修 士 論 文 概 要 書

Master's Thesis Summary

Date of submission: 01/22/2024 (MM/DD/YYYY)

専攻名 (専門分野) Department	情報理工・ 情報通信専攻	氏名 Name	Jichen Ma	指 導 教 員 Advisor	渡辺 裕 ㊞
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	5122F070-2		
研究題目 Title	超解像拡散モデルの中間ステップ予測による加速手法の研究 Research on Acceleration Method for Super-Resolution Diffusion Models by Intermediate Step Prediction				

1. まえがき

近年，SNS の普及により写真撮影の注目が高まっているが，データ圧縮による画像品質の低下が問題となっている．この背景から，単一画像超解像技術の重要性が高まっており，特にニューラルネットワークに基づく超解像手法が注目されている．しかし，低解像度画像が複数の高解像度画像に対応する可能性の問題，従来手法の高周波部分の処理の欠陥などが挙げられる．拡散モデル (Diffusion Model)は画像生成において顕著な進歩を遂げているが，生成速度の遅さが課題である．

本研究では，Diffusion Model に基づく単一画像超解像モデルの生成速度を向上させる新しい手法を提案する．提案手法は，高品質な画像を維持しつつ，他のサンプリング加速方法との互換性も高く，超解像タスクの効率と品質の向上を実現する．

2. 関連技術

2.1. DDPM

Denoising Diffusion Probabilistic Models (DDPM)は画像生成タスクにおいて顕著な進歩を遂げたニューラルネットワークモデルで, U-Net[4]モデルを基にしたアーキテクチャを採用している[1]．DDPM は拡散過程と逆拡散過程の二つの段階から成る．拡散過程では，画像x_0にガウスノイズを T ステップ加え，一連の画像x_1，x_2，…，x_Tを生成し，x_Tはガウスノイズ画像となる．逆拡散過程では，x_Tから始め，拡散モデルを通じて逆拡散処理を行い，最終的に元の画像x_0を復元する．

2.2. SR3

SR3[2]は Diffusion Model に基づく超解像技術である．ネットワーク構造は U-Net に基づき，訓練過程では低解像度の画像を高解像度に補間し，ノイズを加えた高解像度画像と連結してモデルの入力とする．モデルは前の時刻に加えられたノイズを予測し，これによって超解像が実現される．生成段階では，ガウスノイズと補間された低解像度画像を連結してモデルに入力し，逆拡散プロセスを経て最終的な高解像度画像を生成する．このプロセスにより，SR3 は効果的に超解像を行う．

3. 提案手法

3.1. 提案手法概要

本研究では，Diffusion Model に基づく超解像タスクの効率向上を目指し，新たな加速手法を提案する．この手法は，中間ステップx_tの予測に重点を置き，生成プロセスの時間短縮を目指す．従来の Diffusion Model における生成プロセスのガウスノイズからの開始を修正し，予測された中間ステップx_tから開始することで，生成時間が大幅に削減される．図 1 に本研究で提案する加速手法を示す．

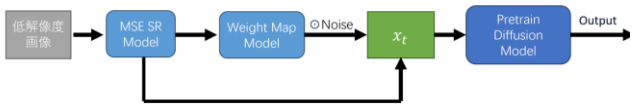

図 1　提案手法の概要図

3.2. 中間x_tの予測

中間x_tの予測モデルは，平均二乗誤差 (MSE) 損失関数で訓練された MSE SR Model と，Weight Map Model の二つの部分から構成される．MSE SR Model は, 超解像画像の生成に用いられ, Weight Map Model は，MSE SR Model の出力と実画像との差異を予測する役割を担う．低解像度の画像を MSE SR Model によって出力された超解像画像をx_{sr}とする．Weight Map Model の出力をWとする．βはWと同じ形状で要素値が同じ行列である．予測x_tの計算方法を式(1)に示す．

$$x_t = \sqrt{\bar{\alpha}_t}x_{sr} + clamp(\beta + W)\sqrt{1-\bar{\alpha}_t}\epsilon, \epsilon \sim \mathcal{N}(0, I)$$
$$, clamp(\beta + W) = \max(0, \min(1, \beta + W)) \quad (1)$$

4. 評価実験

4.1. 実験内容

本研究では，Flickr-Faces-HQ（FFHQ）[3]データセットを用いて MSE SR Model と Weight Map Model の訓練を行う. Weight Map Model は 97M のパラメタを持つニューラルネットワークで, FFHQ データセットで訓練する. また, コントラスト強化技術を導入し, MSE 損失関数を用いて訓練を行う．図 3 に Weight Map の可視化出力の例を示す．提案手法の有効性を評価するため，DDPM のサンプリング技術および DDIM (Denoising Diffusion Implicit Models)[5]加速サンプリング技術と組み合わせて，テストセットで超

解像画像を生成し，FID スコアにより評価する．

実験では，訓練済 SR3 モデルを用い，$x_{t=300}$の中間ステップ画像を直接予測し，300 ステップのイテレーション実験を行う．また，ハイパラメータβの異なる値が生成画像の品質と FID に与える影響を調査し，Weight Map を使用せず，MSE SR Model の出力に直接ノイズを加えたテストも行う，評価結果を表 1 に示す．

