

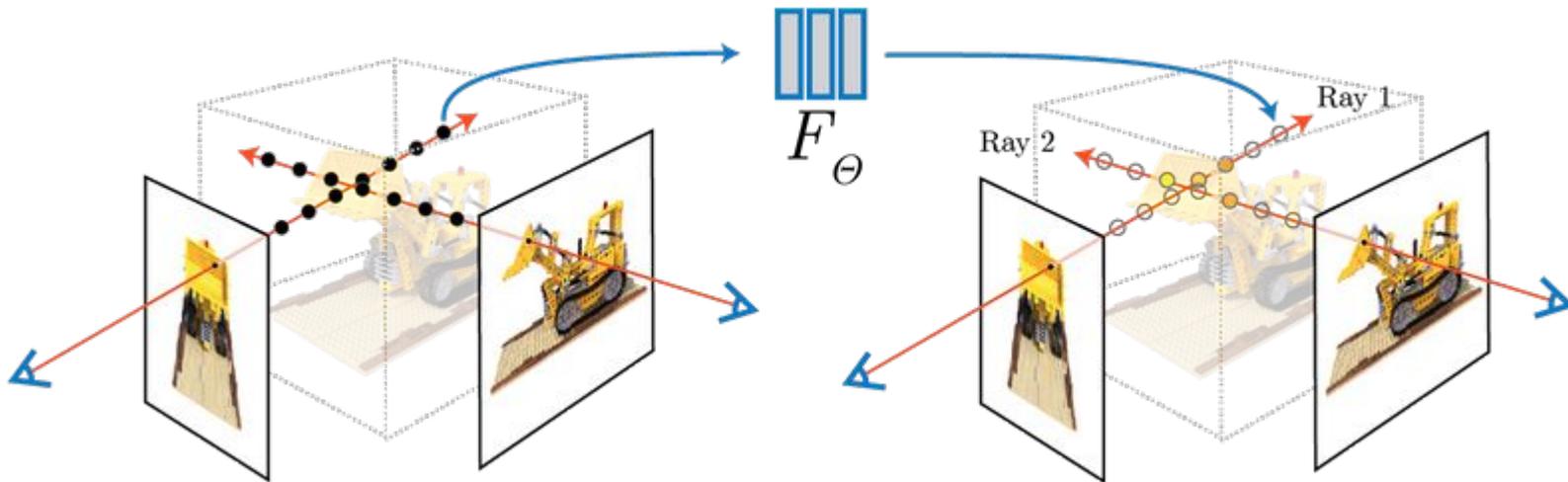
# neuralfieldsメタサーベイ

---

<http://xpaperchallenge.org/cv>

# Neural Fieldsとは

- Neural Networkにより、(主に)3次元の場を再構成する方法論に関する小研究分野
- 代表例: NeRF (Neural Radiance Fields)

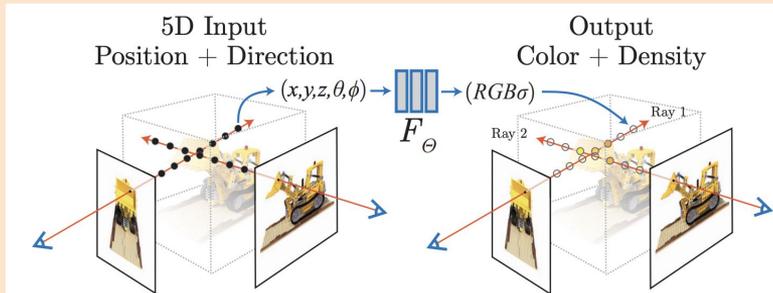


# NeRF

密度及び色を記述する **neural field** と微分可能な **volume rendering** の組み合わせ

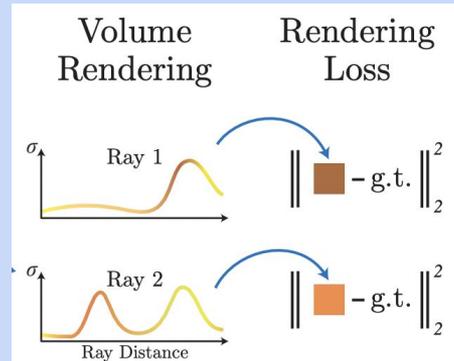
## neural field

- 3次元座標や方向を入力に密度や色を出力する関数をMLPなどのネットワークで回帰する表現方法
- coordinate network、Implicit Neural Representationと呼ばれることも



## volume rendering

- 密度と色情報をレイに沿ってサンプリングし、レンダリング画像を生成する
- 微分可能な方程式によって、3次元形状表現とRGB画像を接続し、測光誤差 (Photometric Error) による最適化が可能になる



# Diffusionモデルの使用

# Diffusionモデルの使用

neural fieldが撮影条件の差や観測の不安定さによらず利用可能な性質から、条件付き拡散モデルと併用し、text-to-3Dタスクの出力として利用する試みが流行している



An astronaut is riding a horse.

# text-to-3D

text-to-3Dの品質は半年で急速に発展している

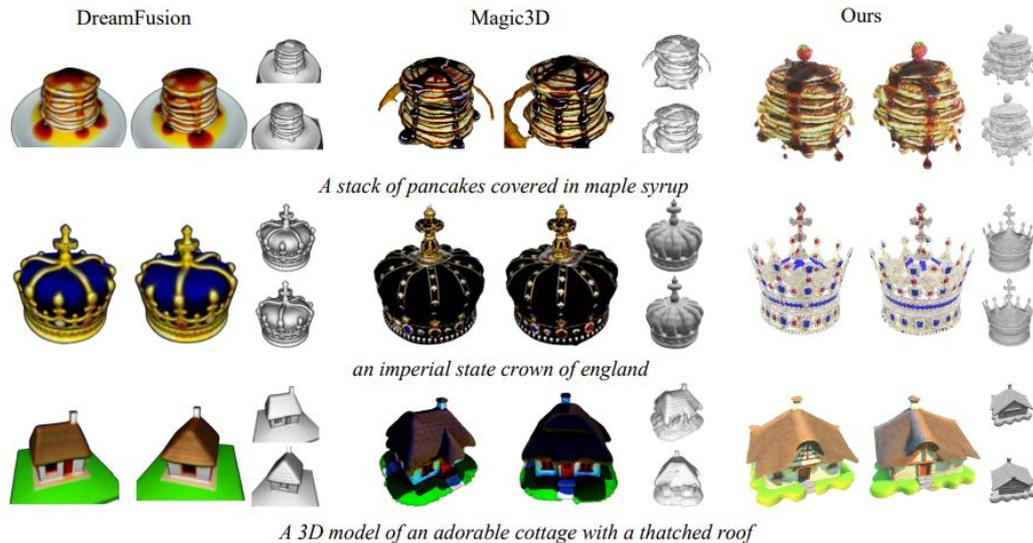
[DreamFusion\(2022/9/29\)](#)

[Magic3D\(2022/11/18\)](#)

[Fantasia3D\(2023/3/24\)](#)

