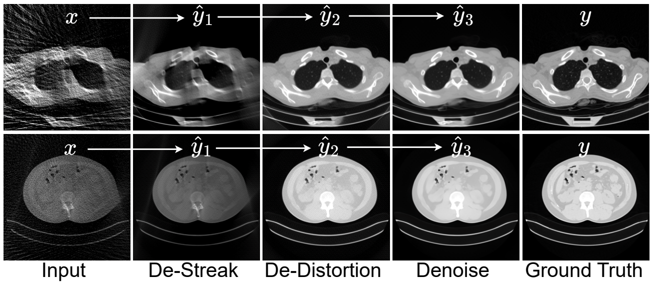

Figure 5: Intermediate results at each processing step.

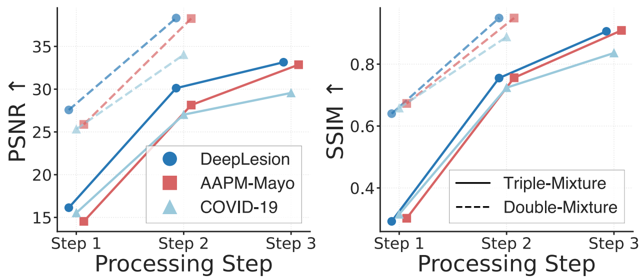

Figure 6: Quantitative analysis across processing steps.

4.6 Intermediate Results

Figure 5 visualizes the intermediate outputs along the planned trajectory. As processing proceeds, mixed artifacts are progressively decoupled. The intermediate results show that each processor makes a targeted contribution and that improvements accumulate rather than conflict, consistent with the non-commutative nature of mixed-degradation reconstruction. Figure 6 provides quantitative support. Metrics increase monotonically with processing steps across all datasets, indicating that the benefit comes from stepwise decomposition rather than dataset-specific fitting. In addition, more complex mixtures rely more on multi-step execution: early steps yield the largest gains by removing the most harmful components, while later steps provide smaller but stable refinements, reflecting diminishing returns after major artifacts are suppressed.

4.7 Processing Order

Mixed-degradation reconstruction is order-sensitive because reconstruction operators are generally non-commutative. A suboptimal schedule can prematurely suppress high-frequency cues needed by later corrections, leading to over-smoothing. We therefore enumerate candidate execution orders and use the average performance over all possible orders as a reference. As shown in Table 3, the Planner-selected order consistently outperforms this average, with larger gains on triple mixtures than on double mixtures, indicating that harmful orders become more common as mixture complexity increases. Figure 1 provides qualitative confirmation. Simultaneous correction and suboptimal orders either blur boundaries or leave residual structured artifacts in the highlighted regions, whereas our planned sequential execution preserves edges while better suppressing mixed artifacts.

# of Degradations	Double Mixture		Triple Mixture	
Metrics	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Avg. of All Possible Orders	32.50	0.8665	28.27	0.8052
Order Porvided by Planner	**33.08**	**0.8707**	**29.22**	**0.8167**

Table 3: Effect of processing order on reconstruction performance.

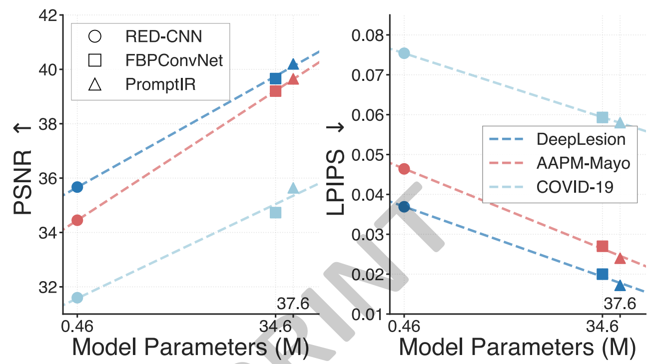

Figure 7: Processor scaling analysis.

4.8 Processor Scaling

A key advantage of AgenticCT is that the Processor Agents act as modular executors and can be instantiated with different backbone architectures. Figure 7 evaluates this flexibility by replacing the Processor model with architectures of increasing capacity while keeping the *Diagnose–Plan–Treat* pipeline unchanged. Across datasets, we observe a consistent scaling trend. As the Processor capacity grows, reconstruction quality improves (higher PSNR), and perceptual score reduces (lower LPIPS). These results indicate that AgenticCT does not depend on a particular processor design. Instead, the explicit state and trajectory formulation allows the framework to benefit from stronger executors when available. This modularity makes AgenticCT flexible and extensible, and it enables the system to scale with model capacity and to incorporate future advances in reconstruction backbones without modifying the Supervisor or Planner.

5 Conclusion

We presented AgenticCT, an LLM-orchestrated *Diagnose–Plan–Treat* framework for mixed-degradation CT reconstruction. By explicitly estimating the degradation state and choosing a reconstruction path, AgenticCT replaces one-time fixes with a more thoughtful, step-by-step approach. Extensive experiments on DeepLesion and zero-shot evaluations on AAPM-Mayo and COVID-19 demonstrate improved fidelity and stronger robustness under domain shift. Additional analyses further verify the importance of processing order, show monotonic quality gains across intermediate steps, and highlight the modularity of our pipeline through processor scaling. These results suggest a practical direction toward high-fidelity CT reconstruction by combining explicit perception, planning, and specialized execution.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62271335, and in part by Sichuan Science and Technology Program under Grant 2025ZNSFSC0470.

