

# Project AA4 - Text-Guided Latent Diffusion Models - TReg

Imane SI SALAH

March 26, 2024

## Abstract

Recent advances in diffusion models have led to significant progress in image restoration, especially in addressing inverse problems by employing these models as generative priors. However, the subsisting challenge of these problems lies in the ill-posed nature of inverse problems in general. The authors of [2] were inspired by the human ability to interpret visual ambiguities through perceptual biases and introduced an approach that combines latent Diffusion Posterior Sampling (DPS) with text regularization (TReg). This method incorporates textual descriptions reflecting a preconceived notion of desired outcomes during the reverse sampling process. Their approach reinforces these descriptions through null-text optimization dynamically. In this work, an attempt is made to recreate this method and evaluate its effectiveness.

## 1 Introduction

In image processing, solving inverse problems, encompassing deblurring, super resolution, and inpainting amounts to reconstructing the unknown image  $\mathbf{x}_* \in \mathbb{R}^n$  given the observed measurement  $\mathbf{y} \in \mathbb{R}^m$  defined as:

$$\mathbf{y} = A(\mathbf{x}_*) + \eta$$

where the operator  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  represents the forward model and  $\eta \in \mathbb{R}^m$  is the measurement noise. In real world we often fall in the case of ill posed problems where  $m < n$ . They are characterized by an ambiguous relationship between the observed measurement and the desired high-fidelity image, making their solution challenging. Classical approaches to tackle this problem involve adding a regularization constraint to guide the solution toward plausible outcomes based on certain prior knowledge.

Diffusion models are applied to these problems by learning to approximate the gradient of the log prior density of the data  $\nabla_x \log p(x)$ . Combined with the assumption that the forward degradation model is known, this can be used to sample from the posterior distribution  $p(x|y)$ .

However, despite the quality of these results, there still remains a gap between the solution of these models and human perception. To bridge this gap, an approach was developed combining latent Diffusion Posterior Sampling (DPS) and text-driven regularization. This method sets the basis for creating images that are not only visually compelling but also semantically meaningful.

## 2 Background and Related Work

This section will cover key concepts of diffusion posterior sampling, latent diffusion models, and classifier-free guidance, which are pivotal to understanding Regularization by Texts for Latent Diffusion Inverse Solvers.

### 2.1 Diffusion Posterior Sampling (DPS):

Diffusion Posterior Sampling (DPS) extends the concept of Denoising Diffusion Probabilistic Models (DDPMs). DDPM offers a framework to gradually transition from a noise distribution to a data distribution over a sequence of steps. DPS builds upon this framework and seeks to recover an original image from the corrupted observation by sampling from the posterior distribution  $p(x|y)$  [1].

For Gaussian likelihood inverse problems, the relationship between  $y$  and  $x$  is modeled as  $p(y|x) = \mathcal{N}(Ax, \sigma^2 I)$ , where  $A$  is the forward operator and  $\sigma^2 I$  represents the variance of the Gaussian noise.

In standard diffusion models, particularly Denoising Diffusion Probabilistic Models (DDPM), the reverse process is driven by the score function associated with the prior, which iteratively denoises the data by adjusting it in the direction of the gradient of its log probability:

$$x'_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t + \beta_t \nabla_{x_t} \log p(x_t)) + \sigma_t z_t \quad (1)$$

where  $\nabla_{x_t} \log p(x_t)$  is typically approximated by denoising score matching (DSM)

$$\theta^* = \operatorname{argmin}_{\theta} \mathbb{E}_{t,x_t,x_0} [\|s_{\theta}(x_t, t) - \nabla_{x_t} \log p(x_t|x_0)\|_2^2]. \quad (2)$$

Once trained, the plug-in estimate  $\nabla_{x_t} \log p(x_t) \simeq s_{\theta}(x_t, t)$  is used for the reverse diffusion in (1).

Now to adapt this to solving inverse problems, we need to leverage the posterior distribution  $p(x|y)$  instead of the  $p(x)$ . Upon this shift, we modify the score functions using Bayes' Rule:

$$\nabla_{x_t} \log p_t(x_t|y) = \nabla_{x_t} \log p_t(y|x_t) + \nabla_{x_t} \log p_t(x_t) \quad (3)$$

$$\simeq \nabla_{x_t} \log p_t(y|x_t) + s_{\theta^*}(x_t, t) \quad (4)$$

so that this adjustment accounts for knowledge embedded in  $p(x)$  and the new information from the observation  $y$  as encapsulated in  $p(y|x)$ .