また，DDIM サンプリング法を組み合わせた加速実験を実施し，提案手法と DDIM の FID スコアを比較する．本実験では，βを 0.8 に設定し， t=300 の時刻でx_tを予測し，加速実験を行った．Number of Function Evaluations (NFE)は，ニューラルネットワーク計算の総回数を指し，NFE が低い場合の提案手法の優位性を評価する．結果を表 2 に示す．

図 2 は，NFE が 8 の場合，提案手法を用いて予測時点 t が 300 であるときに生成した超解像画像と DDIM による生成画像の比較である．また Weight Map を可視化した例を図 3 に示す．本研究で示される画像は，Creative Commons Attribution 2.0 Generic ライセンスに従う Flickr-Faces-HQ データセットからのものである．

表 1　評価結果(FID↓)

	β=0	β=0.2	β=0.4	β=0.6	β=0.8	β=1.0
w/ W-Map	15.70	14.28	13.39	**13.22**	13.56	14.74
w/o W-Map	17.20	17.14	16.80	16.38	15.64	14.74

表 2　DDIM と提案手法における FID の比較

NFE	8	10	15	20	25	50
DDIM	34.27	33.42	30.39	28.25	26.12	19.26
Ours	**16.27**	19.27	18.65	18.68	22.76	19.68

図 2 提案手法と DDIM による生成画像の例

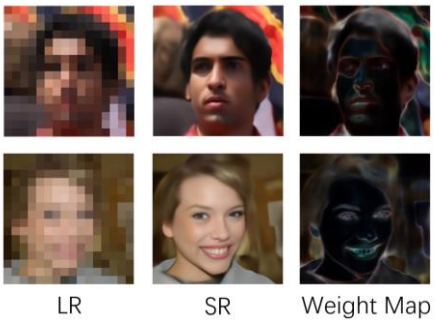

図 3 Weight Map の可視化例

4.2. 考察

表 1 の実験結果から，Weight Map を使用し，β 値を 0.6 に設定した場合に最も低い FID スコアが得られたことが分かる．Weight Map を使用することで，より少ないノイズを加えても低い FID を達成し，生成品質を向上させることができる．

表 2 の評価実験から，DDIM と比較して，提案手法は特に NFE が低い条件下で優れた FID スコアを示し，生成された画像の細部情報の改善が見られる．しかし，NFE が増加すると，提案手法の FID スコアの改善はそれほど顕著ではなく，DDIM との差が縮まる．この結果から，NFE が低い場合に提案手法が DDIM に比べて優れた生成品質を提供することが示されたが，計算時間とリソースが十分にある場合には DDIM の方が優れる可能性がある．

5. むすび

本研究では Diffusion Model に基づく超解像モデルの生成速度問題に対処し，中間ステップの予測による加速手法を提案した．評価実験により，提案手法は DDPM および DDIM サンプリング法と組み合わせて使用可能である．DDIM と組み合わせて，低い NFE の条件下で優れた FID スコアと生成品質を達成することが確認できた．

参考文献

[1] J. Ho, A. Jain, P. Abbeel: "Denoising Diffusion Probabilistic Models," arXiv preprint, arXiv:2006.11239, 2020.

[2] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, M. Norouzi: "Image Super-Resolution via Iterative Refinement," arXiv preprint, arXiv:2104.07636, 2021.

[3] T. Karras, S. Laine, T. Aila: "A Style-Based Generator Architecture for Generative Adversarial Networks," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4217-4228, 2021.

[4] O. Ronneberger, P. Fischer, T. Brox: "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv preprint, arXiv:1505.04597, 2015.

[5] J. Song, C. Meng, S. Ermon: "Denoising Diffusion Implicit Models," arXiv preprint, arXiv:2010.02502v4, 2020.

2023 年度

早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻　修士論文

超解像拡散モデルの中間ステップ予測による加速手法の研究

Research on Acceleration Method for Super-Resolution Diffusion Models by Intermediate Step Prediction

Jichen Ma

(5122F070-2)

提出日：2024.01.22

指導教員：渡辺裕教授　印

研究指導名：オーディオビジュアル情報処理研究

目次

第1章　序論

1.1 研究背景

近年，ソーシャルネットワーキングサービス（SNS）の普及に伴い，写真撮影が広く注目されるようになった．しかし，ネットワーク伝送中のデータ圧縮により，画像の品質が期待に応えられないことがしばしばある．この背景で，画像品質を向上させること，特に単一画像超解像技術の重要性が高まっている．この技術分野は，ニューラルネットワークに基づく超解像手法で特に注目されている．

単一画像超解像技術の主な課題は，同じ低解像度画像が複数の異なる高解像度画像に対応している可能性があることである．従来研究である SRCNN[1]や SRGAN[2]は，この分野で顕著な成果を上げているが，画像の高周波部分を処理する際に一定の欠陥があった．これは，低解像度入力が複数の高解像度出力に対応していることを十分に考慮していなかったことによる．

