

## 872 **A Appendix**

### 873 **Contents**

874	A.1 Model Algorithm and Implementation Details . . . . .	23
875	Implementation Details . . . . .	23
876	Training Algorithm . . . . .	23
877	A.2 Additional Results . . . . .	24
878	A.2.1 Additional Results with Inpainting-Based Editing Models . . . . .	24
879	A.2.2 Additional Comparisons with Inpainting-Based Editing Models . . . . .	25
880	A.2.3 Additional Results with Instruction-Based Editing Model . . . . .	26
881	A.2.4 Additional Robustness Evaluation . . . . .	27
882	A.2.5 Additional Results with Non-ROI Editing . . . . .	29
883	A.3 Imperceptibility Discussion . . . . .	30
884	A.4 Transferability of Immunization Noise . . . . .	31
885	A.5 Robustness to Test-Time Mask Variability . . . . .	32
886	A.6 Video Evaluation . . . . .	33
887	A.7 Dataset Setup . . . . .	34
888	A.8 Prompt-Agnostic Immunization Experiment . . . . .	35
889	A.9 Loss Weight Selection . . . . .	36
890	A.10 User Study . . . . .	37

891 You can find our demo code and the complete immunized videos along with their corresponding  
892 video edits in the provided zip file, located in the ‘supp/code’ and ‘supp/videos’ folders, respectively.

## 893 A.1 Model Algorithm and Implementation Details

894 **Implementation Details** We train our immunizer model for 350 epochs using a batch size of 5 on an  
 895 NVIDIA A100 GPU. We use the Adam optimizer (Kingma and Ba, 2015) with an initial learning  
 896 rate of 0.00001 and set the loss weight parameter  $\alpha = 4$ . Training takes approximately 22 hours  
 897 and leverages 16-bit precision to reduce memory consumption and speed up computation. For the  
 898 editing tools, we use a pre-trained Stable Diffusion v1.5 inpainting model (Rombach et al., 2022) for  
 899 inpainting-based editing, and InstructPix2Pix (Brooks et al., 2023) for instruction-based editing tasks.

900 **Training Algorithm** Algorithm 1 describes the end-to-end training procedure for our immunizer  
 901 model. For each data sample, the model generates an immunized image by injecting noise into the  
 902 masked region. This image is then edited using a black-box editing model. The training objective  
 903 minimizes both the deviation from the original image in the masked region and the effectiveness of  
 904 the edit in the unmasked region.

---

### Algorithm 1 End-to-end Training Framework

---

**Input:** Immunizer model  $f(\cdot; \theta)$ , Editing model  $\text{SD}(\cdot)$ , Dataset  $\mathcal{D}$ ,  
 Dataset size  $N$ , Loss weight  $\alpha$

```

for  $n = 1$  to  $N$  do
   $(\mathbf{I}^n, \mathbf{M}^n, \mathcal{P}^n) \leftarrow \text{sample}(\mathcal{D}, n)$ 
   $\epsilon_{\text{im}}^n \leftarrow f(\mathbf{I}^n; \theta)$ 
   $\mathbf{I}_{\text{im}}^n \leftarrow (\mathbf{I}^n + \epsilon_{\text{im}}^n \odot \mathbf{M}^n). \text{clamp}(0, 1)$ 
   $\mathbf{I}_{\text{im,edit}}^n \leftarrow \text{SD}(\mathbf{I}_{\text{im}}^n, \sim \mathbf{M}^n, \mathcal{P}^n)$ 
   $\mathcal{L}_{\text{noise}} \leftarrow \text{normalize}(\|\mathbf{I}_{\text{im}}^n - \mathbf{I}^n\|_1)$ 
   $\mathcal{L}_{\text{edit}} \leftarrow \text{normalize}(\|\mathbf{I}_{\text{im,edit}}^n \odot (\sim \mathbf{M}^n)\|_1)$ 
   $\mathcal{L} \leftarrow \alpha \cdot \mathcal{L}_{\text{noise}} + \mathcal{L}_{\text{edit}}$ 
   $\theta \leftarrow \text{update}(\nabla_{\theta} \mathcal{L})$ 
end for

```

---

## 906 A.2 Additional Results

### 907 A.2.1 Additional Results with Inpainting-Based Editing Models

908 Figure 6 presents supplementary qualitative results obtained using inpainting-based editing models.  
 909 The examples cover a wide range of scenarios and prompts, demonstrating the effectiveness of  
 910 our immunization method on previously unseen content. Notably, the model performs well even  
 911 on close-up images, maintaining robustness against malicious edits in both broad and fine-grained  
 912 contexts.