HashEncodingの導入や、  
BRDFのモデル化など

NeRFの高品質化のアプローチが素早く導入されて進展している



# Diffusionモデルの使用例

---

CVPR2023では、テキスト以外にも、画像を指示にした条件付き拡散モデルの利用が提案されている

DietNeRF同様に、未観測視点を補う利用法が主流

[DiffusioNeRF](#): シーンの形状・外観の事前分布として拡散モデルを使用. 視点数が不足している場合の品質改善

[NeRD<sub>i</sub>](#): 単一視点画像からNeRFを復元

# DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models

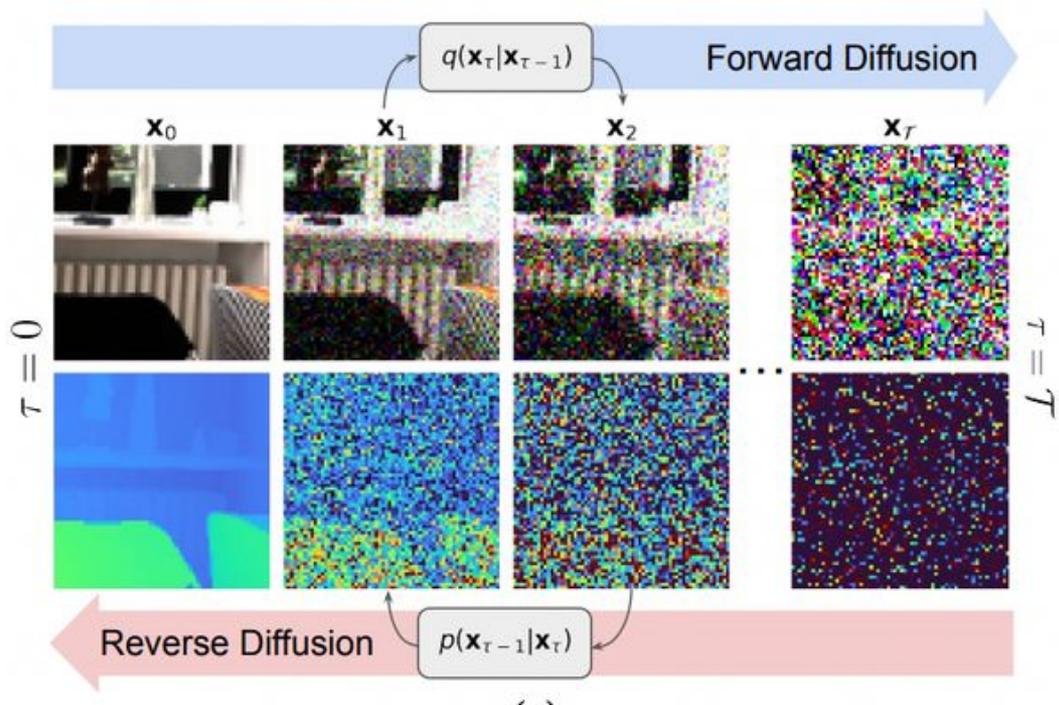
Jamie Wynn, Daniyar Turmukhambetov

## 概要

ノイズ除去拡散モデル (DDM) を用いて、シーンのジオメトリとカラーに関する事前知識を学習し、視点数が不足していても高品質な NeRF モデルを獲得

## 新規性

シーンのジオメトリと色の結合確率分布を学習するアプローチをとっている点がユニーク



# NeRDi: Single-View NeRF Synthesis with Language-Guided Diffusion as General Image Priors

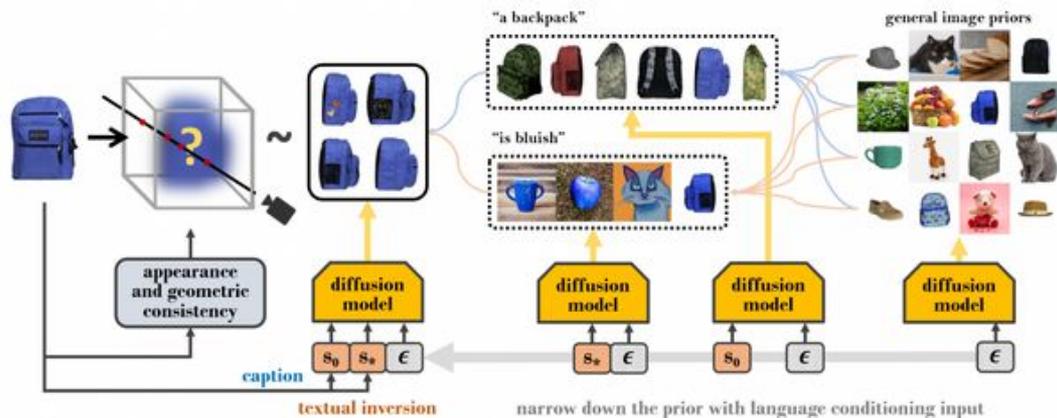
Congyue Deng, Chiyu Max' Jiang, Charles R. Qi, Xinchen Yan, Yin Zhou, Leonidas Guibas, Dragomir Anguelov

## 概要

2次元拡散モデルから得られる一般的な画像事前知識を用い、単一視点画像から NeRF を復元する。vision-languageモデルで得られた特徴を拡散モデルの条件として取り込むことで、入力画像の意味的・視覚的特徴による画像の事前分布が設定可能

## 新規性

2D画像生成をベースとする DreamFusion と異なり、NeRF の潜在空間で拡散モデルを扱う。また GAUDI と異なり、屋外シーンでの利用が可能である

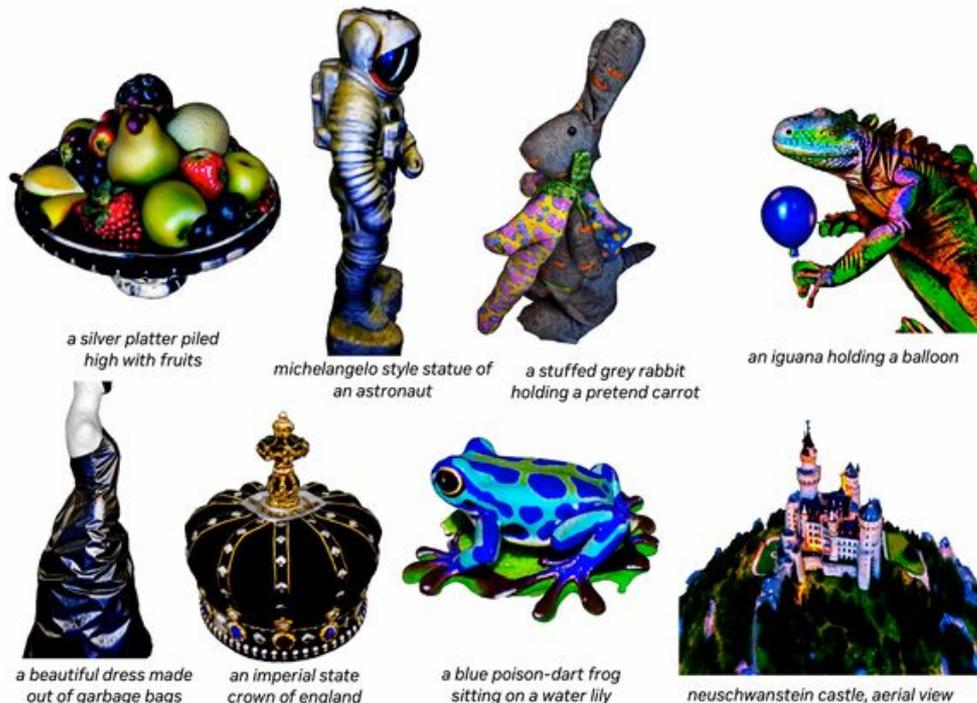


### 概要

Multiscale Hash Encodingを介した疎一密な2段階アプローチによる text-to-3D の性能改善。Text guidedな2D Diffusion Modelを用いて Multiscale Hash Encodingからのレンダリング画像に対してLossを適用。荒いグリッド・細かいグリッドの2段階で訓練する

### 新規性

DreamFusionよりもモデル生成が2倍高速で、高解像度かつ高品質な映像生成が可能



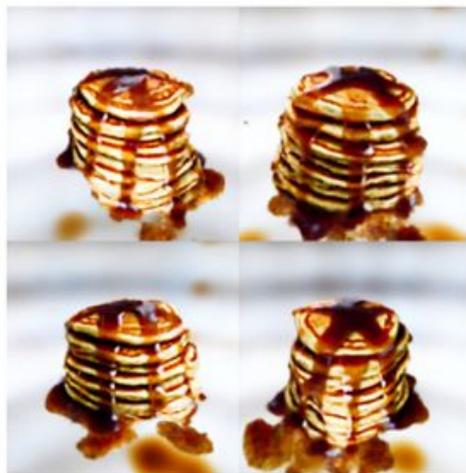
## 概要

NeRFを画像の代わりにオートエンコーダのコンパクトな潜在空間に適用することでText-to-3D生成モデルを実現  
目的のオブジェクトの粗い構造を定義する抽象的な形状スケッチにより3D部分の学習を誘導

## 新規性

3D監理を必要としない包括的なText-to-3D生成の枠組みを構築

“A stack of pancakes covered in maple syrup”



“A highly detailed sandcastle”



