

# Comparison of CoModGans, LaMa and GLIDE for Art Inpainting

## Supplementary Material

### A. Further Qualitative Results

The panel in Figs. 1 to 3 presents additional comparison of the 'failure' cases across different models. It is clear that LaMa and CoModGans are not suited for outpainting tasks. GLIDE on the other hand, performs well across inpainting and outpainting demands but suffers from having the worst resolution output at only 256x256.

### B. Additional Studies on GLIDE's Parameters

The panel in Figs. 4 to 6 shows the summary results of ablation studies performed on GLIDE. We tested the effect of changing the Guidance Scale (which controls the relationship between the prompt and the generated image) and the Upsampling Temperature (which controls the degree of upsampling) for the same image and same mask. The test was performed over 50 samples for each value <sup>1</sup>.

The results show a significant sensitivity of the model outputs to the parameters. As expected, a higher degree of upsampling improves results across all metrics. In fact, the recommended Upsampling Temperature is 0.997. The Guidance Scale controls the content, and thus, does not affect the sharpness or blurriness. In the case of BRISQUE, the sensitivity to the Guidance Scale might be explained by the degree of black color on the image, since this metric is sensitive to large black regions.

Next we analyse the effect of the prompt on the inpainted content. As mentioned, GLIDE is a text-guided model where the incidence of the text is controlled by the parameter "Guidance Scale". Under a low Guidance Scale, GLIDE produces content in consonance with the surrounding objects, which is ideal for art inpainting. On the contrary, a higher Guidance Scale gives more weight to the text prompt on the image generation. The effect of the Guidance Scale over the text guidance is shown on Fig. 7 using a fixed seed and an Upsampling Temperature of 0.997.

<sup>1</sup>prompt: "a gallery with arches wooden windows and arcades floors with tiles"

### C. Detailed Analysis of Inpainted Print Gallery

The panel Figs. 8 and 9 shows a side-by-side comparison of the eight images composing the straight version of Print Gallery with their inpainted counterparts. On each panel, the four images on the right are the inpainted results of the the left images.

On panels Fig. 10 we show a comparative analysis of GLIDE's output under different Guidance Scales, everything else equal. Results indicate that a Guidance Scale value of five results in the best object creation <sup>2</sup>. The prompt used is: "a photograph of a teddy bear using a laptop 1080p 4k."

### D. Analysis on Additional Paintings

This section provides further comparison of the model performance under different scenarios. First in Fig. 11 we provide a further analysis of the Ecce Homo by Elias Garcia Martinez from the main text. This fresco is a challenge since white patches from the degradation are visible all over the surface. In the case of CoModGANS, the model considers as a valid pattern the white areas appearing on the image's surface and tries to replicate them, resulting in a poor inpainting performance. LaMa repeats this behaviour but in a lesser degree, as noted, this model provides the best output resolution among all. Additionally, it performs comparatively well on the face since it has been trained on the faces database Celeb-H. GLIDE, as expected, it is not able to recognize a human face by design. However, the upsampling module produces a nitid result compared to others.

#### D.1. Automatic Generation of Prompts

The Fig. 12 is a work by Torres-Garcia "Composicion constructiva" (1932) [1]. The piece was burnt at the Brazilian Museum of Modern art in 1981 and presents the traces of fire on the wood. This coloration creates a challenge for inpainting models as they deem the burnt area as a valid pattern to reproduce. We see that GLIDE is able to move away from the burnt

<sup>2</sup>Images by artist @litevex reproduced with permission

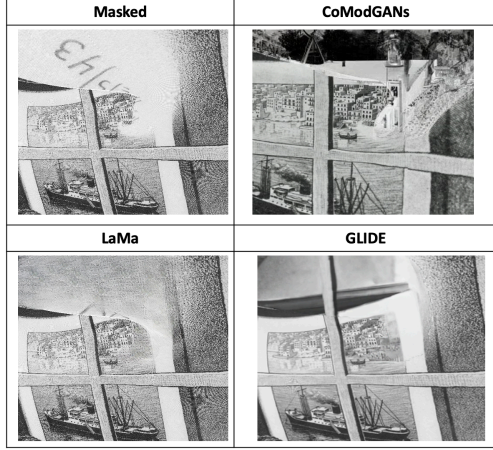


Figure 1. Comparison of outpainting image across different models. GLIDE caption is "window" and Guidance Scale of five.