過去 2 年間で，Diffusion Model は画像生成の複数のタスクで画期的な進歩を遂げた．単一画像超解像タスクを含む，SR3[3]のような超解像 Diffusion Model は，訓練中にノイズを導入することで，Generative Adversarial Networks (GAN)に基づく超解像モデルよりも画像の高周波部分をより良く生成することができる．しかし，Diffusion Model の主な欠点は生成速度が遅いことであり，通常は数千回の自己回帰が必要である．そのため，画像生成タスクにおいて，Diffusion Model の生成速度を最適化することが研究の焦点となっている．

1.2　研究目的

Diffusion Model は画像生成の分野，特に単一画像超解像タスクにおいて顕著な進歩を遂げている．Diffusion Model に基づく単一画像超解像モデルである SR3 は画像生成において成果を上げているが，生成にかかる時間が長いという大きな欠点がある．これは，Diffusion Model が生成過程で多量の計算を必要とするためである．

この問題に対処するため，近年 Diffusion Model の生成プロセスを加速する技術が開発されている．その中で，Denoising Diffusion Implicit Models (DDIM)[4] は，サンプリング回数を減らすことで Diffusion Model の生成速度を大幅に向上させる重要な進歩を遂げている．しかし，これらの加速方法は主に一般的なタスク向けであり，単一画像超解像モデルには特化していない．

この問題に対処するため，本研究では Diffusion Model に基づく単一画像超解像モデルに特化した新しい加速技術を提案する．この方法は生成速度を大幅に向上させるだけでなく，高品質な画像を維持することが可能である．さらに，他の Diffusion Model のサンプリング加速方法との高い互換性を持ち，超解像タスクの生成速度を向上させることができ，効率を高めると同時に画像生成の品質を維持できる．

1.3 本論文の構成

以下に本章以降の構成を示す．

第 1 章では，本研究の背景および目的について述べている．

第 2 章では，本研究の関連技術について述べる．

第 3 章では，予備実験について述べる．

第 4 章では，本論文で提案する手法について述べる．

第 5 章では，評価実験について述べる．

第 6 章では，本論文の結論と考察および今後の課題について述べる．

第2章　関連技術

2.1 まえがき

本章では，Denoising Diffusion Probabilistic Models (DDPM) [5]と Diffusion Model のサンプリング加速手法 Denoising Diffusion Implicit Models (DDIM) について述べた後．単一画像超解像技術において代表的な先行研究である Image Super-Resolution via Repeated Refinement (SR3)について述べる．

2.2 DDPM

2.2.1 DDPM とは

DDPM は，画像生成タスクの分野で顕著な進展を遂げたニューラルネットワークモデルである [5]．DDPM のニューラルネットワークモデルのアーキテクチャ設計は，U-Net [6]モデルを参考にしている．このモデルの動作原理は主に二つの段階に分かれる．それらは拡散過程と逆拡散過程である．

拡散過程では，まずデータセットから画像x_0をサンプリングし，次にこの画像に T ステップのガウスノイズを追加し，一連の画像$x_1, x_2, \ldots, x_T$を生成する．十分な T ステップの拡散過程後，最終的に得られるx_Tはガウスノイズ画像となる．

逆拡散過程では，ガウスノイズ画像x_Tからスタートし，拡散ニューラルネットワークを通じて順序に逆拡散処理を行い，x_{T-1}，x_{T-2}，… を得て，最終的に元の画像x_0を復元する．このプロセスでは，拡散過程を逆転させて，徐々にクリアな画像を再構築する．

2.2.2 拡散過程

データセットから画像x_0をサンプリングする．次に，この画像に T ステップのガウスノイズを逐次的に追加する．この拡散過程はマルコフ連鎖の原則に従う．この処理を，式(2.1)に示す．

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \tag{2.1}$$

再パラメータ化トリックを用いて，x_0から直接 t ステップのx_tが計算できる．計算方法を式(2.2)に示す．β_tはステップ t 時点の時入れるノイズの分散であり，$\alpha_t = 1 - \beta_t$．$\overline{\alpha}_t = \prod_{i=1}^{T} \alpha_i$ である．

$$q(x_t|x_0) = \mathcal{N}\left(x_t; \sqrt{\overline{\alpha}_t}x_0, (1-\overline{\alpha}_t)I\right) \tag{2.2}$$

2.2.3 損失関数

再パラメータ化トリックを利用することで，完全なマルコフ連鎖を計算することなく，x_0から特定の時刻 t の画像x_tを直接得ることができる．ϵ_θは予測モデルである．画像x_tと時刻 t を入力することで，x_{t-1}からx_tへ加えられたノイズを予測する．訓練損失関数は以下式(2.3)に示す．

$$\mathcal{L}_{\text{simple}} = E_{t,q(\epsilon_0),\epsilon}\left[|\epsilon - \epsilon_\theta\left(\overline{\alpha}_t x_0 + \sqrt{1-\overline{\alpha}_t}, t\right)I^2\right], \quad t \sim \mathcal{U}(1,T), \quad \epsilon \sim \mathcal{N}(0,I) \tag{2.3}$$

