論文タイトル: Blue noise for diffusion models

著者: XINGCHANG HUANG, MPI Informatics, VIA Center, Germany

CORENTIN SALAÜN, MPI Informatics, Germany

CRISTINA VASCONCELOS, Google DeepMind, UK

CHRISTIAN THEOBALT, MPI Informatics, VIA Center, Germany

CENGIZ ÖZTIRELI, Google Research, University of Cambridge, UK

GURPRIT SINGH, MPI Informatics, VIA Center, Germany

1.はじめに

Sohl-Dicksteinら[2015]、Hoら[2020]、SongとErmon[2019]による画期的な研究以来、拡散モデルに関する研究が盛んに行われています。これらのモデルは、画像生成の質やトレーニングの安定性において、生成的対向ネットワーク（GAN）[Dhariwal and Nichol 2021]と比較して優れた性能を示しています。さらに、拡散モデルは、テキストから画像への合成、画像の補完、画像の超解像、画像編集など、さまざまなタスクを実行するように訓練できます。

通常、拡散モデルは前処理と後処理の2つのプロセスで構成されています。前処理では、元のデータポイント（例：画像）に徐々にノイズが加えられ、それをランダムなノイズパターンに変換します。後処理では、このノイズから元のデータを再構築することを学習するために、デノイジングニューラルネットワークが使用されます。デノイジングネットワークは、初期段階では低周波成分（粗い形状や構造）を再構築し、時間が経つにつれて詳細な高周波成分を徐々に洗練していきます。この動作は、拡散モデルが粗から細へとデータを生成し、周波数成分との隠れた関係を持っていることを示唆しています。

これらの進展にもかかわらず、この動作と前処理および後処理で使用されるノイズとの関係についての研究は限られています。既存の拡散モデルのほとんどは、ガウシアンノイズ（相関のないガウシアンノイズまたはガウシアンホワイトノイズ）にのみ依存しており、その周波数パワースペクトラムは全ての周波数にわたります（白色の色に似ています）。相関ノイズは、拡散モデルでは十分に検討されていないものの、関連する研究があります。例えば、Rissanenら[2023]は、熱の散逸に着想を得た拡散プロセスを提案し、周波数を明示的に制御しています。同様に、Voletiら[2022]は、スコアベースの拡散モデルで等方性ガウシアンノイズの代わりに非等方性ノイズを使用することを提案しています。これらの手法には可能性があるものの、生成される画像の品質に関する制限があり、主流のモデルでの採用は限られています。

本論文では、時間変化ノイズを用いた新しい拡散モデルを提案します。私たちの目標は、ブルーノイズ[Ulichney 1987]のような相関ノイズを活用して生成プロセスを強化することです。ブルーノイズは、そのパワースペクトラムが低周波領域にエネルギーを持たないことが特徴です。私たちの焦点は、ブルーノイズマスク[Ulichney 1999]にあり、これによりブルーノイズ特性を持つノイズプロファイルが提供されます。これらのブルーノイズマスクを使用して、拡散ベースの生成モデルのための時間変化ノイズを設計することを提案します。このような相関ノイズマスクを生成することは、数千から数百万のマスクをリアルタイムで生成する必要があるため、時間がかかるプロセスです。この問題に対処するために、私たちは低次元および高次元画像の両方に対して、リアルタイムでガウシアンブルーノイズマスクを生成する効率的な方法を提案します。まとめると、私たちの貢献は以下の通りです：

- 相関ノイズとトレーニング画像間の相関が生成拡散モデルに与える影響を調査するフレームワークを提案します。

- このフレームワークに基づいて、各ステップでモデルに導入される相関を制御できる、時間変化ノイズを用いた決定論的拡散プロセスを導入します。

- 相関ノイズマスクの生成における計算上の課題を克服するために、リアルタイムマスク生成アプローチを導入します。

- 提案する時間変化ノイズモデルを使用してガウシアンノイズとガウシアンブルーノイズを補間することにより、IADB[Heitzら 2023]やDDIM[Songら 2021a]のような既存の決定論的モデルよりも優れた性能を示します。

