

cvpaper.challenge

# NeuralFields

---

上田樹, 井手康允, 山田亮佑,  
鳥見晃平, 大塚大地

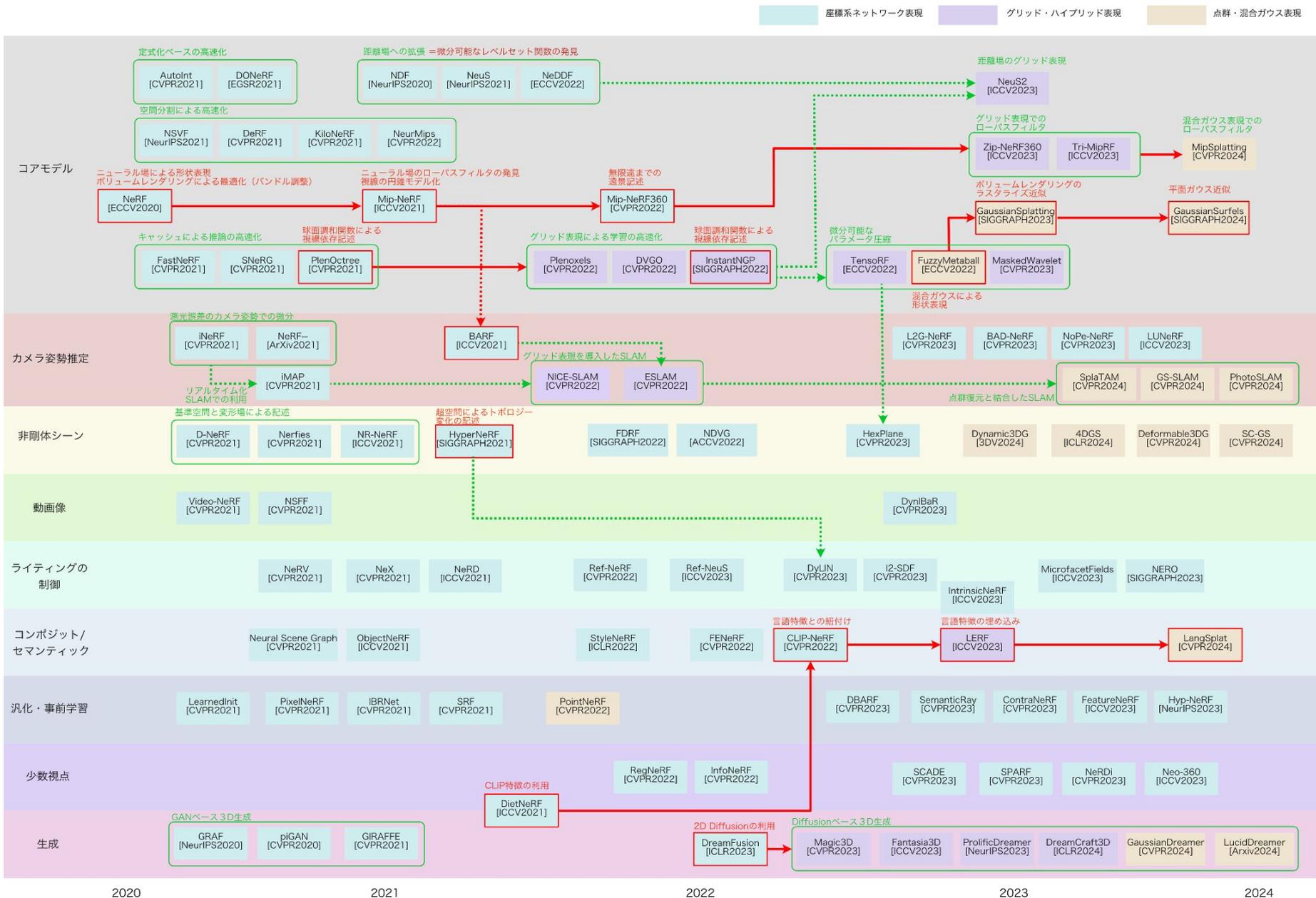
<http://xpaperchallenge.org/cv>

# 本資料の構成

---

- **論文サマリ**
  - 各論文について簡単に紹介
- **メタサーベイ**
  - 論文サーベイ&メンバー間の議論から見えたメタ要素

# 分野全体図



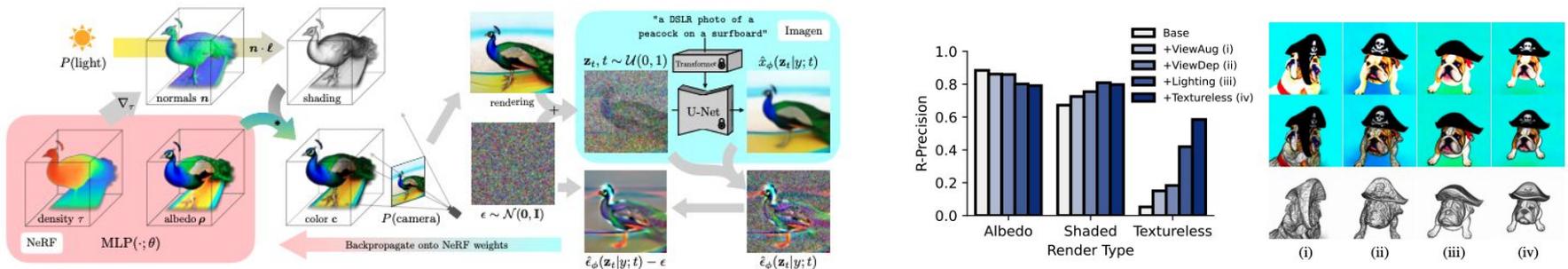
# 論文サマリ

# DreamFusion: Text-to-3D using 2D Diffusion

会議 : ICLR2022

著者 : Ben Poole, Ajay Jain, Jonathan T. Barron, Ben Mildenhall

- Diffusion modelを活用したZero-shot Text-to-3Dモデルの提案
  - Diffusion modelによる生成画像を目標としてNeRFモデルを学習
  - NeRFにより画像レンダリング → Diffusion modelに入力し画像を生成 → 生成画像に基づきNeRFを学習
  - 上記の学習戦略を実現するためにSDSを提案
  - テキストによるガイド/視点情報込みプロンプト/光源変化/テクスチャをなくすことでより高精細な3Dモデルを生成可能
  - 定量的な評価が難しい. これはタスク全体の課題. 初期NeRFじょうたいどうなってる? プロンプトにつきNeRFモデルを学習する

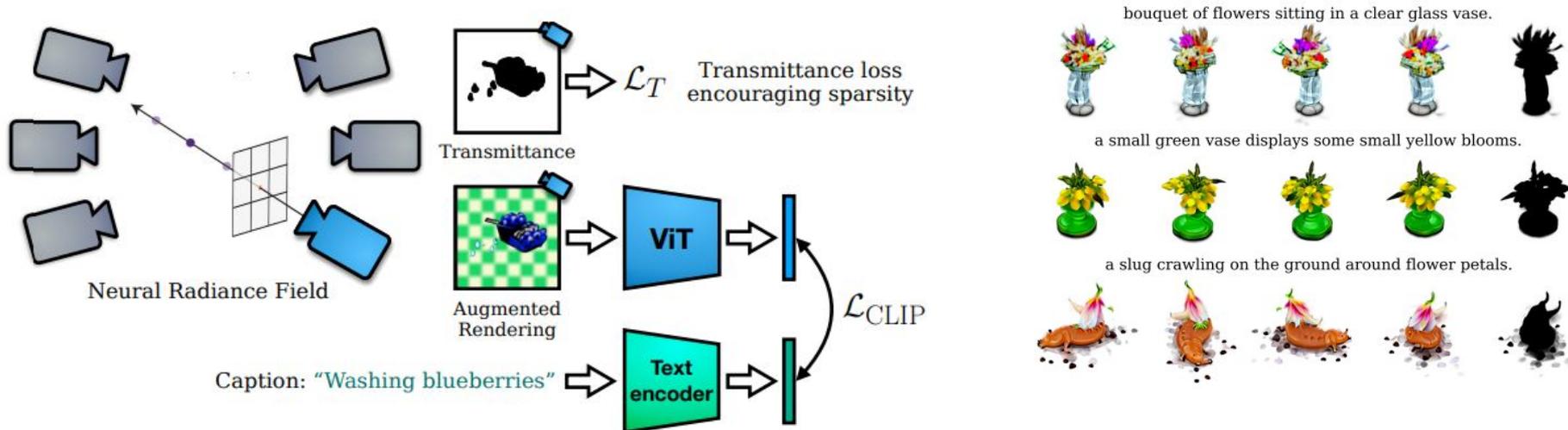


# Zero-Shot Text-Guided Object Generation with Dream Fields

会議 : CVPR2022

著者 : Ajay Jain, Ben Mildenhall, Jonathan T. Barron, Pieter Abbeel, Ben Poole

- 背景: 従来手法では3Dデータセットのlimitationに依存しており一部のカテゴリしかデータを生成できない
- 目的: ニューラル・レンダリングとマルチモーダル画像およびテキスト表現を組み合わせ、自然言語記述のみから多様な3Dオブジェクトの合成を可能にする
- 手法: ウェブからのキャプション付き画像の大規模なデータセットで事前に訓練された画像-テキストモデルを用いて生成
- 結論: Dream Fieldsと呼ばれるText to 3Dモデルを提案. 3Dデータセットのクラスラベルに関係なく様々なオブジェクトの形状と色を生成

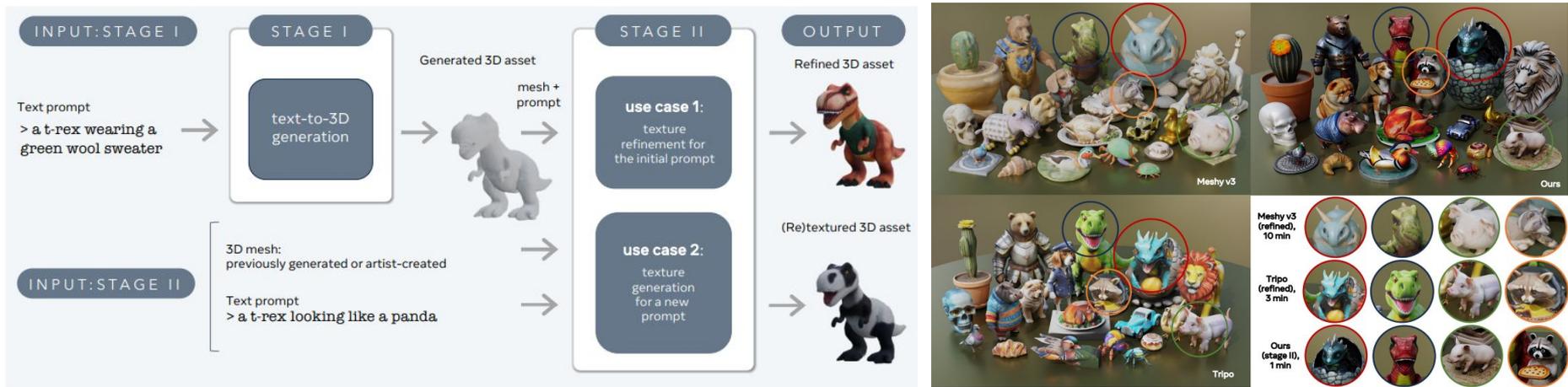


会議 : arXiv2024

著者 : Raphael Bensadoun, Tom Monnier, Yanir Kleiman , Filippas Kokkinos, Yawar Siddiqui, Mahendra Kariya, Omri Harosh, Roman Shapovalov, Benjamin Graham, Emilien Garreau, Animesh Karnewar, Ang Cao, Idan Azuri, Iurii Makarov, Eric-Tuan Le, Antoine Toisoul, David Novotny, Oran Gafni, Natalia Neverova, Andrea Vedaldi

- view space と UV space, テクスチャとシェイプの生成のend-to-endといった技術を基礎としている

## 高いプロンプト忠実度、高品質な3Dシェイプとテクスチャを持つ3Dアセットを1分以内に生成するモデルMeta 3D Genを提案



Method	Generation capabilities				Generation time	
	Mesh	Texture	PBR materials	Clean topology	Stage I only	Stages I+II
CSM Cube 2.0 (CSM, 2024)	✓	✓	✗	✗	15* min	1* h
Tripo3D (TripoAI, 2024)	✓	✓	✗	✗	30* sec	3* min
Rodin Gen-1 V0.5 (Deemos, 2024)	✓	✓	✓	✓	-	3*.† min
Meshy v3 (Meshy, 2024a)	✓	✓	✓	✗	1* min	10* min
Third-party T23D generator	✓	✓	✓	✗	10* sec	10* min
<b>Meta 3D Gen</b>	✓	✓	✓	✗	30 sec	<b>1 min</b>

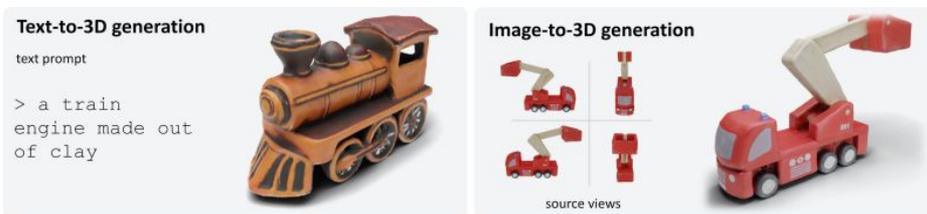
# Meta 3D AssetGen: Text-to-Mesh Generation with High-Quality Geometry, Texture, and PBR Materials <https://assetgen.github.io/static/AssetGen.pdf>