Figure 6: *Additional qualitative results with DiffVax*. Each row displays a different prompt and input image, illustrating DiffVax’s ability to consistently disrupt harmful edits. Despite varying and challenging prompts, the edited outputs from the protected images show clear signs of disruption, emphasizing the robustness of our method.

### A.2.2 Additional Comparisons with Inpainting-Based Editing Models

Figure 7 shows extended qualitative comparisons between DiffVax and various baseline immunization methods, including Random Noise, PhotoGuard-E, PhotoGuard-D, and DiffusionGuard. These results are produced using inpainting-based editing models. The comparison highlights how DiffVax consistently achieves better performance in visually disrupting malicious edits while preserving the semantic integrity of the original image.

We note that other defense methods such as AdvDM (Liang et al., 2023), SDS (Xue et al., 2024), and Mist (Liang and Wu, 2023) have also been proposed in the literature. However, these techniques are tailored for specific editing pipelines like SDEdit (Meng et al., 2022) and are not directly applicable in our inpainting-based setup, thus making direct comparison beyond our experimental scope.

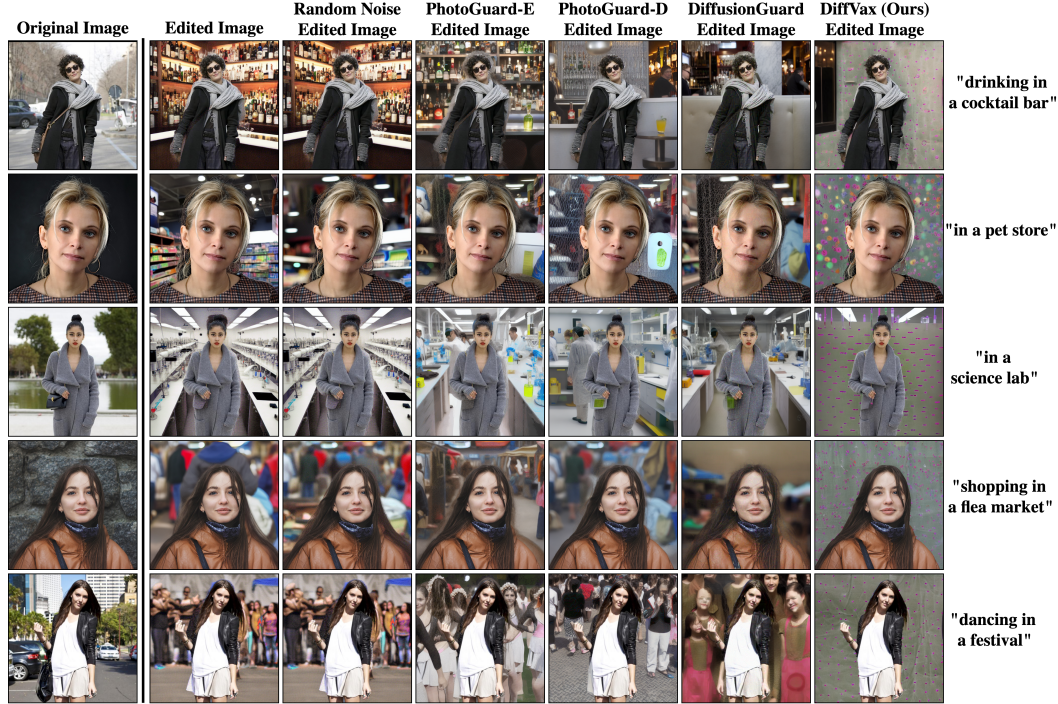


Figure 7: *Additional qualitative comparison between baselines and DiffVax.* Each row represents a unique prompt-image pair, while the columns show outputs for different immunization methods. DiffVax consistently produces better results, effectively disrupting edits while preserving image quality.