2.関連研究

### ブルーノイズ

ブルーノイズは、高周波信号と低周波の欠如を特徴とするノイズの一種です。コンピュータグラフィックスの分野で多くの応用が見られます。その一つの応用例として、Ulichney [1993] によって最初に導入されたブルーノイズマスクを用いた画像のディザリングがあり、これにより画像の知覚品質が向上します。ブルーノイズマスクは、Georgiev と Fajardo [2016]、および Heitz と Belcour [2019] によって示されたように、モンテカルロレンダリングにおける誤差の分布を改善するためにも使用されます。レンダリングにおけるブルーノイズとデノイジングの関係は、Chizhov ら [2022] と Salaün ら [2022] によってさらに探求され、ブルーノイズとローパスフィルターの組み合わせが知覚エラーを減少させることが明らかになっています。ブルーノイズマスクの優れたデノイジング特性を活用するために、我々はそれらを拡散ベースの生成モデリングのためにデータを汚染する加法ノイズとして使用することを提案します。

### 拡散モデル

拡散モデルによる画像生成には、確率的なもの [Ho ら 2020; Song と Ermon 2019; Song ら 2021b] と決定論的なもの [Song ら 2021a; Heitz ら 2023] を含むさまざまな定式化があります。拡散モデルは画像生成を超えて、動画生成 [Ho ら 2022] や3Dコンテンツ生成 [Poole ら 2023] にも拡張されています。より包括的なレビューは、Cao ら [2024] と Po ら [2023] の調査に見られます。拡散モデルは、トレーニングが遅く、生成プロセスも遅いことで知られています。生成プロセスを数ステップで画像を生成するようにスピードアップする方法は、ますます重要な研究テーマとなっています [Lu ら 2022; Liu ら 2023a,b; Salimans と Ho 2022; Karras ら 2022, 2023; Song ら 2023; Luo ら 2023]。推論ステップの数を減らすことと並行して、さまざまな種類のノイズ追加や画像汚染操作をサポートするより一般的なフレームワークを開発する研究もあります [Jolicoeur-Martineau ら 2023; Bansal ら 2024]。一部の研究は、生成プロセスを粗から細への方法でモデル化するために、画像コンテンツの周波数を明示的に考慮しています [Rissanen ら 2023; Phung ら 2023]。しかし、画像汚染に使用されるノイズの周波数がデノイジングプロセスにどのように影響するかについての研究は限られています。この問題を理解するために、我々は相関ノイズを利用してデノイジングプロセスを改善するフレームワークを提案します。

3.提案手法

拡散に基づく生成モデルは、前向きプロセスと後向きプロセスの2つの主要なプロセスで構成されます。前向きプロセスでは、ノイズ（𝝐）が初期画像x0を汚染するために導入され、この際のスケーリングは離散時間パラメータ𝑡によって決定されます。ここで、x0はトレーニングデータ分布𝑝0からサンプリングされた実画像を表します。時間ステップ𝑡は0から𝑇−1までの範囲で、𝑇は離散時間ステップの総数です。汚染された画像と対応する時間ステップ𝑡は、ニューラルネットワーク𝑓𝜽(x𝑡, 𝑡)をトレーニングするための入力として使用されます。

後向きプロセスでは、トレーニングされたネットワークを使用して純粋なノイズをデノイズし、新しい画像を生成します。図2はこのプロセスを示しています。ガウスノイズ（青い分布）から始まり、画像はネットワークを通過するたびに完全にデノイズされた画像（目標分布である赤い分布）に変わります。このプロセスの中間ステップでは、ノイズと画像の混合が見られます。図には3つの例が示されています。時間ステップが0に近づく（𝑡が小さくなる）につれて、画像の品質は向上し、より多くのディテールが現れます。この例では、中間のノイズがガウスノイズから時間変化スケジュールに従ってガウスブルーノイズに移行します（セクション3.2を参照）。

このセクションでは、2つの異なる軸にわたる相関について探ります。ノイズ内のピクセル間の相関と、ミニバッチ内の画像間の相関です。ノイズマスクと画像間の相関の影響を示すために、Heitz ら [2023] の IADB（時間変化ノイズを用いた決定論的拡散プロセス）に基づいて手法を構築しました。簡潔さと公平な一対一の比較のために、我々の手法はIADBの上に開発されており、新しい要素以外の特性とハイパーパラメータは保持されています。しかし、我々の手法は一般的であり、他の既存の生成拡散プロセスの上に展開することも可能です。