会議 : arXiv2024

著者 : Yawar Siddiqui, Tom Monnier, Filippos Kokkinos, Mahendra Kariya, Yanir Kleiman, Emilien Garreau, Oran Gafni, Natalia Neverova, Andrea Vedaldi, Roman Shapovalov, David Novotny

- 入力テキストに忠実な高品質テクスチャとPBR(Physically-Based Rendering)マテリアルを持つ3Dメッシュを生成

テクスチャとマテリアルを制御し忠実で高品質なメッシュを生成するext to 3Dモデル  
Meta 3D AssetGenを提案



Text PromptあるいはImage Promptを入力することで  
詳細なジオメトリと忠実なテクスチャを持つメッシュを生成

Material decomposition

> a cat made of silver



albedo metalness roughness

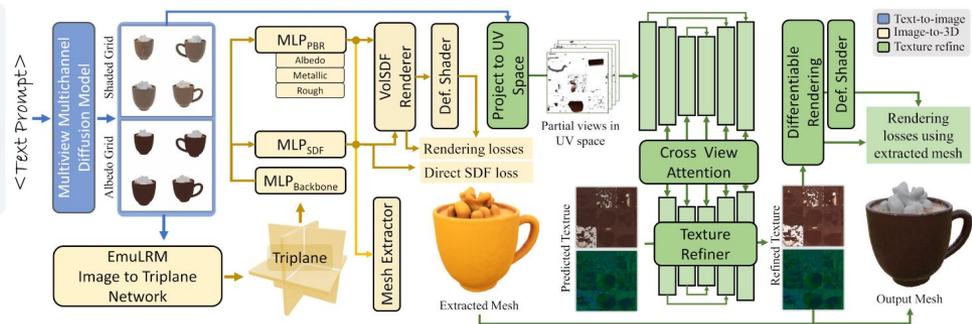
> a cat made of rock



Relighting in different environments



(左)アルベド, メタルネス, ラフネスに素材を分解  
(右)新しい環境でもオブジェクトを現実的に照らす



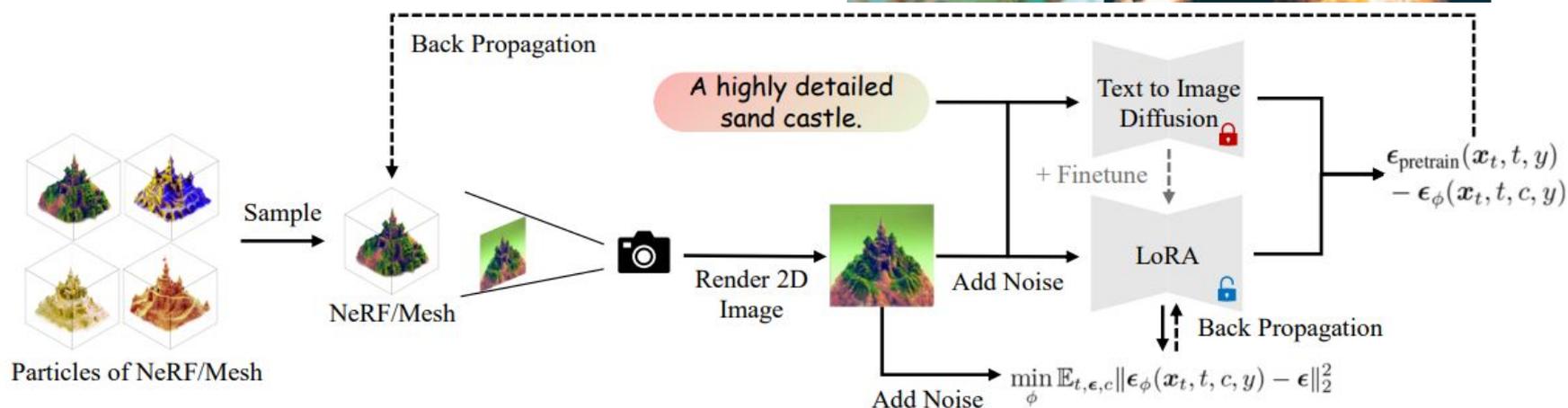
1Step: テキストから画像へ変換  
2Step: 画像から3Dに変換

# ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation

会議 : arXiv2023

著者 : Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, Jun Zhu

- 背景 : Score distillation sampling (SDS) はText to 3Dで大きな成功を収めているがover-saturation, over-smoothing, low-diversityといった課題が存在
- 目的 : variational score distillation (VSD)を提案することでSDSの課題を解決
- 手法 : 3DパラメータをSDSのような定数ではなく、確率変数としてモデル化
- 結論 : 高いレンダリング解像度と、煙や雫などのエフェクトを表現可能な高忠実度のNeRFを生成するProlificDreamerを開発

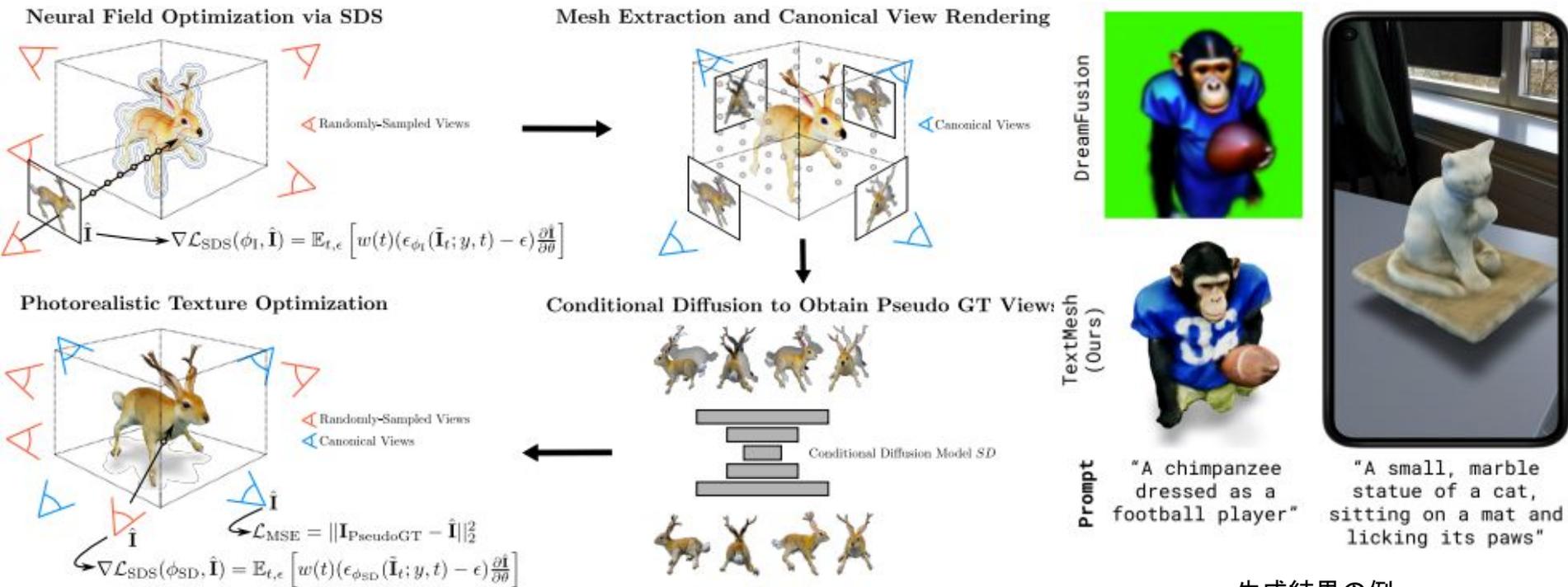


# TextMesh: Generation of Realistic 3D Meshes From Text Prompts

会議 : arXiv2023

著者 : Christina Tsalicoglou, Fabian Manhardt, Alessio Tonioni, Michael Niemeyer, Federico Tombari

- 背景: 多くのText to 3DモデルではNeRFを生成するため実用性に乏しい
- 目的: Text入力, 3Dメッシュ出力のモデルを構築
- 手法: NeRFを活用し, SDFをバックボーンとして使用
- 結論: Stable Diffusion, NeRF, SDFを組み合わせることでText to 3DMeshを構築



概要図

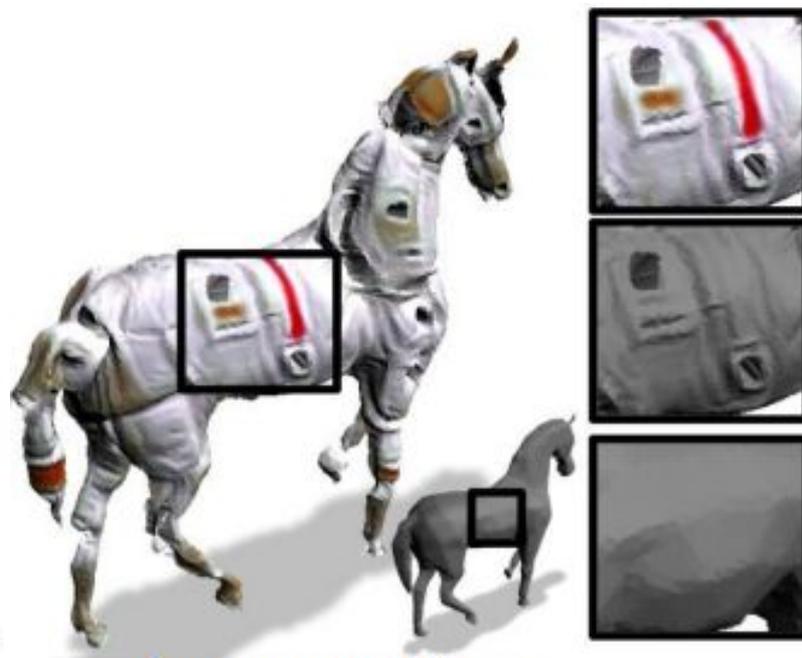
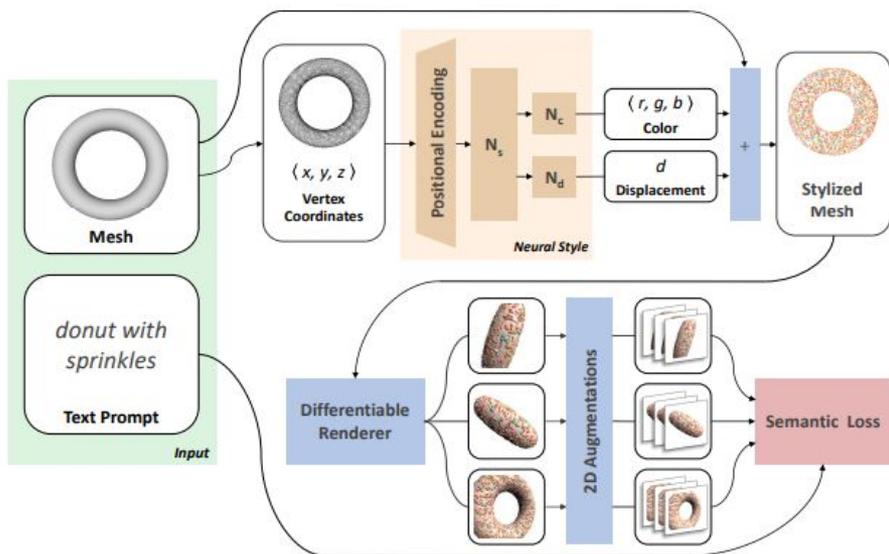
生成結果の例

# Text2Mesh: Text-Driven Neural Stylization for Meshes

会議 : CVPR 2022

著者 : Oscar Michel, Roi Bar-On, Richard Liu, Sagie Benaim, Rana Hanocka

- 背景: 3Dデータの編集技術において表現を修正するための構成要素の適切な定式化が課題
- 目的: 3Dデータの編集を制御するためのText to 3Dモデルを構築
- 手法: MeshデータとTextを入力とし, MeshデータをTextに従うように修正
  - CLIPを利用
- 結論: GANの事前学習や3Dデータセットに依存せず, メッシュの修正を可能とした



Astronaut Horse

Figure 4. Text2Mesh modifies an **input mesh** to conform to the **target text** by predicting color and geometric details. The weights of the **neural style network** are optimized by **rendering** multiple 2D images and applying **2D augmentations**, which are given a similarity score to the target from the CLIP-based **semantic loss**.

