

パッチ分割による拡散確率モデルのメモリ消費量削減の検討

荒川 深映¹

綱島 秀樹¹

堀田 大地²

田中 啓太郎¹

森島 繁生³

¹⁾ 早稲田大学

²⁾ 東京大学

³⁾ 早稲田大学理工学術院総合研究所



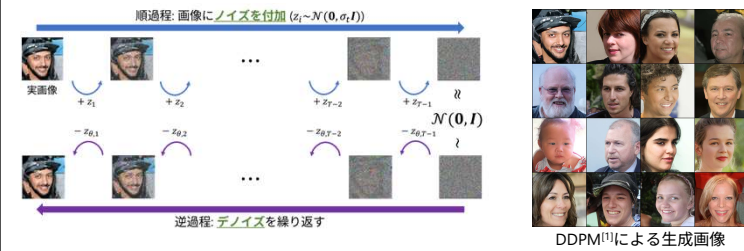
はじめに

- 拡散確率モデル (DDPM) ^[1] は高品質・多様な画像の生成に成功
- ノイズ画像から繰り返しデノイズを行うことで画像を生成

高解像度の画像を生成する場合に大きなメモリを要求する問題

我々の研究では画像の『**パッチ分割**』を行うことで空間的に小さいパッチ上でネットワークの推論を行う

💡 課題: **全体画像の情報とパッチの位置**



関連研究

COCO-GAN^[2]

- GANを用いたパッチ単位の画像生成
- 生成品質が改善
- メモリ効率/並列性に優れる

⇒ **識別器**によりシームレスなパッチ間の特徴を学習

Patch Diffusion^[3]

- ランダムな位置で切り抜いたパッチを用いて拡散確率モデルを学習
- 位置による条件付けを追加
- メモリ効率/損失の収束性が向上

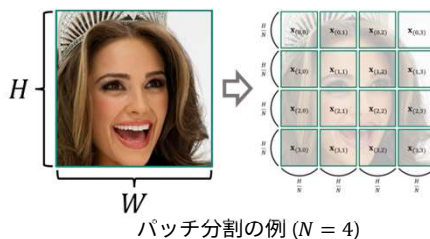
⇒ **学習時のみに有効** (推論時はフルサイズ画像を用いる)

手法

パッチ分割法

- 全体画像を縦横に **N 分割**
⇒ 計 N^2 枚のパッチへ分割
- オーバーラップなし

入力テンソルサイズを $O(1/N^2)$ で小さくする



全体画像情報の条件付け

- パッチ間で全体画像の情報を共有
- **Global Content Conditioning (GCC)**

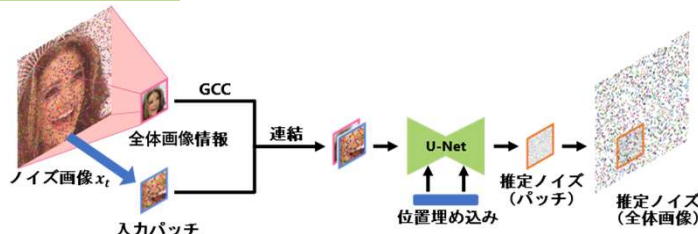


位置の条件付け

- パッチの位置をネットワークへ伝える
- インデックス $s = i \times N + j$ を定義
- s を one-hot 表現に変換し MLP へ入力



サンプリング手法

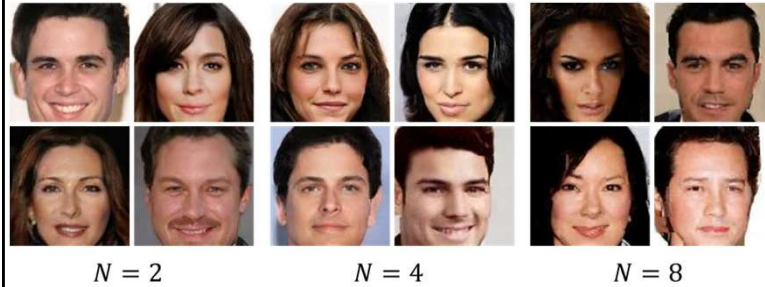


結果

実験設定

- ベースライン: DDPM^[1]
- CelebA^[4]とLSUN bedroom^[5]データセットを用いて学習
- **全体画像サイズは128 (パッチサイズ: 128/ N)**
- $N = 1, 2, 4, 8$ のそれぞれで学習
- 最大メモリ消費量はネットワークの計算部だけで測定

CelebA^[4]



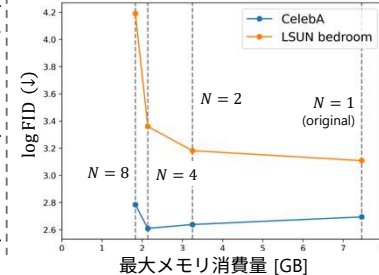
LSUN bedroom^[5]



FID scores

データセット	分割数 N	FID (↓)
CelebA	1 (ベースライン)	14.8
	2	14.0
	4	13.6
	8	16.2
LSUN bedroom	1 (ベースライン)	22.4
	2	24.1
	4	28.8
	8	66.1

最大メモリ消費量 vs FID score



考察

- $N = 2, 4$ の場合はいずれのデータセットでも自然な画像を生成
 - $N = 8$ の場合にパッチ間の境界線が目立つ (特にLSUN bedroom)
 - $N = 8$ の場合に、LSUN bedroomでのFIDが急激に悪化した
- ⇒ LSUN bedroomには**様々な構図**の画像が含まれており学習が困難であることが要因か

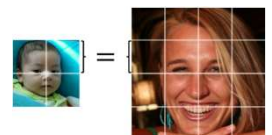
結論 & 展望

結論

- 推論時におけるメモリ消費量を削減するパッチ単位画像生成
- **Global Content Conditioning**と**位置の条件付け手法**の提案
- ベースラインのDDPMと比較し、生成品質を維持しながら最大メモリ消費量を半分以上抑える

今後の展望

- 任意解像度の画像を生成するためにランダムな位置/大きさのパッチで学習



参考文献

- [1] Ho et al., "Denoising diffusion probabilistic models," NeurIPS, 2020.
- [2] Lin et al., "COCO-GAN: Generation by parts via conditional coordinating," ICCV, 2019.
- [3] Wang et al., "Patch Diffusion: Faster and More Data-Efficient Training of Diffusion Models," arXiv:2304.12526, 2019.
- [4] Liu et al., "Deep Learning Face Attributes in the Wild," ICCV, 2015.
- [5] Yu et al., "LSUN: construction of a large-scale image dataset using deep learning with humans in the loop," arXiv:1506.03365, 2015.
- [6] Chai et al., "Any-resolution Training for High-resolution Image Synthesis," ECCV, 2022.

My website