IADBの前向きおよび後向きプロセスと目的関数は以下のように定義されます。ここで、x0は目標画像、𝝐はランダムなガウスノイズ（𝝐 ∼ N(0, 𝑰)）、𝛼𝑡と𝛼𝑡−1は2つのブレンド係数です。ネットワークモデルは𝑓𝜽と呼ばれ、2つの入力パラメータ（汚染された画像x𝑡と時間ステップ𝑡）を取ります。IADBの確率的な定式化も存在しますが、本研究ではその安定性のために決定論的バリアントに焦点を当てます。

3.1 相関ノイズ

決定論的拡散プロセスでは、ノイズマスクが後向きプロセスの初期化および各トレーニングステップでターゲット画像を汚染するために使用されます。トレーニング中のマスク生成は重要な要素であり、特定の要件を満たす必要があります。マスク生成プロセスは、各イテレーションで異なるマスクを生成するために確率的でなければならず、これにより過剰適合を減らし、生成結果の多様性を増加させます。また、マスク生成は各トレーニングステップで使用されるため、計算が高速でなければなりません。

ガウスノイズはこれらの要件を自然に満たしますが、すべての相関マスク手法が同様に適しているわけではありません。特に、IADBはゼロ平均と単位共分散行列を持つ多変量ガウス分布から生成されたマスクを使用します。ブルーノイズのような相関ノイズを生成する方法では、非単位共分散行列が必要です。ブルーノイズマスクの共分散行列は、複数のマスクから推定することができます。Ulichney [1993] の目的関数を使用したシミュレーテッドアニーリングを用いて、一万個のブルーノイズマスクを生成しました。この方法は高品質なマスクを生成しますが、最適化には多くの時間が必要です。その後、例示されたマスクのそれぞれの相関行列を平均することで、ブルーノイズの相関行列Σを計算できます。

指定された共分散行列Σを持つノイズマスクを作成するには、Σに対してコレスキー分解を適用し、下三角行列𝐿（𝐿𝐿𝑇 = Σ）を得ます。最後に、ランダムベクターに𝐿を乗じて、目的のノイズマスクを効率的に生成します。

b=𝐿𝝐

ここで、𝝐 ∼ N(0, 𝐼) は単位分散のガウス分布です。図3は、式(4)を使用して生成されたガウスブルーノイズマスクの一例を示しています。𝐿の各行または列はノイズマスクのピクセルインデックスを表します。𝐿の各セルはノイズマスク内のピクセル間の相関強度を表します。正の値は正の相関を表す明るいセルに対応し、負の値は負の相関を表す暗いセルに対応します。各ピクセルに対して、隣接するピクセルのみがゼロから大きく離れた値を持ち、その他の非隣接ピクセルはゼロに近い値を持つことが、白と黒の線で示されています。本論文では、このノイズマスクをガウスブルーノイズと呼びますが、Ahmed ら [2022] のガウスブルーノイズ手法とは異なります。

高次元の場合、行列とベクトルの乗算は計算コストが高くなります。特に行列 L が高次元である場合、直接行列のサイズを増やして高次元のノイズを生成することは、PyTorchなどの現代の機械学習フレームワークを使用して（相関のない）ガウスノイズを生成するよりも遅くなります（Paszke et al., 2017）。ノイズ生成はトレーニングの各ステップで使用されるため、オーバーヘッドは最小限に抑える必要があります。このため、我々は高次元のノイズマスクを生成するために、KolligとKeller（2002）の手法を適応しました。この手法では、低次元のマスクセットにパディングを適用します。