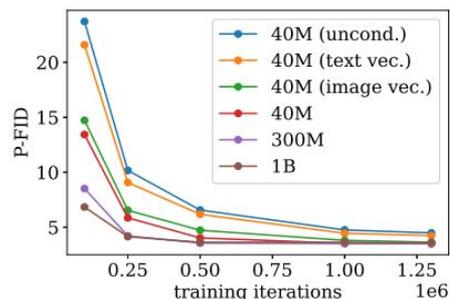
# Point-E: A System for Generating 3D Point Clouds from Complex Prompts

会議 : arXiv

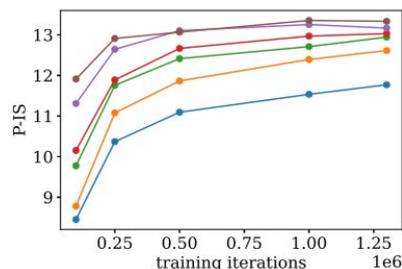
著者 : Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, Mark Chen

- テキストプロンプトを入力することで3Dのオブジェクトデータが生成されるモデルを提案
  - 一つのGPUで1オブジェクト当たり数秒から数分で生成可能
- [CLIP R-Precision](#), P-ISとP-FIDという指標で評価した結果, 複雑なプロンプトにも適切な3D点群が生成されていることを確認
  - P-IS (inception score): ISを点群データに適応させたもの
  - P-FID (Frechet Inception Distance): FIDを点群データに適応させたもの

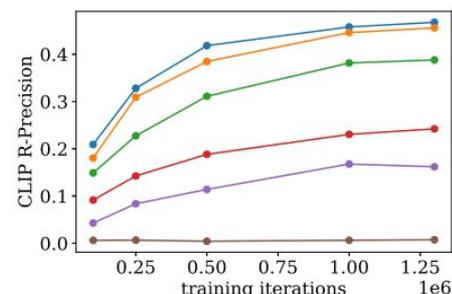
## 従来の3Dデータ生成モデルよりも生成速度の高速化を実現



P-FID

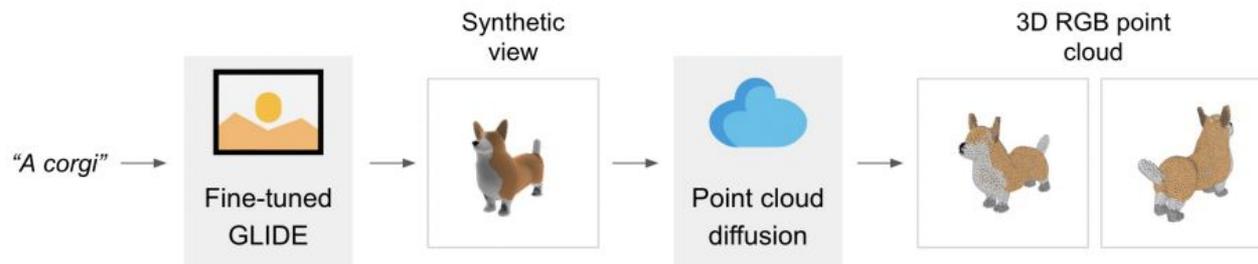


P-IS



CLIP R-Precision

3D RGB point cloud



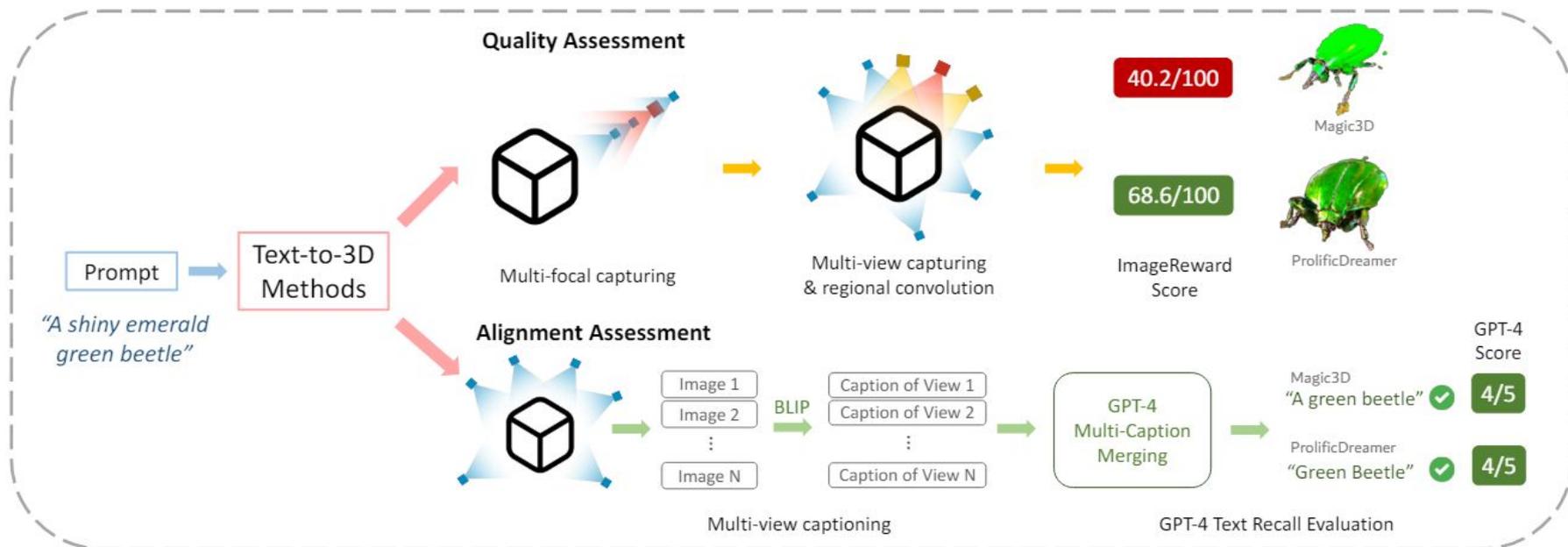
モデルの概要図

# T3Bench: Benchmarking Current Progress in Text-to-3D Generation

会議 : arXiv2024

著者 : Yuze He, Yushi Bai, Matthieu Lin, Wang Zhao, Yubin Hu, Jenny Sheng, Ran Yi, Juanzi Li, Yong-Jin Liu

- Text to 3Dにおける新たな評価指標を提案
- 背景: Text to 3Dの生成データは主観的なcase study/user studyによって結果を評価しており, 定量的な評価が行えていない
- 手法: 生成データの品質とPromptの忠実度合を定量的に評価する手法を提案
- 結論: 生成データを多視点からレンダリングすることで従来の評価指標よりもText to 3Dモデルの適切な評価を実現

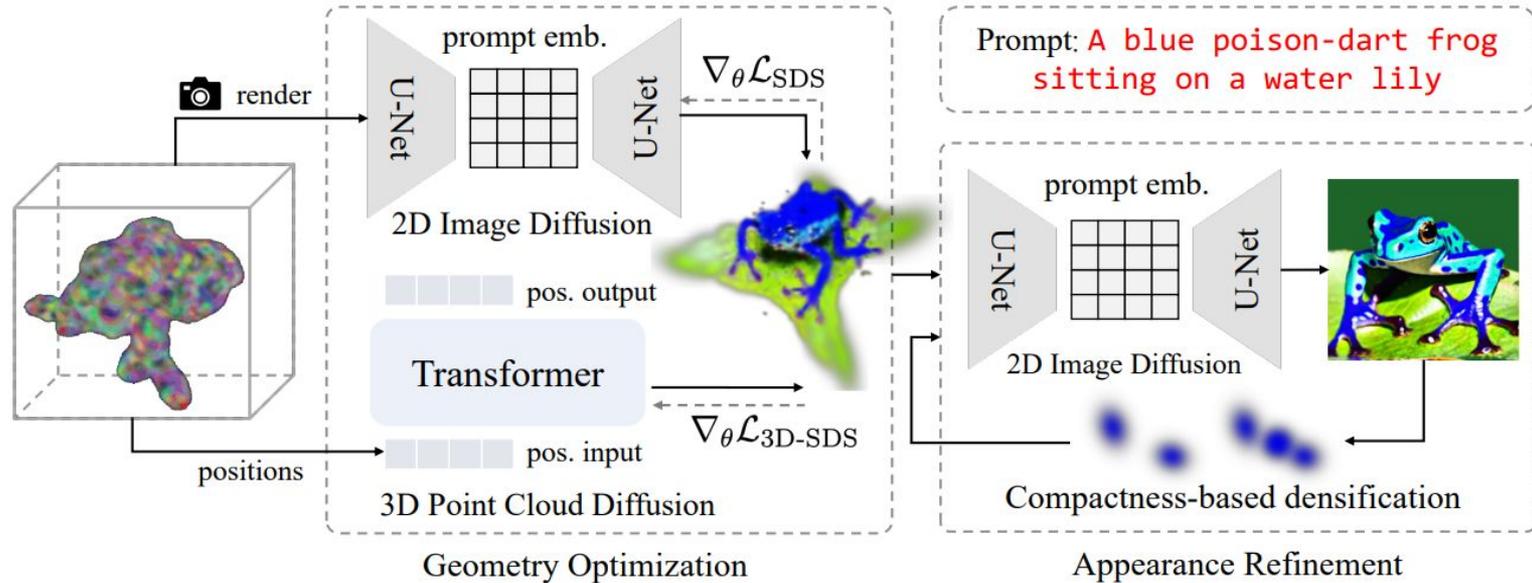


# Text-to-3D using Gaussian Splatting

会議 : CVPR2024

著者 : Zilong Chen, Feng Wang, Yikai Wang, Huaping Liu

- 3DGSを用いたText to3DモデルであるGSGENを提案
- 正確なジオメトリと繊細なアピランスを持つ3Dアセットを生成することが目的
- 3次元SDS損失と事前に学習された点群拡散モデルを統合し、一貫したジオメトリを形成
- Stable diffusionをベースにモデルを構築



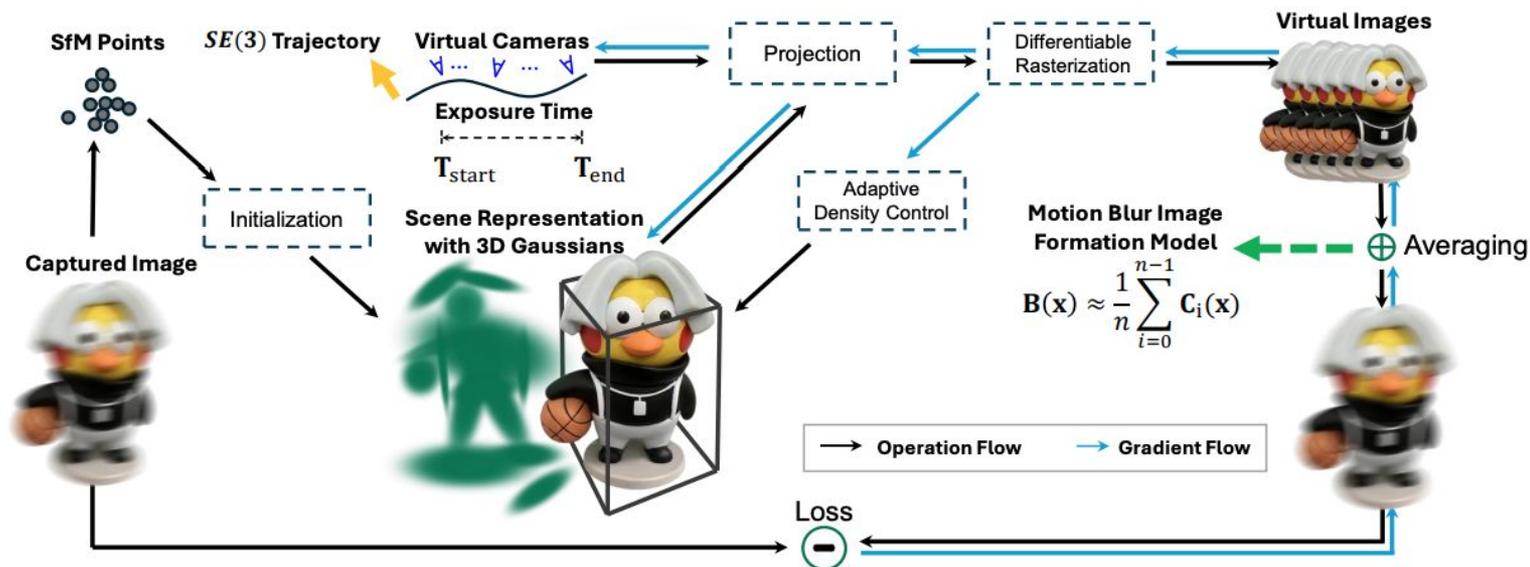
1. ガウシアンを初期化するために Point-E を利用
  - a. 生成された色を直接利用すると有害な影響があることが判明しているため、経験的な観察に基づいてランダムな色の初期化を選択

# BAD-Gaussians: Bundle Adjusted Deblur Gaussian Splatting

会議 : ECCV 2024

著者 : Lingzhe Zhao, Peng Wang, Peidong Liu

- **背景:** カメラモーションによる不鮮明な学習画像が、高精細な3Dシーン再構築を妨げる
- **目的:** 3DGSに基づいたカメラモーションによるぼけを除去するBAD-Gaussianの提案
- **手法:** カメラモーションによるぼけの物理的プロセスを3D-GSの学習に導入
  - カメラ軌跡の最適化: 露出時間内における初期姿勢と最終姿勢からカメラ姿勢を推定
  - 仮シャープ画像の生成: 軌跡に沿ってガウス関数から画像に投影し平均化
  - ガウス関数の最適化: 合成ぼけ画像と入力ぼけ画像間の誤差を最小化
- **結論:** 動きのぼけた画像からリアルタイムで高品質な3Dシーン表現を取得