### A.2.3 Additional Results with Instruction-Based Editing Model

To further evaluate the generalizability of DiffVax, we apply it to edits generated using Instruct-Pix2Pix (Brooks et al., 2023), a widely adopted text-guided diffusion-based editing tool. This setting differs significantly from inpainting models, as edits are applied based on high-level natural language instructions. As shown in Figure 8, DiffVax consistently disrupts a broad range of editing intents across various image types. The examples illustrate the model’s robustness across:

- **Human attribute edits** (e.g., "add a hat to her head", "add bowtie to person", "make him wear a small scarf"): DiffVax suppresses the addition of these features, effectively neutralizing changes to facial and clothing attributes.
- **Background edits** (e.g., "make the background a chapel", "change him to a statue"): Despite significant changes to the scene, the edits fail to render properly on immunized images, showcasing DiffVax’s ability to neutralize edits in large non-focal areas.
- **Style transfer edits** (e.g., "change the style to starry nights", "make the style cubism", "van gogh style"): DiffVax prevents global transformations from taking effect, demonstrating its efficacy in blocking even abstract stylistic alterations.
- **Non-ROI edits** (e.g., "add hot-air balloons to back", "add necklace to person", "add headphones"): These involve subtle object insertions in the background or around the subject. Even though the modification targets are not directly in the immunized region, DiffVax still effectively disrupts the edit.

These results validate the model-agnostic and instruction-resilient nature of DiffVax, confirming its applicability to both local and global edit intents.

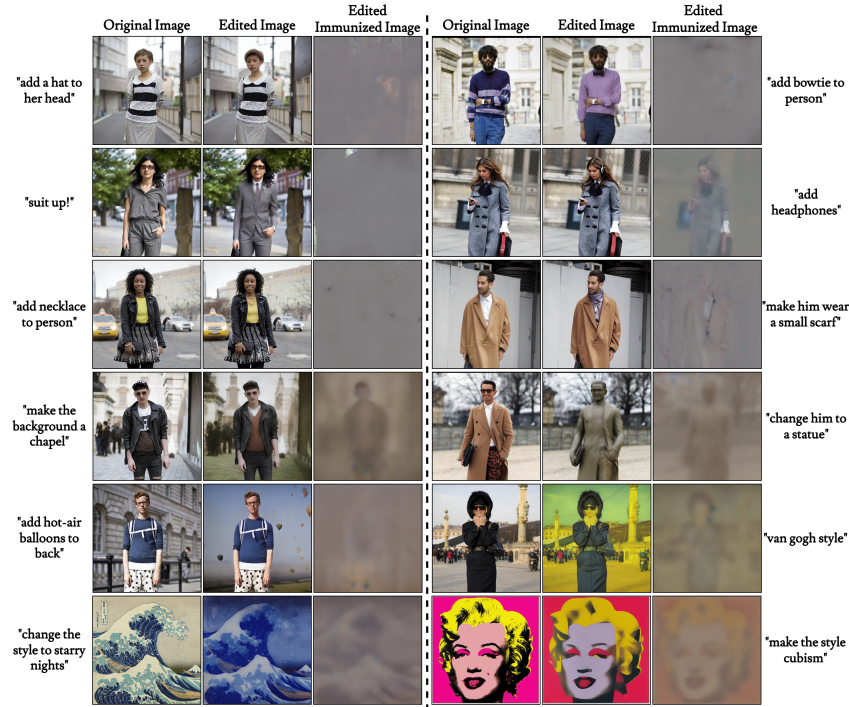


Figure 8: *Qualitative results using the InstructPix2Pix (Brooks et al., 2023) editing model with DiffVax.* Each triplet shows an original image, its edited counterpart, and the result after immunization. DiffVax successfully prevents a diverse set of edits, including background replacement, style transfer, object insertion, and attribute modification, further demonstrating its generalizability across editing types.

## A.2.4 Additional Robustness Evaluation

JPEG compression and denoising techniques are typically designed to remove high-frequency components from images. Since our immunizer model introduces primarily low-frequency perturbations—due to the design of our noise loss—it becomes inherently more robust against such counterattacks.

Table 4 reports results under various JPEG compression ratios and when using IMPRESS (Cao et al., 2023), a model specifically developed for adversarial purification and denoising. Across all configurations, *DiffVax* consistently outperforms PhotoGuard-D and DiffusionGuard in terms of SSIM, SSIM (Noise), and CLIP-T metrics. These results suggest that *DiffVax* maintains its protective efficacy even when subjected to aggressive counterattack scenarios.

Figure 9 presents qualitative results of two counterattack strategies: (a) applying a denoiser and (b) applying JPEG compression. The edited image, along with its attacked counterpart, is shown for both PhotoGuard-D and *DiffVax*. While the visual changes for PhotoGuard-D are significant—indicating its vulnerability to counterattacks—*DiffVax* retains its robustness, preventing successful malicious edits.

To further explore robustness, Figure 13 presents additional qualitative comparisons under varying JPEG compression ratios (from 0.85 to 0.55) and under the IMPRESS purification attack. Even at high compression levels, *DiffVax* continues to disrupt the edits, showcasing its superior generalization and resistance to counter-editing.

