

---

# Additional Appendix for StyleGuard: Preventing Text-to-Image-Model-based Style Mimicry Attacks by Style Perturbations

---

Anonymous Author(s)

Affiliation

Address

email

## 1 B. Additional Demonstration of Visual Examples

### 2 B.1 Compare the Clean, Protected and Upscaled Images

3 To demonstrate the effectiveness of our method, Figure 1 presents a comparison of original clean  
4 images without any processing, protected versions after applying our protection methods. and the  
5 noise up-scaled images that the adversary applies Noise Upscale with a different version of the  
6 upscale model (SD-x2-latent-upscaler) from the training stage to the protected images. We have  
7 made some findings as follows. First, it is shown that the noise introduced by our method is very  
8 small and does not affect the image quality. Second, Noise Upscale can better restore the details of  
9 the image, such as the face in the sixth row. We think this may be because Van Gogh’s image appears  
10 in the Upscaler training dataset. However, for some parts related to the image style, Noise Upscale  
11 cannot be restored well, such as the sky in the first row and the grass in the fifth row, which become  
12 more blurred after Upscale. We think this is because these images may not in the training images of  
13 the Upscale model.

### 14 B.1 Compare the Images Trained on Clean Images and Protected Images

15 Figure 2 and Figure 3 compare the results of style mimicry on clean and protected images with  
16 the StyleGuard protection. For the StyleGuard, we generate perturbations using SD1.4 and SD x4  
17 upscaler. During the test, we first apply the Noise Upscale using the SD x2 upscaler and then train the  
18 SD1.5 model on the protected paintings. With protection, the quality of the protected image decreases  
19 significantly and the style also changes considerably from the original images.

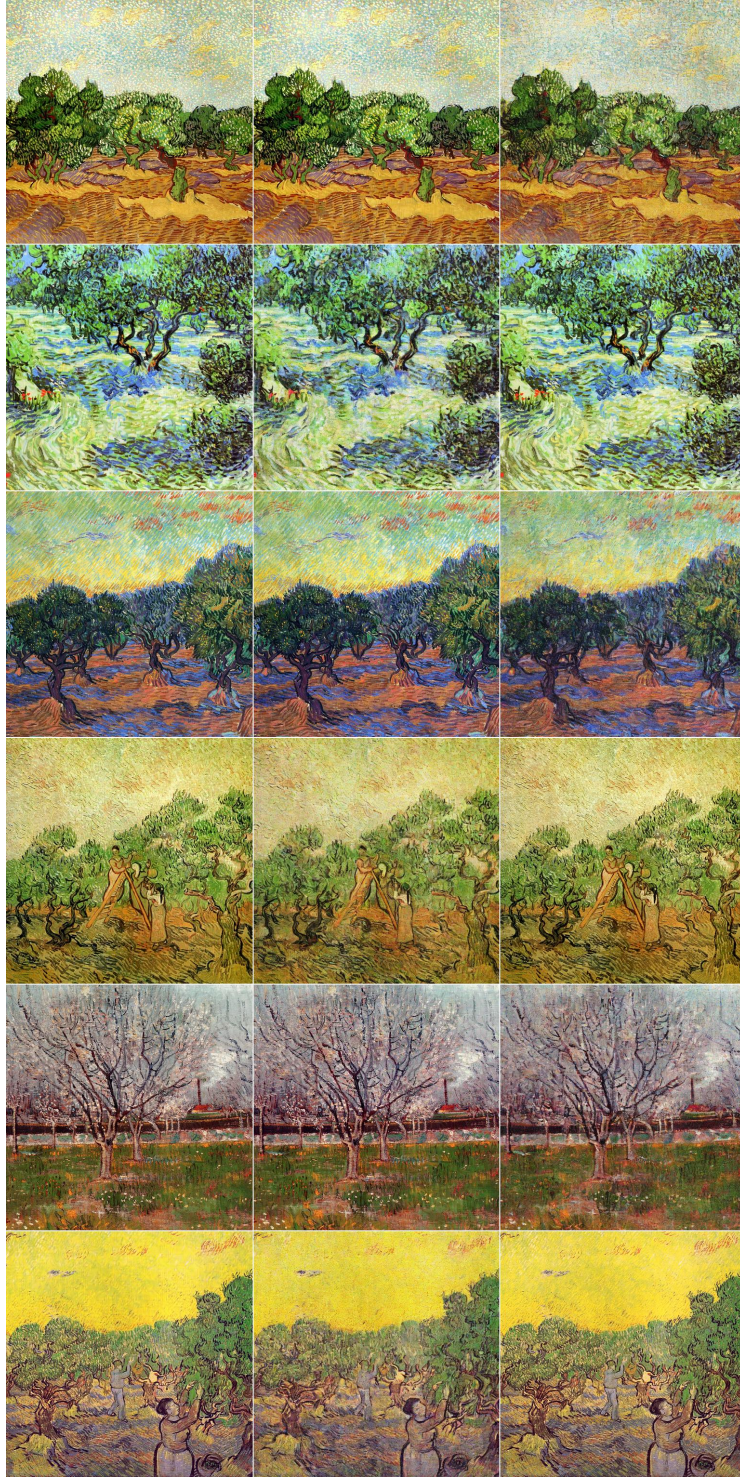


Figure 1: Visual comparison between (a) clean/original images (left column), (b) protected images (middle column), and (c) Noise Upscaled results (right column). Each row shows the same image processed through different pipeline stages.





Figure 2: The results of style mimicry on clean images that are without any protection. We train the SD v1.5 model on Van Gogh's paintings.





Figure 3: The results of style mimicry on protected images that with the StyleGuard protection. For the StyleGuard, we generate perturbations using SD1.4 and SD x4 upscaler. During the test, we first apply the Noise Upscale using the SD x2 upscaler and then train the SD1.5 model on the protected paintings.