

You Only Need One Step: Fast Super-Resolution with Stable Diffusion via Scale Distillation

論文ソース

- [You Only Need One Step: Fast Super-Resolution with Stable Diffusion via Scale Distillation](#)

概要

- SF(Scale Factor)が大きくなると計算速度が遅くなる問題に取り組んだ
- Scale Distillationという手法を提案
- これは教師モデルの出力を正解とする蒸留の手法
- 生徒モデルのタスクよりもSFが小さいタスクの教師モデルを利用する
 - SFを N とする(縦横それぞれ N 倍にする超解像タスク)
 - このとき、 $\frac{N}{2}$ 倍のタスクで学習済みのモデルをもってきて教師とする
 - 正解画像に対して、解像度を $\frac{1}{N}$, $\frac{2}{N}$ 倍に落としたものを用意
 - 前者の画像を生徒モデルに入力して、後者の画像を教師モデルに入力
 - この出力結果の誤差をとって生徒を学習

前提

条件付きLatent Diffusion ModelによるSR

- 低画質画像と高画質画像のペアを $(\mathbf{x}_h, \mathbf{x}_l) \sim p(\mathbf{x}_h, \mathbf{x}_l)$ とする
- Encoderを通して得られる潜在変数を $\mathbf{z}_h = \mathcal{E}(\mathbf{x}_h)$, $\mathbf{z}_l = \mathcal{E}(\mathbf{x}_l)$ とする
- このとき、拡散過程は以下

$$q(\mathbf{z}_t | \mathbf{z}_h) = \mathcal{N}(\mathbf{z}_t; \alpha_t \mathbf{z}_h, \sigma_t \mathbf{I})$$
$$\arg \min_{\theta} \mathbb{E}_{\epsilon, t} [\omega(\lambda_t) \|\hat{\mathbf{z}}_{\theta}(\mathbf{z}_t, \mathbf{z}_l, \lambda_t) - \mathbf{z}_h\|_2^2]$$

- ただし $\lambda_t = \log \frac{\alpha_t^2}{\sigma_t^2}$
- ω は t に依存する係数

提案手法

- N 倍の超解像タスクのために、Teacherモデルとして $\frac{N}{2}$ 倍のモデルを持ってくるというもの
- x_h に対して解像度を $\frac{1}{N}$, $\frac{2}{N}$ に落としたものをそれぞれ x_l, x'_l とする
- N 倍タスクのためのStudentモデルを \hat{z}_θ として $\frac{N}{2}$ 倍タスクのためのTeacherモデルを \hat{z}_ϕ とする
- このとき、distillation lossを以下で定義する

$$\arg \min_{\theta} \mathbb{E}_{\epsilon, t} [\omega(\lambda_t) \|\hat{z}_\theta(z_t, z_l, \lambda_t) - \hat{z}_\phi(z_t, z'_l, \lambda_t)\|_2^2]$$

英単語メモ

- drastically : 劇的に
- exacerbate : 悪化させる
- prohibitive : 法外な
- alleviate : 軽減する、緩和する
- catastrophic : 壊滅的な
- differentiating : 区別する
- peculiarity : 特殊性
- dubbed : 吹き替えられた
- progressive : 進歩的な
- supervisory : 監督上の
- threshold : 三重の
- degradation : 劣化
- shortcoming : 欠点、短所
- tandem : 縦に並んだ