Metric	<i>DiffVax</i> (JPEG .85)	DG (JPEG .85)	PG (JPEG .85)	<i>DiffVax</i> (JPEG .65)	DG (JPEG .65)	PG (JPEG .65)	<i>DiffVax</i> (JPEG .55)	DG (JPEG .55)	PG (JPEG .55)	<i>DiffVax</i> (IMPRESS)	DG (IMPRESS)	PG (IMPRESS)
SSIM ↓	<b>0.517</b>	0.646	0.640	<b>0.530</b>	0.696	0.692	<b>0.534</b>	0.706	0.693	<b>0.489</b>	0.605	0.578
SSIM (Noise) ↑	<b>0.968</b>	0.955	0.961	<b>0.951</b>	0.946	0.950	<b>0.944</b>	0.940	0.944	<b>0.644</b>	0.636	0.640
CLIP-T ↓	<b>25.76</b>	30.83	32.00	<b>26.83</b>	31.80	32.15	<b>27.67</b>	31.93	32.20	<b>24.67</b>	30.71	31.35

Table 4: *Additional counterattack experiments.* The SSIM, SSIM (Noise), and CLIP-T metrics are reported for JPEG compression with ratios of 0.85, 0.65, and 0.55, as well as for the adversarial purification model IMPRESS. The metrics demonstrate that *DiffVax* consistently outperforms PG-D and DG, even when counterattacks are applied to all methods.



Figure 9: *Qualitative results of counter-attacks on immunization methods.* The first row shows results when an off-the-shelf denoiser is applied to the immunized image, while the second row displays results under JPEG compression. Columns 2–3 correspond to PhotoGuard-D, while columns 4–5 show results for DiffVax. PhotoGuard-D is visibly more susceptible to counterattacks, whereas DiffVax maintains strong protection.

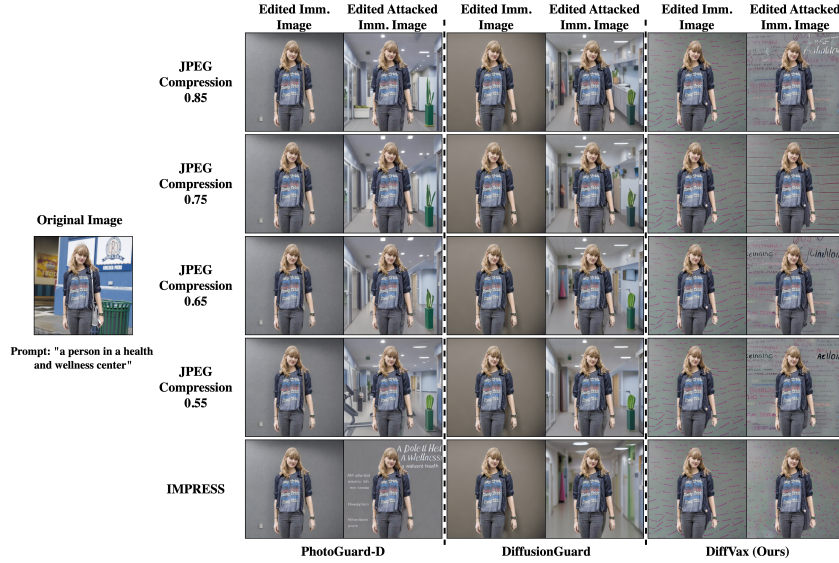


Figure 10: *Additional qualitative results of counter-attacks on immunization methods.* Each row corresponds to a different JPEG compression ratio or the IMPRESS model. DiffVax shows robust behavior across all levels, continuing to suppress harmful edits even under heavy degradation or purification.

### 963 A.2.5 Additional Results with Non-ROI Editing

964 To evaluate the generalizability of DiffVax beyond human-centric content, we conduct experiments  
 965 on non-human subjects, such as animals and other inanimate objects. As illustrated in Figure 11,  
 966 DiffVax effectively immunizes these non-person regions, preventing malicious edits while preserving  
 967 the visual fidelity of the original image. These results further demonstrate the versatility and zero-shot  
 968 capabilities of DiffVax across diverse object domains.



Figure 11: *Qualitative results for non-human objects edited using DiffVax.* These examples show that DiffVax extends effectively to domains beyond human subjects, maintaining its edit-resistance and imperceptibility.

### A.3 Imperceptibility Discussion

To evaluate the imperceptibility of the perturbations introduced by DiffVax, we present qualitative comparisons against PhotoGuard in Figure 12. Our method generates noise that is concentrated in the low-frequency components of the image, making it visually more subtle and less disruptive. In contrast, PhotoGuard introduces high-frequency noise that appears scattered across broader regions.

