

SuperGaussian: 3D 超解像のためのビデオ モデルの再利用

ユアン・シェン・ニ ^{1,2}、ドウィグ・セイラン・シェン ²、ポール・ゲレロ ²、ゼシアン・シュウ・ア ²、
ロイ・J・ミトラ ^{2,3}、ロン・ワン ¹、ンナ・フリーシュテュック ²
¹ イリノイ大学アーバナシャンペーン校
² アドビリサーチ
³ ユニバーシティ・カレッジ・ロンドン

概要 :本稿では、粗い3Dモデルに幾何学のおよび外観の詳細を追加することでアップサンプリングする、シンプルでモジュール化された汎用的な手法を提示する。
生成3Dモデルは既に存在しますが、画像や動画の領域における同等の品質にはまだ及ばない状況です。本稿では、既存の（学習済み）動画モデルを3D超解像用に直接再利用することで、高品質3D学習モデルの大規模リポジトリ不足という問題を回避できることを示します。3D一貫性のない動画アップサンプリングモデルを再利用し、3D統合と組み合わせることで3D一貫性のある結果を生成する方法について説明します。出力として、オブジェクト中心で効果的な高品質ガウスプラットモデルを生成します。本手法はカテゴリに依存せず、既存の3Dワークフローに簡単に組み込むことができます。提案するSuperGaussianを、複雑さと表現の両面で多様な3D入力（ガウスプラットやNeRFなど）に対して評価し、本手法が既存の生成3Dモデルの忠実度を大幅に向上させることを実証します。

詳細については、プロジェクトの Web サイト(supergaussian.github.io) をご覧ください。

キーワード: 3D超解像度 · ビデオアップサンプリング · カテゴリ非依存
· 3D生成 · ガウスブラッキング · 3D一貫性

1 はじめに

生成型3Dモデルは現実のものとなりました。テキストプロンプトや画像を用いて条件付けされた3Dモデルを生成するための複数の手法 [2-46] が開発されています。これらの手法は、画像と3Dデータの組み合わせを教師として用いており、高速で多様な結果を生成します。

残念ながら、現在の手法で生成された3Dモデルは、画像[26,40]や動画[6]を対象とした最先端の生成モデルが達成するレベルの詳細さと精度を依然として欠いています。この限界には、複数の課題が寄与しています。

まず、3D表現の選択です。グリッドベースのモデルは、（生成される）形状に関する事前知識を必要としないため最も人気がありますが、その規則的な構造（例えば、ボリュームグリッド[25, 34]、三面体[2, 5])は、忠実度に限界を設けます。

[?] このプロジェクトは、Yuan が Adobe Research でインターンシップをしていたときに行われました。

406.00609v4

2 Y. Shen 他



図1: 既存のビデオを再利用する新しい手法であるSuperGaussianを紹介します
 3D超解像タスクのためのアップサンプリング手法。SuperGaussianはNeRF、ガウスブラット、ノイズの多い画像から得られた再構成画像など、様々な入力タイプを扱うことができます。
 スキャン画像、最近のテキストから3Dへの手法[31]によって生成されたモデル、または低ポリゴンメッシュ（例：SuperGaussianは、Sim-on-Wheels[45]で使用されるアセットです。SuperGaussianは、ガウスブラットの形式で、豊富な幾何学およびテクスチャの詳細を備えた高解像度の3D出力を生成します。

生成結果の。第二に、高品質でありながら大量の
 多様な3Dデータの処理は依然として困難です。最先端の画像や動画は
 モデルは数十億のトレーニングサンプルでトレーニングされ、最も広範な3D
 トレーニング データセットには、せいぜい数百万個のオブジェクトしか含まれません。

