# **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**

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

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

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

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

$$\mu_{\theta}(\boldsymbol{x}_{t},t) = \frac{1}{\sqrt{\alpha_{t}}} \left( \boldsymbol{x}_{t} - \frac{\beta_{t}}{\sqrt{1-\alpha_{t}}} \boldsymbol{\epsilon}_{\theta}(\boldsymbol{x}_{t},t) \right)$$
 (5)

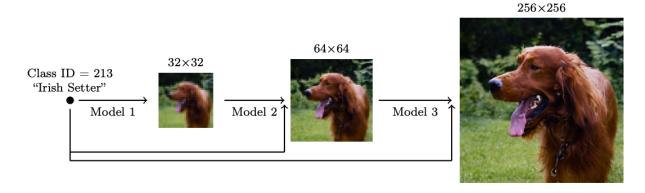
$$\Sigma_{\theta}^{ii}(\boldsymbol{x}_{t},t) = \exp\left(\log \tilde{\beta}_{t} + (\log \beta_{t} - \log \tilde{\beta}_{t}) \cdot v_{\theta}^{i}(\boldsymbol{x}_{t},t)\right)$$
 (6)

$$L_{ ext{simple}}(oldsymbol{ heta}) = \mathbb{E}_{oldsymbol{x}_0, oldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), t \sim \mathcal{U}(\{1, \dots, T\})} \left[ \left\| oldsymbol{\epsilon} - oldsymbol{\epsilon}_{ heta} \left( \sqrt{ar{lpha}_t} oldsymbol{x}_0 + \sqrt{1 - ar{lpha}_t} oldsymbol{\epsilon}, t 
ight) 
ight\|^2 
ight] \ \left( \therefore oldsymbol{x}_t = \sqrt{ar{lpha}_t} oldsymbol{x}_0 + \sqrt{1 - ar{lpha}_t} oldsymbol{\epsilon} 
ight)$$

#### **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

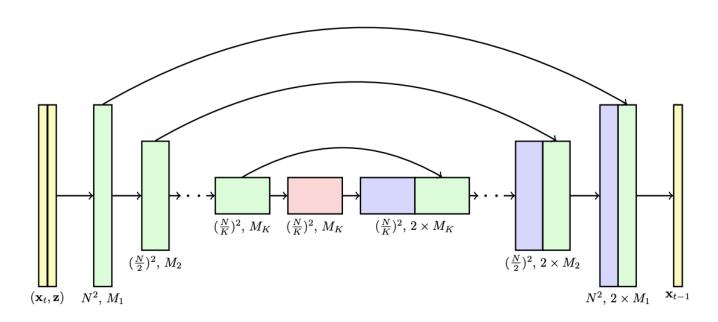
• 入力は3つ

。  $\boldsymbol{x}_t$ : tステップ目のノイズ画像

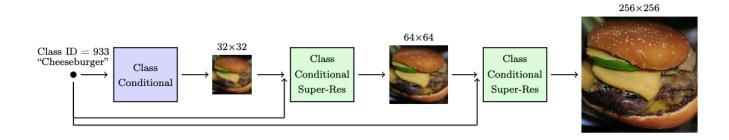
。 y: クラス

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