This low-frequency characteristic of DiffVax offers two key advantages. First, it enhances the perceptual quality of the immunized images by producing smoother perturbations that minimally interfere with semantic content. Second, it contributes to robustness against counterattacks such as JPEG compression or denoising—these techniques are typically designed to suppress high-frequency information, which is assumed to correspond to noise. Since DiffVax avoids relying on high-frequency artifacts, its perturbations are more likely to survive such transformations, preserving the protective effect.

We further examine the role of the loss norm in shaping the visual quality of the immunization. As shown in Figure 13, using  $L_2$  or  $L_\infty$  norms leads to less perceptible perturbations than the default  $L_1$  formulation. However, this comes at the expense of reduced edit resistance, underscoring a critical trade-off between imperceptibility and robustness.

Future work will explore more principled approaches to navigating this trade-off, such as incorporating perceptual similarity metrics or frequency-domain regularization directly into the optimization objective.

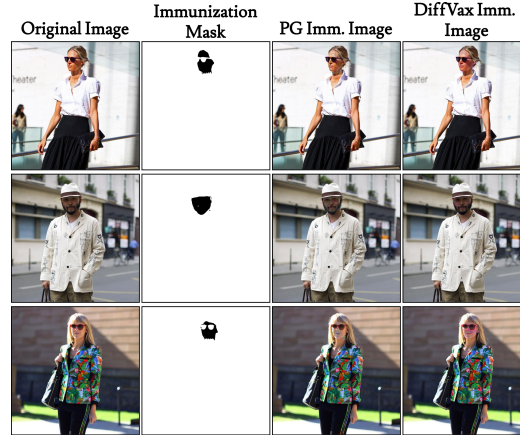


Figure 12: *Comparison of immunization noise.* Visual comparison of immunized images generated by PhotoGuard and DiffVax using a face mask. PhotoGuard produces scattered and higher-frequency noise, while DiffVax generates smoother, low-frequency perturbations.



(a) Immunization with different norms

Figure 13: *Additional comparison of immunization noise under different norms.* This figure compares immunized images generated using different norm constraints:  $L_1$ ,  $L_2$ , and  $L_\infty$ , as well as results from PhotoGuard and DiffusionGuard.

#### 988 A.4 Transferability of Immunization Noise

989 While no existing method has been shown to generate immunization perturbations that reliably gener-  
 990 alize across different editing models, our approach demonstrates a promising level of transferability.  
 991 Specifically, when the immunization noise is trained on Stable Diffusion v1.5 and evaluated on Stable  
 992 Diffusion v2, DiffVax retains its effectiveness in disrupting edits, unlike PhotoGuard (see Figure 14).

993 This result highlights the ability of DiffVax to generalize across unseen editing models without  
 994 retraining. In contrast, the perturbations generated by PhotoGuard are highly model-specific and  
 995 fail to transfer effectively across versions. We believe this generalization capability is crucial for  
 996 real-world deployment, where the specific editing model may not be known in advance or may evolve  
 997 over time.

998 Improving the robustness and transferability of immunization remains an open research challenge,  
 999 and our findings suggest that DiffVax provides a strong foundation for future work in this direction.



Figure 14: *Transferability of perturbations across editing models.* Red labels indicate the immunization training model, and blue labels denote the editing model. The results show how well each immunized image resists edits across different model configurations. When trained on Stable Diffusion (SD) v1.5, DiffVax successfully prevents edits even when tested on SD v2. In contrast, PhotoGuard’s perturbations trained on SD v1.5 do not generalize to SD v2. These results illustrate the superior cross-model generalizability of DiffVax.

## 1000 A.5 Robustness to Test-Time Mask Variability

1001 Most existing state-of-the-art (SOTA) methods assume that the same mask is used during both the  
 1002 immunization (training) and editing (testing) phases. While this assumption aligns with standardized  
 1003 deepfake pipelines—where masks are often fixed to cover specific regions such as the head or full  
 1004 body—it limits the robustness of these methods to real-world scenarios involving unpredictable or  
 1005 mismatched editing masks.

1006 To evaluate this limitation, we conduct an experiment where the editing mask during test time  
 1007 differs from the mask used during immunization. As shown in Figure 15, when the test-time mask  
 1008 diverges from the training mask, existing methods such as PhotoGuard (PG) and DiffusionGuard  
 1009 fail to maintain their edit-disrupting behavior. In contrast, DiffVax remains effective, successfully  
 1010 disrupting the malicious edits even when significant changes are made to the mask size or region.