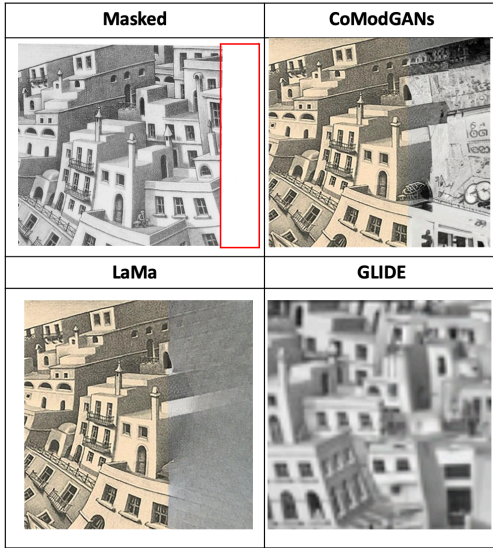


Figure 2. Comparison of outpainting image across different models. GLIDE caption is "buildings" and Guidance Scale of five.

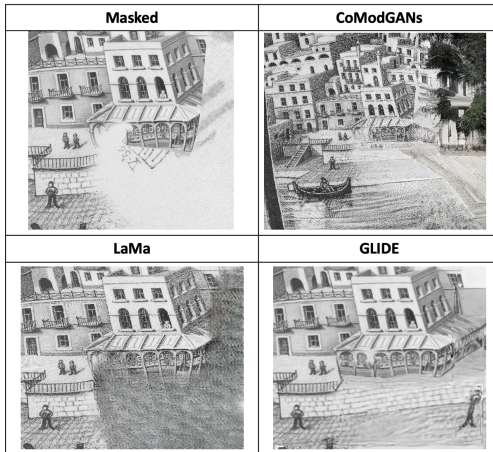


Figure 3. Comparison of outpainting/inpainting image across different models. GLIDE caption is "buildings" and Guidance Scale of five.

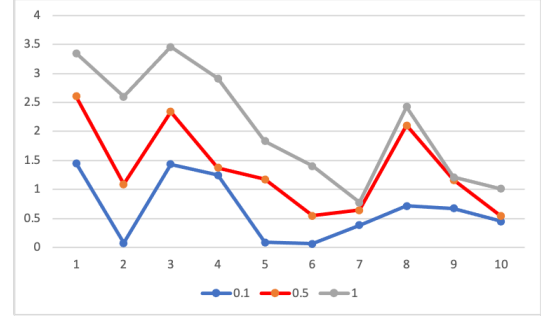


Figure 4. Plot of BRISQUE values. The three categories represent the Upsampling Temperature and the horizontal axis represent the Guidance Scale

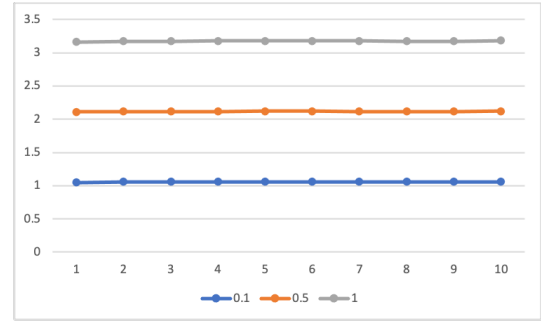


Figure 5. Plot of DOM values. The three categories represent the Upsampling Temperature and the horizontal axis represent the Guidance Scale