Incorporating the modified score function gives an updated reverse process rule for DPS:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t + \beta_t \nabla_{x_t} \log p(x_t)) + \sigma_t z_t + \frac{\beta_t}{\sqrt{\alpha_t}} \nabla_{x_t} \log p(y|x_t) \quad (5)$$

$$= x'_{t-1} + \frac{\beta_t}{\sqrt{\alpha_t}} \nabla_{x_t} \log p(y|x_t) \quad (6)$$

where the first term represents the original DDPM reverse process (prior-driven) and the second term introduces a correction based on the observed data  $y$  (data-driven).

The challenge lies in the term  $\nabla \log p(y|x_t)$ , which directly depends on  $p(y|x_t)$ —a term that cannot be directly accessed. To address this, the relationship between  $x_0$  (the original data) and  $x_t$  (the data at timestep  $t$ ) known from the diffusion process, along with the knowledge of the forward model, is considered. This allows the expression of  $p(y|x_t)$  in terms of  $p(x_0|x_t)$  and  $p(y|x_0)$ , using the fact that the relationship between  $y$  and  $x_0$  is known.

Now, the measurement model is factorized as:

$$p(y|x_t) = \int p(y|x_0, x_t)p(x_0|x_t)dx_0 = \int p(y|x_0)p(x_0|x_t), \quad (7)$$

where the fact that  $y$  and  $x_t$  are conditionally independent given  $x_0$  was utilized. This integral, as convenient as it appears, is intractable, so a closed-form approximation for  $p(x_0|x_t)$  is adopted.

Then, the fact that for DDPM sampling,  $p(x_0|x_t)$  has a unique posterior mean centered at

$$\hat{x}_0 := \mathbb{E}[x_0|x_t] = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t + (1 - \bar{\alpha}_t)s_{\theta^*}(x_t, t)), \quad (8)$$

so the approximation can be used:

$$\nabla_{x_t} \log p(y|x_t) \simeq \nabla_{x_t} \log p(y|\hat{x}_0(x_t)). \quad (9)$$

Since Gaussian noise is involved, the likelihood function can be expressed as:

$$p(y|x_0) = \frac{1}{\sqrt{(2\pi)^n \sigma^{2n}}} \exp \left[ -\frac{\|y - \mathcal{A}(x_0)\|_2^2}{2\sigma^2} \right], \quad (10)$$

where  $n$  denotes the dimension of  $y$ . Differentiating with respect to  $x_t$ , the following is obtained:

$$\nabla_{x_t} \log p(y|x_t) = -\frac{1}{\sigma^2} \nabla_{x_t} \|y - \mathcal{A}(\hat{x}_0(x_t))\|_2^2, \quad (11)$$

which represents the guidance term driving the DPS process. Integrating this into the previously established DDPM framework leads to the DPS update rule:

$$x_{t-1} = x'_{t-1} - \eta_t \nabla_{x_t} \|y - \mathcal{A}(\hat{x}_0(x_t))\|_2^2, \quad (12)$$

where  $\eta_t = \frac{\beta_t}{\sqrt{\bar{\alpha}_t}}$ .

## 2.2 Latent Diffusion Models

Instead of performing the diffusion process in pixel space, Latent Diffusion Models (LDMs) operate in latent space, which is a lower-dimensional and more compact representation of the data. This is achieved by involving a Variational Autoencoder (VAE), enhancing the computational efficiency of the model.

Latent diffusion begins with the encoding process, represented by

$$z = E_\phi(x) = E_{\phi_\mu}(x) + E_{\phi_\sigma}(x) \odot \epsilon, \epsilon \sim \mathcal{N}(0, I)$$

where  $E_{\phi_\mu}$  and  $E_{\phi_\sigma}$  denote the encoder's output for the mean and standard deviation, respectively. Then, the latent representation  $z$  is subject to a series of diffusion steps formulated under the VP-SDE framework as:

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon$$

where  $z_0$  is the latent representation of the clean image, and  $\bar{\alpha}_t$  is the cumulative product of  $\alpha_t$ , accounting for the variance schedule of the diffusion process.