1011 This robustness can be attributed to our model’s design, which does not overfit to the spatial shape or  
 1012 scale of the mask used during training. Instead, it learns to encode more generalizable perturbations  
 1013 that degrade editing attempts across a range of editing contexts. These findings suggest that DiffVax  
 1014 offers better real-world applicability where attackers may alter masks to evade immunization.

1015 Moreover, we observe that large discrepancies between the training and test masks can introduce  
 1016 semantic inconsistencies or visual artifacts—particularly in inpainting-based editing models that  
 1017 rely heavily on localized structure. This further motivates the use of instruction-based models such  
 1018 as InstructPix2Pix (IP2P) for evaluating robustness under complex or localized edit conditions.  
 1019 In supplementary Figure 8, we demonstrate that DiffVax remains effective against stylistic and  
 1020 localized edits, which often fall outside the scope of inpainting approaches.

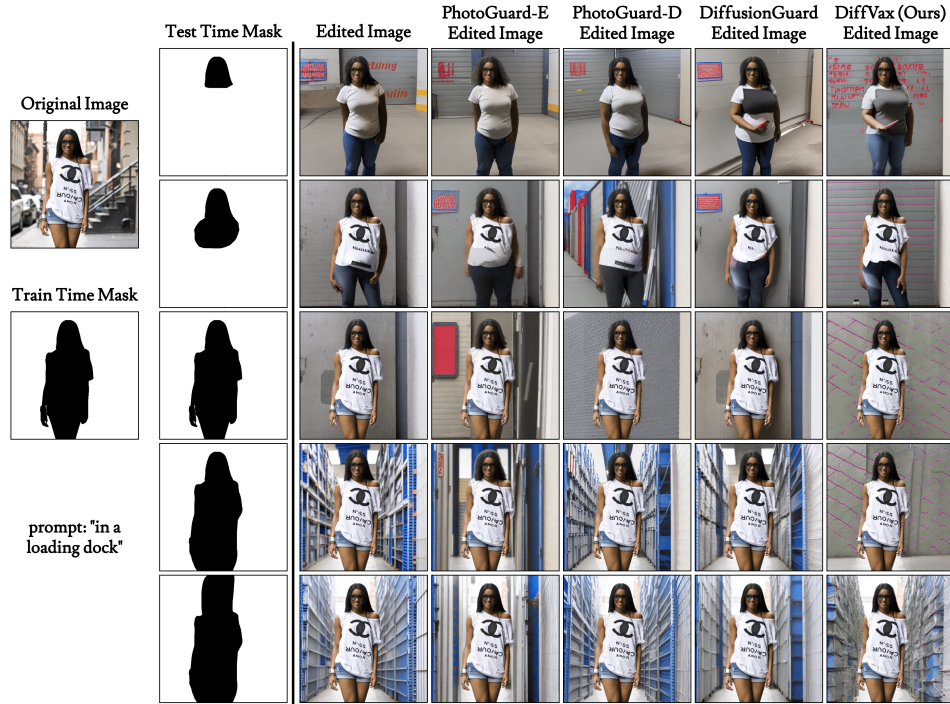


Figure 15: *Comparison of edited immunized images with different immunization and editing masks.* PhotoGuard uses the same mask for both training and testing, making it highly sensitive to changes in the editing mask. DiffVax, by contrast, is trained with a fixed immunization mask but remains robust even when the test-time editing mask significantly deviates. The results show consistent disruption of edits by DiffVax despite large mask variability.

## A.6 Video Evaluation

To our knowledge, this is the first immunization-based video evaluation using a diffusion model for editing. We construct a video benchmark consisting of 4 human activity videos, each containing 64 frames and paired with 4 unique prompts. Since no prior method directly supports training-free video immunization using inpainting-based diffusion models, we adopt a naive per-frame editing pipeline to extend our approach to video. Despite not incorporating any explicit temporal modeling, our method yields strong results.

As reported in Table 5, DiffVax outperforms all baselines across multiple metrics, including PSNR, SSIM (Noise), CLIP-T, and runtime. Notably, it achieves a dramatic reduction in runtime—processing the full dataset in just **0.739 seconds**—compared to PhotoGuard-D’s 64-hour runtime. These results emphasize the efficiency and practicality of our approach in real-time or large-scale settings.