2.3 DDIM

2.3.1 DDIM とは

DDIM は，DDPM に基づく新しいサンプリング技術である [4]．DDPM の訓練プロセスでは，主に周辺分布$q(x_t|x_0)$と再パラメータ化トリックが使用され，結合分布$q(x_{1:t}|x_0)$は関与しない．したがって，この方法で訓練されたモデルは，非マルコフ連鎖の実現と見なすことができる．この基礎の上に，respacing 技術を組み合わせることで，Diffusion Model の生成プロセスを加速することができる．この respacing 技術は，WaveGrad [7]の研究で効果的であることが証明されている．

2.3.2 サンプリング手法

DDIM は生成モデルであり，DDPM と同じ$q(x_t|x_0)$を保持したまま，DDPM と同じ訓練関数を使用する非マルコフ連鎖に基づく生成プロセスを持っている．これは，DDIM のサンプリング方法を既に訓練済みの DDPM モデルに直接適用できることを意味し，ニューラルネットワークに何の変更も加える必要がないことを示す．

DDIM が提案する新しいサンプリング方法を，式(2.4)に示す．式(2.4)において，新たなハイパーパラメタσ_tが導入されている．σ_tが 0 に等しい場合，サンプリング生成プロセスは決

定論的なプロセスとなる．この特性により，DDIM は生成品質を維持しつつ，サンプリング効率を高めることができる．

$$x_{t-1} = \sqrt{\overline{\alpha}_{t-1}} \frac{x_t - \sqrt{1-\overline{\alpha}_t}\epsilon_\theta^{(t)}(x_t)}{\sqrt{\overline{\alpha}_{t-1}}} + \sqrt{1-\overline{\alpha}_{t-1}-\sigma_t^2}\epsilon_\theta^{(t)}(x_t) + \sigma_t\epsilon_t \tag{2.4}$$

サンプリング生成プロセスを加速するために，DDIM は respacing 技術を採用している．この技術により，サンプリングプロセスは，通常大きな数である T ステップを生成に必要とせず，より短いサブシーケンスで実行することができる．この方法により，生成される画像に必要な時間を大幅に削減することができる．

2.4 SR3

SR3 は，Diffusion Model を使用して超解像を実現する技術であり，新しい条件付き画像生成方法である [3]．SR3 のネットワーク構造は U-Net に基づいている．

訓練過程では，SR3 は低解像度の画像を高解像度に補間し，それとノイズを加えた高解像度画像と連結してモデルの入力とする．モデルは，前の時刻に加えられたノイズを予測することで，超解像タスクを実現する．生成段階では，モデルにガウスノイズと補間された低解像度画像を連結して入力し，逐次的な逆拡散プロセスを通じて最終的に対応する高解像度画像を生成する．

2.5 むすび

本章では, Diffusion Model の原理と DDIM による加速サンプリング法及び, Diffusion Model に基づく超解像モデルである SR3 につい述べた．

第3章　予備実験

3.1 まえがき

本章では，SR3 を用いて超解像画像を生成する予備実験と DDIM 用いて SR3 の生成速度を加速する予備実験について述べる．

3.2 データセット

本研究で使用するデータセットは，Flickr-Faces-HQ（FFHQ）[8]であり，切り取られた多様性に富んでいる高品質な人物の顔の画像 70,000 枚を含んでいる． FFHQ データセットの中から 60,000 枚を訓練データとして使用し，残りの 10,000 枚はテストデータとして用いた．また，低解像度化処理を行い，16x16 の低解像度（LR）画像データセットと 128x128 の高解像度（HR）画像データセットを生成した．

3.3 評価指標

画像生成品質の評価は FID（Fréchet Inception Distance）を用いて評価する．FID は，実際の画像と生成された画像間の分布距離を測定する指標である．生成された画像の特徴ベクトルと実際の画像の特徴ベクトル間の距離を比較することで，生成モデルの生成品質を評価する．距離が小さいほど，生成された画像は実際に近く，多様性が豊かであることを意味する．

FID の計算は，Inception v3[10]モデルに依存している．このモデルを，入力画像から特徴ベクトルを抽出するために使用し，特徴ベクトルをモデルのエンコーディング層から得る．その後，実際の画像と生成された画像のこれらの特徴ベクトルを使用して，Fréchet 距離を計算し，二つの画像特徴分布の類似度を測定する．

3.4 SR3 の再現実験

パラメータ数が 97M である 16x16 から 128x128 への超解像度モデル SR3 を構築し，FFHQ データセットを使用して訓練を実行した．総ステップ数 T = 2000，バッチサイズ 32