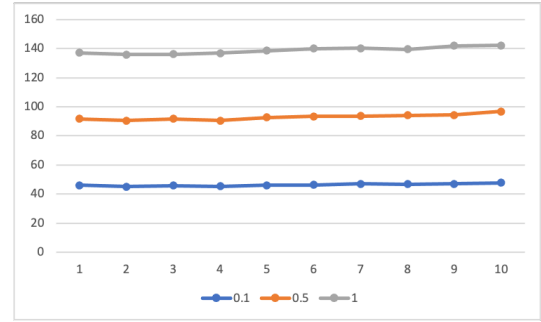


Figure 6. Plot of KONIQ values. The three categories represent the Upsampling Temperature and the horizontal axis represent the Guidance Scale

colorization, either by ramping the Guidance Scale or by trimming the area with a clever prompt. For the left-bottom image, the prompt was generated using a image-to-text bot by the developers EleutherAI<sup>3</sup>, the generated prompt is "Paul Klee's rectangular piece of wood is a rare example of early Christian art.". The images with Guidance Scales two and 20 have used the

<sup>3</sup>[www.eleuther.ai](http://www.eleuther.ai)

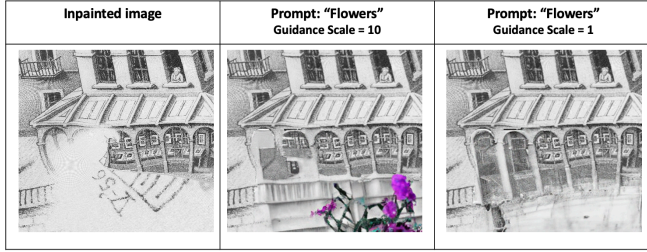


Figure 7. GLIDE’s outputs for different Guidance Scales and same prompt.

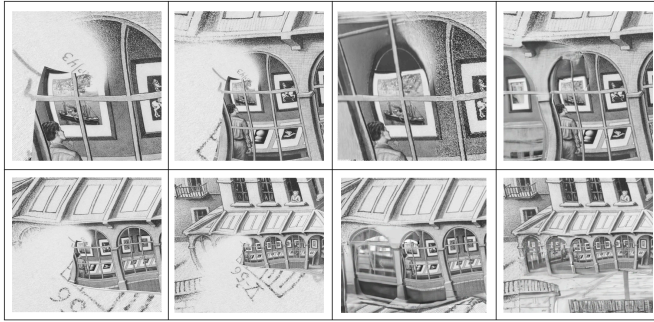


Figure 8. Inpainting of the first four images composing the blank center. GLIDE with Guidance Scale of five and prompt "Escher".

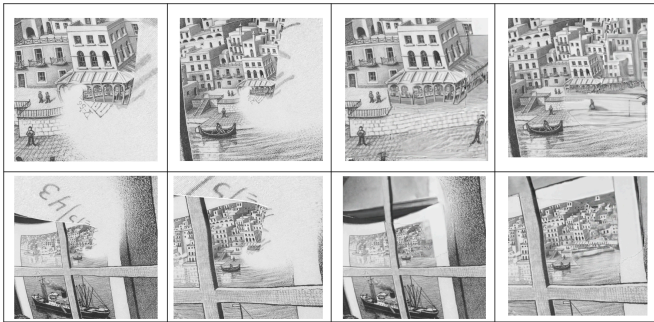


Figure 9. Inpainting of the second four images composing the blank center. GLIDE with Guidance Scale of five and prompt "Escher".

simple prompt "patterns".

## D.2. Crowd-sourcing of Prompts By Experts

The panels on Fig. 13 show Cezanne’s unfinished "Turning Road" (1905), which has whole sections of the canvas bare. The inpainting of this work is more open-ended given the style of the painter. For this reason, all models present equivalent performance, to the casual eye. To generate GLIDE’s prompt we did a crowd-sourcing experiment and relied on the expertise of ten visual artist from EleutherAI who suggested



Figure 10. Different guidance scale settings over the same prompt. "a photograph of a teddy bear using a laptop 1080p 4k".



Figure 11. Different outputs for the Ecce Homo by Elias Garcia Martinez. Image from Wiki Commons. Glide was used with a Guidance Scale of three and six, where indicated.

by consensus "the fog of the valley painting from last century". The two GLIDE images on the bottom were generated by using different seeds on the same prompt.