Importantly, we make no architectural or training modifications for video data. The strong results achieved without temporal modeling suggest that our method generalizes well across sequential data, capturing consistent patterns in human identity, pose, and structure across frames. This robustness is further demonstrated in Fig. 1 and Fig. 4 (c), where the model effectively adapts to changes in body motion and facial expressions.

Our work targets general-purpose editing protection and is evaluated on diverse, open-domain video data. The effectiveness of our approach under such settings demonstrates its promise as a scalable and general immunization strategy for future video editing systems.

Table 5: **Results on video editing.** We report the average PSNR, SSIM, FSIM, SSIM (Noise), CLIP-T, and total runtime for Random Noise, PhotoGuard-D, DiffusionGuard, and DiffVax on a video dataset consisting of 4 videos, each with 4 prompts and 64 frames. Best results per column are **bolded**.

Method	SSIM ↓	PSNR ↓	FSIM ↓	SSIM (Noise) ↑	CLIP-T ↓	Runtime ↓
Random Noise	0.774	21.09	0.547	0.786	29.62	N/A
PhotoGuard-D	0.738	17.31	0.448	0.965	26.52	64 hours
DiffusionGuard	0.750	17.43	0.478	0.922	25.41	10 hours
DiffVax	<b>0.681</b>	<b>16.78</b>	<b>0.374</b>	<b>0.974</b>	<b>22.51</b>	<b>0.739 seconds</b>

## 1040 **A.7 Dataset Setup**

1041 Our dataset consists of 1,000 images, each associated with two prompts, resulting in a total of 2,000  
1042 prompts. We split the dataset into 80% for the training set (seen) and 20% for the validation set  
1043 (unseen). The prompt set was constructed using ChatGPT (OpenAI, 2024), specifically by generating  
1044 prompts designed for background editing. A total of 1,000 prompts were collected and subsequently  
1045 split into 80% for the training set (seen) and 20% for the validation set (unseen). Finally, we sampled  
1046 two random prompts for each image in the dataset, ensuring the prompts corresponded to whether the  
1047 image was categorized as seen or unseen.

## 1048 A.8 Prompt-Agnostic Immunization Experiment

1049 We conduct additional experiments to demonstrate that the noise produced by our DiffVax (and  
 1050 consequently the immunized images) is prompt-agnostic. To achieve this, we train DiffVax three  
 1051 times, using a different image for each training setup. In each experiment, we use a single image  
 1052 with 100 seen prompts for training and evaluate it on 75 seen prompts and 75 unseen prompts (not  
 1053 included in the training set). The results are then averaged across all images for each prompt. As  
 1054 shown in Fig. 16, the quantitative results for seen and unseen metrics are highly similar, and the low  
 1055 variances further confirm that the noise generalizes effectively across diverse prompt conditions.

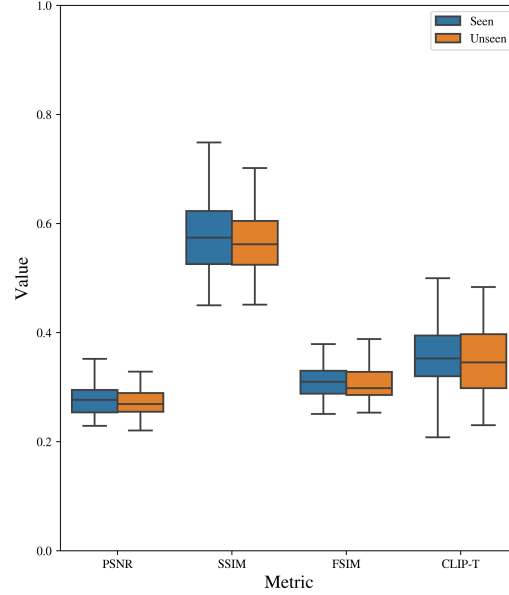


Figure 16: **Experiment results for prompt-agnostic noise.** We present our performance metrics between prompts for 75 prompts seen in training (blue color) and 75 prompts unseen in training (orange color). PSNR and CLIP-T values are divided by 50 for visualization purposes. We can see that the two distributions are almost identical, suggesting that our method performs similarly across all prompts, suggesting the prompt-agnostic nature of our DiffVax.

## 1056 A.9 Loss Weight Selection

1057 The hyperparameter  $\alpha$  in DiffVax’s loss function controls the balance between imperceptibility  
 1058 and edit disruption. It is defined in the overall loss as  $\mathcal{L} = \alpha \cdot \mathcal{L}_{\text{noise}} + \mathcal{L}_{\text{edit}}$ , where a larger  $\alpha$   
 1059 emphasizes minimizing visible noise, potentially at the cost of reduced editing resistance, and a  
 1060 smaller  $\alpha$  enhances robustness to edits but may introduce more perceptible perturbations.

