

Cascaded Diffusion Model

ソース

- [Cascaded Diffusion Models for High Fidelity Image Generation](#)

概要

- クラス条件付き画像生成においてBigGAN-deepやVQ-VAE-2を上回る性能
- diffusion modelで画像生成をした後に、Super Resolution Diffusion Modelで解像度をあげる
 - SRの処理を重ねて段階的に解像度を上げている
 - これをCascadeと名付けた
- さらにdata augmentationによる学習で精度向上させた

前提

Diffusion Model

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad (1)$$

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \quad (2)$$

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \quad (3)$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)) \quad (4)$$

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) \quad (5)$$

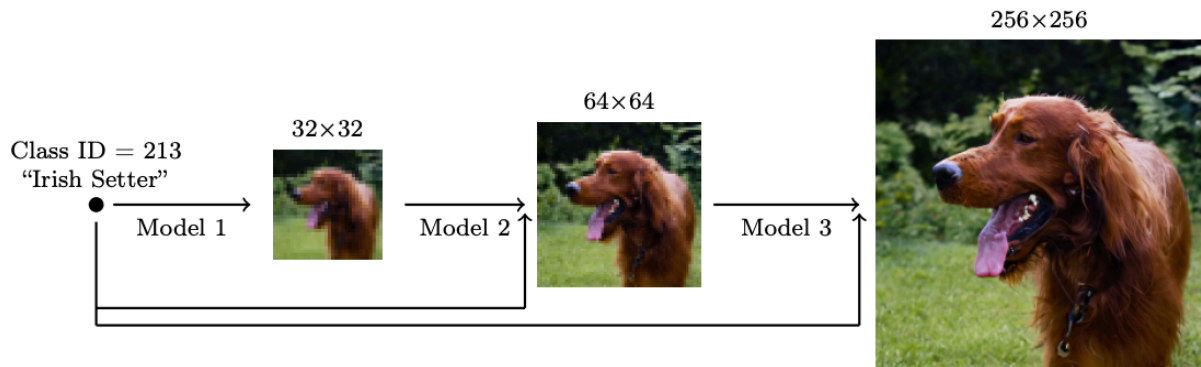
$$\boldsymbol{\Sigma}_\theta^{ii}(\mathbf{x}_t, t) = \exp \left(\log \tilde{\beta}_t + (\log \beta_t - \log \tilde{\beta}_t) \cdot v_\theta^i(\mathbf{x}_t, t) \right) \quad (6)$$

$$L_{\text{simple}}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}_0, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), t \sim \mathcal{U}(\{1, \dots, T\})} \left[\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta \left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, t \right) \right\|^2 \right] \\ \left(\because \mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon} \right)$$

Conditional Diffusion Model

提案手法

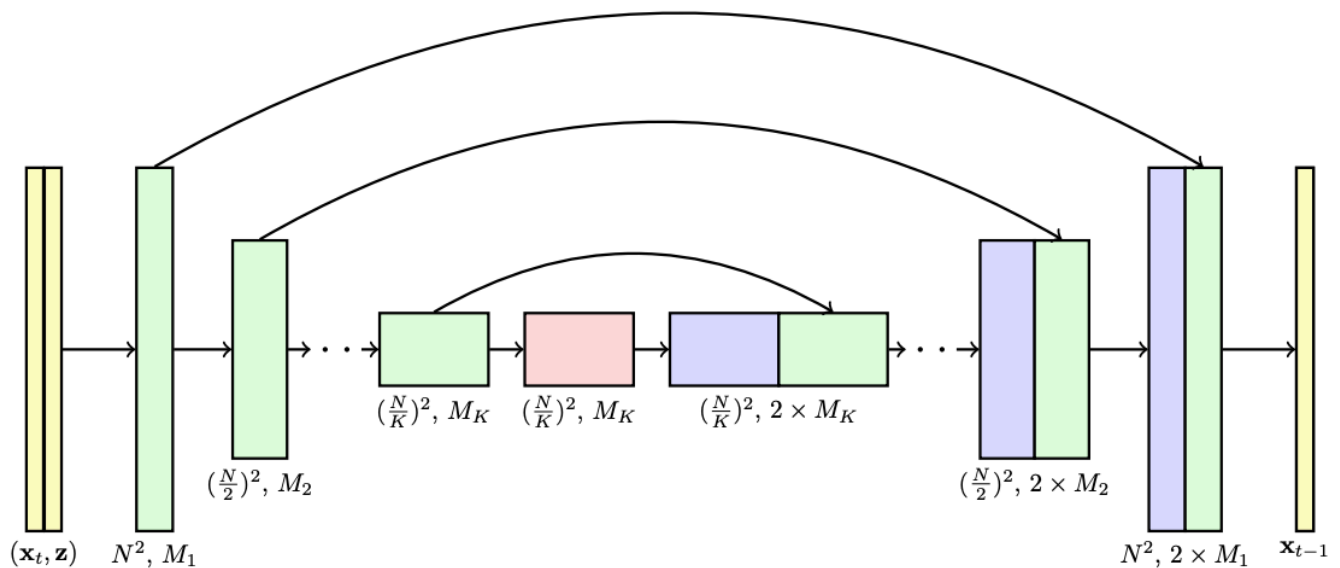
- 以下全体図



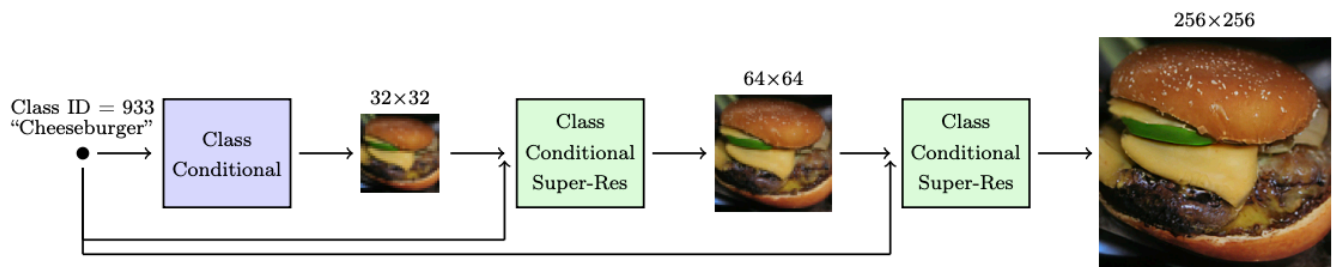
- Model1がconditional diffusion model
- Model2,3がclass conditional Super Resolution Diffusion Model
 - これの元は、Image super-resolution via iterative refinementという論文のモデル

class conditional super resolution diffusion model

- 入力は3つ
 - \mathbf{x}_t : tステップ目のノイズ画像
 - y : クラス
 - \mathbf{z} : 低画質画像
- 全体図はU-Net



- クラスラベルは以下のように与えられる



Data Augmentation

- Blurring Augmentation
- Truncated Conditioning Augmentation
- Non-truncated Conditioning Augmentation