# GeNVS: Generative Novel View Synthesis with 3D-Aware Diffusion Models

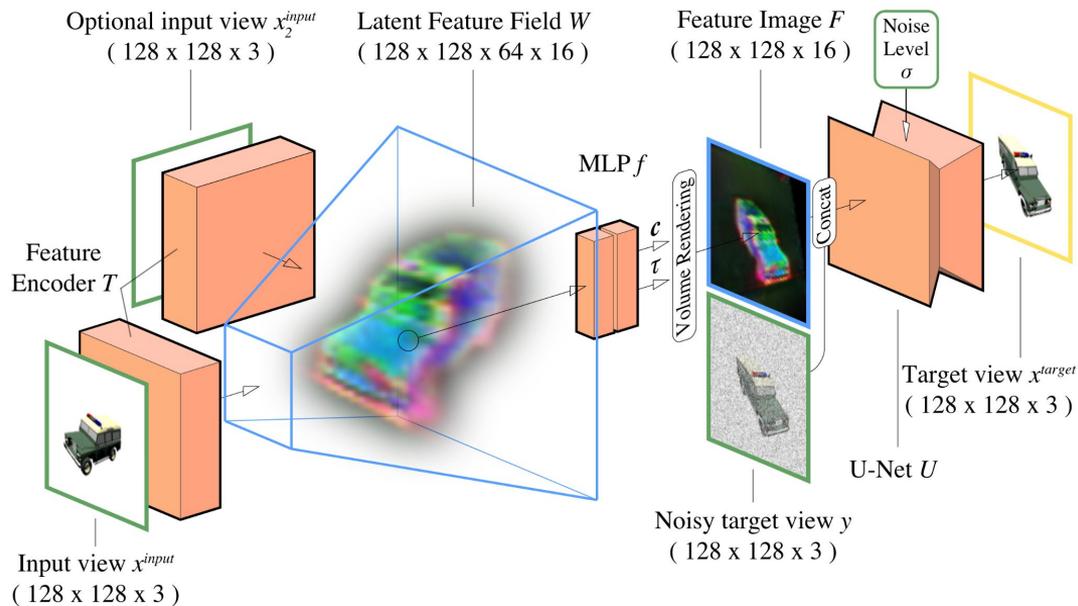
Eric R. Chan, Koki Nagano, Matthew A. Chan, Alexander W. Bergman, Jeong Joon Park, Axel Levy, Miika Aittala, Shalini De Mello, Tero Karras, Gordon Wetzstein

## 概要

1枚の入力画像から3次元的な新規視点映像を生成する画像条件付き拡散モデルを提案  
 入力画像と一致するレンダリングの可能性のある分布からサンプリングし、曖昧さがある場合でも、多様でもっともらしい新規ビューをレンダリングすることが可能

## 新規性

画像そのものを条件付けにした拡散モデルを扱い、単一画像から写実的な映像の生成が可能である

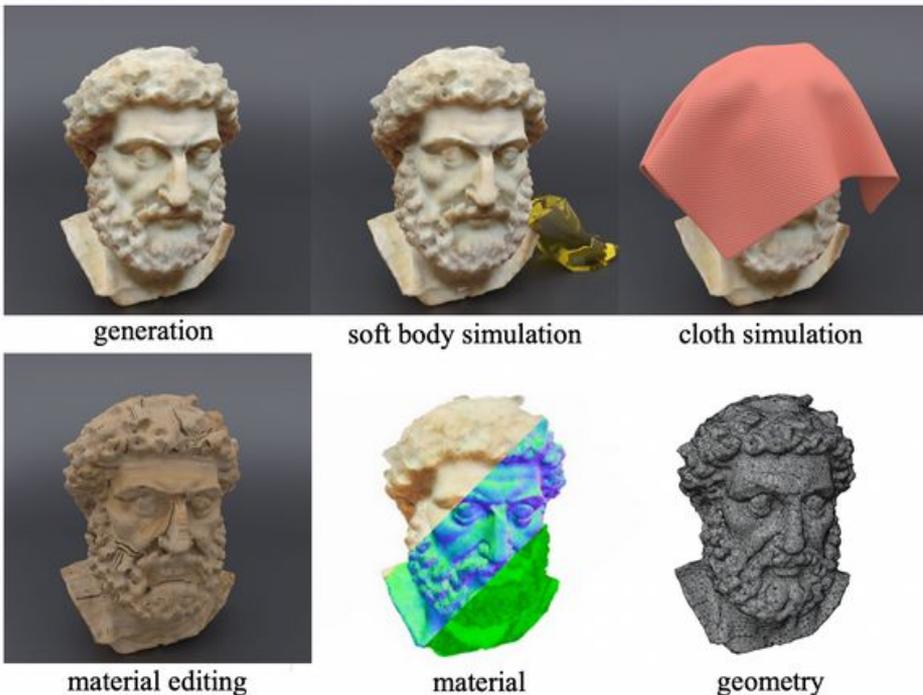


## 概要

2次元拡散モデルから得られる一般的な画像  
事前知識を用い、単一視点画像から NeRFを  
復元する。vision-languageモデルで得られた  
特徴を拡散モデルの条件として取り込むこと  
で、入力画像の意味的・視覚的特徴による画  
像の事前分布が設定可能

## 新規性

2D画像生成をベースとする DreamFusionと異  
なり、NeRFの潜在空間で拡散モデルを扱う。  
またGAUDIと異なり、屋外シーンでの利用が  
可能である



高速化

# 高速化の経緯① 推論時の高速化

---

NeRFでは1視点のレンダリング(500x500)で数分程度要求

→ 2021年頃、推論の高速化が流行

NeRFの推論では、1画素あたり数百のサンプリング点でNNを実行

→ **NNの順伝搬時間**と**サンプリング点数**が支配的

# 高速化の経緯① 推論時の高速化

## ○サンプリング点の削減

[AutoInt\(CVPR2021\)](#): ボリュームレンダリングの積分構造を埋め込んだネットワークを作成し、レイ両端の2点のサンプリングのみで定積分を計算可能

[DONeRF\(CVPR2021\)](#): 単眼デプス推定に基づいてサンプリング点配置の事前確率を設定し、最大48倍の高速化

## ○ネットワークの分割

[NSVF\(CVPR2021\)](#): Sparse Voxel Octreeでモノのある箇所にはのみNNを配置. NeRFより10倍高速

DeRF (CVPR2021): Voronoi空間分割でシーンの分解方法を含め学習. NeRFより3倍高速

KiloNeRF(ICCV2021): 数千オーダーの小規模なMLPに分割。NeRFより3000倍高速

# 高速化の経緯② NeRFのキャッシュ

## ○学習結果のキャッシュ

視線依存成分をネットワークではなく**基底をとったパラメータ表示**にすることで色場ネットワークを省略し、**訓練結果をグリッドなどにキャッシュ**することで推論時間を削減するアプローチが発生。

これによりNeRFのリアルタイムレンダリングが実現

[PlenOctree\(ICCV2021\)](#): 視線依存成分を球面調和関数でモデル化し、8分木構造にキャッシュをとる

[SNeRG\(ICCV2021\)](#): 視点非依存特徴を事前に計算し2Dキャッシュを取る

[FastNeRF\(Arxiv2021\)](#): 視線依存成分を分離したネットワークをとり基底を設定

[MobileNeRF \(CVPR2023\)](#): メッシュテクスチャにキャッシュすることでエッジ端末でのリアルタイムレンダリングが可能

# 高速化の経緯③ 訓練時の高速化

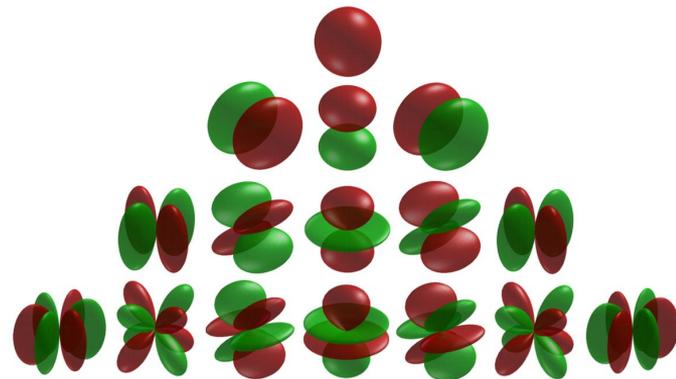
ボリュームレンダリングが微分可能であるなら、訓練済みのモデルをグリッドにキャッシュするのではなくキャッシュ先のパラメータを直接訓練する発想が流行