生成される3Dオブジェクトの忠実度を高める方法を探ります。
 一般的な粗い3D表現ではなく、私たちの目標は（粗い）3Dを「アップサンプリング」することです。
 カテゴリー固有のトレーニングを必要としない入力モデル。私たちの主な観察は
 3D表現は複数の視点からレンダリングできるということです
 滑らかな軌跡を描き、中間的な汎用ビデオ表現にマッピングする。これにより、既存のビデオモデル
[\[9, 69\]](#)を再利用して、
 3Dアップサンプリングや超解像タスクなど。このようなモデルは大規模な
 ビデオデータセットを学習することで、一般的なシナリオに適用できる強力な事前分布が得られます。
 重要な課題は、3Dの一貫性を確保することです。
 ビデオモデルは時間的に滑らかですが、3Dの一貫性が保証されるわけではありません。画像ベースの
 モデルとは異なり、ビデオを使用するという私たちのアプローチは、
 事前分布と各フレームを独立して扱うことで、（初期）
 時間を超えた一貫性。

私たちはこの課題を、シンプルでモジュール化された汎用的なアプローチで解決します。
 既存のワークフローに統合できます。手順は以下のとおりです。まず、
 サンプリングされた視点の軌跡が与えられた粗い 3D 入力からのシーンのビデオ。
 第二に、入力のドメイン固有のアーティファクトを処理するためにオプションで微調整された、事前学習済みのビデオ
 オベースのアップサンプラーを使用してレンダリングされたビデオをアップサンプリングする。
 モダリティ。3D統合では、出力表現としてガウスブラッティング[\[27\]](#)を採用する。ガウスブラッ
 ティングはオブジェクト中心の表現であるため、
 個々のオブジェクトをエンコードするのに最適で、ローカルの
 細部まで表現できる。ガウスブラットは、シンプルさと忠実度のバランスも優れている。
 エンコードされたモデルの精度とレンダリングの効率性を示した。図1は、提案された

ビデオベースの 3D 超解像度フレームワークは、さまざまな粗い 3D モダリティのセットをアップスケールするためにうまく適応できます。

さまざまな低解像度アセットでアルゴリズムを評価し、生成された高解像度ガウスプラットの品質と多様性。本アルゴリズムを他のアルゴリズムと比較し、アブレーション研究を報告する。
私たちのデザインの選択。要約すると、私たちの主な貢献は次のとおりです。

- (i)事前学習済みのビデオアップサンプリングモデルを3D超解像度用に再利用する。
- (ii)一般的な3Dデータを使用してアーティファクトを処理する前に、ビデオのアップサンプリングを微調整する低解像度のガウスプラットの特徴であり、
- (iii) 効率的でクラスに依存しないモジュール型3DモデルとしてSuperGaussianを提案する多様な3Dシーンをアップサンプリングする超解像処理
ガウス プラット、NeRF などの表現。

2 関連研究

画像の超解像。画像の詳細と解像度を向上させることは、コンピュータビジョンにおける長年の課題です。これは本質的に不適正な問題であり、
生成的手法は、失われた幻覚に対処するのに適している。
簡潔にするために、最近の生成的アプローチに焦点を当てる。
読者にはこの分野に関する包括的なレビュー[11]を参照していただきたい。

初期のCNNベースのアプローチ[14, 15, 28, 66]では、超解像のための深層ネットワーク。画像合成モデルによって学習された強力な事前分布を利用して画像の詳細を推測する手法がいくつかある。生成的敵対的ネットワーク[17]は、このタスクに広く利用されている。

最適化ベースの技術[36]または条件付きフィードフォワードネットワーク設計[3, 24, 29, 53–55]。他のアプローチとしては、正規化フロー[35]やトランスフォーマーベースの方法[12, 13]が検討されている。非常に人気のあるノイズ除去拡散確率法は、
モデル[22, 41]も最近では画像のアップサンプリングに利用されている[30, 42, 63]。
ビデオ超解像。学習ベースの単一画像超解像の成功に続き、
超解像法では、この成功を拡張するための様々な取り組みが行われてきた。
ビデオ領域への拡張を実現するために、いくつかの戦略が検討されてきた。時間軸を跨いで情報を集約するために時間集約モジュールを導入する研究[33]がある一方で、ユニフィケーションとロバスト性を備えたリカレントニューラルネットワーク[8, 60]は、
双方向の情報共有もアーキテクチャの選択肢の一つとなっている。変形可能な畳み込み[48, 52]は、フレームごとの特徴を画像と音声の両方に整列させるために利用されている。
ビデオのアップサンプリングと復元タスクに使用されます。最近では、拡散ベースのビデオアップサンブラーは、ベース拡散に時間層を注入することによって提案されている。
モデル[69]を用いている。本研究では、最先端の生成ビデオアップサンブラーであるVideoGigaGAN [57]を活用している。しかし、本手法はビデオモデルに依存しない。
使用されており、ビデオアップサンプリングの進歩により将来的にさらに改善される可能性があります。
拡散ベースのビデオアップサンブラーなどの技術。
画像ベースの3D超解像。最近いくつかのアプローチが、
暗黙の放射輝度場、特に神経放射輝度場 (NeRF)の解像度向上に取り組んでいます。これは、より高い詳細度と品質が求められるキャプチャと合成の両方のシナリオにおいて、重要かつ実用的なアプリケーションです。