具体的には、高次元のガウス青色ノイズマスクのバッチを生成するために、解像度 64×64 で式（4）を用いてを使用して大きなバッチを生成します。次に、これらの 64^2 個のマスクをパディングして、より高次元のガウス青色ノイズマスクを得ます。このようにして、解像度のガウス青色ノイズを生成するための計算オーバーヘッドは無視できる程度です（約 0.0002 秒）。図1は、パディングによって生成された解像度 \( 128 \times 128 \) のガウス青色ノイズの例を示しています。パディングによって高次元のマスクにはタイル間の継ぎ目が生じることがありますが、このアーティファクトは実際にはほとんど目に見えず、手法の低オーバーヘッドによって補償されています。我々はさまざまな解像度でマスクを提供し、補足文書セクション3でガウス青色ノイズの対応する周波数パワースペクトルを示すことで、パディング手法を使用して異なる解像度で青色ノイズの特性が保存されていることを示しています。

3.2時間変化ノイズを使う拡散モデル

単一の行列では1つの相関しか生成できません。拡散モデルでは、各時間ステップで導入される相関の量を制御する必要があります。時間変動する L は、2つの相関タイプをエンコードする固定された2つの行列から計算できます：

ここで、とはそれぞれ異なる行列を表し、はブレンディング係数です。これに基づき、前向きのプロセスは次のように定義されます：

この前向きプロセスにより、時間ステップt に基づいてガウスノイズとガウス青色ノイズが補間されます。一般的に、このモデルはに基づいて2種類のノイズをスムーズに補間することができます。

図4は、ガウスノイズからガウス青色ノイズへの線形補間の例を示しています。離散フーリエ変換によって計算された対応する周波数パワースペクトルは、低周波領域のエネルギーが左から右に減少していることを示しています。

次に、前向きプロセスを逆に定義して、逆向きのプロセスを定義することができます。L と前向きプロセスの定義に基づいて、逆のステップを次のように導出できます：

ここで、はガウス（白色）ノイズを表す単位行列であり、は式（4）で定義された行列です。の場合、我々のモデルはIADBに退化します。の場合、時間変動するノイズを持つより一般的なモデルを得ます。

IADBでは、彼らの場合にはが単位行列であるため、ネットワークは学習するのはのみです。ここでは、式（7）の両方の項を学習するようにネットワークを設計できます。これを達成するための一つの方法は、2つのニューラルネットワークを使用することですが、これはIADBよりもはるかに多くの計算を導入するため実用的ではありません。私たちは、との2つの項をそれぞれ3チャンネルの画像として表す6チャンネルの画像を出力することを選択しました。したがって、目的のネットワーク出力はおよび となります。したがって、損失関数は次のようになります：

省略

われわれのモデルは時間変動するノイズで訓練されていますが、逆向きのプロセスでは依然として確定的です。逆向きのプロセスは初期のガウスノイズから始まり、中間の時間ステップでは追加のノイズは必要ありません。代わりに、ネットワークは時間変動するノイズをガイドするように学習します。

前向き、逆向き、およびノイズ生成の手順は、Algorithm 1からAlgorithm 3にまとめられています。Algorithm 2では、Heitzら[2023]に従って線形関数（𝛼𝑡 = 𝑡/𝑇 ）として get\_alpha（𝛼スケジューラ）を考慮していますが、非線形関数でも構いません。

次に、式（9）で一般的なシグモイドベースの関数 get\_gamma を定義します。式（8）の重み付け項（𝛾𝑡 − 𝛾𝑡−1）/（𝛼𝑡 − 𝛼𝑡−1）は、𝛼スケジューラと𝛾スケジューラの違いを自動的に考慮します。𝛾𝑡 – 𝛾(𝑡−1) が小さい場合、𝑓′′𝜽(x𝑡, t) の寄与が減少します。これは式（7）で述べられた逆向きプロセスと一致し、𝛾𝑡 − 𝛾𝑡−1 が小さいときには 𝑓′′𝜽(x𝑡, t) が重要ではなくなります。

ノイズスケジューラ。Chen [2023]の研究に触発され、特に画像の解像度が高い場合には、スケジューラの選択が重要な影響を与えます。我々は、2つのノイズ間の補間を制御するために、𝛾スケジューラをシグモイドベースの関数としてパラメータ化します。具体的には、Chen [2023]に従って、𝑠𝑡𝑎𝑟𝑡、𝑒𝑛𝑑、𝜏の3つのパラメータでパラメータ化されます：

