Existing methods utilize a guidance function to adjust a diffusion model’s noisy estimate $z_t$ at each time step.

We compute the denoised estimate $\hat{z}_t$ at each time step and compute the steering loss using this estimate.

For certain tasks, a looping scheme helps: For each $t$, we perform the implicit steered diffusion step multiple times.

Differences from existing diffusion-based methods

- Conditional generative models typically require large annotated training sets to achieve high-quality synthesis.
- We propose Steered Diffusion: a diffusion-based solution for zero-shot conditional image generation.
- Our method incorporates off-the-shelf models and performs conditional sampling without the need for any task-specific training.
- Steered Diffusion combines a diffusion model trained for unconditional generation with a pre-trained/predefined steering function to perform zero-shot conditional generation.

Widespread applications of Steered Diffusion

(a) Inpainting  (b) Colorization  (c) Super-resolution

(d) Semantic Generation  (e) Identity Replication  (f) Text-based editing

Quantitative results and comparisons:

- For steered diffusion, we need an inverse mapping from the output image space to the input space of the condition. This can be a neural network or a predefined inverse mapping function.
- In complex conditional mappings (such as identity replication and semantic generation), we utilize off-the-shelf neural networks to perform the inverse mapping from a denoised image to the condition.
- For linear inverse problems (such as inpainting and super-resolution), the inverse mapping is a known linear degradation function $D$. We replace the degraded part by the conditioning input $c$ and proceed with sampling process.

Implicit steering control corrects (steers) the implicit prediction according to a loss function. The loss function is chosen according to the condition, $c$.

Implicit steering control corrects (steers) the implicit prediction according to a loss function. The loss function is chosen according to the condition, $c$.

For steered diffusion, we need an inverse mapping from the output image space to the input space of the condition. This can be a neural network or a predefined inverse mapping function.

In complex conditional mappings (such as identity replication and semantic generation), we utilize off-the-shelf neural networks to perform the inverse mapping from a denoised image to the condition.

For linear inverse problems (such as inpainting and super-resolution), the inverse mapping is a known linear degradation function $D$. We replace the degraded part by the conditioning input $c$ and proceed with sampling process.

Implicit steering control corrects (steers) the implicit prediction according to a loss function. The loss function is chosen according to the condition, $c$.

For steered diffusion, we need an inverse mapping from the output image space to the input space of the condition. This can be a neural network or a predefined inverse mapping function.

In complex conditional mappings (such as identity replication and semantic generation), we utilize off-the-shelf neural networks to perform the inverse mapping from a denoised image to the condition.

For linear inverse problems (such as inpainting and super-resolution), the inverse mapping is a known linear degradation function $D$. We replace the degraded part by the conditioning input $c$ and proceed with sampling process.