で，500,000 イテレーションの訓練を行った．超解像画像の生成には，DDPM のサンプリング方法と DDIM のサンプリング方法を用いて生成した．DDPM サンプリング手法については総ステップ数 2000 である．DDIM については respacing 技術と組み合わせて異なるステップ数に設定して実行した．テストデータセットの評価には FID [9]を使用した．評価結果を表 3.1 に示す．また、DDIM サンプリング法を用いて，Step 数を 100 に設定し，20step ごと生成した画像の例を図 3.1 に示す．本研究で示される画像は，Creative Commons Attribution 2.0 Generic (CC BY 2.0) ライセンスに従う Flickr-Faces-HQ（FFHQ）データセットからのものである．

表 3.1　DDPM と DDIM による生成画像の FID 評価結果

サンプリング法	Step 数	FID
DDPM	2000	10.82
DDIM	100	14.62
DDIM	75	16.06
DDIM	50	19.26
DDIM	25	26.12
DDIM	20	28.25
DDIM	15	30.39
DDIM	10	33.42
DDIM	8	34.27

図 3.1 DDIM による SR3 の中間生成画像

3.5 考察

DDIM サンプリング法を用いて加速する過程で，ステップ数の増加に伴い，FID スコアが徐々に下がることが分かった．これは，生成される画像の品質が DDPM のサンプリングレベルに近づいていることを意味する．さらに，生成過程の可視化分析において，SR3 の中間生成される画像の多くが視覚的にほぼノイズであることが確認された．しかし，サンプリングプロセスの後半になると，画像の輪郭が明確になり，より多くのディーテル情報が現れ始めたことが分かった．

3.6 むすび

本章では，SR3 を用いて超解像画像を生成する実験と DDIM 用いて SR3 の生成速度を加速する実験と評価結果について述べた．

第4章　提案手法

4.1 まえがき

本章では，Diffusion Model を基にした超解像手法において，生成プロセスの前半段階での効率の低さという問題に対処するために，新たな加速手法を提案する．提案手法では，中間ステップx_tを予測することで，生成プロセスの効率を向上させることができると考えられる．

4.2 提案手法の概要

超解像タスクにおける Diffusion Model の効率を高めるため，新たな加速手法を提案する．中間ステップx_tを予測するための専用のニューラルネットワークを提案する．生成プロセスをガウスノイズから始めるのではなく，予測された中間ステップから始めることで，生成時間を大幅に削減できると考えられる．

x_t予測モデルは，二つの部分で構成される．第一部分は，MSE（平均二乗誤差）損失関数を用いて訓練された MSE SR Model である．第二部分は Weight Map Model であり，このモデルは MSE SR Model の出力と真実の画像との絶対差異を予測する役割を担っている．

生成プロセスでは，まず MSE SR Model を使用して超解像画像を出力する．超解像画像を Weight Map Model に入力し Weight Map を得る，次に，予測する時間点 t に基づいて，Weight Map と対応するガウスノイズの積を超解像画像に加え，その出力をx_tとして使用する．最後に，x_tを Diffusion Model に基づく超解像モデル SR3 に入力し，t 時間点から生成プロセスを開始する．この方法により，生成プロセスに必要な時間を大幅に削減できると考えられる．図 4.1 に本研究で提案する加速手法を示す．

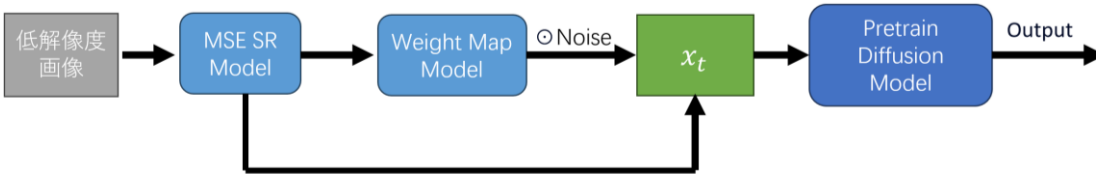

図 4.1　提案手法の概要図

4.3 MSE SR Model

MSE SR Model は，MSE Loss（平均二乗誤差損失）に基づいて訓練された超解像モデルである．モデルの構造は SR3 と同様であり，唯一の違いは入力チャネルが 3 個であることである．モデルは U-Net 構造を採用し，低解像度の画像を入力として受け取り，すべての超解像画像の可能な平均を予測するために使用する．

4.4 Weight Map Model

Weight Map Model は MSE SR Model の出力と実際の高解像度画像との差の絶対値を予測する．モデルは SR3 と構造が似ており，U-Net アーキテクチャを採用する．

Weight Map Model の訓練には，MSE SR Model の出力を入力として使用する．モデルは，MSE SR Model の出力と実際の高解像度画像との差の絶対値を学習する．訓練時では平均二乗誤差（MSE）損失関数が用いられる．この方法により，Weight Map Model は MSE SR Model の出力の精度と実際の画像との差異を効果的に学習し，予測することができると考えられる．