いずれも**球面調和関数**で視線依存性に基底をとっている

[Plenoxels\(CVPR2022\)](#): PlenOctreeを直接訓練する

[DVGO\(CVPR2022\)](#): ボクセルと事後活性化関数の組み合わせで鮮明な表面を表現可能

[InstantNGP\(SIGGRAPH2022\)](#): 境界の異なる多層ハッシュグリッドでパラメータ化



# 高速化の経緯④ モデルサイズの削減

グリッドを利用したモデルでは時間計算量が圧縮できる一方で、ボクセル表現と同様に空間全域を離散化するためパラメータ数が多くなる

→微分可能なパラメータの圧縮方法を用いてモデル化することで空間計算量も削減

Plenoxels(CVPR2022): 8分木構造で表層付近の解像度を高くパラメータ化

TensoRF(ECCV2022): ボクセル表現をテンソル分解した形で低次元近似してパラメータ化

MaskedWavelet(CVPR2023): ウェーブレット変換によりパラメータ効率を高め、また学習可能なマスク表現を実現

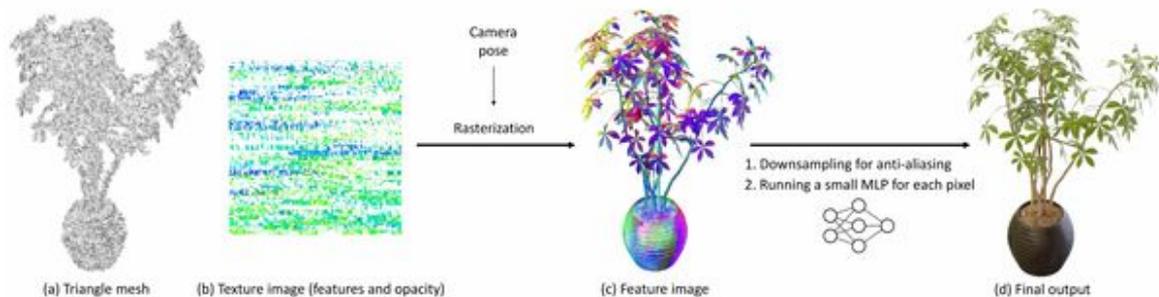
VQRF (CVPR2023): Codebookを用いてパラメータを離散的に圧縮

## 概要

テクスチャ付きポリゴンに基づく、エッジ端末でリアルタイムレンダリングが可能な NeRF 表現を提案。通常の NeRF を訓練後、密度を 2 値化してメッシュを抽出し、不透明度と特徴量をテクスチャに bake する。

## 新規性

volume ではなく mesh texture を用いるためメモリ使用量が小さく、また SNeRG と同等品質で 10 倍高速

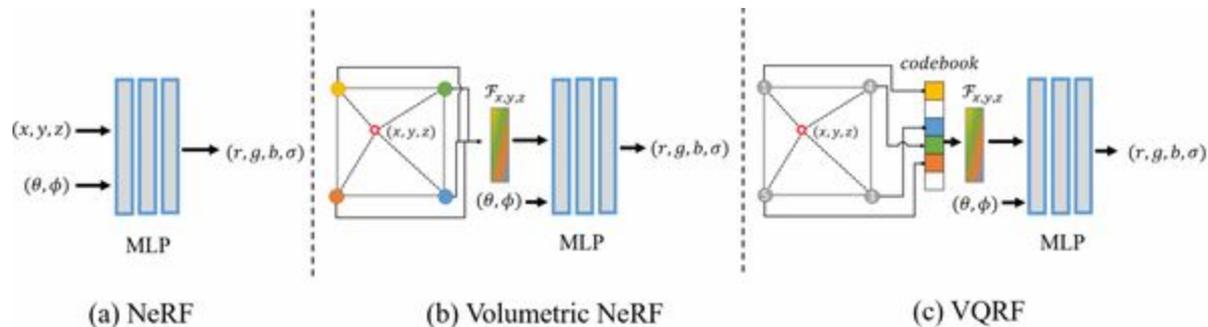


## 概要

vector quantized radiance fields (VQRF)で量子化し、グリッドベース NeRFのモデル容量を1MBまで圧縮

## 新規性

グリッドベースでNeRFを高速化する場合モデル容量が課題になる。提案手法では視覚的な品質低下を無視できる範囲でモデル容量を1MBまで圧縮



# Masked Wavelet Representation for Compact Neural Radiance Fields

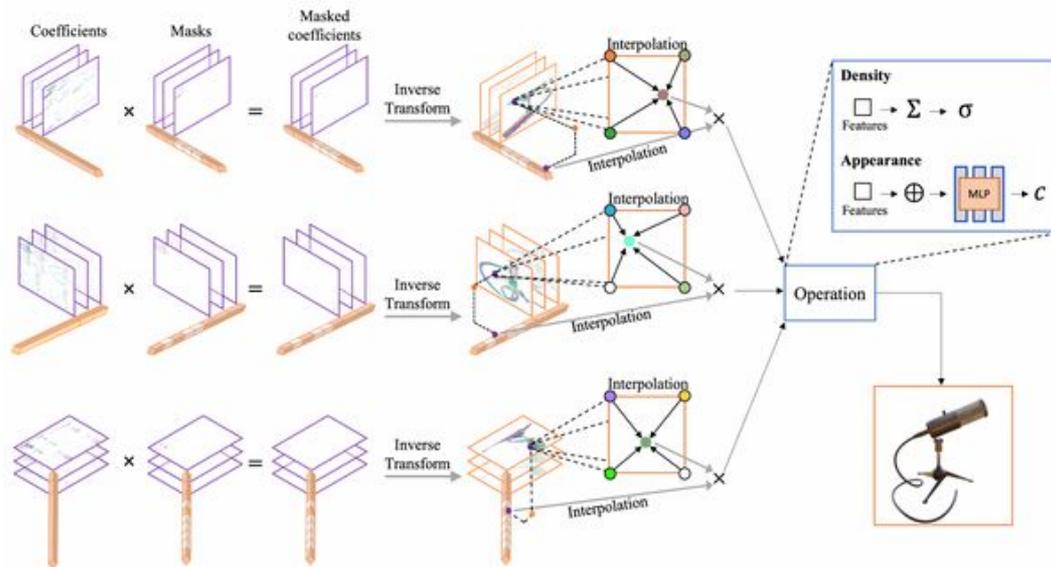
Lingzhi Li, Zhen Shen, Zhongshu Wang, Li Shen, Liefeng Bo

## 概要

NeRFでグリッドなど付加的なデータ構造を持つことの利点を損なうことなく、メモリサイズを縮小する方法を提示。グリッドベースのニューラル場にウェーブレット変換を適用し、グリッドのパラメータ効率を向上。またグリッド係数のスパース性を実現するための学習可能なマスキングアプローチを提案した。

## 新規性

パラメータのスパース性が高く、TensorRFよりもさらに小さい2MBを達成



# Few shot環境でのNeuralFields

# SPARF: Neural Radiance Fields from Sparse and Noisy Poses

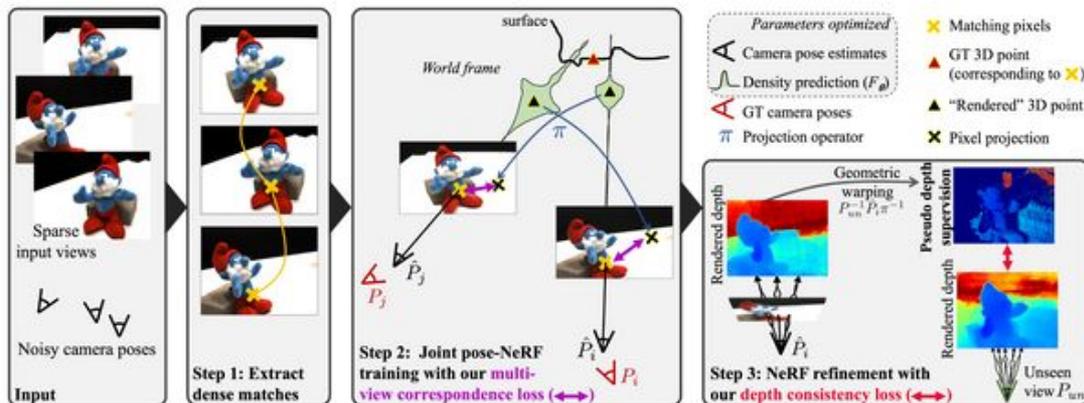
Prune Truong, Marie-Julie Rakotosaona, Fabian Manhardt, Federico Tombari

## 概要

入力ビュー間で抽出されたピクセルマッチに依存するジオメトリ制約を利用し、カメラポーズにノイズを含む三枚の画像から高品質な NeRF を復元

## 新規性

従来のソリッドベースの制約を導入することで NeRF では困難なシーンに適合可能

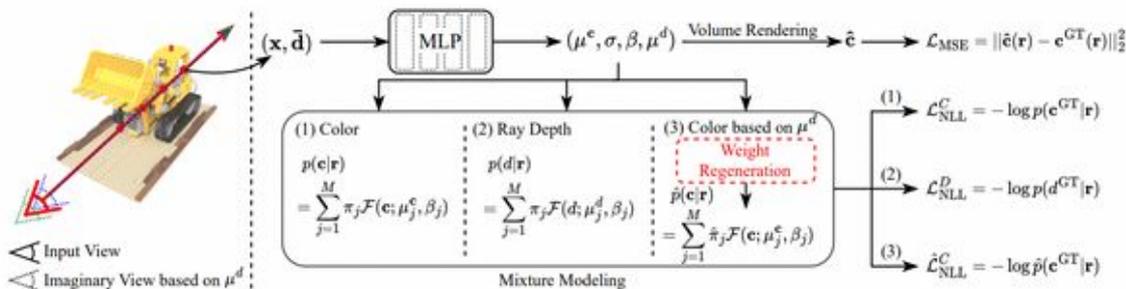


## 概要

光線を混合密度モデルでモデル化することで、疎な入力からNVSを行うための効果的な学習方法を提案。光線に沿ったRGBの色の分布を混合分布でモデル化することにより、光線に沿ったRGBの色の共同分布を推定。測光誤差だけでなく色と奥行きを目的関数に含める