低解像度の放射輝度フィールドを使用して、各反復で、この一連の研究[19,23,49]では現在の表現から画像をレンダリングします。次に、事前学習済みの画像アップサンプラーを使用してレンダリングをアップサンプリングします。次に、現在のレンダリングとアップサンプリングされた画像の差から得られる勾配フローを使用して、3D表現を最適化します。同様に、CROP [61]も単一画像アップサンプラーを活用しますが、マルチビューの一貫性を改善するために不確実性ボリュームを導入します。ただし、これらの方法とは異なり、単一画像アップサンプリングネットワークの代わりにビデオアップサンプラーを採用します。時間的一貫性のより適切な処理のおかげで、ビデオアップサンプラーを使用すると最終結果の視覚的な品質が向上することを示しています。さらに、NeRF 表現のアップサンプリングに主に焦点を当てた以前の研究とは異なり、SuperGaussian は汎用的なアプローチであり、さまざまな 3D 入力タイプでそれを実証します。

3D対応GANの文脈では、三面体に基づく3D表現と2D画像のアップサンプリングモジュールを統合学習する研究がいくつか提案されている[7, 16]。アップサンプリングモジュールは、生成器によって学習された特定の三面体特徴にのみ適用可能であり、各ビューが独立してアップサンプリングされるため、最終出力は完全に3D的に整合しているわけではない。

フィードフォワード3D超解像。いくつかの3D生成および再構築法では、直接的な3D超解像モジュールが提案されています。Smithら[47]は、3D形状をマルチビュー正投影深度マップとして表現し、それをアップサンプリングします。最近の拡散ベースの3D生成法であるLASDiffusion [68]は、占有マップから符号付き距離フィールドを予測して3D形状の解像度を向上させる方法を提示しています。ただし、両方の方法はジオメトリの改良のみに焦点を当てているのに対し、私たちの研究は外観とジオメトリのアップサンプリングを共同で実行しています。Rodin [51]は三面ベースの拡散法であり、三面のアップサンプリング戦略を提示しています。このアプローチは、学習した特定の三面の特徴にのみ適用できます。最後に、私たちのアプローチは大規模な3Dデータセットを必要とせず、代わりにビデオデータでトレーニングされた既存のビデオモデルを活用します。

深層暗黙表現。近年、NeRF [38]、VolSDF [59]、NeuS [50]などのニューラル暗黙表現が、3D形状の効率的かつコンパクトな表現として普及している。しかし、低次元のボリューム構造と計算集約型のボリュームレンダリングプロセスのため、高解像度の詳細を効率的にキャプチャすることは、このような表現では依然として課題となっている。最近、ガウスブラッティング[27]が、高速レンダリング機能を備えたオブジェクト中心の表現を提供する代替手段として提案された。シーンは、微分可能レンダリングを通じてプロパティが最適化された多数の3Dガウスプロブで表現される。本研究では、効率的な最適化とレンダリング機能を備えたガウスブラッティングを最終的な3D表現出力として採用している。

