PhySe-RPO: Physics and Semantics Guided Relative Policy Optimization for Diffusion-Based Surgical Smoke Removal

Supplementary Material

6. Additional Experimental Results

6.1. Results on Synthetic Datasets

To verify that our model has basic desmoking capabilities at cold start, we tested it on a synthetic dataset. The experimental results are shown in Table 6,our method achieves the best overall performance across all four evaluation metrics, demonstrating strong desmoking ability even at cold start. In terms of distortion-oriented metrics, our approach attains the highest PSNR (35.2552 dB) and SSIM (0.9776), indicating that the restored images are closest to the ground truth in both pixel accuracy and structural fidelity. For perceptual quality, our method also achieves the lowest LPIPS (0.0227), substantially outperforming recent deep models including PFAN, Dehamer, and LightDiff. In addition, our model attains the best FID (30.5633) among all competitors, which is nearly 40–90 points lower than traditional and GAN-based baselines, indicating that the generated images are not only perceptually plausible but also distributionally closer to clean images.

Table 6. Comparison of image desmoking methods on the synthetic dataset in cold start.

Method	PSNR↑	SSIM↑	LPIPS↓	FID↓
DCP[10]	29.0136	0.8529	0.1758	121.6056
Desmoke_LAP[24]	30.9505	0.9144	0.0975	81.8183
SelfSVD[37]	31.2658	0.8780	0.1237	150.4745
PFAN[43]	30.5614	0.9030	0.1068	111.8937
Dehamer[8]	31.6271	0.9627	0.0406	67.8821
Tap[7]	31.5969	0.9622	0.0472	74.9391
LightDiff[5]	30.5844	0.9431	0.0874	49.9105
Noise-DA[15]	31.2110	0.9568	0.0505	66.1669
DGFDNet[45]	30.7754	0.9677	0.0452	46.5128
Ours	**35.2552**	**0.9776**	**0.0227**	**30.5633**

In addition to the quantitative results, qualitative comparisons on the synthetic dataset are presented in Figure 6. Classical prior-based methods, such as DCP [10], tend to produce over-saturated colors and residual haze, while learning-based approaches like PFAN [43] and Dehamer [8] achieve cleaner results but often suffer from texture loss or halo artifacts around structural boundaries. Diffusion-based methods (e.g., LightDiff [5]) restore global illumination effectively yet sometimes generate overly smooth surfaces due to weak structural regularization. In contrast, our method reconstructs fine anatomical textures and preserves surgical structures more faithfully.

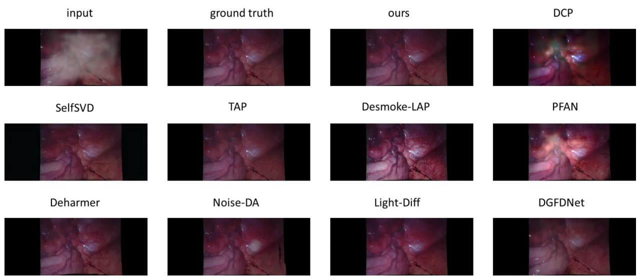

Figure 6. Qualitative comparison on synthetic images,compared to other methods, PhySe-RPO exhibits better desmoking capabilities in the cold start phase.

6.2. Ablation Study in cold start

To further understand the contribution of each component in our framework, we conducted ablation experiments focusing on the Semantic Feature Fusion (SF) and the proposed Temporal Adapter (TA). As shown in Table 7, both SF and the TA provide clear benefits under the cold-start setting. Adding SF yields a notable gain over the baseline , showing that semantic cues effectively guide the diffusion process toward structure-preserving restoration. TA also improves performance by enhancing cross-step consistency. Combining both modules leads to the best results for PSNR(35.2552 dB) and SSIM(0.9776), demonstrating their complementary roles in strengthening the model's early-stage desmoking capability.

Table 7. Ablation results of cold start on the synthetic dataset.

Model	PSNR↑	SSIM↑
Baseline	34.0958	0.9748
Baseline+SF	34.9618	0.9764
Baseline+TA	34.2361	0.9752
Ours(Baseline+SF+TA)	**35.2552**	**0.9776**

6.3. Ablation on the Physics-Guided Reward Weight

To analyze the influence of the physics-guided term, we reformulate the overall reward as:

$$R = \alpha R_{PG} + R_{RF} + R_{VC}, \tag{18}$$

where α controls the strength of the physics-guided prior. As shown in Figure 7,we observe that setting α=1 overemphasizes the color prior and leads to reddish artifacts in the desmoked outputs. This effect arises because an excessively strong physics-guided constraint encourages overcorrection of illumination imbalance, unintentionally amplifying the red channel in surgical scenes. By changing α, the model achieves more natural color reproduction while preserving dehazing performance. The results demonstrate that properly tuning α is essential to avoid color shifts and maintain visually balanced reconstruction quality.

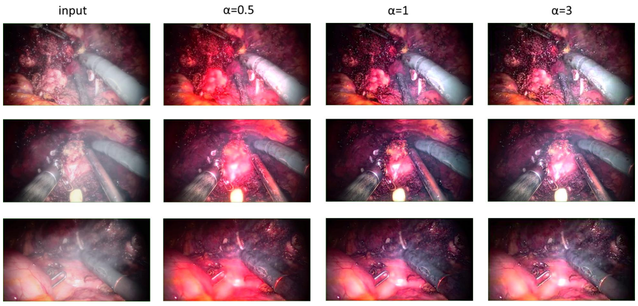

Figure 7. Output results for different α values

6.4. Hyperparameter experiments

We further investigate two key hyperparameters in our framework: the length of the learning token embedding and the number of groups G used in the PhySe-RPO. The learning token length determines the expressiveness of the learned visual concepts. As shown in Table 8, varying the token length has only a marginal impact on performance. Shorter embeddings already capture sufficient semantic cues to guide the diffusion process, while increasing the token length yields only slight improvements in PSNR and SSIM. This indicates that the model is relatively robust to the choice of token dimension, and the semantic concepts can be effectively encoded even with compact embeddings. Therefore, a moderate token length provides a good balance between representation capacity and computational cost, without significantly affecting overall performance.

Table 8. Hyperparameter study on token length.

Length of token	PSNR↑	SSIM↑
10	35.2440	0.9762
25	35.2552	**0.9776**
50	**35.2558**	0.9774

The group number G controls the granularity of the PhySe-RPO. As shown in Table 9, small G values reduce exploration diversity and may lead to unstable reward estimation, limiting the benefits of PhySe-RPO. As G increases, the model receives more reliable group-level comparisons, enabling more stable and effective policy updates. However, an excessively large G value will introduce additional computational load and memory overhead. Our experiments show that moderate group sizes strike the best balance between exploration, stability, and computational cost, providing the strongest performance.

Table 9. Hyperparameter study on number of groups G.

Number of G	SSEQ↓	MANIQA↑	PI↓	FADE↓	MUSIQ↑	IS↑	NIQE↓
2	4.036	0.346	3.296	0.279	53.751	2.796	4.777
4	**3.443**	**0.378**	**3.125**	**0.216**	54.911	2.797	**4.608**
8	3.517	0.374	3.235	0.232	**55.125**	**2.806**	4.657