省略

時間ステップ t によって、これらのパラメータは調整されます。これらのパラメータを事前にどのように設定するかはわかりませんが、ネットワークパラメータに加えて、最適化を考慮します。ここで、𝑠𝑡𝑎𝑟𝑡∈[−3,0)𝑠𝑡𝑎𝑟𝑡 \in [-3, 0)start∈[−3,0)、𝑒𝑛𝑑∈(0,3]𝑒𝑛𝑑 \in (0, 3]end∈(0,3]、𝜏∈[0.01,1000.0]𝜏 \in [0.01, 1000.0]𝜏∈[0.01,1000.0] の範囲で最適化します。初期の実験では、𝑠𝑡𝑎𝑟𝑡 と 𝑒𝑛𝑑 が安定して0と3に収束し、𝜏 が画像の解像度に応じて0.2に収束するか、増加し続けることがわかりました。また、これらの3つのパラメータを最適化することで収束に追加のエポックが必要であり、エポックを通じて変化するため、ネットワークのトレーニングを困難にします。

3つのパラメータの選択を実用的にするために、次のように決めました： start=0、end=3、そして画像の解像度に基づいて 𝜏 を設定します： 𝜏=0.2 は 128×128 画像用、𝜏=1000 は 64×64 画像用です。異なる 𝜏 値を持つ 𝛾-スケジューラの曲線は、付録文書の挿入画像に示されています。

我々は、すべての実験で使用する3つのパラメータの値を付録文書セクション2にまとめます。

3.3 リクティファイド・マッピングを用いたデータサンプル間の相関

前述の段落では、ピクセル間の相関ノイズを使用して拡散プロセスを向上させる方法が示されています。相関は、単一のミニバッチ内でノイズとターゲット画像のマッピングを改善するためにも使用できます。

Rectified flow [Liu et al., 2023a] や Instaflow [Liu et al., 2023b] にインスパイアされ、相関はノイズ画像ペアを修正するために利用されます。図5は、トレーニングイテレーション中の単一のミニバッチにおけるペアデータサンプル x0​（赤い分布）とノイズ b（青い分布）を視覚化し、我々のリクティファイド・マッピングを示しています。以前の研究（図5(a)）では、x0​ と b の間にランダムなマッピングが適用されていました。ノイズデータのマッピングは、これらを前向きプロセスに供給する前にインコンテキストの層別化を適用することで改善できます（図5(b)）。このリクティファイド・マッピングにより、各ノイズとそのターゲット画像との距離が縮小され、より直接的な軌道が生まれます。このマッピングを見つけるために、L2ノルムを使用して個々のピクセルレベルでノイズと画像間の二乗距離を計算します。その後、各 b について、まだ使用されていない最も距離が短い x0​ が選択されます。この改善されたマッピングにより、特定の画像がトレーニングプロセス中に一貫して同じタイプのノイズに関連付けられることが保証され、時間ステップ間での滑らかな勾配フローが実現されます。

4.実験

4.1 実装の詳細

我々は、無条件/条件付き画像生成のために、異なる解像度の CelebA [Lee et al. 2020]、AFHQ-Cat [Choi et al. 2020]、LSUN-Bedroom [Yu et al. 2015] データセットを使用します。実験のセットアップの詳細については、付録文書セクション2を参照してください。我々のフレームワークは、PyTorch [Paszke et al. 2017]で実装されており、Song et al. [2021a]、Heitz et al. [2023] の公式実装に基づいています。我々は diffusers ライブラリ [von Platen et al. 2022] で実装された2D U-Net [Ronneberger et al. 2015] を使用します。ネットワークアーキテクチャやトレーニングの詳細、式（9）の 𝜏 の値などの詳細については、付録文書セクション2を参照してください。拡散モデルのハイパーパラメータに関しては、訓練には 𝑇 = 1000 を使用し、テストには 𝑇 = 250 を使用します。ネットワークパラメータの最適化には、AdamW オプティマイザー [Loshchilov and Hutter 2017] を学習率 0.0001 で使用します。すべてのデータセットでのトレーニングおよびテストには、NVIDIA Quadro RTX 8000 (48 GB) GPU を4台使用します。 評価には、すべてのモデルの生成品質を測定するために、FID [Heusel et al. 2017]、Precision および Recall [Kynkäänniemi et al. 2019] を使用します。これらのメトリクスは、Inception-v3 ネットワーク [Szegedy et al. 2016] をバックボーンとして、Stein et al. [2024] の実装を使用して計算されます。FID、Precision、およびRecall を計算するために、30,000 枚の画像を生成します。