## References

- [1] Joaquin Torres García. Torres García. Obras Destruídas En El Incendio Del Museo De Arte Moderno De Río De Janeiro. Fundación Torres García, 1981. 1



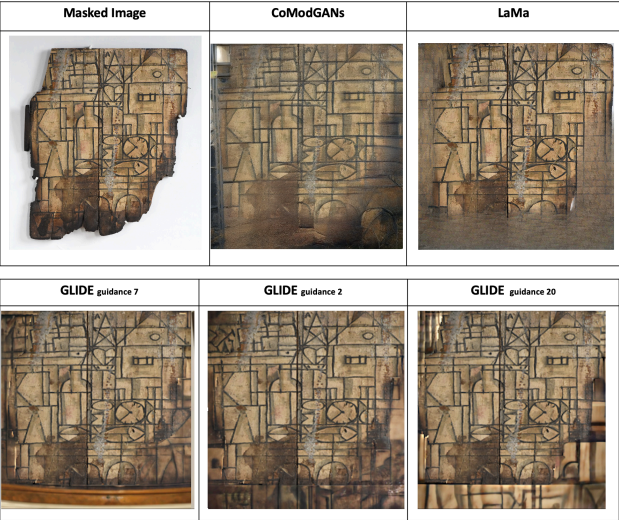


Figure 12. Inpainting outputs for the work of Torres-Garcia. Composicion constructiva (1932). GLIDE with different Guidance Scales and prompts generated by image-to-text bots. Image from Wiki Commons.

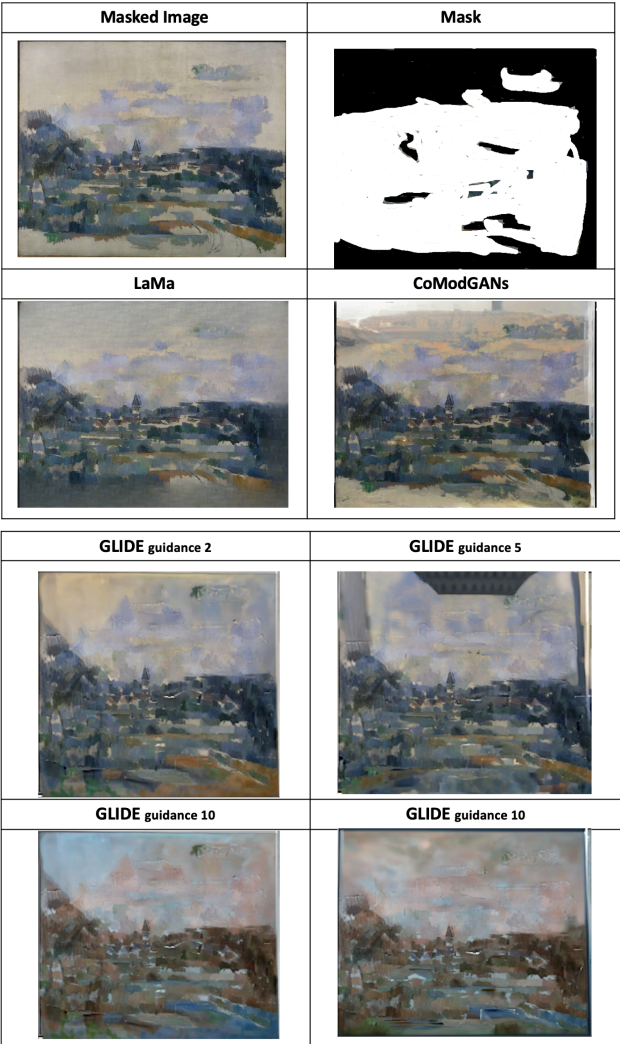


Figure 13. Comparison of inpainting models for Cezanne's unfinished "Turning Road" (1905). GLIDE's output is presented with several values for the Guidance Scale. Image from Wiki Commons.