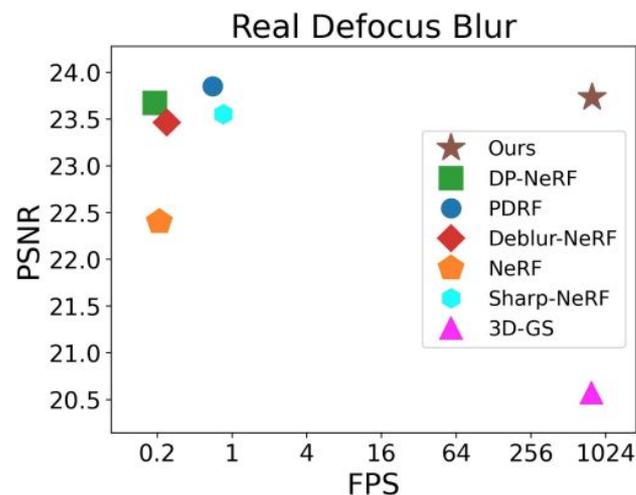
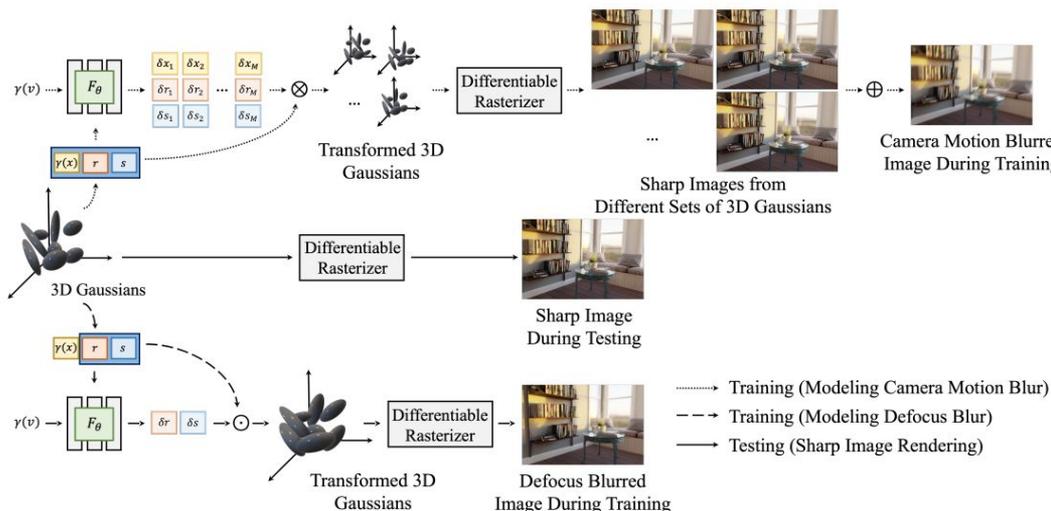


# Deblurring-3D-Gaussian-Splatting

会議 : ECCV 2024

著者 : Byeonghyeon Lee, Howoong Lee, Xiangyu Sun, Usman Ali, and Eunbyung Park

- **背景:** カメラモーションやオブジェクト動き等による不鮮明な学習画像が、高精細な3Dシーン再構築を妨げる
- **目的:** リアルタイムでぼやけ除去を可能とするDeblurring 3D-GSを提案
- **手法:** MLPを使用して、各3Dガウシアンを操作し、3Dシーンのぼやけをモデル化  
ぼやけた画像によるスパース点群を補うため、有効な色特徴を持つ追加点を追加し、シーンの遠方平面&ぼかしの激しい領域により多くの点を配置
- **結論:** 既存の最先端モデルと同等かそれ以上のレンダリング品質を達成

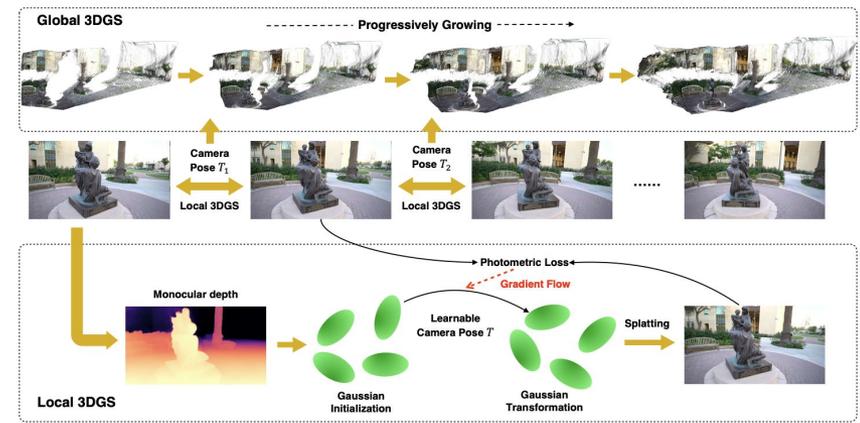


# COLMAP-Free 3D Gaussian Splatting

会議 : CVPR 2024

著者 : Yang Fu, Sifei Liu, Amey Kulkarni, Jan Kautz, Alexei A. Efros, Xiaolong Wang

- **背景:** Neural RenderingはSfMなどの正確なカメラポーズ推定に大きく依存
  - 処理時間, 特徴抽出の正確性, テクスチャ問題
- **目的:** 動画を活用することでSfMの前処理なしに3DGSを実現する
- **手法:** 各フレームにてLocal 3DGSを生成し, GLocal 3DGSに順次統合
  - フレーム間のガウス分布を用いアフィン変換を經由し、カメラ動きやシーン変化を学習
- **結論:** 先行研究では大きなカメラの動きを扱うことが困難かつ非常に長い学習時間が必要であることを明らかにし、動画と3DGSを組み合わせることで解決



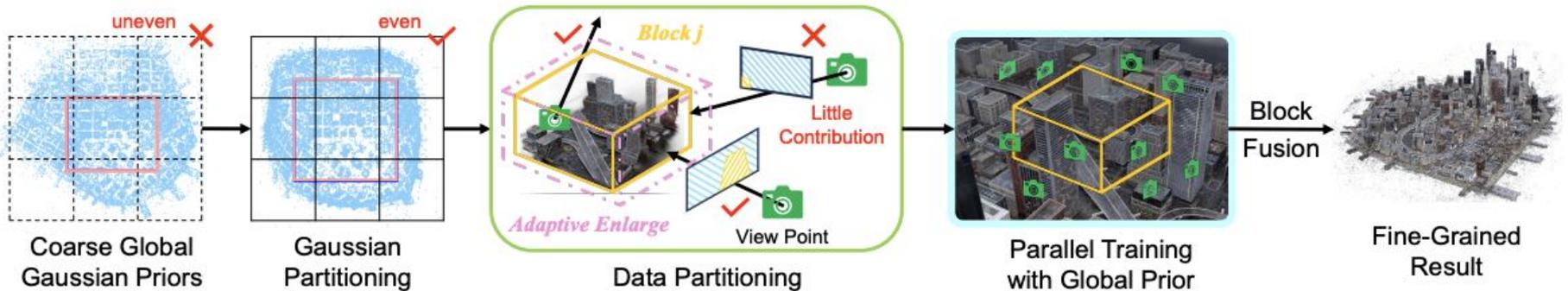
# CityGaussian: Real-time High-quality Large-Scale Scene Rendering with Gaussians

会議 : ECCV 2024

著者 : Yang Liu, He Guan, Chuanchen Luo, Lue Fan, Junran Peng, Zhaoxiang Zhang

- **背景:** 巨大3Dシーン( $1.5\text{km}^2$ )に3DGS適用時, 20Mガウシアンが必要でGPUメモリが不足
- **目的:** 1. 分割統治法と2. Level-of-Detailを組み合わせた学習戦略により, 巨大3Dシーンに対する効率的な学習&レンダリングを実現するCityGaussian(CityGS)を提案
- **手法:** 分割統治法では, 巨大3Dシーンを複数に分割処理し統合することでメモリを軽量化. さらに, Level-of-Detailでは unnecessary ガウシアンを削減することで高速化.
- **結論:** MatrixCityにおいて最新手法と比較して良好な性能を達成

How to free unnecessary Gaussians from rasterization is the key to real-time large-scale scene rendering.



# 3D Gaussian Splatting as Markov Chain Monte Carlo

会議 : arXiv2024

著者 : Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Jeff Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, Kwang Moo Yi

## 概要

- 3D Gaussian Splatting (3DGS)をMarkov Chain Monte Carlo (MCMC)サンプリングとして再解釈
- Stochastic Gradient Langevin Dynamics (SGLD)を用いたガウシアンを更新
- 複雑なヒューリスティクスを排除し、理論的に裏付けられたアプローチを実現

## 新規性

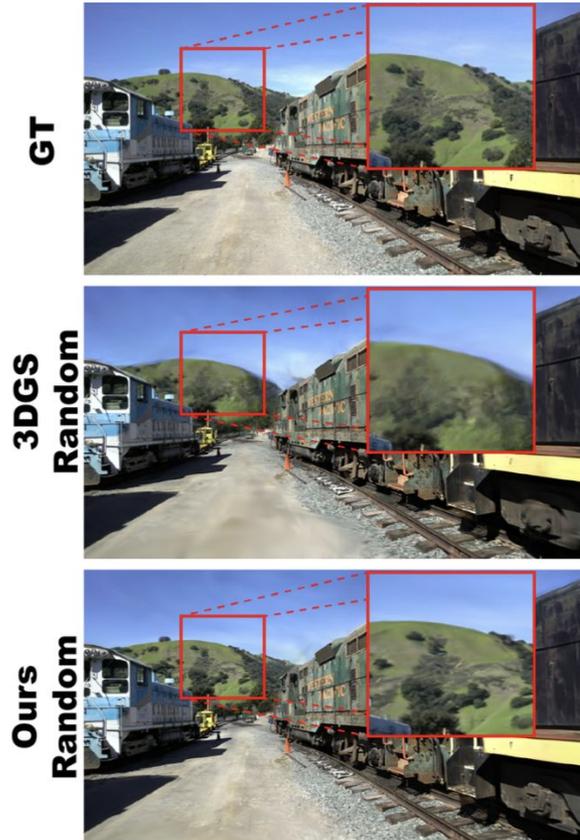
- 3DGSのシーン最適化をMCMCの理論に基づき実現
- 初期化に対するロバスト性が大幅に向上し、ランダムな初期化からでも高品質な再構成

## 結果

- NeRF Synthetic、MipNeRF 360、Tank & Temples、Deep Blendingデータセットで評価
- PSNR、SSIM、LPIPSなどの指標で従来手法と同等以上の性能を達成
- 特にランダム初期化からの学習で従来の3DGSを大きく上回る性能を示す

Table 1: **Quantitative results with same number of Gaussians** – We report results using PSNR, SSIM and LPIPS metrics. Our method significantly outperforms all baselines even when starting from random initialization, with a large gap in performance when compared with 3DGS [9]-random.

	NeRF Synthetic [16]	MipNeRF 360 [2]	Tank & Temples [12]	Deep Blending [8]
	PSNR↑ / SSIM↑ / LPIPS↓			
NeRF [17]	31.01 / - / -	24.85 / 0.66 / 0.43	-	21.18 / 0.78 / 0.34
Plenoxels [25]	31.76 / - / -	23.63 / 0.67 / 0.44	21.08 / 0.72 / 0.38	-
INGP-Big [17]	33.18 / - / -	26.75 / 0.75 / 0.30	21.92 / 0.75 / 0.31	-
MipNeRF [1]	33.09 / - / -	27.60 / 0.81 / 0.25	-	21.54 / 0.78 / 0.37
MipNeRF360 [2]	-	29.23 / 0.84 / 0.21	22.22 / 0.76 / 0.26	-
3DGS [9]-paper	33.32 / - / -	28.69 / 0.87 / 0.18	23.14 / 0.84 / 0.18	-
3DGS [9]-Random.	33.42 / 0.97 / 0.04	27.89 / 0.84 / 0.26	21.93 / 0.79 / 0.27	29.55 / 0.90 / 0.33
Ours-Random.	<b>33.85 / 0.97 / 0.04</b>	<b>29.58 / 0.89 / 0.19</b>	<b>24.47 / 0.86 / 0.19</b>	<b>29.66 / 0.90 / 0.31</b>
3DGS [9]	-	29.30 / 0.88 / 0.21	23.67 / 0.84 / 0.22	29.64 / 0.90 / 0.32
Ours	-	<b>29.83 / 0.89 / 0.19</b>	<b>24.19 / 0.86 / 0.18</b>	<b>29.80 / 0.90 / 0.30</b>



# Mip-Splatting: Alias-free 3D Gaussian Splatting

会議 : CVPR2024 (Best Student Paper)