The reverse process, that aims at reconstructing  $z$  from  $z_T$  uses a neural network to predict the noise  $\epsilon$  added at each step and is given by:

$$\hat{z}_0 = \frac{z_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(z_t, t)}{\sqrt{\bar{\alpha}_t}}$$

where  $\epsilon_\theta(z_t, t)$  is the noise predicted by the model – typically a time-conditioned U-Net, for  $z_t$  at time step  $t$ . Then,

$$z_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \hat{z}_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon_\theta(z_t, t)$$

## 2.3 classifier free guidance:

Exploiting a classifier  $p(c|z)$  has proved to improve the sampling quality of diffusion models, where a classifier  $p_\theta(c|z_t, t)$  is trained on  $z_t$ , and then the gradients  $\nabla_{z_t} \log p_\theta(c|z_t, t)$  are used to guide the diffusion sampling process toward a given label  $c$ . However, this required training an additional classifier on noisy images, which can be computationally expensive. Therefore, the authors introduced a way to achieve this classifier guidance without an explicit classifier by setting  $\nabla_{z_t} \log p(c|z_t) = \nabla_{z_t} \log p(z_t|c) - \nabla_{z_t} \log p(z_t)$ , by jointly training an unconditional denoising model  $p_\theta(z)$  parameterized by a score estimator  $\epsilon_\theta(z) = \epsilon_\theta(z, c = \emptyset)$  along with a conditioned model  $p_\theta(z|c)$  parameterized by  $\epsilon_\theta(z, c)$  where  $\emptyset$  represents the null token. Then, sampling is performed using the linear combination of both models' score estimates as:

$$\epsilon_{\text{guided}}(z, c) = \epsilon_\theta(z, c) + w(\epsilon_\theta(z, c) - \epsilon_\theta(z)).$$

## 3 Methodology

In this section, the methodology adopted in this work is outlined, focusing on the implementation and evaluation of Regularization by Text for Latent Diffusion Inverse Solvers as explained in [2]. The focus is on Latent DPS implementation, then exploring Posterior Sampling with Latent Diffusion Models (PSLD) to finally determine how to incorporate Text Regularization into these models.

### 3.1 Latent DPS

The transition from traditional DPS to its latent counterpart is made because it offers significant computational efficiency and adaptability advantages. This involves applying the approximation:  $p(y|z_t) \simeq p(y|\hat{x}_0 = \mathcal{D}(\mathbb{E}[z_0|z_t]))$  such that

$$\nabla_{z_t} \log p(y|z_t) \simeq \nabla_{z_t} \log p(y|\hat{x}_0 = \mathcal{D}(\mathbb{E}[z_0|z_t])). \quad (13)$$

Here,  $x_0$  is approximated with the decoded version of the conditional expectation of the clean latent  $z_0$  given the noisy  $z_t$ . Consequently, the latent DPS update rule becomes:

$$z_{t-1} = z'_{t-1} - \eta_t \nabla_{z_t} \|y - \mathcal{A}(D(\hat{z}_0(z_t)))\| \quad (14)$$

where  $z'_{t-1}$  comes from the general update rule of DDIM Sampling.

However, as pointed out in [3], the issue with the "vanilla" Latent DPS lies in the fact that the encoding process is a many-to-one mapping, i.e., a single encoded outcome could arise from multiple latent variables  $z_0$ . This introduces ambiguity during the restoration process where the gradient in Eq. (13) may steer the noisy  $z_t$  to various clean latents  $z_0$ , each pointing in different gradient directions.

### 3.2 PSLD

To address this drawback, the authors of [3] introduced Posterior Score Latent Sampling, which incorporates an additional term to constrain the latents to remain as fixed points when subjected to the successive applications of the decoder and encoder. Thus, the approximation of the intractable  $\nabla_{z_t} \log p(y|z_t)$ :

$$\nabla_{z_t} \log p(y|z_t) \approx \nabla_{z_t} \log p(y|\hat{x}_0 = D(E[z_0|z_t])) + \gamma_t \nabla_{z_t} \|E[z_0|z_t] - E(D(E[z_0|z_t]))\|^2, \quad (15)$$

where the first term is the unattainable gradient from the vanilla Latent DPS extension, and the second term, scaled by  $\gamma_t$ , assesses the 'goodness' of  $z_0$  by imposing penalties on deviations from the fixed-point condition.