3 私たちのアプローチ

3.1 概要

粗い3D表現を与えられた場合、私たちの目標は超解像処理を実行して3D表現の忠実度を向上させ、より多くの局所的な詳細を捉えることです。

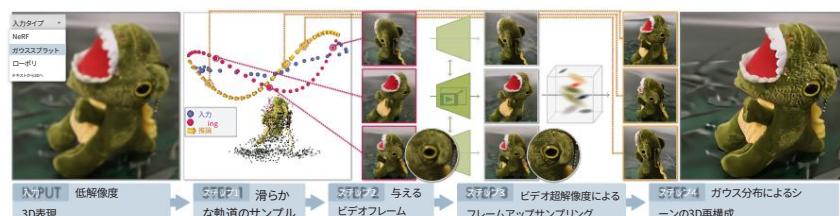


図2: SuperGaussianパイプライン。入力の低解像度3D表現が与えられ、様々な形式で提供できるため、まず滑らかなカメラの軌跡をサンプリングし、低解像度の中間ビデオ。次に、既存のビデオを使用してこのビデオをアップサンプリングします。アップサンプラーを使用して、より鮮明で高解像度の3D表現を実現します。より鮮明なディテール。私たちの手法であるSuperGaussianは、最終的な3D表現を生成します。高解像度のガウスプラットの形式で。

3Dコンテンツは、複数の視点から3Dシーンを作成する場合、私たちの作業の前提は、3Dアップサンプリングのタスクにおける既存のビデオアップサンプリングの事前分布。図示のように図2に示すように、我々の手法SuperGaussianは主に2つのステップから構成されています。まず、粗い3D表現からレンダリングされたビデオをアップサンプリングして解像度を高め、鮮明な結果を得ます。次に、3D再構成を行い、一貫性のある3D表現を生成します。

大量のビデオデータで訓練されたビデオアップサンプラーを使用して、ドメイン固有の低解像度ビデオ（低解像度の3Dからレンダリングされたビデオ）に対して微調整を実行します。そのため、SuperGaussianは、様々な3Dキャプチャおよび生成プロセスによって引き起こされる複雑な劣化を処理できます。

私たちのフレームワークのコンポーネントは高度にモジュール化されており、簡単に置き換えることができます。その他の最先端のビデオ手法と組み合わせて使用します。

次に、このセクションを、入力と出力を定式化して展開します。

方法（第3.2節）、フレームワークにおける重要なビデオアップサンプリング（第3.3節）を紹介し、3D最適化の詳細について説明します（第3.4節）。

3.2 問題の定式化

私たちのフレームワークは、静的な3D表現の多様なセットを扱うことができます。

シーンをlowと表記する。例えば、lowはガウスプラットなどである。

NeRF、低ポリゴンメッシュ、低品質のキャプチャビデオ、または生成された3Dオブジェクト

最近のテキストから3Dへの手法[31]によって、このような3D表現は、低い

複数の視点からレンダリングすることで、一般的な中間表現である動画が得られる。実際には、各3D入力をN個の滑らかな軌跡からレンダリングする。

視点 $1 \dots N$ $1 \dots T$ RGB画像のシーケンス $1 \dots N$ が生成されます $1 \dots T$ 。ここで、下付き文字は各軌跡上の視点または姿勢のインデックスであり、上付き文字は軌跡IDを表します。カメラの動きは

隣接するフレームは十分に小さいため、標準的なビデオアップサンプラーで十分な時間的アライメントを活用します。ビデオのアップサンプリングシーケンスを実行した後、出力として、高精細な3D表現を生成します。

ガウスプラットの形式（カメラのビューはセットアップで既知であり、推定する必要はありません）。この最終的な3D最適化により、真の3D

6 Y. Shen 他

出力し、その過程で、洗練されたビデオ表現に残っている時間的な矛盾を除去します。

3.3 初期アップサンプリング

まず、空の領域内のターゲットシーンの近くの軌跡を手動でサンプリングします。

各ビデオのカメラパスを表す軌跡 $I_{1...T} \{R, W, H, 3\}$ が与えられると、ビデオアップサンプラーは r アップサンプリング（実験では $r=4$ ）の軌跡を出力する。数学的には、

$$I_{1...T} = f(I_{1...T}) \otimes R \otimes W \otimes H \otimes 3, \quad (1)$$

ここで、 f はビデオアップサンプラー、 $I_{1...T}$ はアップサンプリングされたビデオを表します。初期レンダリング解像度は、入力3D表現の粗いレベルによってレンダリング忠実度がボトルネックにならない程度に十分に高いものと仮定します。

