

ADM(Ablated Diffusion Model)

ソース

- [Diffusion Models Beat GANs on Image Synthesis](#)

概要

- diffusion modelのablation studyを行って最適なモデルアーキテクチャを探した
- 評価指標としてFIDを採用した
- 当時の画像生成のSOTAであるBigGANをFIDで上回った
 - diffusionがGANを超えたと言われるきっかけになった論文がこれ

提案手法

ablation study

- 以下のことを検証
 - モデルサイズを固定した時にdepthとwidthのどちらを増やすべきか
 - attention headsを増やす
 - attentionの計算時に 16×16 だけでなく、 32×32 , 16×16 , 8×8 についても検証する
 - アップサンプリングやダウンサンプリングでBigGANの残差ブロック(residual block)を使う
 - 残差接続時に $\frac{1}{\sqrt{2}}$ でrescaleする
- その結果、以下のことがわかった
 - headsは多い方が良い
 - channelsは少ない方が良い
 - 学習時間が増えるがresidual blocksを増やした方が良い
- 次のラベルを用いた条件付き生成では上記のことを採用した上で実験

classifier guidance

- ラベルを用いた条件付き生成においてBigGANを上回った
- DDPM, DDIMそれぞれにおける生成過程(reverse noising process)のアルゴリズムは以下

DDPMの場合

- diffusion model : $(\mu_\theta(x_t), \Sigma_\theta(x_t))$
- classifier : $p_\phi(y|x_t)$
- class label : y
- gradient scale : s
- $x_T \sim \mathcal{N}(0, I)$
- このとき $t = T, \dots, 1$ まで以下を繰り返す

$$\begin{aligned}\mu &:= \mu_\theta(x_t) \\ \Sigma &:= \Sigma_\theta(x_t) \\ x_{t-1} &\sim \mathcal{N}(\mu + s\Sigma\nabla_{x_t} \log p_\phi(y|x_t), \Sigma)\end{aligned}\tag{1}$$

- 最後に x_0 を出力

DDIMの場合

- diffusion model : $\epsilon_\theta(x_t)$
- classifier : $p_\phi(y|x_t)$
- class label : y
- gradient scale : s
- $x_T \sim \mathcal{N}(0, I)$
- このとき $t = T, \dots, 1$ まで以下を繰り返す

$$\begin{aligned}\hat{\epsilon} &:= \epsilon_\theta(x_t) - \sqrt{1 - \bar{\alpha}_t} \nabla_{x_t} \log p_\phi(y|x_t) \\ x_{t-1} &\sim \sqrt{\bar{\alpha}_{t-1}} \left(\frac{x_t - \sqrt{1 - \bar{\alpha}_t} \hat{\epsilon}}{\sqrt{\bar{\alpha}_t}} \right) + \sqrt{1 - \bar{\alpha}_{t-1}} \hat{\epsilon}\end{aligned}\tag{2}$$

- 最後に x_0 を出力