著者 : Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, Andreas Geiger

## 概要

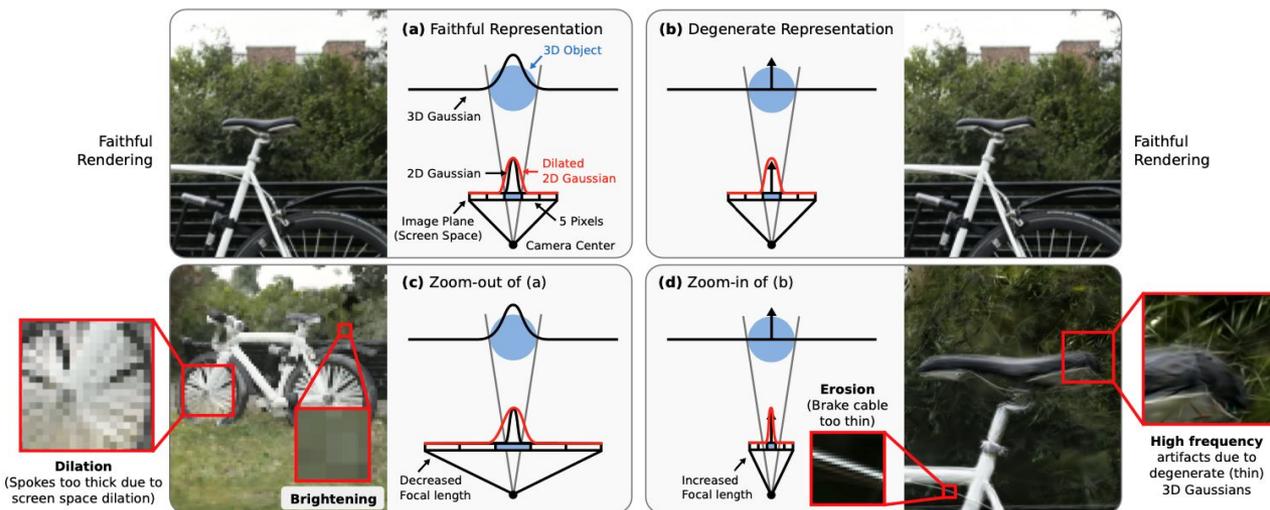
- 3D Gaussian Splatting (3DGS)を拡張し、マルチスケールのレンダリングを実現
- 3D表現の最大周波数を制限する3Dスムージングフィルターと物理的な撮像プロセスを模倣するボックスフィルターを近似した2D Mipフィルターを導入

## 新規性

- 3DGSの単一スケール最適化の限界を克服
- トレーニング時と異なるスケール(視点や倍率)でのレンダリングを可能に
- ズームインやズームアウト時のアーティファクトを削減

## 結果

- Blenderデータセットと Mip-NeRF 360データセットで評価
- トレーニングと同じスケールでのテストで最先端の手法と同等の性能
- 異なるスケールでのレンダリングで既存手法を大幅に上回る性能



# RadSplat: Radiance Field-Informed Gaussian Splatting for Robust Real-Time Rendering with 900+ FPS

会議 : arXiv2024

著者 : Michael Niemeyer, Fabian Manhardt, Marie-Julie Rakotosaona, Michael Oechsle, Daniel Duckworth, Rama Gosula, Keisuke Tateno, John Bates, Dominik Kaeser, Federico Tombari

## 概要

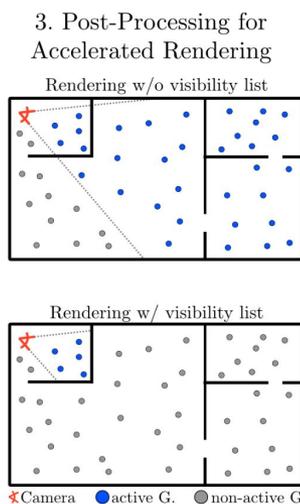
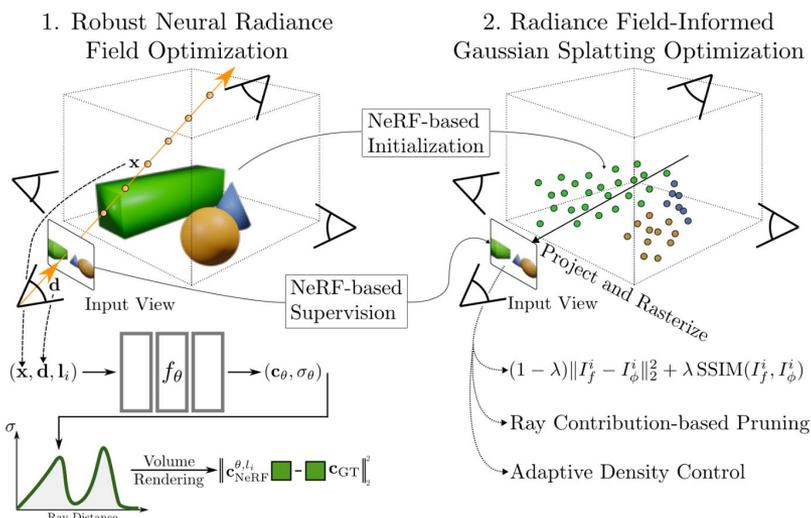
- Neural Radiance Field (NeRF)と3D Gaussian Splatting (3DGS)の利点を組み合わせた手法RadSplatを提案
- 実世界シーンにおける3DGSの最適化を不安定なNeRF表現を初期値として使用することで解決
- 複雑な実世界シーンを900+ FPSでリアルタイムレンダリングを実現

## 新規性

- NeRFの品質と3DGSの速度を組み合わせた手法
- ガウス点の重要度に基づく新しいプルーニング手法の実装
- 大規模シーンに対応するための視点ベースの可視性フィルタリング手法の考案

## 結果

- 3D Gaussianの数はオリジナルと比較して最大10倍削減
- 大規模シーンでも高品質なレンダリングを維持



	SSIM↑	PSNR↑	LPIPS↓	FPS↑	#G(M)↓
INGP [35]	0.705	25.68	0.302	9.26	-
BakedSDF [73]	0.697	24.51	0.309	539	-
MERF [56]	0.722	25.24	0.311	171	-
SMERF [10]	0.818	27.99	0.211	228	-
CompactG [23]	0.798	27.08	0.247	128	1.388
LightG [11]	0.799	26.99	0.25	209	1.046
EAGLES [14]	0.809	27.16	0.238	137	1.712
3DGS [21]	0.815	27.20	0.214	251	3.161
<b>Ours Light</b>	0.826	27.56	0.213	907	0.370
<b>Ours</b>	0.843	28.14	0.171	410	1.924
Zip-NeRF [2]	0.836	28.54	0.177	0.25	-

(a) Mip-NeRF360 dataset [1]

# StopThePop: Sorted Gaussian Splatting for View-Consistent Real-time Rendering

会議 : arXiv2024

著者 : Lukas Radl, Michael Steiner, Mathias Parger, Alexander Weinrauch, Bernhard Kerbl, Markus Steinberger

## 概要

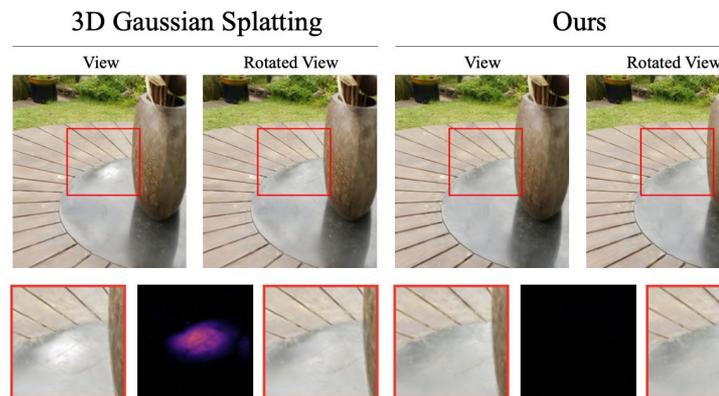
- 3D Gaussian Splattingのポッピング問題を解決する階層的レンダリング手法
  - ポッピング: オブジェクトの一部が突然現れたり消えたりしたように見え、レンダリングの連続性が損なわれる現象。
- 視点に依存しない一貫したレンダリングを実現
- レンダリング/学習速度を維持しつつ、アーティファクトを大幅に削減

## 新規性

- ガウシアン深度を正確に計算し、視点に依存しない並べ替えを実現
- 階層的なレンダリングパイプラインの実装
- ポッピングアーティファクトの定量的評価手法の提案

## 結果

- オリジナルの3DGSと比較して4%のオーバーヘッドで視点一貫性を改善
- Opacity Decayを用いると、3D Gaussianを50%削減しつつ同等の品質を維持



Dataset Metric	Deep Blending				Mip-NeRF 360 Indoor				Mip-NeRF 360 Outdoor				Tanks & Temples			
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$\uparrow$ LIP $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$\uparrow$ LIP $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$\uparrow$ LIP $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$\uparrow$ LIP $\downarrow$
Mip-NeRF 360 $\dagger$	29.40	0.900	0.245	0.138	31.57	0.914	0.182	0.088	24.42	0.691	0.286	0.170	22.22	0.758	0.256	0.232
Instant-NGP (base) $\dagger$	23.62	0.797	0.423	0.258	28.65	0.840	0.281	0.120	22.63	0.536	0.444	0.203	21.72	0.723	0.330	0.245
Instant-NGP (big) $\dagger$	24.96	0.817	0.390	0.222	29.14	0.863	0.241	0.114	22.75	0.567	0.403	0.200	21.92	0.745	0.304	0.241
Plenoxels $\dagger$	23.09	0.794	0.425	0.244	24.84	0.765	0.366	0.182	21.69	0.513	0.467	0.229	21.09	0.719	0.344	0.262
3DGS	29.46	0.900	0.247	0.131	30.98	0.922	0.189	0.094	24.59	0.727	0.240	0.167	23.71	0.845	0.178	0.199
Ours	29.86	0.904	0.234	0.127	30.62	0.921	0.186	0.099	24.60	0.728	0.235	0.167	23.21	0.843	0.173	0.216
3DGS (Opacity Decay)	28.94	0.894	0.262	0.134	30.57	0.918	0.198	0.097	24.45	0.718	0.261	0.169	23.52	0.839	0.194	0.205
Ours (Opacity Decay)	29.84	0.905	0.241	0.126	30.03	0.917	0.194	0.103	24.46	0.722	0.254	0.169	23.18	0.839	0.184	0.214

# Scaffold-GS: Structured 3D Gaussians for View-Adaptive Rendering

会議 : CVPR 2024 (Highlight)

著者 : Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, Bo Dai

## 概要

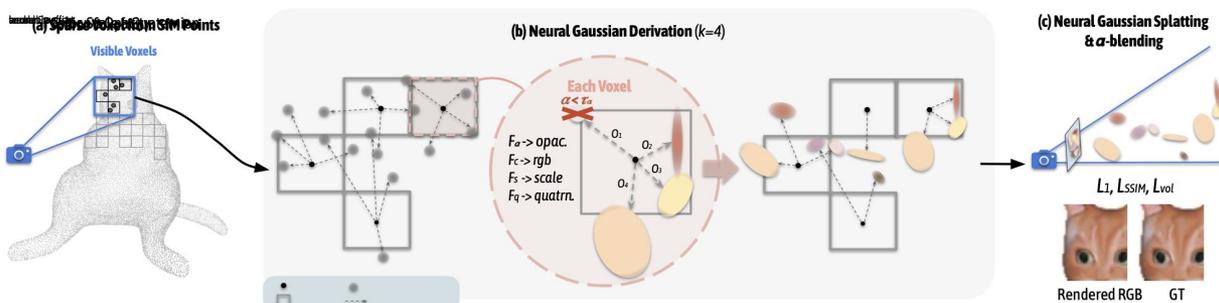
- シーンの構造を考慮した階層的な3D Gaussianを提案
- Anchor pointsを中心にneural Gaussiansを組織化し、視点に応じて適応的にレンダリング
- SfMから得られた点群をもとに疎な格子状のanchor pointsを構築
- Neural Gaussiansのパラメータをanchor pointの特徴量と視点位置に基づいて動的に予測

## 新規性

- 3D Gaussian Splatting (3DGS)の課題であるGaussianの過剰な拡大と視点変化へのロバスト性の欠如を解決
- Anchor pointsを用いた階層的なシーン表現により、シーンの幾何学的構造を考慮したコンパクトなモデルを実現
- Neural Gaussiansのパラメータを動的に推論することで、視点変化に対する適応性を向上

## 結果

- Mip-NeRF360、Tanks & Temples、Deep Blending、Synthetic Blender、BungeeNeRF、VR-NeRFで3D-GSと同等以上のレンダリング品質を達成
- モデルサイズを大幅に削減しつつ、3D-GSと同等の高速レンダリングを実現
- マルチスケールのシーンや新しい視点 (Novel-view) での優れた性能を示す



Dataset	Mip-NeRF360		Tanks&Temples		Deep Blending	
	FPS	Mem (MB)	FPS	Mem (MB)	FPS	Mem (MB)
<b>3D-GS</b>	97	693	123	411	109	676
<b>Ours</b>	<b>102</b>	<b>156</b> (4.4× ↓)	<b>110</b>	<b>87</b> (4.7× ↓)	<b>139</b>	<b>66</b> (10.2× ↓)

Dataset	BungeeNeRF		VR-NeRF		Synthetic Blender	
	PSNR	Mem (MB)	PSNR	Mem (MB)	PSNR	Mem (MB)
<b>3D-GS</b>	24.89	1606	28.94	263	33.32	53
<b>Ours</b>	<b>27.01</b>	<b>203</b> (7.9× ↓)	<b>29.24</b>	<b>69</b> (3.8× ↓)	<b>33.68</b>	<b>14</b> (3.8× ↓)