### 3.3 Reverse Sampling with Text Regularization

After exploring Latent-DPS, text guidance is added as explained in [2], primarily consisting of three steps: adaptive negation, latent optimization, and finally DDPM type forward sampling to return to the intermediate reconstruction of the correct noisy manifold.

For adaptive negation, the approach relies on the concept of negations, which involves using the opposite concept of a given prompt. This is achieved by optimizing the null-text token of Classifier-Free Guidance (CFG), initialized as an empty placeholder devoid of any semantic information. This optimization minimizes a cost function measuring the similarity between the intermediate image outputs and these null-text tokens in the CLIP embedding space. The objective function is given by:

$$\mathcal{L}_\phi = \text{sim}(\mathcal{T}_{img}(\hat{x}_0(y)), c_\emptyset), \quad (16)$$

where  $\mathcal{T}_{img}$  is the CLIP image encoder,  $\hat{x}_0(y)$  is the decoded denoised estimate,  $c_\emptyset$  are null-text embeddings, finally  $\text{sim}()$  is the inner product and gives the similarity measure.

Then there is latent optimization is used to integrate the text driven regularization along with the process of solving the inverse problem, this is done by solving the following minimization problem:

$$\hat{x}_0(y) = \min_x \frac{\|y - A(x)\|_2^2}{2\sigma^2} + \lambda \|x - \mathcal{D}_\phi(\hat{z}_0|t)\|_2^2, \quad (17)$$

which combines the measurement fidelity of the reconstructed image with the encoded latent expectations and is solved using conjugate gradient. Finally, DDIM sampling is performed in the latent space as explained in section 2.2.

Finally, the algorithm as presented by the authors is:

## 4 Experiments and Results

### 4.1 Data and Blurring process

A subset of the Flickr-Faces-HQ (FFHQ) dataset was utilized, aligning with the authors' validation process in their study. This dataset consists of high-quality portrait images. For Gaussian blurring, a kernel size of 61 and a standard deviation of 5.0 were employed, followed by the addition of measurement noise characterized by a variance of  $\sigma^2 = 0.01$ . These forward operations are notably more pronounced than those typically found in related literature. The rationale provided by the authors for these rigorous parameters is to ensure there is sufficient null space.

---

**Algorithm 1** Inverse problem solving with TReg

---

**Require:**  $\epsilon_\theta, \phi, c, \mathcal{E}, \mathcal{D}, \mathcal{A}, y, \mathcal{T}_{img}$

$$z_T \sim \mathcal{N}(0, I)$$

**for**  $t \in [T, 1]$  **do**

$$\hat{\epsilon}_\theta = \epsilon_\theta(z_t, \phi) + \omega(\epsilon_\theta(z_t, c) - \epsilon_\theta(z_t, \phi))$$

$$\hat{z}_{0|t} = (z_t - \sqrt{1 - \bar{\alpha}_t} \hat{\epsilon}_\theta) / \sqrt{\bar{\alpha}_t}$$

**if**  $t \in \Gamma$  **then**

$$\hat{x}_0 \leftarrow \mathcal{D}(\hat{z}_{0|t})$$

$$\hat{x}_0(y) \leftarrow \arg \min_x \frac{\|y - \mathcal{A}(x)\|_2^2}{2\sigma^2} + \lambda \|x - \hat{x}_0\|_2^2$$

$$\hat{z}_0(y) \leftarrow \mathcal{E}(\hat{x}_0(y))$$

$$\hat{c}_\emptyset \leftarrow c_\emptyset - \eta \nabla_\phi (\text{sim}(\mathcal{T}_{img}(\hat{x}_0(y)), c_\emptyset))$$

$$\hat{z}_0^{ema} \leftarrow \bar{\alpha}_{t-1} \hat{z}_0(y) + (1 - \bar{\alpha}_{t-1}) \hat{z}_0(z_t)$$

$$\tilde{\epsilon}_t \leftarrow \frac{\sqrt{1 - \bar{\alpha}_{t-1}} \eta^2 \beta_t^2}{\sqrt{1 - \bar{\alpha}_{t-1}}} \hat{\epsilon}_\theta + \frac{\eta \beta_t}{\sqrt{1 - \bar{\alpha}_{t-1}}} \epsilon \text{ State}$$

$$z'_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}} \hat{z}_0^{ema} + \sqrt{1 - \bar{\alpha}_{t-1}} \tilde{\epsilon}_t$$