4.2 画像生成

我々は、無条件の画像生成において、既存の決定輸送拡散モデル DDPM [Ho et al. 2020]、DDIM [Song et al. 2021a]、および IADB [Heitz et al. 2023] と我々の方法を比較します。公平な比較を確保するために、全ての方法で生成過程全体で同一の初期ガウスノイズを使用します。なお、DDIM は diffusers ライブラリ [von Platen et al. 2022] を使用してトレーニングされ、IADB および我々の方法と同じトレーニングセットアップが適用されます。さらに、同じ時間ステップ数を使用して完全性を持たせるために、確率的拡散モデルである DDPM [Ho et al. 2020] と IHDM [Rissanen et al. 2023] とも比較します。我々の方法は、これら2つの方法に対して一貫した改善を示します。

図11には、同じ解像度（64^2）での AFHQ-Cat、LSUN-Church、Celeba の結果が示されています。我々の方法は、約 𝑡 = 75)から青色雑音効果を示し、視覚的に他の方法と区別されます。時間ステップ 𝑡 = 0 における生成された画像に関して、我々の方法は建物の柱周りの歪みが少なく、窓やドア周りのより詳細な内容を持つ画像を生成します。視覚的比較に加えて、表1に示されている定量的評価は、64^2の解像度のデータセットにおいても、我々の方法が IADB および DDIM よりも一貫した改善を示しています。

より高解像度の結果では、図12に示すように、約 𝑡 = 75 から生成される内容の差異を観察します。逆方向のプロセスの終わり近く、約 𝑡 = 25 で、我々は 𝜏 = 0.2 (式（9）) を使用することで青色雑音効果の出現を始めることに注目します。詳細をより詳しく見るには、拡大してください。また、補足的な HTML ビューアを提供し、逆方向のプロセス中のさまざまな時間ステップで生成された中間画像をインタラクティブに視覚化できるようにしています。リアリズムの観点から、我々の生成された画像は、特にテーブル1で示されるような領域で改善された品質を示しています。テーブル1では、異なるデータセット間での IHDM [Rissanen et al. 2023]、DDPM [Ho et al. 2020]、DDIM [Song et al. 2021a]、IADB [Heitz et al. 2023]、および我々の方法に対する定量的な FID スコアの比較が示されています。特筆すべきは、我々のアプローチがすべての評価されたデータセットで IADB よりも改善を示している点です。我々の方法が DDIM には1つのデータセットで劣っている一方、同じデータセットで IADB も低い性能を示していることに注意すべきです。追加のメトリクスについては、付録文書セクション3で提供しています。

図12に示すように、IADBと比較して、我々の方法は髪、口、目の領域で改善された視覚的品質を実現しています。DDIMと比較すると、我々の方法は類似した視覚品質を達成しています。表1に示されているCelebA（1282）データセットの定量的結果は、DDIMがIADBおよび我々の方法を上回っていることを示しています。この結果は、DDIMが式（1）の中で異なる表現を使用していることに起因しており、これはHeitz et al. [2023]で議論されています。データセットによっては、DDIMが異なる𝛼𝑡の選択によりIADBを上回ることがあります。ただし、これはIADBまたは我々の方法のいずれにとっても制限ではありません。

また、我々のフレームワークではトレーニング中に画像間での修正されたマッピングも考慮できます。表2では、AFHQ-Cat（642）でテストされた拡散ステップ数に応じたデータ相関の有無によるFIDスコアを提供しています。ミニバッチ内の修正されたマッピングは、ステップ数が少ない場合に低いFIDを達成しますが、ステップ数を増やすとやや高いFIDとなります。

我々は、付録文書セクション3でガウス青色雑音生成と逆方向プロセスの詳細なタイミング、そして最近傍テストを含む追加の結果を提供しています。これにより、我々の方法が訓練データに過剰適合していないことを確認しています。