4.5 中間x_tの予測

中間ステップにおけるx_tを予測するために，以下のプロセスを定義する．まず，低解像度の画像を MSE SR Model によって出力された超解像画像をx_{sr}とする．Weight Map Model の出力をWとする．予測x_tの計算方法を(式 4.1)に示す．この式において，ハイパーパラメタ β を導入しており，βはWと同じ形状で要素値が同じ行列である．計算の安定性と有効性を確保するために，$(\beta + W)$に clamp 処理を行い，処理後の値が 1 を超えることも 0 未満になることもないように制限する．

$$x_t = \sqrt{\bar{\alpha}_t} x_{sr} + clamp(\beta + W)\sqrt{1 - \bar{\alpha}_t}\epsilon,\ \epsilon \sim \mathcal{N}(0, I), clamp(\beta + W) = \max(0, \min(1, \beta + W)) \tag{4.1}$$

4.6 むすび

本章では， 提案手法の概要について述べた．また，多くの Diffusion Model に基づく超解像モデルは，訓練時で再パラメータ化トリックを採用したため，マルコフ連鎖に厳密に従う必要はなくなる．したがって，MSE SR Model と Weight Map Model を使用して，中間ステップx_tを予測することで，中間ステップから生成プロセスを開始することができ，ガウスノイ

ズから始める必要はなくなると考えられる．この加速手法は，Diffusion Model に基づく超解像モデルの生成速度を効果的に加速し，全体の効率を向上させると考えられる．

第5章 実験と考察

5.1 まえがき

本章では，第 4 章で提案した手法の有効性を確認するための実験を行う．まず，FFHQ データセットを用いて MSE SR Model と Weight Map Model を訓練する．また，FFHQ データセットで訓練された SR3 超解像モデルも利用し，DDPM のサンプリング技術および DDIM 加速サンプリング技術を用いて，テストセットで高解像度の画像を生成する．加速手法のパフォーマンスを評価するために FID スコアを計算する．

5.2 Weight Map Model

Weight Map Model は SR3 と同じアーキテクチャを採用し，97M のパラメータを持つニューラルネットワーク構成とする．Weight Map Model は FFHQ データセットを使用して訓練する．さらに，より効果的に訓練するために，実際の高解像度画像と MSE SR Model の出力との差の絶対値を計算する際に，コントラスト強化技術を導入する．差異画像のコントラストを強化し，これら強化された差の絶対値を訓練の目標として，均方誤差（MSE）損失関数を用いて訓練を行う．図 5.1 に，Weight Map の可視化出力の例を示す．

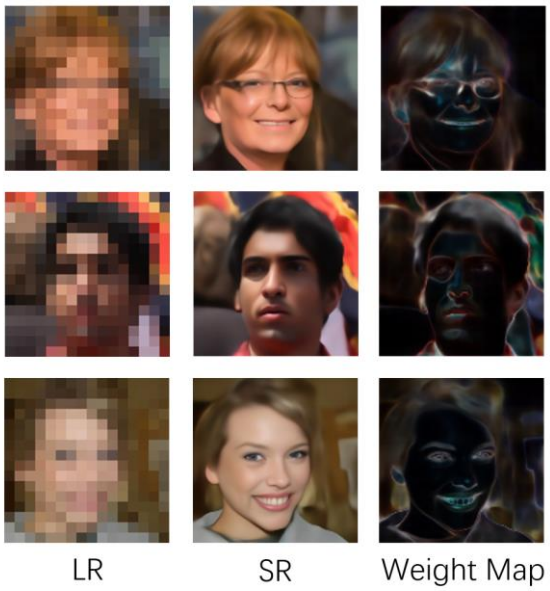

図 5.1　Weight Map の可視化，左から低解像度画像，MSE SR Model により生成された超解像画像，Weight Map.

5.3 DDPM サンプリング法を組み合わせた加速実験

5.3.1 実験内容

事前訓練された SR3 モデルは通常, 2000 ステップの自己回帰が必要である. 本実験では, x_t=300 の画像を直接予測し，この点から 300 ステップのイテレーション実験を行う．さらに，ハイパーパラメタβに関する実験を行い，異なるβ値が生成画像の品質と FID にどのような影響を与えるかを調べる．また，Weight Map を使用せず，MSE SR Model の出力に直接ノイズを加えたテストも行う．実験結果を表 5.1 に示す．また，異なるβ値による生成した画像の例を図 5.4 に示す.

図 5.2 は，Weight Map の有効性を確認する実験である。青線は Weight Map のβ値を 0.6 に設定した際に，異なるx_tから超解像された画像の FID を示す．オレンジ色の線は，Weight Map を使用せず MSE SR Model の出力に直接ノイズを加えた場合超解像された画像で計算された FID を示す.