私たちのフレームワークは、最先端の事前学習済みビデオアップサンプラーを簡単に統合できます。今回のケースでは、生成ビデオアップサンプラーである VideoGigaGAN [57] を使用します。ベースラインと公平な比較を行うためにビデオアップサンプラーをどのように設定したかについては、セクション 4.1 を参照してください。重大なドメインバイアスのある入力表現を処理する場合（図6を参照）、追加の微調整が必要です。たとえば、ズームイン後に縞模様またはブロブ状のアーティファクトがある場合、ガウスブラットからのレンダリングは、最先端のビデオアップサンプラーで展開される標準的な拡張とは異なる劣化に従います。したがって、ビデオアップサンプラーを微調整するには、モデル化したい特定の劣化を示す低解像度と高解像度のビデオのペアが必要です。この目的のために、さまざまな3Dオブジェクトとシーンを示すマルチビューデータセット MVImgNet [62] を使用します。まず、データセットのオリジナル画像を8分の1に、つまり 64×64 ピクセルの解像度に双線形ダウンサンプリングして、低解像度画像のセットを取得する。次に、セクション 3.4 でより詳細に説明されているように、これらの画像に低解像度のガウスブラット [27] を適合させる。データセットによって提供されたオリジナルのカメラ軌跡で最適化された低解像度ガウス分布を、ビデオアップサンプラーへの入力としてレンダリングする。ターゲットの真実として、データセットのオリジナルビデオを使用し、入力の解像度の4倍にリサイズする。外れ値ピクセルに対する堅牢性のためにシャルボニエ回帰損失 [4]、知覚レベルの改善のためにLPIPS損失 [64]、生成動作のためにGAN損失 [18] を使用して、モデルを微調整する。

3.4 ガウスブラットによる3D最適化

公式のガウスブラットコードベース4を用いて、アップサンプリングされた動画にガウス分布を当てはめる3D最適化を実行しました。特に、本実験では、標準的なガウスブラット最適化プロセスに従い、2,000ステップで実行しました。

我々はカメラ情報を完璧に取得しているので、SfMを用いて推定するのではなく、最適化に直接入力することに注意してください。損失関数については、原論文 [27] の慣例に従い、L1損失とLSSIM損失を使用します。

⁴ <https://github.com/graphdeco-inria/ガウスブラット>

ガウス分布の最適化。3Dガウス分布スプラッティングを採用する利点は、オブジェクト中心の表現であることと、学習およびレンダリングにおける効率性にあります。さらに、アップサンプリングされたフレームにおいて、視点依存効果を良好に捉えることができます。しかし、SuperGaussianは、Blender合成データセット[38]での実験で使用したNeural Radiance Fieldsなど、他の種類の3D表現とも容易に統合できます。

4つの実験

4.1 セットアップ

データセット :MV-ImageNetデータセット[62]を用いて、本手法のビデオアップサンプラーコンポーネントを微調整する。このデータセットは、多数の高品質で実物体中心のビデオから構成される。このデータセットは、前節で説明したように、 64×64 ピクセルのポーズ画像から再構成された低解像度のガウススプラットに適合するように処理されている。合計20万シーンが学習に、5千シーンが検証に使用されている。

定量評価には、別途500シーンのデータセットを用いた。MVImgNetを用いて、高解像度画像を中央切り取り、 256×256 ピクセルの正方形画像にリサイズすることで、真の画像を得た。本手法の一般化可能性を示すため、合成3DオブジェクトからなるBlender合成データセットでの結果も示す。評価には、先行研究[32, 49]と同様に、元のデータ分割を用いた。

メトリクス :MVImgNetデータセットでは、定量的な比較を行う際に、知覚品質に焦点を当てたメトリクス、すなわちLPIPS [65]、NIQE [39]、インセプションスコア[43]、およびFID [21]を使用しています。FIDの計算には、MVImgNetが提供する実画像を実分布として使用しています。NeRF Syntheticデータセットでは、FastSR-NeRF [32]およびNeRF-SR [49]と同様に、LPIPS、PSNR、およびSSIMという同じメトリクスを報告しています。定量的なメトリクスとは別に、読者の皆様には、補足資料に掲載されている定性的な結果と動画をぜひご確認いただきたいと思います。