## 新規性

CLIPのような外部モジュールや余分な監視信号の推論を必要とせずに、疎な視点からの品質改善が可能



## 概要

周波数正則化を用いて Fewshot な NeRF の性能向上を試みた研究。Positional Encoding の高周波成分をマスクし、学習時に徐々に高周波成分を追加していくフローをとる

## 新規性

追加の計算コストなしで非常にシンプルな実装ながら、従来の複雑な Fewshot 手法に匹敵する性能を達成



Turning the left to the right by adding *one* line of code: `pos_enc[int(t/T*L)+3:] = 0`

## FlexNeRF: Photorealistic Free-viewpoint Rendering of Moving Humans from Sparse Views

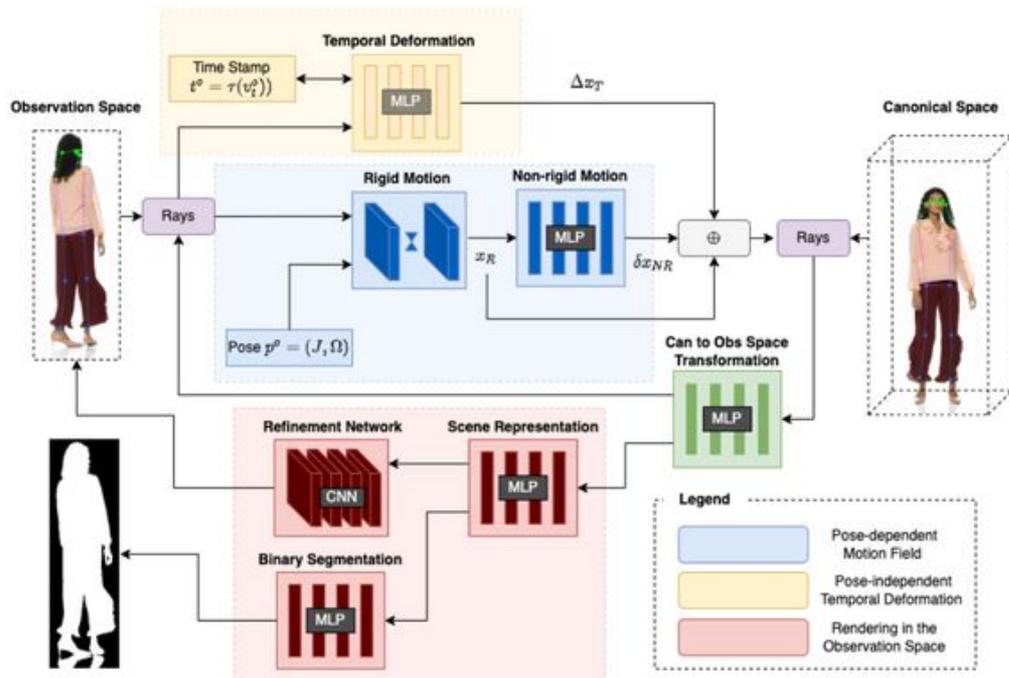
Vinoj Jayasundara, Amit Agrawal, Nicolas Heron, Abhinav Shrivastava, Larry S. Davis

## 概要

単眼動画から動き回る人間の NeRF を構成。ポーズに依存する MotionField とポーズに依存しない時間的変形が互いに補完し合い、正統的な時間とポーズの構成を共同で最適化する。

## 新規性

被写体が十分に移動することを仮定することで、単一視点から写実的な人物モデルの構築が可能

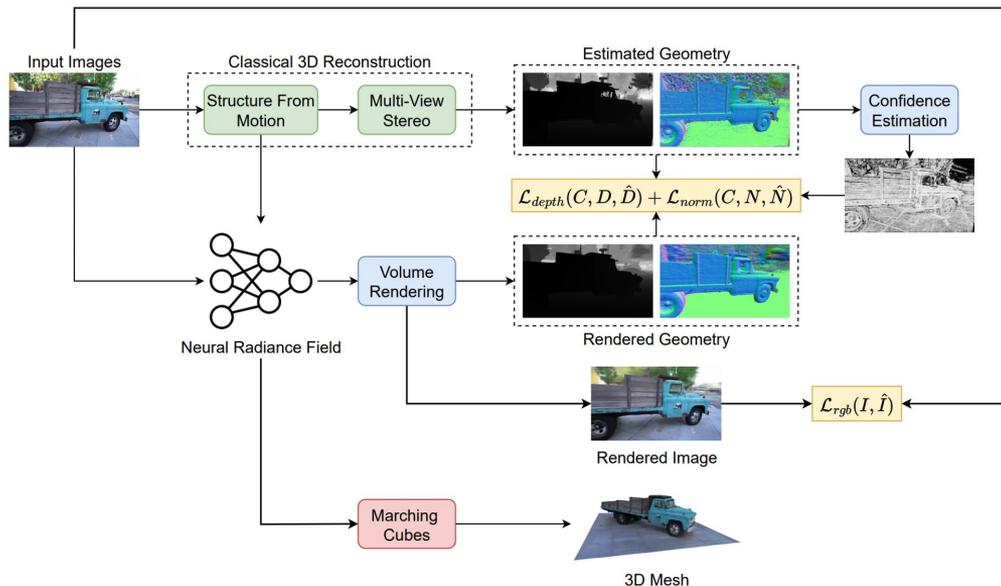


## 概要

SfM, MVSから得たデプスとノーマルを事前分布として擬似的GTとし、NeRFのdensity fieldから得られたデプスと法線で損失関数を設計することで、表面再構成の品質を向上させた研究。古典的 3D再構成アプローチで取得した値を用いて信頼度推定を行うことで、信頼性の高い擬似的 GTとした。

## 新規性

NeRFは形状に関して明示的に制約を受けないため、マーチングキューブでメッシュ化する際にノイズが多くなってしまいうらしいがこれを解決。また密度の勾配で法線が計算できることを利用した初めてのNeRFである。



## 概要

未知の光源位置から照らされた非ランバート物体を撮影した多視点画像が与えられた時、マルチビューフォトメトリックステレオ (MVPS) から表面法線マップを推定し、NeRFから得られる法線を正則化。次に法線やBRDF、照明を共同で最適化。入力画像が少数であっても正確、高品質な形状再構成を実現。

## 新規性

密度場の勾配により得られる法線情報を利用している点。NeRFを用いたMVPSの手法は存在するが、表面と光方向を分離することが出来ないが、提案手法はこの問題に対処。

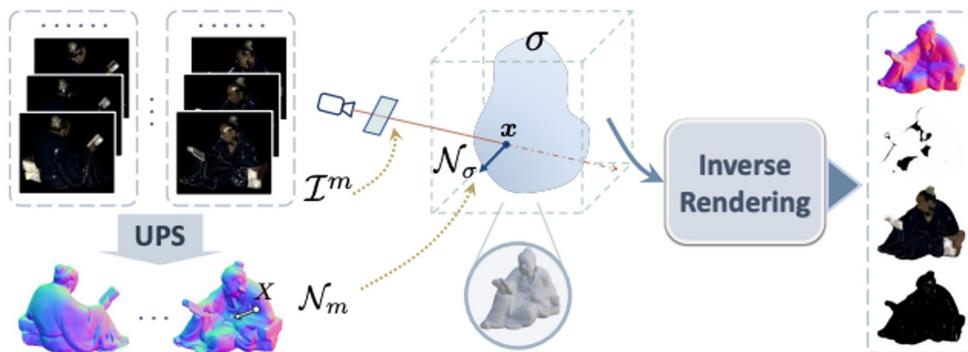


Fig. 2. Given multi-view multi-light images, we first obtain guidance normal maps via uncalibrated photometric stereo (UPS) to regularize the neural density field, which encourages accurate surface reconstruction. We then perform neural inverse rendering to jointly optimize surface normals, BRDFs and lights based on the initial shape.