**else**

$$z'_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}} \hat{z}_0(z_t) + \sqrt{1 - \bar{\alpha}_{t-1}} \tilde{\epsilon}_t$$

**end if**

$$z_{t-1} \leftarrow z'_{t-1} - \rho_t \nabla_{z_t} \|\mathcal{A}(\mathcal{D}_\varphi(\hat{z}_{0|t}, \hat{c}_\emptyset)) - y\|$$

**end for**

---

Figure 1: Algorithm

## 4.2 Experimental Setup and Observations

Initially, a stable diffusion pipeline was adapted to facilitate the implementation of DPS in latent space. A pretrained diffusion model "runwayml/stable-diffusion-v1-5" was used. The first challenge encountered at this stage was the memory limitation when attempting deblurring images of 512x512 pixels. In response to this, the images were scaled down to 256x256 pixels, allowing the model to perform deblurring on a smaller scale. The results obtained are shown in Figure 2.



Figure 2: Latent DPS on 256x256 image, blurred, restored, and original images from left to right respectively,  $T=1000$

It is observed that despite the reduction in image size, the results on the portrait image are suboptimal, where some deblurring is evident, but the image is not properly restored. Therefore, instead of samples from the FFHQ dataset, low-resolution images generated by the stable diffusion model were used to ensure compatibility with the VAE model's decoder field. Two images were generated with the prompt "The white house", one with resolution 256 and the other 512, then blurred with the same kernel as previously. The results are displayed in Figure 3.

Once again, it is observed that the images are not completely restored, as in the previous case.

Based on these observations, an adjustment was made to the parameter  $\eta$  in the DPS update rule (equation 14). This parameter controls the extent to which the computed gradient influences the reverse diffusion process. Therefore, variations of  $\eta$  ranging from 0.5 to 2.5 were tested to explore their impact on the deblurred images. The results are shown in Figure 5.



Figure 3: Latent DPS on 256x256 generated image, blurred, restored, and original images from left to right respectively