ベースライン: 最近、ドメイン固有の 3D アップサンプリング ソリューションがいくつかありますが[23, 49, 58]、私たちは一般的な 3D 表現に適用できるベースラインとの比較に焦点を当てています。一般的な 3D アップサンプリングに関するこれまでの研究は、アップサンプリングの事前分布を利用して選択された 3D 表現を反復的に最適化するという同様のパラダイムに従っています。事前分布としては、私たちが提案するビデオ ベースのアップサンプラーではなく、画像ベースのアップサンプリング モデルが一般的に選択されます。ビデオ ベースのアップサンプラーは、アップサンプリングされた品質に大きな影響を与える、一見些細な設計選択の 1 つです。これにより、最先端の手法を 3 つに絞込みます。(1) Instruct-G2G (3D 形式として Gaussian Splats を使用したInstruct-NeRF2NeRF [20])、(2) 3D 形式として Gaussian Splats を使用した Super-NeRF [19]、(3) FastSR-NeRF [67]。完全性を期すため、ドメイン固有のベースラインとしてNeRF-SR [49] を1つ含める。さらに、後述するように、カスタマイズされたベースラインであるPre-hoc imageも導入する。

公平な比較のため、すべてのベースラインに最先端の画像アップサンプラーであるGigaGAN [24]を使用する。特にGigaGANのチェックポイントでは、

表1: ビデオおよび画像アップサンプラー事前分布のベンチマーク。
アップサンプリングされたフレームの事前ホック画像と、適用後の我々の画像のパフォーマンス
対応するアップサンプラー。注: これらの結果に対して再構成は実行されません。

モデル	LPIPS # NIQE # FID #	は "
画像アップサンプラー	0.1533 7.06 14.67	11.67 ± 1.14
デオアップサンプラー	0.1511 6.04	14.61 12.45 ± 1.19

我々は、元の論文で提案されたのと同じ戦略に従って、LAION [44]上で4の超解像モデルを訓練し、64 から64に画像をアップサンプリングすることを学習する。256 256pxで非常に詳細な画像が得られます。SuperGaussianではVideoGigaGAN [57]を使用しています。ビデオアップサンプラーとして、時間的特徴抽出のための追加のBasicVSR++レイヤーを除いて、以前のGigaGANアーキテクチャを再利用しています。低解像度のビデオ入力フレームを処理するプロパゲーション[9]。このモデルはビデオ事前分布のサイズは画像事前分布よりもわずかに小さいことがわかった特徴の寸法を調整した後、画像とMVMgNetのビデオアップサンプラーを収束まで実行する。両方の事前分布のベンチマーク表1に報告されている。どちらの事前分布もアップサンプリング後に生成的な挙動を示している。また、ビデオの事前分布は、アップサンプリングされたフレーム間で優れた時間的一貫性を示しています。SuperGaussianは、例えば、より強いビデオ事前分布と簡単に統合できることに注意してください。アップスケールビデオ[69]またはOpenAIのSora [1]による低解像度ビデオの微調整条件として、Instruct-G2Gでは、まずウォームアップのために2kステップを最適化します。低解像度の表現を準備し、さらに2kを最適化します20ステップごとにアップサンプリングしながらステップアップする。SuperNeRFでは、著者はコードベースを公開していなかったため、私たちは彼らのメソッドを独自に実装しました。我々はGigaGANを一貫性の強制で含むモジュール[3]。同様に、低解像度表現のために2kステップを最適化し、さらに2kステップ最適化する。CCLC埋め込みの学習率を5 105 とすると、ぼやけた結果を避けるためには小さくすることが重要になります。実験では、画像を用いた反復最適化を実行すると、アップサンプラー、すなわちSuperNeRF [19]とInstruct-G2Gは、よりぼやけた画像を生成することが多い。事前にすべてのフレームを個別にアップサンプリングした場合と比較した結果最適化を行う。そこで、別のベースラインとなる事前画像を定義する。この画像では、すべてのフレームを個別に一度アップサンプリングし、次に 3D 表現を最適化します。最後に、実験結果における表現の影響を除去するために、MVMgNetでの比較のために、ベースラインをすべて3Dガウスブラッティングでグローバル表現するように変更しました。Blender合成データセットでは、ニューラルネットワーク放射輝度フィールドは、失敗する方法を横断するグローバル表現として選択される。NeRF に表示される他のベースラインと比較します。