図5.3 は、異なる予測されたx_tとステップ数に基づいて生成された超解像画像の例を示す。画像を 3 行に分けて示す。1 行目は、MSE SR Model と Weight Map を使用して計算された異なるx_tから生成された高解像度画像である。2 行目は、Weight Map を使用せず、MSE SR Model だけで得られたx_tに基づいて生成された高解像度画像の結果を示す。

表 5.1　評価結果(FID)

	β=0	β=0.2	β=0.4	β=0.6	β=0.8	β=1.0
w/ Weight Map	15.70	14.28	13.39	**13.22**	13.56	14.74
w/o Weight Map	17.20	17.14	16.80	16.38	15.64	14.74

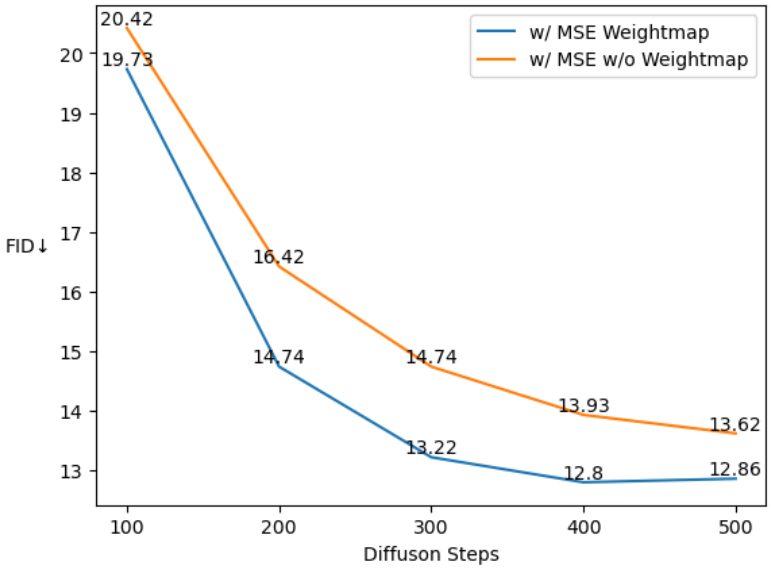

図 5.2　Weight Map の有効性

図 5.3 生成された超解像画像の例の比較

図 5.4 異なるβ値による生成した画像の例

5.3.2 考察

表 5.1 から、Weight Map を使用し，β値を 0.6 に設定した場合に FID スコアが 13.22 になることが分かった．また、Weight Map を使用しない場合、FID の最小値は、βが 1 の時、FID スコアが 14.74．実験結果から Weight Map を使用した場合生成された画像の品質がより高いことが確認できた，Weight Map を使用することで，より少ないノイズを加えながらも，より低い FID を実現し，生成品質を向上できることが分かった．

図 5.2 の Weight Map 有効性の結果から， 500 ステップ以内では，Weight Map を使用する場合の FID スコアが常に Weight Map を使用しない場合よりも低いことがわかる．Weight Map の有効性を確認できた．

5.4 DDIM サンプリング法を組み合わせた加速実験

5.4.1 実験内容

提案方法と DDIM を組み合わせた加速サンプリングの実験を行った.実験では，提案手法の中間x_t予測モデルを用いて特定の時刻x_tを直接予測し，x_tから DDIM サンプリング方法を適用して超解像プロセス加速し，生成された超解像画像の FID を計算する．本実験では，βを 0.8 に設定し，t=200 および t=300 の時刻でx_tを予測し，加速実験を行った．また，提案

手法の有効性を評価するために，ガウスノイズから始める DDIM 加速サンプリング方法と比較し, 両手法の FID スコアを計算した. 表 5.2 に実験結果を示す. 本実験において, Number of Function Evaluations (NFE)は、タスクを完了するために必要なニューラルネットワーク計算の総回数を指す．通常，NFE の数値が高いほど、必要とされる実行時間が長くなる．

図 5.5 と図 5.6 は，NFE が 8 の時，提案手法を用いて予測時点 t が 300 であるときに生成した超解像の一部の結果画像を示す。また，比較のために DDIM 生成方法による結果も含める．これらの図において，「LR」は低解像度の入力画像を，「HR」はデータセット内の対応する高解像度画像を指す．

NFE	8	10	15	20	25	50
DDIM	34.27	33.42	30.39	28.25	26.12	19.26
Ours(t=200)	18.1	21.92	20.61	21.59	17.76	18.84
Ours(t=300)	**16.27**	19.27	18.65	18.68	22.76	19.68

表 5.2　DDIM と提案手法の比較

図 5.5　提案手法と DDIM による生成した画像の例(1)

図 5.6　提案手法と DDIM による生成した画像の例(2)

5.4.2 考察

提案手法では，DDIM 加速サンプリング法を適用する際，　NFE が 8 で，ガウスノイズから始めて全ステップ（同じく NFE=8）を完了する DDIM と比較して，提案手法は FID スコア上で顕著な差を示し，明らかな優位性を示すことが分かった．

しかし，実験結果からはまた，NFE が増加するにつれて，提案手法で計算された FID には大きな変化がないこともわかる．NFE が 50 未満の場合，提案手法は常に DDIM よりも優れていた．NFE=8 かつ t=300 の条件下で，提案手法は最低の FID スコアを達成した．しかし，NFE が 8 を超えると，DDIM によるサンプリングの FID 値は徐々に安定して低下し，提案手法の FID スコアには顕著な変化が見られなかった．