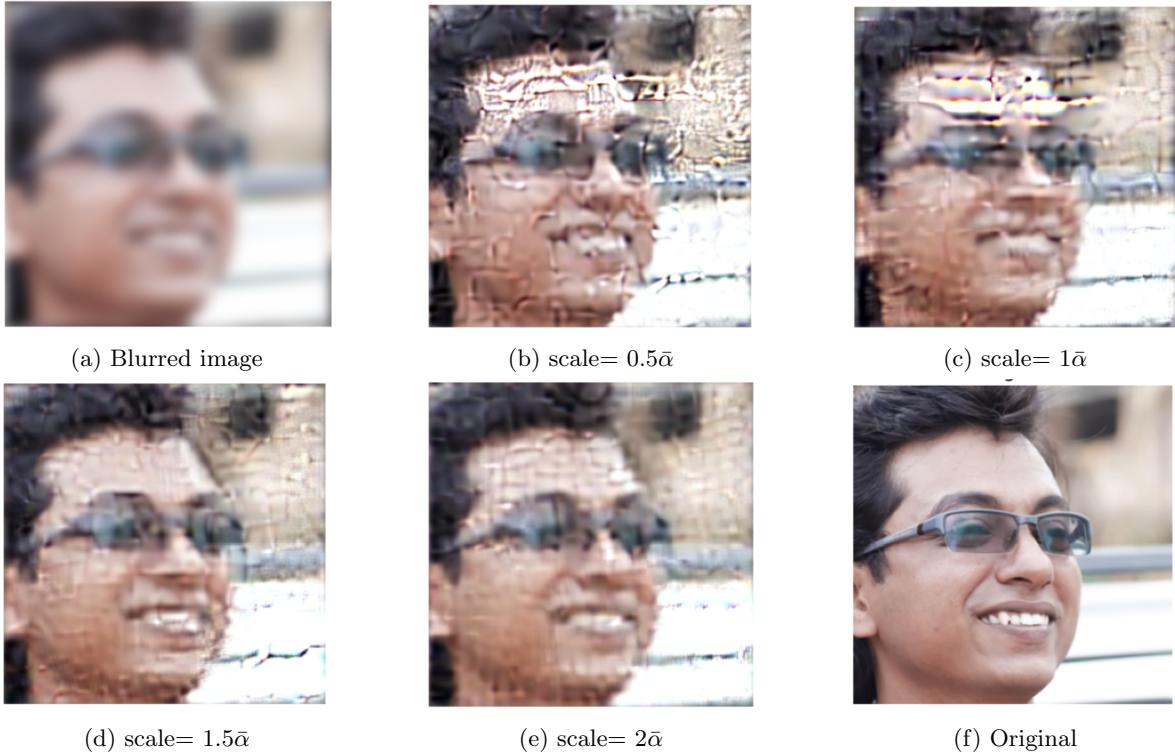


Figure 4: Different reconstructions at different scales of  $\eta$ ,  $T=1000$

It is observed that at smaller scales, the images are noisy with some sharp edges detected, inconsistent with the content of the image. As the scale increases, this noise decreases, although the resulting image becomes more blurry. It is evident that even with modifying the scale in the latent DPS update, the overall resulting images are not quite restored.

Following this, an attempt was made to further improve the quality of the generated images by implementing the other guidance term introduced by PSLD to address the uncertainty in the gradient direction of DPS in latent space. The results obtained from this implementation are displayed in Figure 5.

Here, a slight improvement in the restored images is observed, especially at small values of the scale.

Finally, it was possible to run one instance of DPS on images of 512x512 pixels to determine if the issue actually stems from the low resolution of the images themselves. The result obtained is displayed in Figure 6.

The next phase involved implementing Text Regularization (TReg). The initial attempt to generate images of 256x256 pixels was unsuccessful due to GPU memory limitations and the computational complexity associated with the numerous gradient calculations required for updating  $z_t$ . This complexity further led to errors related to the size of the computational graph. Consequently, the resolution was further decreased to 128x128 pixels, which is considered very mediocre.

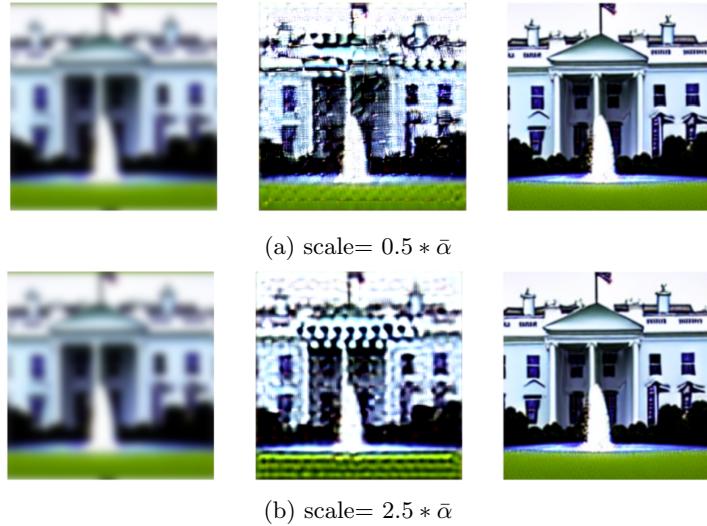


Figure 5: Different reconstructions at different scales of  $\eta_t$  usign PSLD, T=1000



Figure 6: Latent DPS on 512x512 generated image, blurred, restored, and original images from left to right respectively, T=1000

Following the implementation details provided in the article, the Gradient Conjugate algorithm was first used to solve the latent optimization given in equation (17) for 5 iterations. To update the null-text embeddings for performing adaptive negation, the null embedding was set as a learnable parameter. Then, the inner product of the image embedding and these null-text embeddings was minimized using the AdamW optimizer with a learning rate of 1e-2.

In addition, since the prompt "The white house" was used to generate the image from stable diffusion, it was decided to use the same prompt for text guidance in TReg for semantically meaningful results while deblurring. The result is shown in Figure 7



Figure 7: TReg 128x128 image, blurred, restored, and original images from left to right respectively, T=1000

The output of the Text Regularization (TReg) at this resolution was observed to be very poor.