実装の詳細: 微調整中、すべてのモデル パラメータは学習率 5 105 で微調整されます。LPIPS、シャルボニエ回帰損失、敵対的損失の重みはそれぞれ 15、10、0.05 になります。

GigaGAN [24]に従って、識別器にR1正則化[37]を適用する。重み0.02048、間隔16。バッチサイズは64で、ビデオフレームは12です。それぞれに80GBのVRAMを搭載した64個のA100 GPUでモデルをトレーニングしました。

表2: MVI_{img}Netによる低解像度画像のアップサンプリングの定量的比較
ガウスブラッティング。知覚的特徴を用いて、ベースラインとの比較を行った。
メトリクス。私たちの手法は汎用的であるにもかかわらず、一貫して最高の定量的結果を生み出します。Instruct-G2G
は、3D Gaussian SplatsをベースとしたInstruct-N2Nの略です。
3D表現。読者の皆様には、当社のウェブサイトで視覚的な結果をご確認いただくことをお勧めします。
これは、私たちの方法の視覚品質が基準を上回っていることを強調しています。

方法	LPIPS # NIQE # FID # IS "			
インストラクトG2G [20]	0.1867	スーパー	8.33	32.56 10.52 ± 1.06
NeRF [19]	0.2204		8.84	37.54 10.40 ± 1.03
事前画像 私たちの	0.1524		7.65	27.04 11.27 ± 0.99
	0.1290	6.80 24.32	11.69 ± 1.08	

表3 :Blender合成データセットを用いて低解像度のポーズ付きRGB画像をアップサンプリングした場合の定量的な比較。ここでは、200 200から 4にアップサンプリングした場合を比較している。
800 800pxまで。公式テストセットのベースラインとの比較では、
先行研究で報告された指標と比較すると、本手法は同等の定量的な結果をもたらす。さらに、本手法は、参考文献では捉えられていない、より生成的な詳細をもたらす。
公平な比較のために、神経放射場、すなわちTensorRF [10]を次のように使用する。
3D表現です。その他のベースライン結果は、彼らの論文から直接引用したものです。

方法	LPIPS # PSNR " SSIM "		
高速SR-NeRF [32]	0.075	30.47	0.944
NeRF-SR [49]	0.076	28.46	0.921
私たちの	0.067	28.44	0.923

30フレームのセットの場合、ビデオのアップサンプリングには 1秒かかります
64 64入力。高解像度ガウススプラット最適化には約30分かかります。
解像度256 256の画像をフィットさせるのに2kステップで数秒かかります。平均すると、私たちのパイプラインは
完了まで141.33秒かかり、ベースラインと比較して最も効率的です。
コードベースは効率性をさらに高めるために最適化できる
I/O とパイプラインの改善の余地。

4.2 比較研究

実験プロトコル :MVI_{img}Netでは、定量評価のために、
低解像度の3D表現では、滑らかなカメラ軌道を用意し、
提供された元のカメラポーズから乱されたカメラポーズのセットを通じて
対応するデータセットによって、得られたカメラ軌跡からレンダリングしたビデオをビデオアップサン
プラーの入力として用いる。姿勢摂動を行うために、
シーン間の線分に沿ってカメラの中心をランダムにサンプリングする
中心と各ポーズを元のカメラ軌道に当てはめる。Bスプライン曲線を当てはめる。
サンプリングされたカメラの中心にカメラの位置を再サンプリングし、結果として得られる
最後に、上方向に沿って正弦波状の摂動を加えます。
カメラの位置が決まったら、各カメラの向きを定義して
視線ベクトルをシーン中心に向ける。結果として得られるカメラ軌道は、微調整に使用した元のカメラ
軌道とは異なる。
MVI_{img}Netデータセットのビデオアップサンプラー。高解像度の3D画像を取得したら、
表現を定量的に評価するために、同じ摂動戦略を使用して、従来のものとは異なる新しい評価軌道をサン
プリングします。