**他の拡散モデルへの拡張** 我々の方法は、付録文書セクション1、3での導出と予備的な結果を通じて、DDIMにも拡張できます。また、我々の方法は高解像度画像生成のためにLDM [Rombach et al. 2022]に組み込むことができます。図9に示すように、我々の方法はAFHQ-Cat（512^2）においてIADBと比較してよりリアルな目を生成し、より良いFID（11.45 < 12.19）を示します。ただし、他のモデルに基づく新しいフレームワークの開発には追加の努力が必要であり、これは将来の作業に先送りされます。

**4.3 条件付き画像生成**

ノイズからの無条件生成に加えて、我々のモデルは画像の条件付き生成（例：画像超解像）にも対応しており、単に条件となる低解像度画像をノイズ画像に連結して入力します。

図6は、LSUN-Churchデータセットにおける画像超解像のIADBと我々の方法との比較を示しています。解像度が32^2から128^2に増加する過程で、我々の方法はSSIM [Wang et al. 2004]、PSNR、平均二乗誤差（MSE）による定量的評価でIADBを上回ります。我々の方法は、MSEが低いことからも参照画像に対する忠実度でIADBを上回ります。視覚的には、IADBは特に最初の画像の下部に過剰な詳細を導入する傾向があります。一方、我々の方法は画像全体で直線を効果的に保持します。画像超解像のすべての定量的結果は、付録文書セクション3で見ることができ、我々の方法が一貫してIADBを上回っていることを示しています。

4.4 要因分析と解析

高周波特性によってガウス青色ノイズが機能することを確認するために、低周波ノイズであるガウス赤色ノイズにガウス青色ノイズを置き換えます。ガウス赤色ノイズは図10で視覚化されています。赤色ノイズは、最小化ではなく最大化することによって同じ方法[Ulichney 1993]で生成されます。その後、対応する共分散行列と下三角行列を計算し、我々のフレームワークのためにガウス赤色ノイズを生成することができます。図7に示されるように、我々のフレームワークでガウス赤色ノイズを使用すると、その低周波特性のために細部を回復することができませんでした。表3によると、ガウス赤色ノイズをガウス青色ノイズで置き換えると、Precisionが著しく低下しますが、Recallは比較可能です。これは図7での視覚的観察と一致しています。Precisionは生成された画像のリアリズムを主に測定するため、この結果は一貫しています。

もう一つの選択肢はガウス青色ノイズのみを使用することですが、これは図1（2行目）で示されています。最終的に生成された画像はIADBおよびOursと比較して少し現実的でないです。視覚的品質も表3に示される定量的指標と一致しています。ただし、ガウス青色ノイズのみを使用する場合、画像の内容が早期のステップでより速く、クリアに表示されることが観察されます（図1参照）。追加実験である早期停止を行いました。これは、早期ステップで停止した場合、ガウス青色ノイズのみを使用すると、ガウスノイズのみを使用するよりも良い結果が得られることを示しています。補足文書の図3に示されるように、早期停止したガウス青色ノイズのみを使用した結果（2行目）は、より鮮明な詳細を示しています。早期停止の定量的評価は補足文書のセクション3で見ることができますが、青色ノイズは低周波領域にエネルギーがないため、拡散プロセスが限られた方向の範囲内に制限されます。このため、後の時間ステップで中間生成コンテンツを洗練するのが難しく、結果的に品質が悪化することがあります（表3参照）。それに対して、我々の方法は、低周波成分が既に見えており、ネットワークが高周波の詳細を洗練することに焦点を合わせる中間または後の時間ステップから青色ノイズを考慮に入れます。

**異なるノイズの大きさでの拡散**

すべての時間ステップでガウス青色ノイズを使用すると品質が低下するため、初期の時間ステップを明示的に無視して後の時間ステップでの拡散を分析します。特定のノイズの大きさ（例えば30％）まで拡散モデル（IADB）を訓練し、ガウスノイズとガウス青色ノイズを比較します。テストフェーズでは、真の画像が提供されるため、ノイズ除去後の画像と真の画像を比較することができます。ノイズの100％を使用すると、実験は標準的な拡散生成プロセスに戻ります。図8に基づくと、ガウス青色ノイズを使用することで、ガウスノイズを使用した場合と比べて、さまざまなノイズの大きさでより詳細で内容を保持した画像を生成できます。これは、低周波成分が見えるようになるとガウス青色ノイズがノイズ除去に適していることを示しています。これは、ガウス青色ノイズを中間または後の時間ステップから混合するという我々のアイデアと一致しています。