1061 To determine an optimal value for  $\alpha$ , we conduct an ablation study on a subset of 100 images,  
 1062 evaluating three values:  $\alpha = 2, 4$ , and  $6$ . The results are summarized in Table 6. We observe that  
 1063 while increasing  $\alpha$  improves imperceptibility—as indicated by slightly higher SSIM (Noise) and  
 1064 PSNR scores—the edit disruption becomes weaker, reflected in a deterioration of the SSIM and  
 1065 PSNR metrics.

1066 We select  $\alpha = 4$  as the optimal configuration. It provides a strong balance between imperceptibility  
 1067 and disruption: the gain in SSIM (Noise) from  $\alpha = 4$  to  $\alpha = 6$  is marginal, while the drop in editing  
 1068 robustness is more pronounced. Furthermore, qualitative inspection confirms that the perturbations at  
 1069  $\alpha = 4$  are already imperceptible, making further increase in  $\alpha$  unnecessary.

Table 6: *Ablation study on the loss weight  $\alpha$  in  $\mathcal{L} = \alpha \cdot \mathcal{L}_{\text{noise}} + \mathcal{L}_{\text{edit}}$ .* Metrics demonstrate the trade-off between imperceptibility and edit disruption. Best values for SSIM (Noise) are bolded, while lower SSIM and PSNR indicate stronger editing disruption.

Configuration	SSIM ↓	PSNR ↓	SSIM (Noise) ↑
DiffVax w/ $\alpha = 2$	<b>0.536</b>	<b>14.47</b>	0.987
DiffVax w/ $\alpha = 4$	0.588	15.38	<b>0.993</b>
DiffVax w/ $\alpha = 6$	0.625	16.23	<b>0.996</b>

## 1070 A.10 User Study

1071 To assess the human-perceived quality and effectiveness of each immunization method, we conducted  
 1072 a user study with 67 participants recruited via Prolific. Participants were asked to rank edited images  
 1073 based on how unrealistic or misaligned they appeared.

1074 Each participant was shown a set of five edited images derived from the same input image and  
 1075 text prompt (see Figure 17). These five outputs corresponded to different immunization strategies:  
 1076 Random Noise, PhotoGuard-E, PhotoGuard-D, DiffVax, and an unprotected baseline. For each  
 1077 prompt-image pair, participants were instructed to rank the edits from **least aligned** to **most aligned**  
 1078 with the editing prompt. A lower ranking indicates better disruption of the intended edit (i.e., more  
 1079 effective immunization), as participants found the result less realistic or aligned with the prompt.

1080 We randomly shuffled the order of methods in each trial to avoid position bias. In total, the study  
 1081 included 20 image-prompt pairs covering both seen and unseen examples, ensuring a fair and  
 1082 comprehensive evaluation.

Table 7: *User Study Rankings*. Lower values indicate better perceived editing failure prevention, imperceptibility, and alignment with the original content.

Immunization Method	Average Ranking ↓
Random Noise	3.74
PhotoGuard-E	3.33
PhotoGuard-D	2.63
DiffVax (Ours)	<b>1.64</b>

1083 As shown in Table 7, DiffVax significantly outperforms prior methods, receiving the best average  
 1084 ranking of **1.64**. This demonstrates the effectiveness of our method in fooling editing models in a  
 1085 way that is perceptually convincing to human observers. The next-best method, PhotoGuard-D, trails  
 1086 behind with a score of 2.63, while other methods rank even lower.

Within this User-Study Experiment form, you will encounter **one input image, one text prompt and multiple edited images for each section**. For each question, please **rank** the edits from **least aligned to most aligned** with the **intended background change described in the prompt**.


**Note:** We randomly shuffle the image editing approaches' order in every section for fairness.

Example: ★

Each **edited image** below shows a **change** made to the **background** of the original image based on the **text prompt** written below. Please **rank** these edits from **least aligned to most aligned** with the intended **background change** described in the **prompt**.

**Text Prompt:** "A person in a horse racing track"

Input Image



Edited Images

Image 1




Image 2




Image 3




Image 4




Image 5




	Image 1	Image 2	Image 3	Image 4	Image 5
Least aligned	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Second least aligned	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Middle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Second most aligned	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Most aligned	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 17: *Instructions provided to user study participants.* Users were asked to rank edited images from least to most aligned with the text prompt. Lower alignment suggests more successful immunization.