図 5.5 と図 5.6 を観察すると，NFE が 8 の場合，提案手法で生成された画像は，DDIM と比較して，特に細部において顕著な改善が見られる．特に髪の毛や目などの部分では，DDIM で生成された画像が比較的ぼやけているのに対し，提案手法ではより多くの細部情報が表現されている．

全体として，DDIM 加速法と組み合わせた場合，NFE が非常に少ない条件では，提案手法は DDIM よりも大きな優位性を示す．NFE が 8 の場合，提案手法は DDIM サンプリング

NFE=50 の生成品質を上回ることが確認できた．しかし，NFE が増加するにつれて，提案手法は不安定性を示す．したがって，実験は，NFE が低い場合には提案手法が DDIM 加速サンプリング法に比べて優れた生成品質を提供することを示すが，計算時間とリソースが十分にある場合には，提案手法が DDIM より劣る可能性がある．

5.5 むすび

本章では，提案手法の有効性を確認するための評価実験の結果を示した．実験結果から，NFE が低い場合，提案手法は DDIM よりも低い FID スコアを達成し，生成された画像の品質がより高いことを示している．しかし，NFE が高い場合では，提案手法の DDIM に対する改善はそれほど顕著ではなく，場合によっては DDIM よりも性能が劣ることがある．

第6章 結論・今後の課題

6.1 結論

本研究は，Diffusion Model に基づく超解像手法に存在する生成効率の問題に対処し，改善を図った．現在の多くの Diffusion Model は再パラメータ化技術に基づいており，これにより生成プロセス中にマルコフ連鎖の原則を厳密に遵守する必要がなくなる．この点を踏まえ，本研究では生成速度を最適化するための方法を提案した．提案手法では，MSE SR Model と Weight Map Model を使用して中間時刻のx_tを予測し，完全なガウスノイズから始めてx_0まで逆拡散する必要性を回避している．さらに，提案手法は他の加速サンプリング手法とも組み合わせて使用することができる．DDPM サンプリング法および DDIM サンプリング法と組み合わせた実験では，提案手法の有効性が確認できた．特に DDIM サンプリングと組み合わせた場合，非常に低い NFE の条件で，提案手法は生成品質の測定である FID 値において DDIM を大きく上回った．しかし，NFE が高い条件では，提案手法は DDIM 加速サンプリングに比べてそれほど優れていないことがわかる．

6.2 今後の課題

Diffusion Model に基づく超解像技術は，訓練過程で大量のノイズを導入し，生成品質の向上を実現できた．しかし，生成効率が低いため，Diffusion Model の加速方法が重要な研究分野となっている．

本研究では，中間ステップx_tを直接予測し，DDIM などのサンプリング加速方法と組み合わせることで加速を試みた．非常に限られた NFE で，提案手法は DDIM よりも優れた生成品質を達成できることがわかった．しかし，NFE が増加するにつれ，提案手法の生成品質は不安定になり，高 NFE では DDIM に劣るようになった．これは，本研究で提案された方法において，予測されたx_tと実際に可能なx_tの分布との間にまだ一定の差が存在することを示しており，予測された中間x_tの精度をさらに向上させる余地があることが必要と考えられる．

謝辞

本研究の遂行にあたり，先導的な洞察力と研究指導を賜りました渡辺教授に心からの敬意と感謝を表します.

研究活動を通じて，研究室の皆様から得た多大なる助力と知恵は，私の研究進行において不可欠な要素でした．研究室の皆様方には，心から感謝の意を申し上げます.

最後に，常に心の支えとしており，日々の生活におけるご支援を賜ります家族に，心から深い感謝の意を表させていただきます.

参考文献

[1] C. Dong, C. Change Loy, K. He, X. Tang: "Image Super-Resolution Using Deep Convolutional Networks", arXiv preprint, arXiv:1501.00092v3, 2015.

[2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi: "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 4681-4690, 2017.

[3] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, M. Norouzi: "Image Super-Resolution via Iterative Refinement", arXiv preprint, arXiv:2104.07636, 2021.

[4] J. Song, C. Meng, S. Ermon: "Denoising Diffusion Implicit Models." , arXiv preprint, arXiv:2010.02502v4, 2020.

[5] J. Ho, A. Jain, P. Abbeel: "Denoising Diffusion Probabilistic Models", arXiv preprint, arXiv:2006.11239, 2020.

[6] O. Ronneberger, P. Fischer, T. Brox: "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv preprint, arXiv:1505.04597, 2015.

[7] N. Chen, Y. Zhang, H. Zen, R. J. Weiss, M. Norouzi, W. Chan: "WaveGrad: Estimating Gradients for Waveform Generation", arXiv preprint, arXiv:2009.00713, 2020.

[8] T. Karras, S. Laine, T. Aila: "A Style-Based Generator Architecture for Generative Adversarial Networks" ,Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 4217-4228, 2021.

[9] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter: "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium", arXiv preprint, arXiv:1706.08500, 2017.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna: "Rethinking the Inception Architecture for Computer Vision", Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2818-2826, 2016.

図一覧

表一覧