**さらに多くのアブレーション**

異なる𝛾値と𝛾スケジューラに使用されるコサインベースのスケジューラ[Nichol and Dhariwal 2021]も比較します。また、パディング/タイルに使用される異なるガウス青色ノイズマスクのサイズも比較します。詳細は補足文書のセクション3に記載されています。

5.結論

私たちは、決定論的生成拡散モデルに相関ノイズを組み込む新しい方法を提案しました。この技術は、行列ベースの方法で生成された非相関ノイズと相関ノイズマスクの組み合わせを使用することに基づいています。異なるノイズ相関を調査することで、ノイズの特性と生成画像の品質との間の複雑な関係を明らかにしました。我々の発見は、高周波ノイズが詳細を保持するのに効果的であるが、低周波成分の生成に苦労する一方で、低周波ノイズは複雑な詳細の生成を妨げることを示しています。最適な画像品質を達成するために、各ノイズ成分の強みを活かして、時間に依存して異なる種類のノイズを選択的に使用することを提案します。提案手法の有効性を検証するために、IADB [Heitz et al. 2023] という有名な手法と組み合わせた広範な実験を行いました。訓練データと最適化ハイパーパラメータを一貫して維持することで、様々なデータセットにおいて画像品質の大幅な向上を一貫して観察しました。これらの結果は、決定論的拡散モデルの画像生成能力を向上させる上で、我々のアプローチの優位性を示しています。

### 制約事項

1. \*\*γスケジューラのパラメータ調整\*\*: 現在、γスケジューラのパラメータ調整は画像解像度に依存しており、異なる解像度に対する追加の複雑性を導入しています。

2. \*\*計算の負荷\*\*: ガウシアンブルーノイズマスクの計算は特に高解像度モデルへの拡張時に計算負荷が高くなります。

3. \*\*GPUメモリの制約\*\*: 特定のミニバッチのすべてのデータは、修正マッピングが機能するために単一のGPUに配置される必要があり、方法のスケーラビリティが制限されます。

4. \*\*分散トレーニング\*\*: 現在のアプローチでは、インターGPU同期を含む分散トレーニングをサポートしていないため、大規模なデータセットを扱う際の効率とスケーラビリティが制限されます。

### 今後の課題

1. \*\*ノイズパターンの設計\*\*: 提案されたモデルは、生成拡散モデルの効率を向上させるためのノイズパターンの設計において新たな研究方向を示唆するでしょう。

2. \*\*複数のノイズの補間\*\*: 興味深い今後の課題として、我々のモデルを複数の種類のノイズを補間するように拡張することが挙げられます。これには、低パスおよびバンドパスノイズなどの異なるタイプのノイズを組み込むことが含まれ、トレーニングとサンプリングの効率を向上させるための自由度が増します。

3. \*\*高度なデータ相関技術\*\*: トレーニング中のデータサンプルを相関させるための高度な技術を設計するためのさらなる研究が必要です。これは相関ノイズの使用とは異なるアプローチです。

4. \*\*確率モデルへの拡張\*\*: 我々のフレームワーク、特に時間変化ノイズモデルを、Hoら（2020）やSongら（2021b）が提案した確率モデル、およびKarrasら（2022）、Songら（2023）、Luoら（2023）のような少ないステップモデルに拡張することがもう一つの有望な方向です。これにより、最先端のデノイジング拡散モデルに対する我々のフレームワークの一般化が可能になります。

5. \*\*2D画像生成以外のアプリケーション\*\*: 我々のモデルは2Dの無条件および条件付き画像生成でテストされましたが、今後の課題としては、我々のモデルをビデオや3Dメッシュなどの他のデータ表現を合成するために一般化することが考えられます。これにより、提案されたアプローチの適用範囲と影響が広がります。