編集可能NeuralFields

# 編集可能NeRF

---

NeRFの外観, スタイル, 形状を学習シーンから編集するタスク

CVPR2023では, 編集の指示方法やUIに着目した研究が多数発表

[SINE](#): 1枚の画像でセマンティック駆動編集

[Ref-NPR](#): 1枚の画像でスタイルを指示してスタイル操作

[PaletteNeRF](#): 拡散反射成分をカラーパレットの線型結合に分解し, パレットの編集により外観操作

[EditableNeRF](#): 学習動画中にキーポイントの影響度を導入し, ユーザーがキーポイントのドラッグ操作で変形操作

# SINE: Semantic-driven Image-based NeRF Editing with Prior-guided Editing Field

Chong Bao, Yinda Zhang, Bangbang Yang, Tianxing Fan, Zesong Yang, Hujun Bao, Guofeng Zhang, Zhaopeng Cui

## 概要

1枚の画像で神経輝度フィールドを編集できる、セマンティック駆動型 NeRF 編集アプローチを提案。



## 新規性

3次元空間における微細な幾何学的およびテクスチャ編集を符号化するガイド付き編集フィールドを用いることでユーザーが簡単に編集可能



# Ref-NPR: Reference-Based Non-Photorealistic Radiance Fields for Controllable Scene Stylization

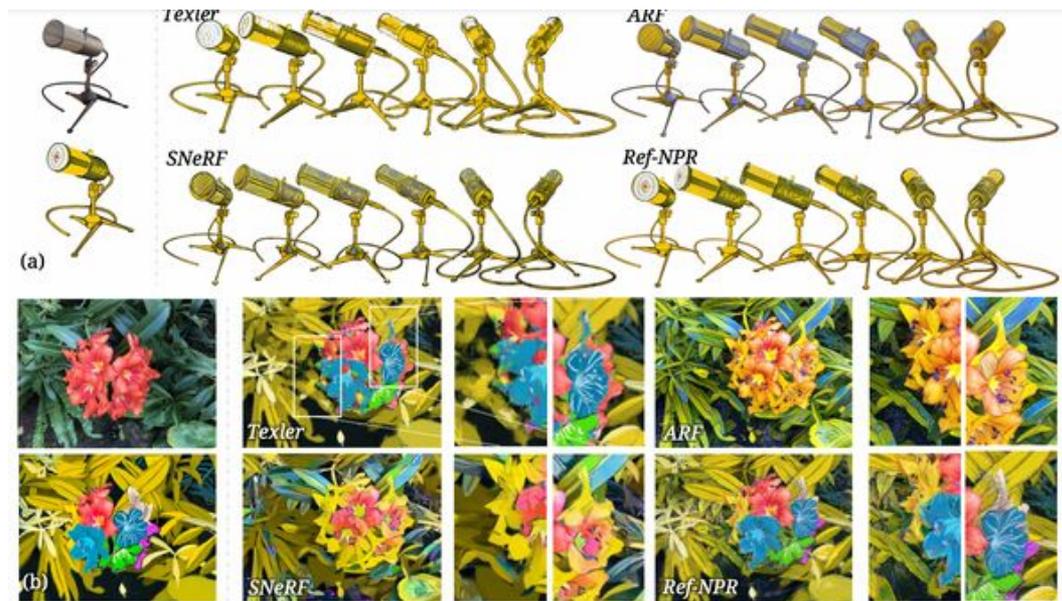
Yuechen Zhang, Zexin He, Jinbo Xing, Xufeng Yao, Jiaya Jia

## 概要

リファレンス画像を用いたノンフォトリアルな放射フィールドを生成し、シーンのスタイルをコントロールする方法を提案。

## 新規性

従来法では、生成された結果が写実的であるため、スタイル化には適さない。この論文では、リファレンス画像を用いることで、よりコントロールしやすく、写実性から外れたスタイル化が可能



# EditableNeRF: Editing Topologically Varying Neural Radiance Fields by Key Points

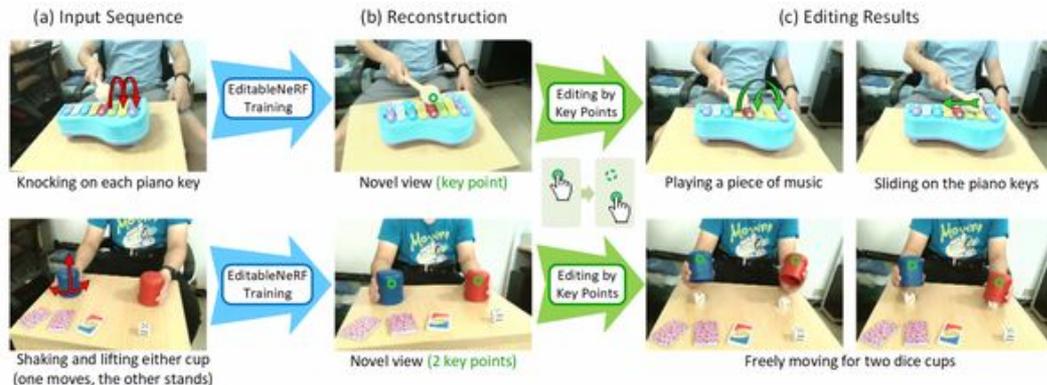
Yuechen Zhang, Zexin He, Jinbo Xing, Xufeng Yao, Jiaya Jia

## 概要

ユーザーがキーポイント操作で編集可能な動的シーンNeRFを提案。キーポイントとその重みの場を導入する戦略により動的シーンをモデル化し、ボクセル中で表面上の点の分散の極大値からキーポイントを初期化することで安定した最適化が可能

## 新規性

従来法と異なり、単眼の映像からアノテーションなしで構築可能である点がユニーク



# PaletteNeRF: Palette-based Appearance Editing of Neural Radiance Fields

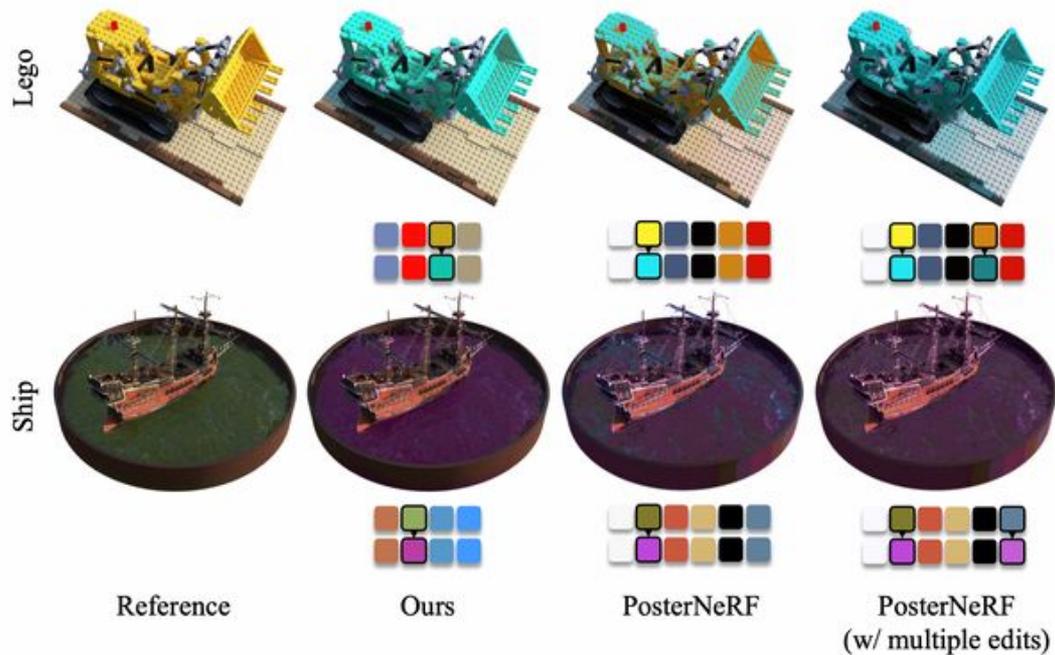
Zhengfei Kuang, Fujun Luan, Sai Bi, Zhixin Shu, Gordon Wetzstein, Kalyan Sunkavalli

## 概要

3次元色分解に基づく写実的な外観編集の手法を提案。Diffuse成分をシーン中で共有するカラーパレットのベースの線型結合に分解し、カラーパレットの編集により外観を操作する。