References

[Afshar *et al.*, 2022] Parnian Afshar, Moezedin Javad Rafiee, Farnoosh Naderkhani, Shahin Heidarian, Nastaran Enshaei, Anastasia Oikonomou, Faranak Babaki Fard, Reut Anconina, Keyvan Farahani, Konstantinos N Plataniotis, et al. Human-level covid-19 diagnosis from low-dose ct scans using a two-stage time-distributed capsule network. *Scientific Reports*, 12(1):4827, 2022.

[Barrett and Keat, 2004] Julia F Barrett and Nicholas Keat. Artifacts in ct: recognition and avoidance. *Radiographics*, 24(6):1679–1691, 2004.

[Chen *et al.*, 2017] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, 36(12):2524–2535, 2017.

[Chen *et al.*, 2024] Hongchi Chen, Qiuxia Li, Lazhen Zhou, and Fangzuo Li. Deep learning-based algorithms for low-dose ct imaging: A review. *European Journal of Radiology*, 172:111355, 2024.

[Cui *et al.*, 2026] Yuning Cui, Wenqi Ren, and Alois Knoll. Focal modulation for image restoration. *International Journal of Computer Vision*, 134(1):6, 2026.

[Hendee and Edwards, 1986] William R Hendee and F Marc Edwards. Alara and an integrated approach to radiation protection. In *Seminars in nuclear medicine*, volume 16, pages 142–150. Elsevier, 1986.

[Hsieh, 2003] Jiang Hsieh. Computed tomography: principles, design, artifacts, and recent advances. 2003.

[ICRP and others, 1977] Ann ICRP et al. Recommendations of the icrp. *Ann. ICRP*, 1(3), 1977.

[Jin *et al.*, 2017] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[Li *et al.*, 2022] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022.

[Liu *et al.*, 2025] Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, et al. Deepseek-v3. 2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*, 2025.

[Ma *et al.*, 2023] C Ma, Z Li, J He, J Zhang, Y Zhang, and H Shan. Universal incomplete-view ct reconstruction with prompted contextual transformer. *arXiv preprint arXiv:2312.07846*, 2023.

[McCollough *et al.*, 2017] Cynthia H McCollough, Adam C Bartley, Rickey E Carter, Baiyu Chen, Tammy A Drees, Phillip Edwards, David R Holmes III, Alice E Huang, Farhana Khan, Shuai Leng, et al. Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge. *Medical physics*, 44(10):e339–e352, 2017.

[Mehri *et al.*, 2021] Armin Mehri, Parichehr B Ardakani, and Angel D Sappa. Mprnet: Multi-path residual network for lightweight image super resolution. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2704–2713, 2021.

[Pedersen *et al.*, 2025] Rasmus Juul Pedersen, Luke Besley, Jakob Sauer Jørgensen, Jens Wenzel Andreasen, Anders Bjorholm Dahl, and Vedrana Andersen Dahl. Adversarially informed neural fields for computed tomography reconstruction. In *Scandinavian Conference on Image Analysis*, pages 131–144. Springer, 2025.

[Potlapalli *et al.*, 2023] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one blind image restoration. In *Advances in Neural Information Processing Systems*, 2023.

[Standley *et al.*, 2020] Trevor Standley, Amir Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? In *International conference on machine learning*, pages 9120–9132. PMLR, 2020.

[Wang *et al.*, 2020] Ge Wang, Jong Chul Ye, and Bruno De Man. Deep learning for tomographic image reconstruction. *Nature machine intelligence*, 2(12):737–748, 2020.

[Würfl *et al.*, 2018] Tobias Würfl, Mathis Hoffmann, Vincent Christlein, Katharina Breininger, Yixin Huang, Mathias Unberath, and Andreas K Maier. Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems. *IEEE transactions on medical imaging*, 37(6):1454–1463, 2018.

[Yan *et al.*, 2017] Ke Yan, Xiaosong Wang, Le Lu, and Ronald M Summers. Deeplesion: Automated deep mining, categorization and detection of significant radiology image findings using large-scale clinical lesion annotations. *arXiv preprint arXiv:1710.01766*, 2017.

[Yang *et al.*, 2018] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, 37(6):1348–1357, 2018.

[Yang *et al.*, 2024] Zhiwen Yang, Haowei Chen, Ziniu Qian, Yang Yi, Hui Zhang, Dan Zhao, Bingzheng Wei, and

Yan Xu. All-in-one medical image restoration via task-adaptive routing. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 67–77. Springer, 2024.

[Yang *et al.*, 2025] Zhiwen Yang, Jiaju Zhang, Yang Yi, Jian Liang, Bingzheng Wei, and Yan Xu. Tat: Task-adaptive transformer for all-in-one medical image restoration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 565–575. Springer, 2025.

[Zhang and Wu, 2019] Xudong Zhang and Pengxiang Wu. Heuristic search for homology localization problem and its application in cardiac trabeculae reconstruction. In *IJCAI*, 2019.

[Zhang *et al.*, 2021] Zhicheng Zhang, Lequan Yu, Xiaokun Liang, Wei Zhao, and Lei Xing. Transct: dual-path transformer for low dose computed tomography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 55–64. Springer, 2021.

[Zhang *et al.*, 2024] Ju Zhang, Weiwei Gong, Lieli Ye, Fanghong Wang, Zhibo Shangguan, and Yun Cheng. A review of deep learning methods for denoising of medical low-dose ct images. *Computers in Biology and Medicine*, 171:108112, 2024.