# Trimming the Fat: Efficient Compression of 3D Gaussian Splats through Pruning

会議 : arXiv2024

著者 : Muhammad Salman Ali, Maryam Qamar, Sung-Ho Bae, Enzo Tartaglione

## 概要

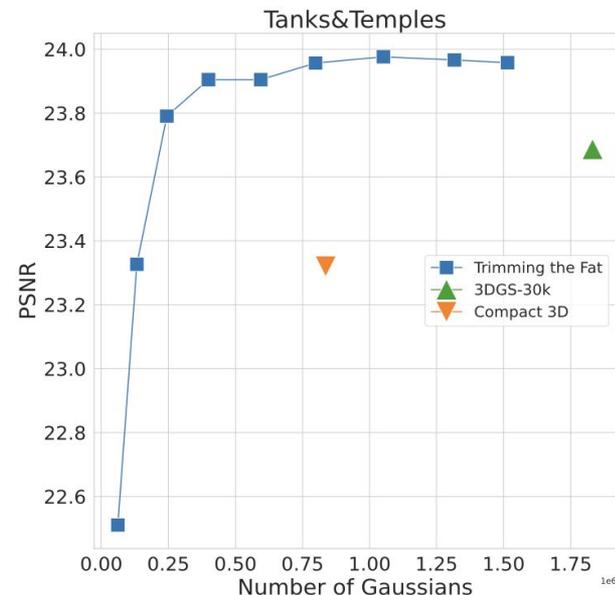
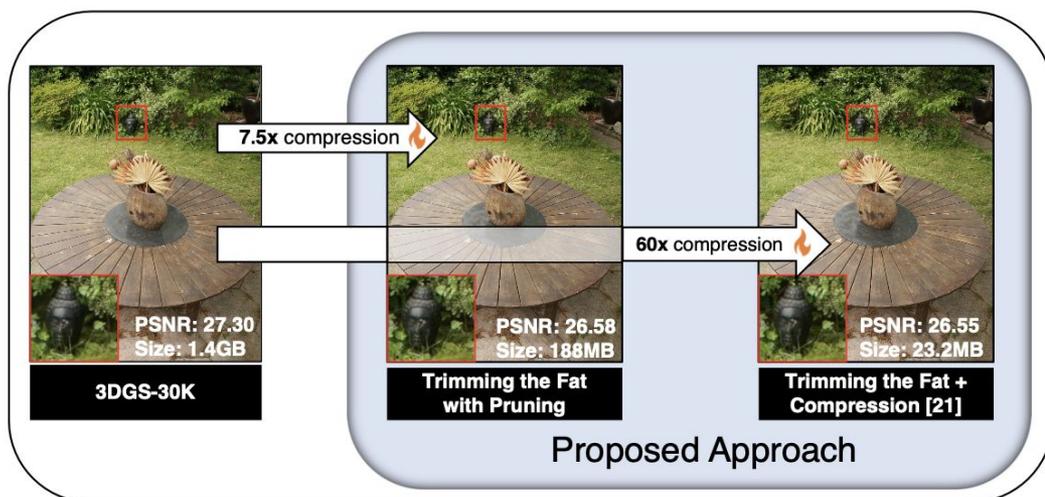
- 3D Gaussian Splatting (3DGS) モデルの効率的な圧縮手法を提案
- 勾配情報を活用した反復的プルーニングとFine-tuningのアプローチを導入
- モデルサイズを削減しながら、レンダリング品質を維持/向上

## 新規性

- 不透明度と勾配の両方を考慮したプルーニング条件の設計
- 段階的なスパース化による局所最適解の回避
- 過剰パラメータ化された大きなモデル(3DGS-30k)から開始、徐々に圧縮

## 結果

- 最大75%のGaussianを除去し、ベースラインと同等以上の性能を維持
- ベクトル量子化と併用すると約50倍のメモリ圧縮率ベースラインモデルと同等の品質
- レンダリング速度が向上、最大600 FPSを達成



# High-quality Surface Reconstruction using Gaussian Surfels

会議 : SIGGRAPH 2024

著者 : Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, Weiwei Xu

## 概要

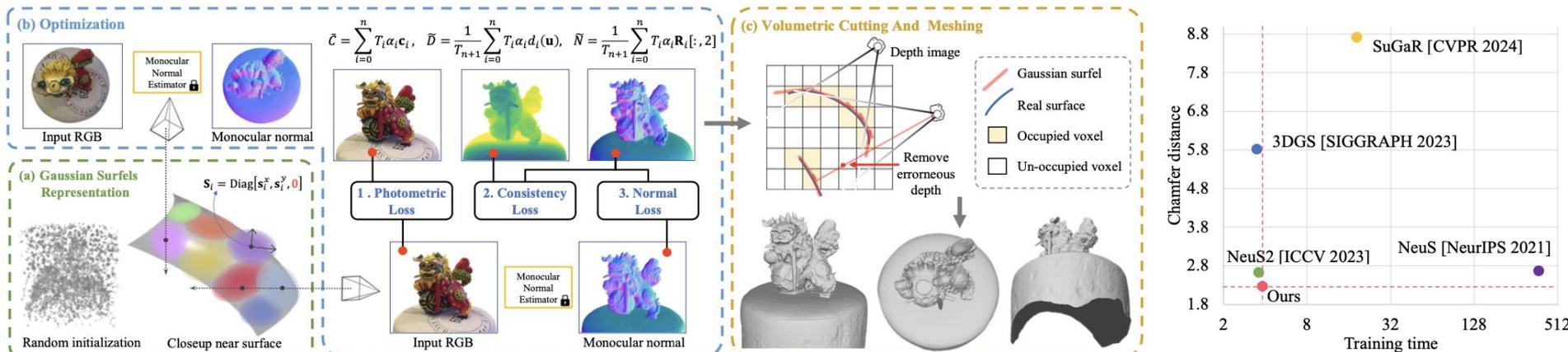
- 3D Gaussian Splatting (3DGS)の最適化手順とsurfelsの表面整合性を組み合わせた新しい表現方法、Gaussian surfelsを提案
- 3D Gaussianのz軸スケールを0に設定し、3D楕円体を2D楕円に平坦化
- 物体形状の詳細な再構成と効率的なレンダリングの両立を目的とする

## 新規性

- 3D Gaussianを2D楕円に平坦化し、明示的に法線方向を定義
- レンダリングされた深度マップから計算される法線と、レンダリングされた法線マップの一貫性を保つself-supervised損失を提案
- Gaussian surfelsの距離情報を用いたvolumetric cuttingを提案し、物体エッジ付近の深度推定精度を向上

## 結果

- BlendedMVSで3DGS、SuGaRよりChamfer distanceが大幅改善
- レンダリング品質において通常設定でNeuS2を上回る
- 学習時間はNeuSより大幅に高速化、3DGS/NeuS2と同等



# DGD: Dynamic 3D Gaussians Distillation

会議 : ECCV 2024

著者 : Isaac Labe, Noam Issachar, Itai Lang, Sagie Benaim

## 概要

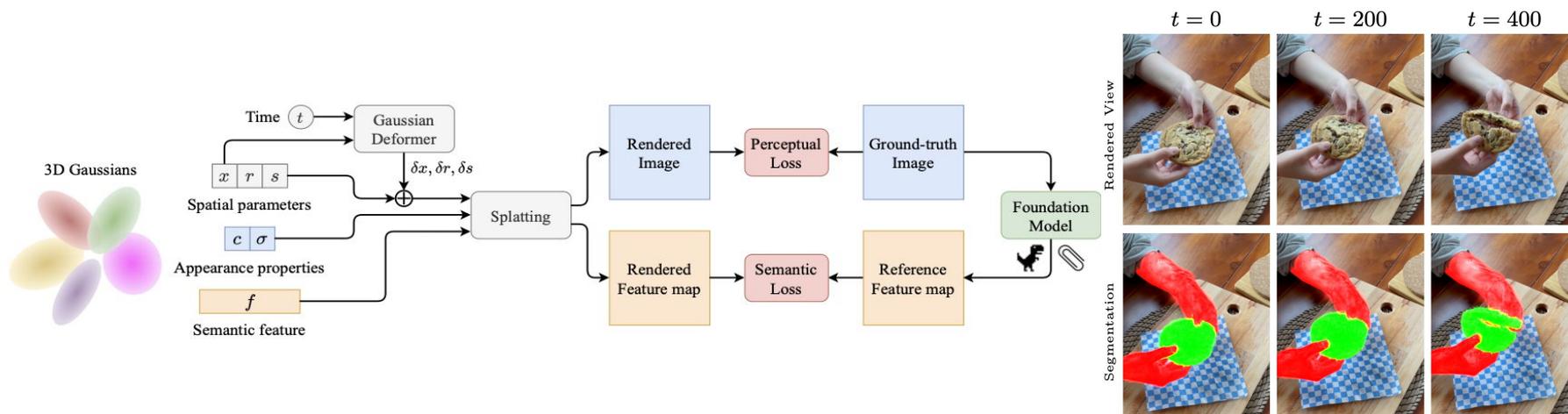
- 動的3Dシーンのセマンティクスと外観を同時に表現する3D Gaussian Splatting (3DGS)ベースの手法
- 単一の単眼ビデオから、動的な3DセマンティックRadiance Fieldを学習
- テキストプロンプトなどによる直感的なインターフェースで物体のセマンティックセグメンテーションと追跡を実現

## 新規性

- 従来の動的3DGS手法では扱えなかったセマンティクスを、外観とジオメトリと同時に表現
- 各3D GaussianにセマンティックLLM特徴量を組み込み、動的シーンでのセマンティックセグメンテーションを実現

## 結果

- セグメンテーション精度で従来の2D手法(LSeg)を大幅に上回る
- ユーザー評価においてベースラインを上回るセグメンテーション品質と視点間の一貫性を達成
- レンダリング速度は3DGSと同等(200Kガウシアン使用時に少なくとも30 FPS)



# Segment Any 4D Gaussians

会議 : arXiv

著者 : Shengxiang Ji, Guanjun Wu, Jiemin Fang, Jiazhong Cen, Taoran Yi, Wenyu Liu, Qi Tian, Xinggang Wang

## 概要

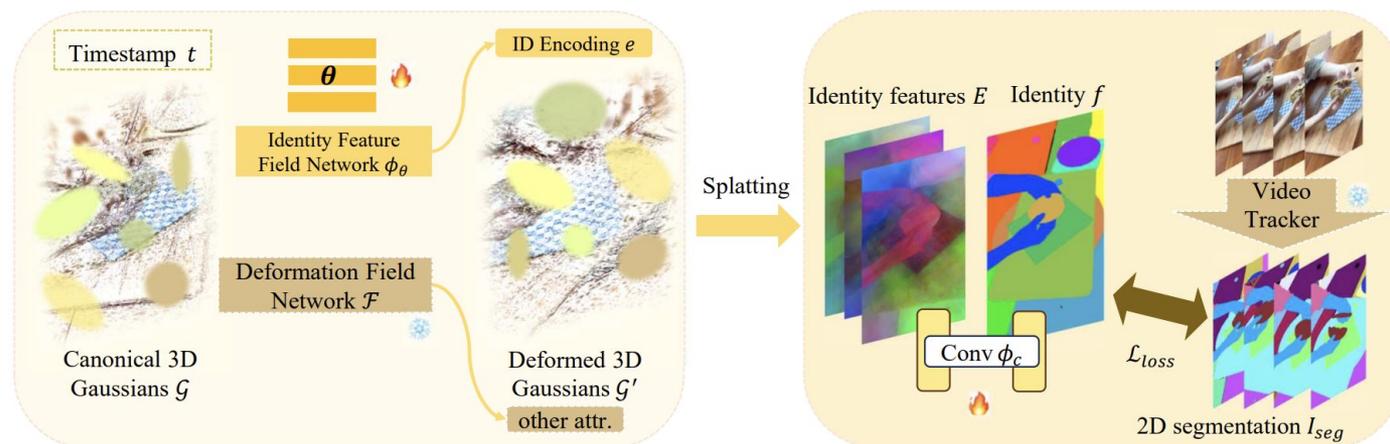
- 4Dガウシアンに基づいて4Dデジタル世界のあらゆるものをセグメンテーション

## 手法

- ノイズの多い疎な2D特徴マップからCanonicalな3D特徴場を学習
- Splatした2D特徴画像を精密化する2DCNNを学習

## 結果

- RTX 3090 GPUを使用して10秒以内の高速なインタラクティブセグメンテーションを実現



# NeRF-RPN: A general framework for object detection in NeRFs

会議 : CVPR2023

著者 : Benran Hu, Junkai Huang, Yichen Liu, Yu-Wing Tai, Chi-Keung Tang

## 概要

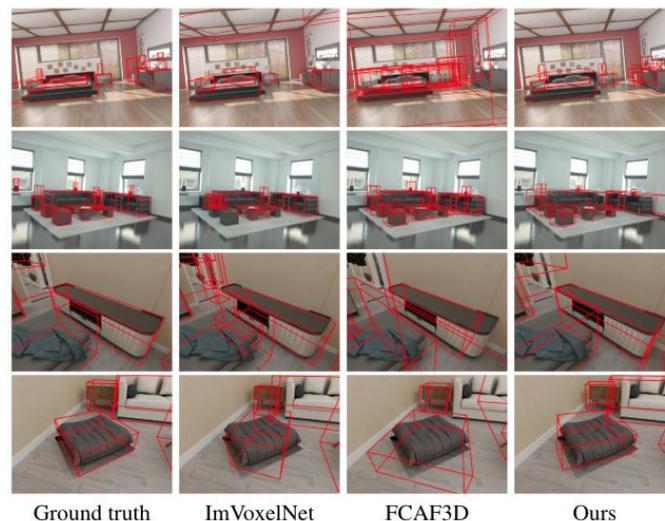
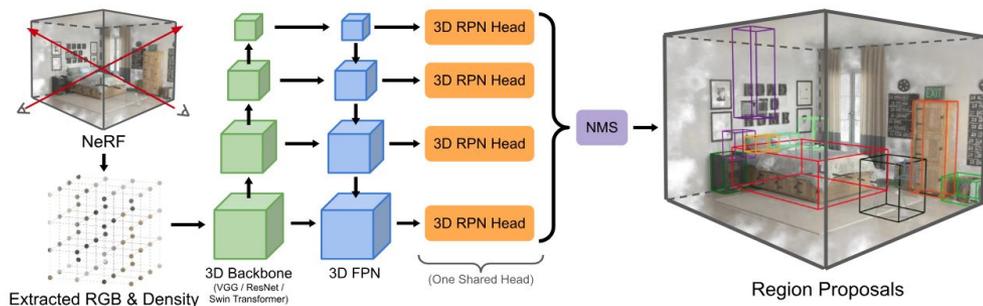
- NeRF上で直接動作する3次元物体検出のフレームワークを初めて提案
- NeRFで3次元物体検出を学習するためのデータセットを構築