## 新規性

ユーザーが直感的かつ制御可能な方法でNeRFをインタラクティブに編集可能なユーザーインターフェースを提供。



# MagicPony: Learning Articulated 3D Animals in the Wild

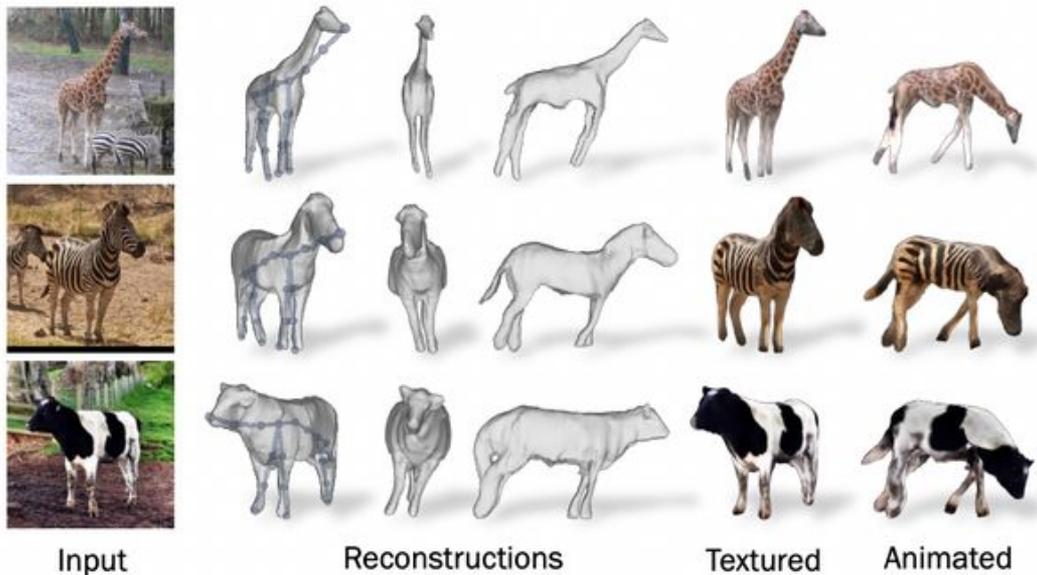
Shangzhe Wu, Ruining Li, Tomas Jakab, Christian Rupprecht, Andrea Vedaldi

## 概要

自然環境下での動物の3Dモデルを高精度に生成する手法を提案。DINO-ViTから3Dモデルに融合する効果的な自己教師特徴を導入し、少ない視点から構成可能。DMTetをJoint機構で拡張した、SDFとメッシュのハイブリッド表現により、変形の逆変換の学習困難性を緩和

## 新規性

従来法ではglobalからcanonicalな空間への座標の逆変換が必要であり、この部分の学習が困難であった。メッシュ表現とハイブリッドにすることでこれを緩和



# カメラ姿勢最適化

# カメラ姿勢最適化

---

NeRFでは、教師視点の正確なカメラ姿勢が既知であることを前提  
一度SfM等を行う必要があり、計算コストの高さや、視線依存が多いなどSfMの  
適用が困難なシーンでは精度の課題があった

→ NeRFの学習過程でのカメラ姿勢の調整に需要

[iNeRF\(IROS2020\)](#): 学習済みNeRFシーンで新規視点のカメラ姿勢推定. ボ  
リュームレンダリング時にカメラ姿勢まで逆伝搬可能なフローを構築

[NeRF--\(Arxiv 2021\)](#): NeRFの学習中にカメラ姿勢を同時に調整

[BARF\(ICCV2021\)](#): PEの高周波成分を抑制することで広域追従が可能

直接法SLAM同様に測光誤差のカメラ姿勢での勾配を算出する方法が主流

# カメラ姿勢最適化

---

CVPR2023では, BARFをベースにより困難な状況設定へ拡張

DBARF: シーンをまたぐ汎化NeRFでカメラ姿勢の自己教師学習を行う目的関数を導入

NoPe-NeRF: 連続した単眼映像を対象に, カメラ姿勢の事前知識無しでNeRFとカメラ姿勢を同時に最適化

BAD-NeRF: 強いモーションブラーを含む不正確なカメラ姿勢から最適化

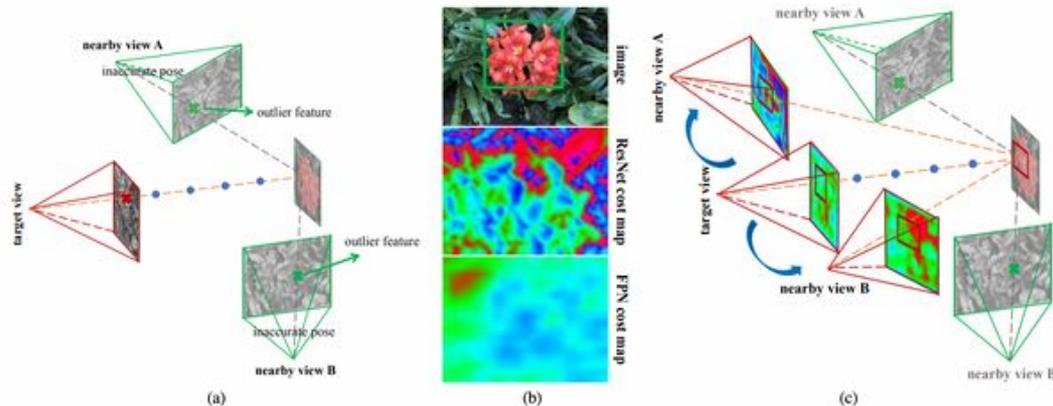
SPARF: ノイズを含む少数視点(3視点)から最適化

# DBARF: Deep Bundle-Adjusting Generalizable Neural Radiance Fields

Yu Chen, Gim Hee Lee

## 概要

コスト特徴マップを暗黙のコスト関数とすることでカメラポーズを一括調整し、自己教師付きで汎化NeRFと同時にカメラ姿勢を学習することで汎化NeRFにBundle Adjustmentを導入



## 新規性

汎化NeRFでは複雑な3次元CNNや変換器アーキテクチャに基づく画像特徴抽出を必要とするためBAの導入が困難であった。提案手法では新しいコスト関数の導入により実現

# NoPe-NeRF: Optimising Neural Radiance Field with No Pose Prior

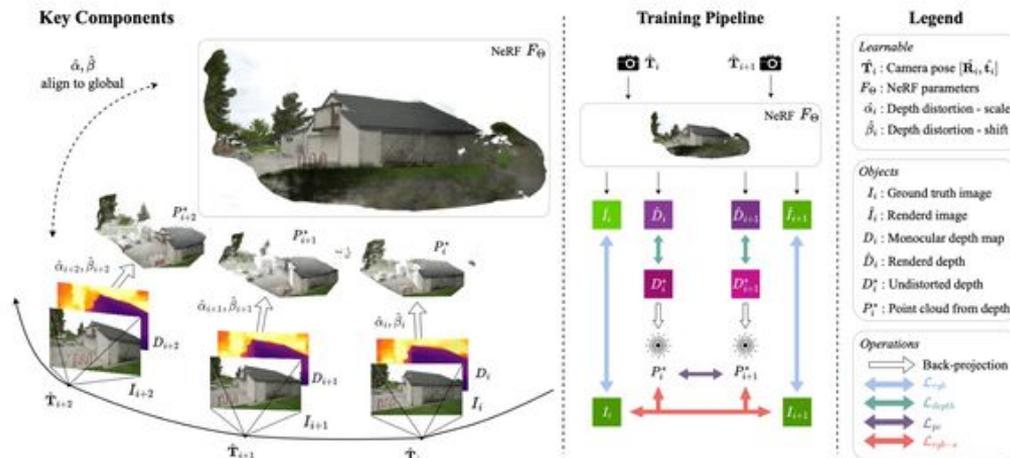
Wenjing Bian, Zirui Wang, Kejie Li, Jia-Wang Bian, Victor Adrian Prisacariu

## 概要

単眼深度推定の結果を用いて連続したフレーム間の相対姿勢に制約を追加し、カメラ姿勢の事前知識なしでNeRFとカメラ姿勢を同時最適化

## 新規性

従来のNeRFとカメラ姿勢の同時最適化はカメラの移動が大きい場合には大局解に到達できなかった。提案手法では深度値による連続したフレーム間の制約を用いることで、姿勢推定が困難なシーンでも高い精度を実現

