Mimic the Diffusion

Forward SDE (data → noise)

$\mathbf{x}(0)$ ———— $\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$ ———→ $\mathbf{x}(T)$

score function

$\mathbf{x}(0)$ ←— $\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\boxed{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})}\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}$ —— $\mathbf{x}(T)$

Reverse SDE (noise → data)

Feature 1.

- High-noised distribution => Low-noised distribution (gradually update)

- High-β VAE => Low-β VAE
- This means "high distortion => low distortion"
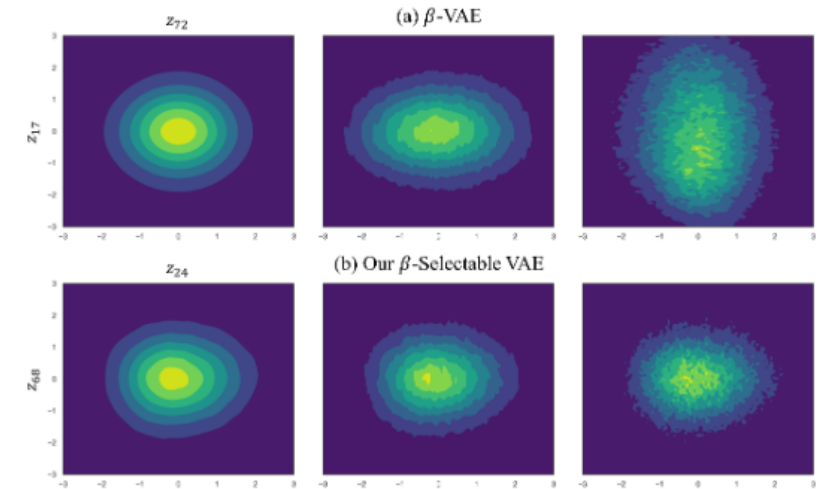- It may not be perfect, but the meaning is similar

Figure 2: Aggregated posteriors for arbitrarily chosen latent channels. Each row shares a model and channels, and distributions are placed in order of decreasing β. The posteriors of our model (b) are *aligned* relative to their counterparts (a).

Feature 2.

- Iterative refinement with "very small step"

- β-Selectable VAE has the power to perform it
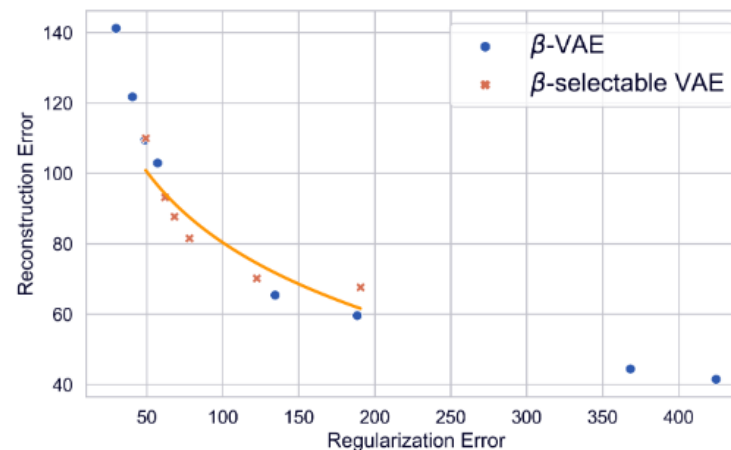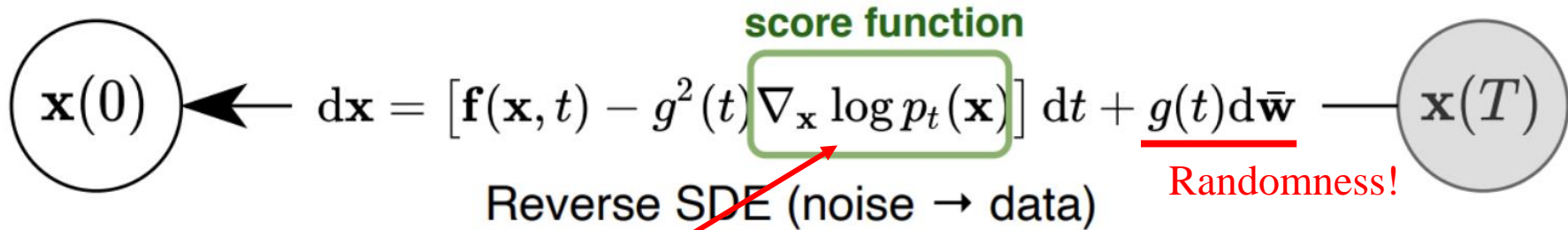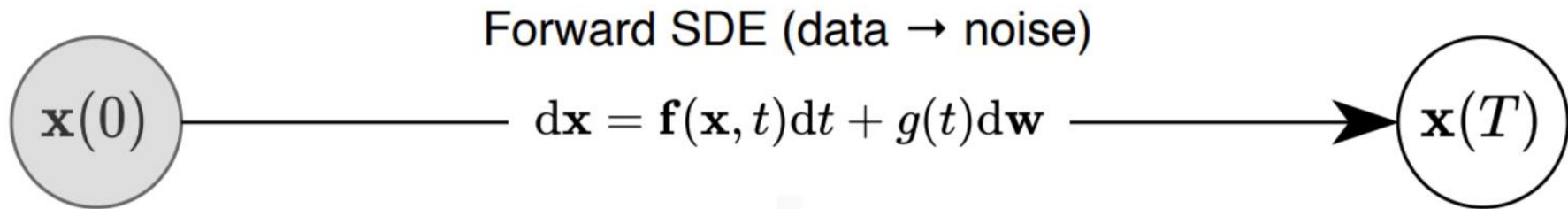- It can represent continuous β intervals,
- With the same model capacity

Figure 3: Reconstruction and regularization errors of ELBO for various βs. For ELBO, i.e. the learning objective, a single β-Selectable VAE is capable of approximate many β-VAEs. This graph can be interpreted as a rate-distortion curve.

Mimic the Diffusion

Forward SDE (data → noise)

$$\mathbf{x}(0) \quad\quad d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w} \quad\quad \mathbf{x}(T)$$

score function

$$\mathbf{x}(0) \quad d\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\boxed{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})} \right] dt + \underline{g(t)d\bar{\mathbf{w}}} \quad \mathbf{x}(T)$$

Reverse SDE (noise → data)

Randomness!

Model's target
Not random

Feature 3.

- "Randomness"

- Recursively re-input to VAE
- I think the latent variables on the posteriors with very close beta values,
- Would maintain its meaning on the other side well
- There will be a small loss of information, i.e. acceptable randomness
- This is a logical leap though

Feature 4.

- Iterative "refinement" with very small step

- I think that the optimization of ELBO for the latent z can take over the role

- ***BUT!!***
- Consider a few steps later:
- x_t (modified x_0) $\rightarrow z_t \rightarrow x'_t$ (reconstructed $x_t = x_{t+1}$)
- ELBO means how similar x_t and x'_t are… not likelihood!
- It loses the meaning as a refinement

Reflection

- This is a huge problem that is *easy to spot*

- I accepted it as the right method

- Because some experimental results came out good (e.g. MNIST)

- Of course, one need to experiment with various variables, but

- There is a complacent attitude because that were good in one or two…

Features (all)

- Iterative "refinement" with "very small step"
- With "gradually noised distribution" and "Randomness"

- If there is a good way to refine a sample, it could be done again