## 手法

- NeRFで学習された空間を一定間隔でグリッド化し, サンプリング
- サンプリングした空間に対して3D FPNをかけることで物体を検出

## 結果

- Hypersimを用いた実験で良好な結果を達成



# NeRF-Det: Learning Geometry-Aware Volumetric Representation for Multi-View 3D Object Detection

会議 : ICCV2023

著者 : Chenfeng Xu, Bichen Wu, Ji Hou, Sam Tsai, Ruilong Li, Jialiang Wang, Wei Zhan, Zijian He, Peter Vajda, Kurt Keutzer, Masayoshi Tomizuka

概要

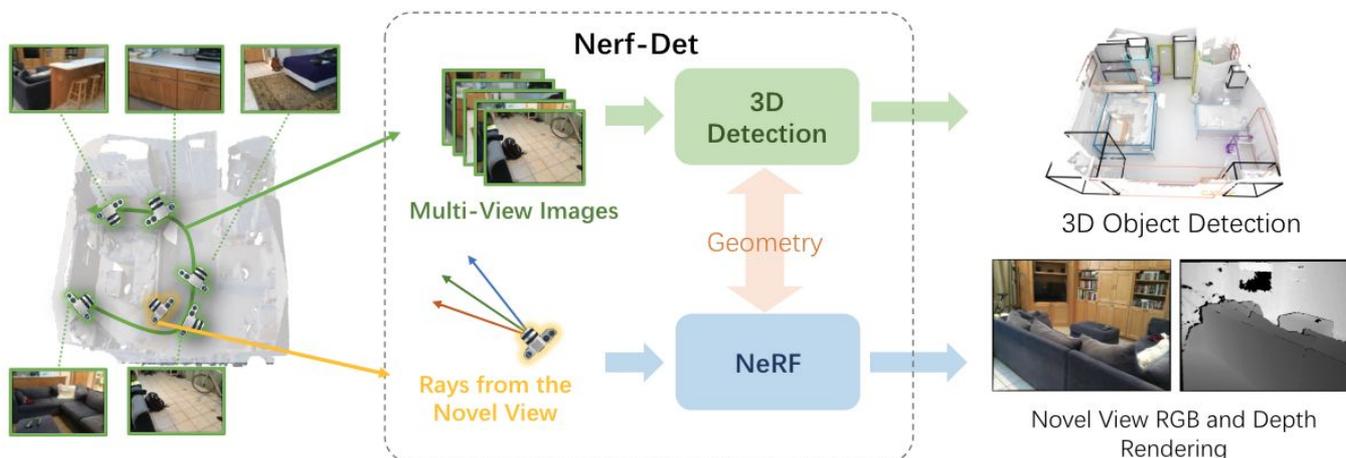
- 複眼の3次元物体検出とNeRFを相互に接続することで3次元物体検出/NeRFの精度が向上

手法

- NeRFの密度フィールドの推定結果と3次元物体検出による密度フィールドの推定結果を相互にマージ
- 3次元物体検出で得られた特徴量をNeRF側に取り入れることで事前分布を獲得

結果

- 既存の3次元物体検出と比較して精度向上, NeRF-RPNよりも格段に高速化



# NeRF-Det++: Incorporating Semantic Cues and Perspective-aware Depth Supervision for Indoor Multi-View 3D Detection

会議 : arxiv (2024)

著者 : Chenxi Huang, Yuenan Hou, Weicai Ye, Di Huang, Xiaoshui Huang, Binbin Lin, Deng Cai, Wanli Ouyang

## 概要

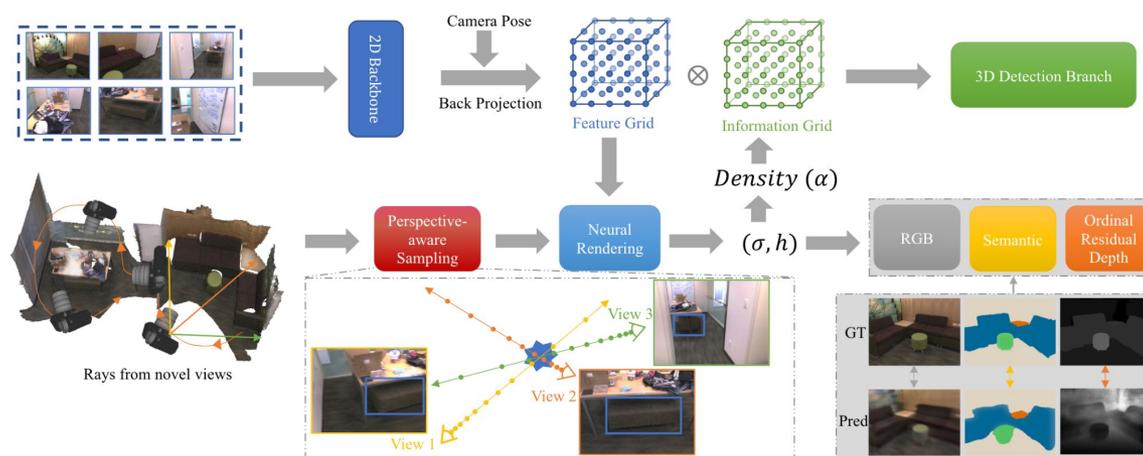
- NeRF-Detにおける意味理解, 空間のサンプリング方法, 深度情報の活用方法それぞれについて改善を行うことで精度を改善

## 手法

- Semantic Segmentationも学習させることで物体の誤検出を抑制
- 3次元空間のサンプリングを一定間隔ではなくカメラ周辺に対して密に行う
- 深度情報を直接回帰で推定せず, 深さのレベルで分類し, その後回帰を行う

## 結果

- ScanNetV2とARKITScenesで既存手法と比較したところ優位性を確認



# NeRF-MAE: Masked AutoEncoders for Self-Supervised 3D Representation Learning for Neural Radiance Fields

会議 : arxiv (2024)

著者 : Muhammad Zubair Irshad, Sergey Zakahrov, Vitor Guizilini, Adrien Gaidon, Zsolt Kira, Rares Ambrus

## 概要

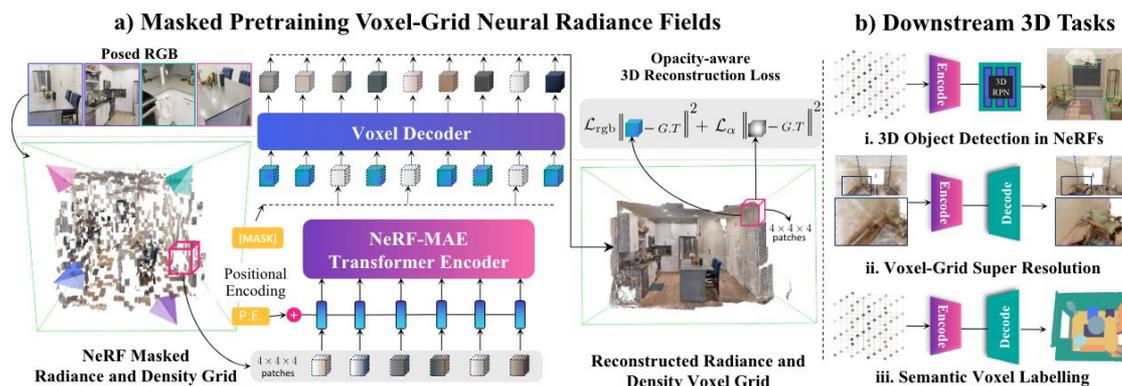
- NeRFで表現された空間に対してマスキング, 再構成を行う自己教師あり学習を用いた事前学習方法を提案
- 学習されたEncoder, Decoderを使用することでいくつかの3次元認識の下流タスクに活用可能

## 手法

- NeRF空間における輝度と密度をEncoderに入力, その位置におけるボクセル表現をDecoderで出力

## 結果

- NeRF-MAEで事前学習したモデルで3次元物体検出, 超解像などのタスクで既存手法を上回る精度を達成



# メタサーベイ

Transferability metricの長所と、それを生かした「これから」の動向

# メタサーベイの目次

---

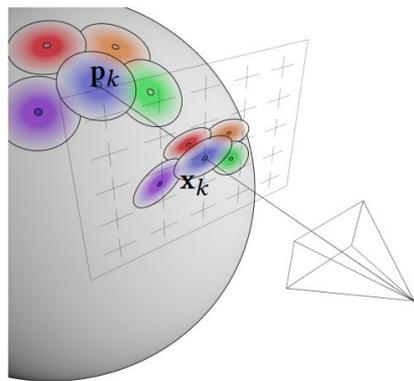
- これまでの知見からわかること
- これまでの知見でこれからの推移として予想できるもの

# 形状表現と微分可能レンダリング

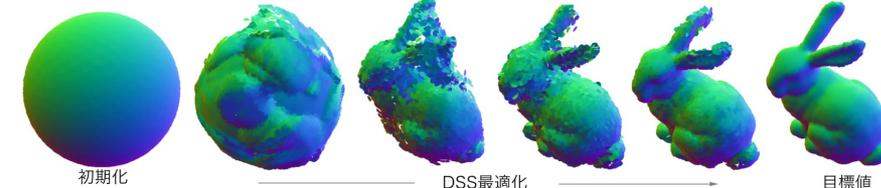
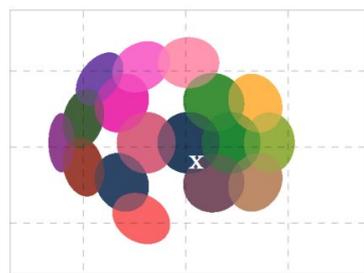
「3D形状とカメラパラメータから2D画像を得る」計算の  
微分可能性に着目

→2Dの観測を教師に3D形状やカメラ姿勢を最適化

Bundle Adjustmentの適用範囲を拡張する研究分野



EWA



GSS

点群、メッシュ、ボクセルの各3D形状表現で模索された



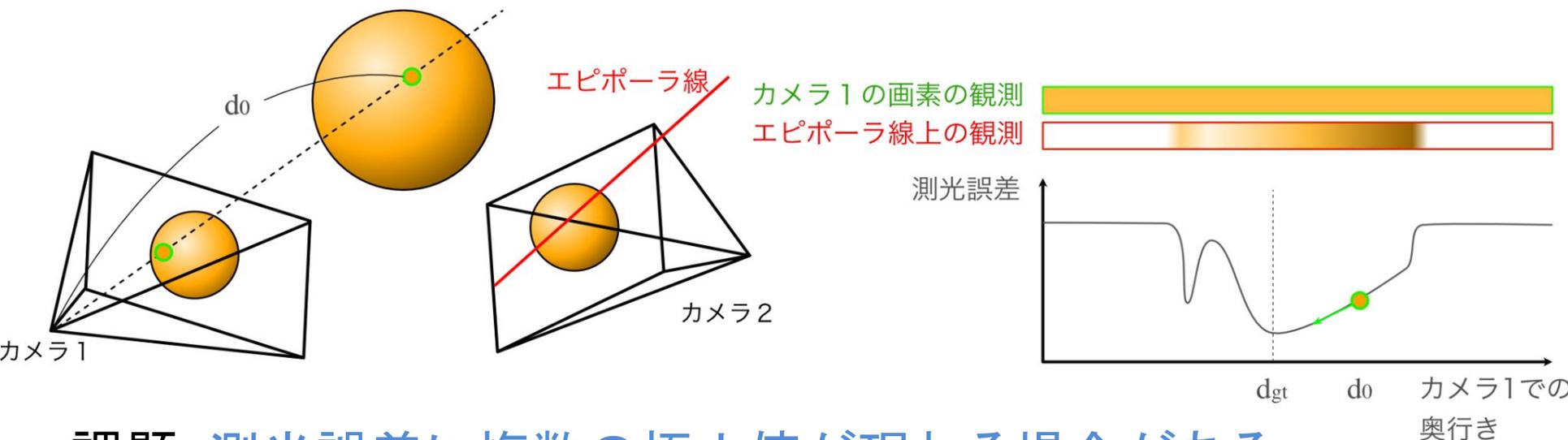
cvpaper.challenge

# レンダラ: NeRF以前の微分可能レンダリング

点群の例(DTAM/DSOなど)