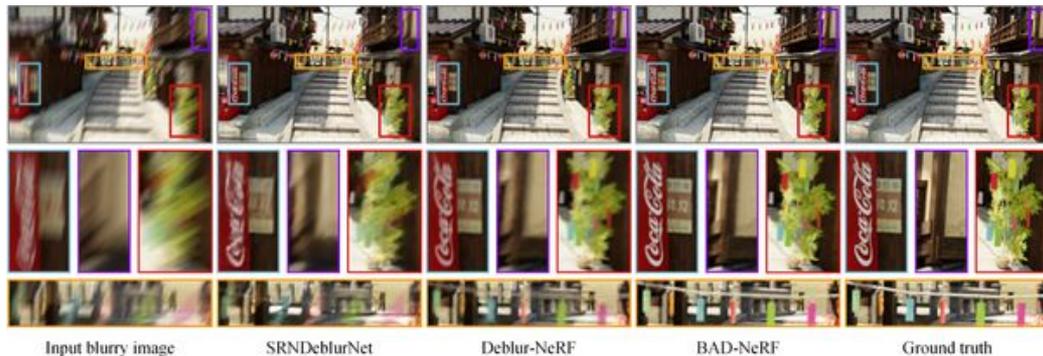


# BAD-NeRF: Bundle Adjusted Deblur Neural Radiance Fields

Peng Wang, Lingzhe Zhao, Ruijie Ma, Peidong Liu

## 概要

激しいモーションブラー画像や不正確なカメラポーズに強いBAを行うNeRFを提案。モーションブラーの物理的な画像形成過程をモデル。カメラ軌跡をパラメータ化し、動作ボケを含めてNeRFと同時に最適化を行う



## 新規性

不正確なカメラ姿勢とモーションブラーの両方を扱える枠組みを提案している

# SPARF: Neural Radiance Fields from Sparse and Noisy Poses

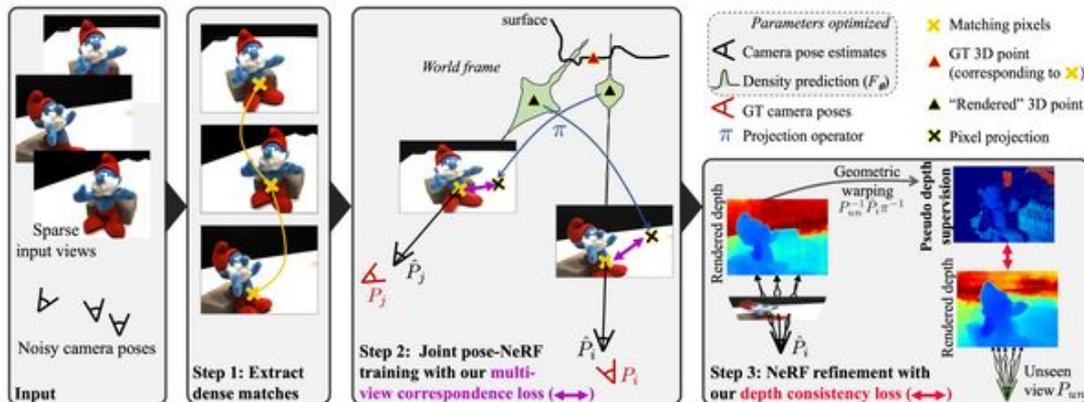
Prune Truong, Marie-Julie Rakotosaona, Fabian Manhardt, Federico Tombari

## 概要

入力ビュー間で抽出されたピクセルマッチに依存するジオメトリ制約を利用し、カメラポーズにノイズを含む三枚の画像から高品質な NeRF を復元

## 新規性

従来のソリッドベースの制約を導入することで NeRF では困難なシーンに適合可能



# 人物のNeuralFields

# FlexNeRF: Photorealistic Free-viewpoint Rendering of Moving Humans from Sparse Views

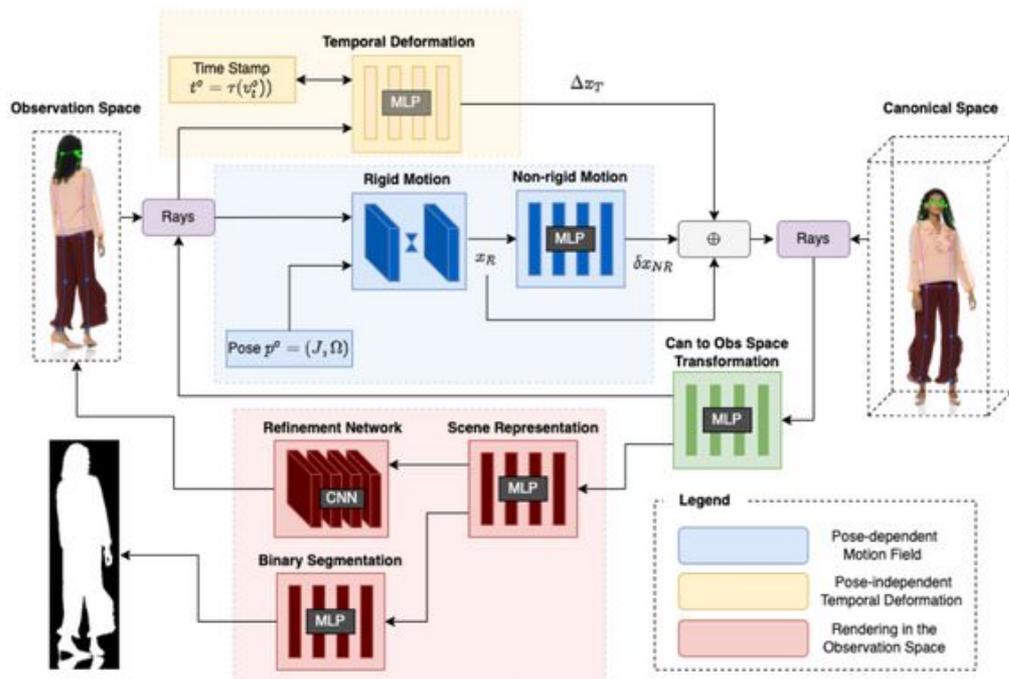
Vinoj Jayasundara, Amit Agrawal, Nicolas Heron, Abhinav Shrivastava, Larry S. Davis

## 概要

単眼動画から動き回る人間の NeRF を構成。  
ポーズに依存する MotionField とポーズに依存しない時間的変形が互いに補完し合い、正統的な時間とポーズの構成を共同で最適化

## 新規性

疎な視点から写実的なモデルの構築が可能

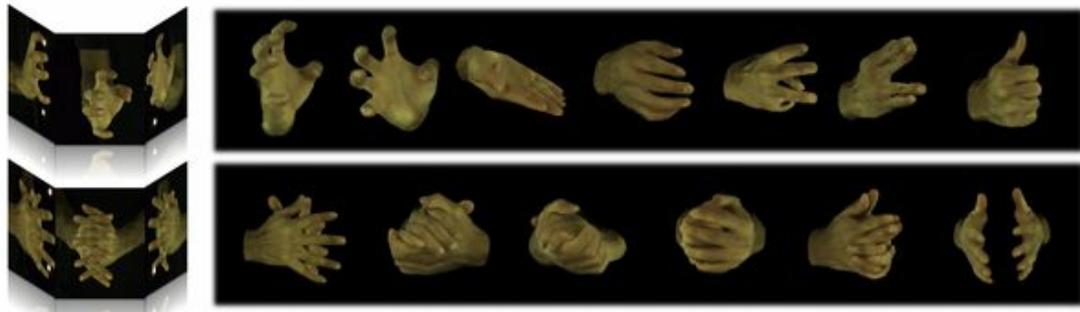


# HandNeRF: Neural Radiance Fields for Animatable Interacting Hands

Vinoj Jayasundara, Amit Agrawal, Nicolas Heron, Abhinav Shrivastava, Larry S. Davis

## 概要

インタラクションを行う手の NeRF を構成変形場を用いて両手のポーズを分離し、NeRF で共有されたベースの手モデルを最適化。両手を合成する際に正しい奥行き関係を確認するため、奥行き監視を確立して密度最適化を行う。



## 新規性

インタラクション可能な写実的な手のモデルを初めて提案

# PersonNeRF: Personalized Reconstruction from Photo Collections

Chung-Yi Weng, Pratul P. Srinivasan, Brian Curless, Ira Kemelmacher-Shlizerman

## 概要

複数年にわたり任意の姿勢で撮影した人物写真から、視点、体型、体裁を任意に組み合わせることでレンダリングできる手法を提案  
 基準ポーズ(Tポーズ)のneural volumetric表現を復元することで、異なる観測で変化する外観を許容しつつポーズに対する一貫性を保持

## 新規性

構造化されていない写真群からモデルの構築が可能