Through these experiments, the delicate balance between model capabilities, image quality, and computational resource constraints is underscored. This emphasizes the need for more efficient computational strategies when using such models.

## 5 Discussion

The series of experiments conducted provided valuable insights into the limitations and challenges faced during the implementation of Text Regularization (TReg) within a stable diffusion framework. This discussion aims to address the issues encountered and provide justifications for the observed results, which did not consistently meet our expectations for image restoration quality.

### 5.1 GPU Memory Constraints

One of the most significant technical constraints encountered was the limitation imposed by GPU memory capacity, only 15MB of GPU RAM were accessible. The ambitious goal of applying TReg to high-resolution images was constrained by the inability of our current computational resources to process images of  $512 \times 512$  pixels. This limitation was particularly impactful due to the intensive memory requirements of the TReg process, which involves maintaining a complex computational graph to handle the multitude of gradient computations necessary for updating  $z_t$ .

The constraints of the GPU memory not only restricted the resolution of the images that could be processed but also limited the scope of the model’s learning capacity. Training on lower-resolution images as a workaround, while necessary, inherently reduced the amount of detail and information that the model could leverage, leading to a trade-off between feasibility and quality.

### 5.2 Challenges with Low-Resolution Image Deblurring

The decision to shift focus to lower-resolution images arose from the necessity to fit within the available computational limits. However, this adaptation introduced a new set of challenges. Low-resolution images contain less detail, which increases the difficulty of deblurring. When resolution is reduced, essential information that could aid in the restoration process is lost, often leading to outputs that are less clear and less detailed.

Furthermore, the use of a large Gaussian blur kernel size and a high standard deviation amplified the null-space of the problem, creating a more challenging test environment for the model’s deblurring capabilities. This also likely contributed to the mediocre restoration quality observed, particularly in the context of low-resolution images where the initial information content is limited.

### 5.3 Results Justification

The suboptimal results obtained from the deblurring experiments can be attributed to a combination of the aforementioned factors. On one hand, the limited GPU memory curtailed the ability to work with high-resolution images that contain rich detail, which is often crucial for effective deblurring. On the other hand, the very nature of low-resolution images implies a reduction in available information, making it inherently challenging to restore them to a high-quality state.

Moreover, the aggressive forward operations that were intended to produce a robust null-space further obscured the images, making it even more challenging for the model to discern and reconstruct the original content accurately. This, paired with the computational limitations, presents a clear justification for the lack of fidelity in the restored images.

In conclusion, while the results of these experiments were not as successful as desired, they offer important lessons on the complexities of image restoration using advanced generative models like TReg, particularly under the constraints of available computational resources. Moving forward, it would be beneficial to explore more efficient models and training techniques that can operate within such limitations while still achieving high-quality results.

## References

- [1] Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems, 2023.
- [2] Jeongsol Kim, Geon Yeong Park, Hyungjin Chung, and Jong Chul Ye. Regularization by texts for latent diffusion inverse solvers, 2023.
- [3] Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. Solving linear inverse problems provably via posterior sampling with latent diffusion models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

## Additional Notes

In our pursuit to enhance the application scope of Text Regularization, an attempt was made to apply this technique to higher-resolution images, specifically those with a resolution of 512x512 pixels. Despite the potential for richer detail representation at this resolution, the output obtained was notably irrelevant. This unexpected outcome suggests a discrepancy in the algorithm's performance when scaling to higher resolutions. Initial investigations indicate that this might be due to an implementation error within our algorithm, which, despite our efforts, remained unresolved. This experience highlights the nuanced challenges of adapting Text Regularization techniques to more complex, higher-resolution images and underscores the need for further refinement and testing of our approach.

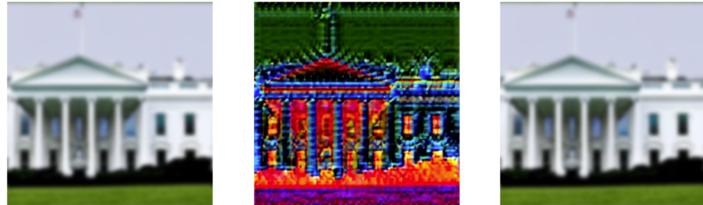


Figure 8: TReg 512x512 image, blurred, restored, and original images from left to right respectively,  $T=1000$