カメラ1でオレンジ色に観測された3次元点は、

カメラ2のエピポーラ線上の色が近い方向へ最適化される



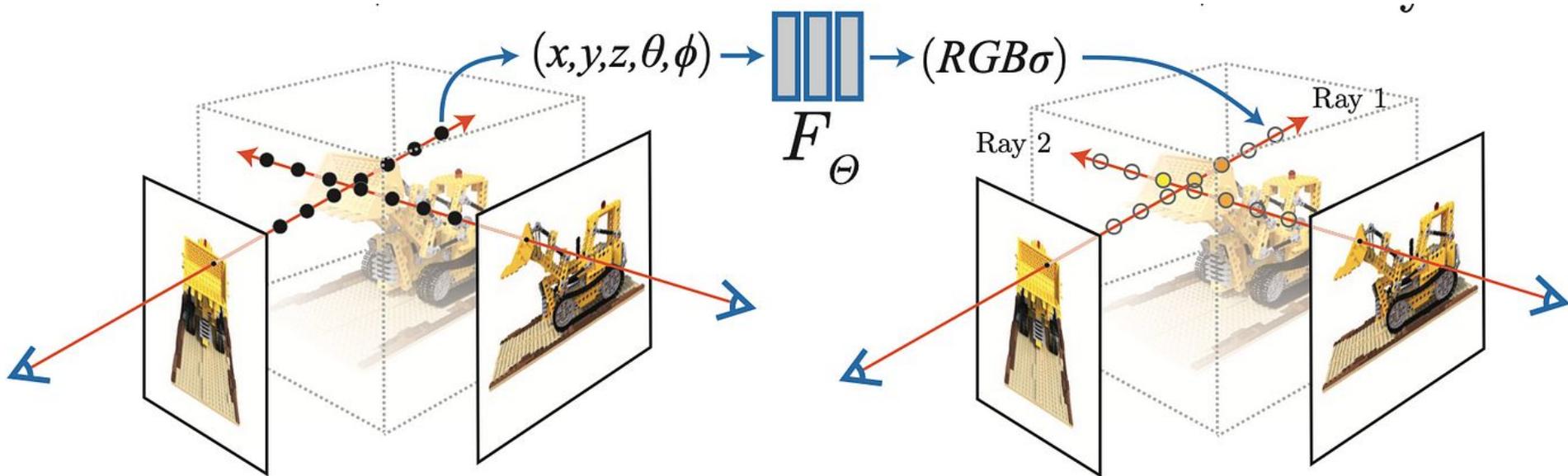
課題: 測光誤差に複数の極小値が現れる場合がある

- 被写体に繰り返しテクスチャパターンがある
- 似た色の被写体が複数ある

# レンダラ: NeRF

3D形状を「物体のある座標(Explicit)」ではなく

「ある座標に物体がある可能性 (Implicit)」で記述する

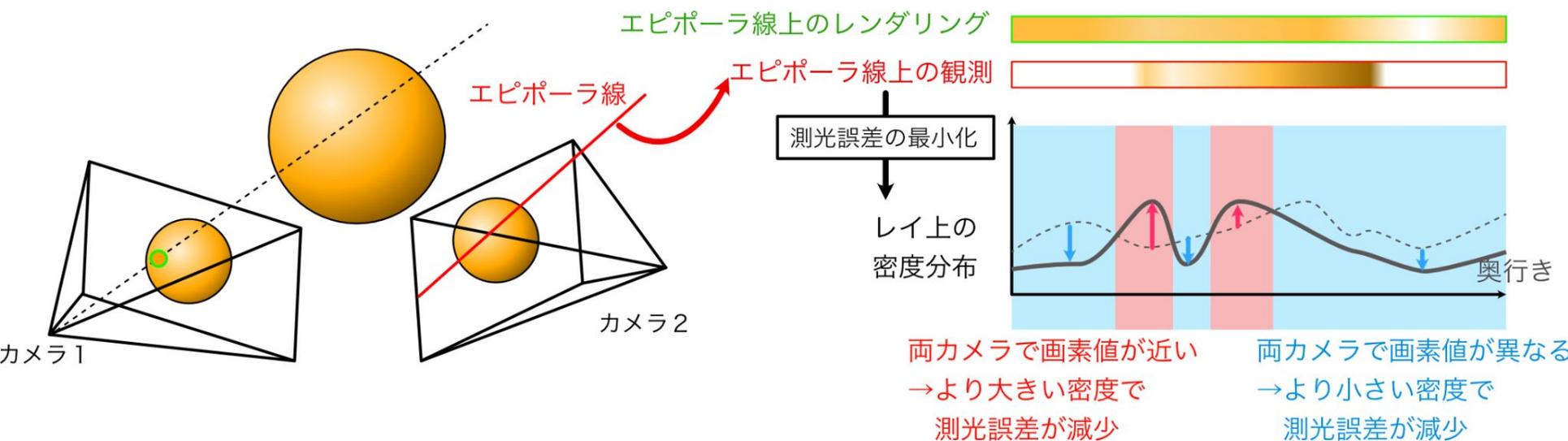


Implicit表現を2Dに集約できる微分可能レンダラとして

**ボリュームレンダリング** を導入

# レンダラ: NeRFはなぜ収束できるのか

あるカメラの観測では複数の局所解が区別できない場合に  
その両方の可能性を残すように最適化

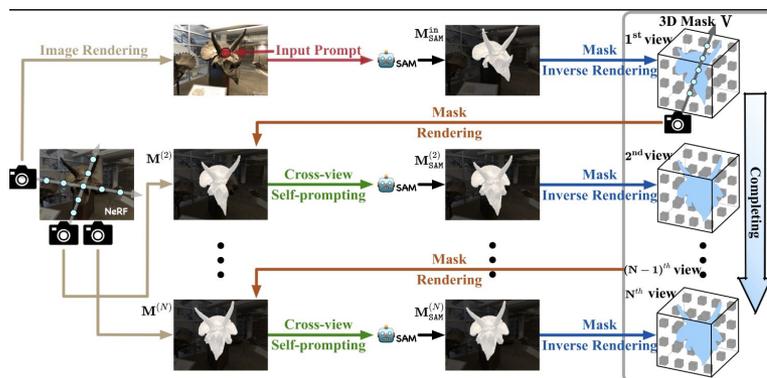


不正確なカメラ姿勢や地形からでも適切な最適化方向が取得可能！

# レンダラ: NeRFの収束性の強み

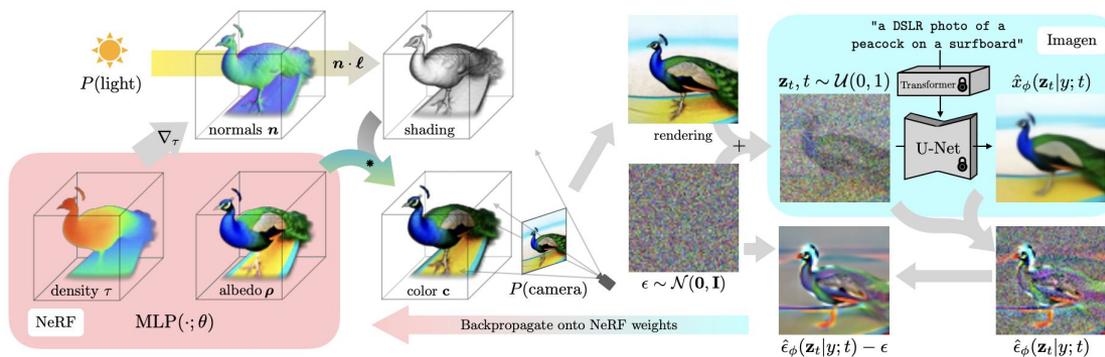
NeRFは乱数で初期化された(事前情報のない)MLPから  
3D形状を最適化する勾配を正しく与えられる

一貫性が完全でない2D特徴を3Dに立ち上げられる (ex: [SA3D](#))



ノイズ状の初期出力画像でも学習が成立

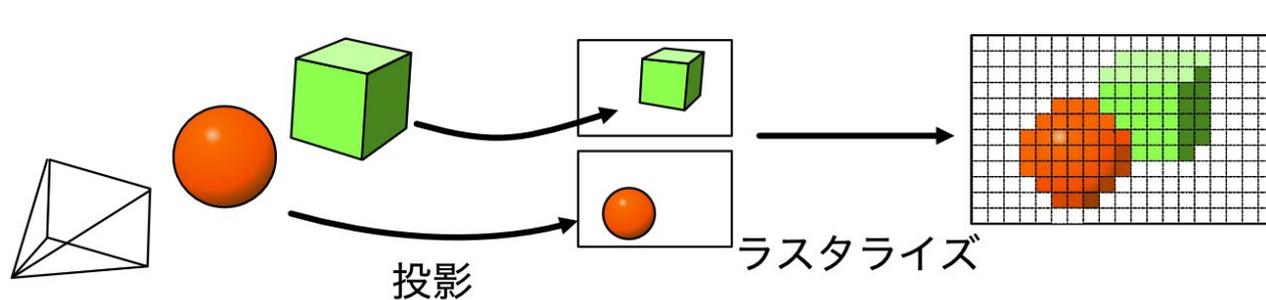
→ 拡散生成モデルのパイプラインと接続できる (ex: [DreamFusion](#))



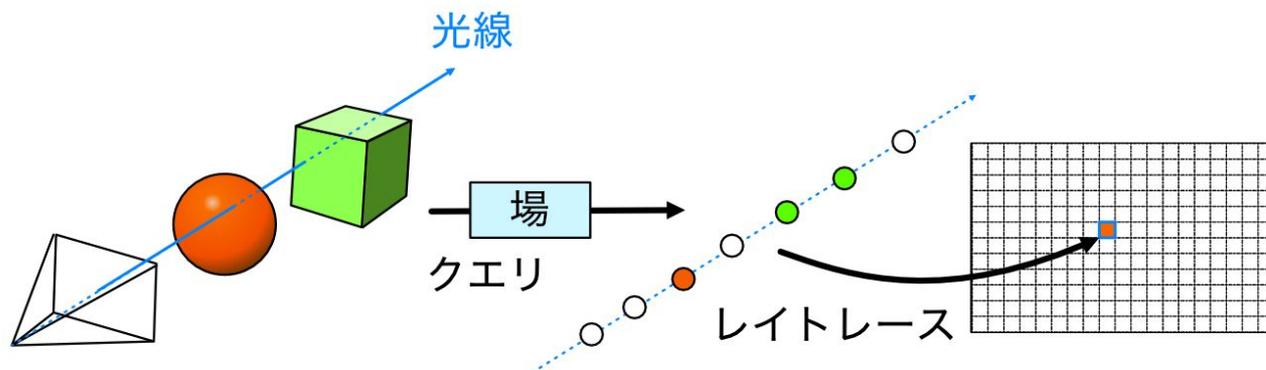
# レンダラ: NeRFの課題

Implicitな表現の集約に必須のボリュームレンダリングは  
レイトレースの形態が必須

→レンダリングコストが高い



プリミティブの数  
(点・メッシュ数)だけ  
ループ



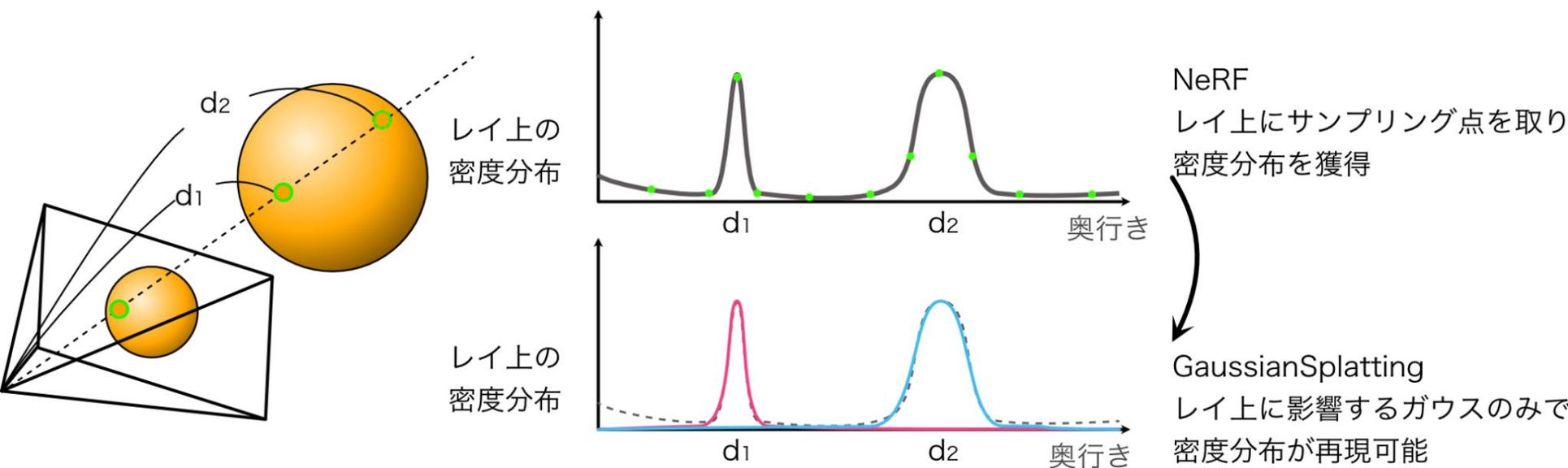
ピクセル数×サンプル数  
ループ

# レンダラ: Gaussian Splatting

Gaussian SplattingではNeRFの密度分布を混合ガウスで記述  
ガウス同士の重複が無視できる時、それぞれの画素値への影響を独立に計算できる

→ボリュームレンダリングをラスタライズで近似可能

→レンダリングの大幅な高速化を実現



# レンダラ: Gaussian SplattingのLimitation

---

## (1)局所解

各ガウス分布の中心を明示的に保持するため、

局所解を脱出できない

→高品質な初期値を与えるか、各局所解に冗長に配置する  
ような初期化が必要

対策研究→[COLMAP-free 3DGS](#)

## (2)近似品質

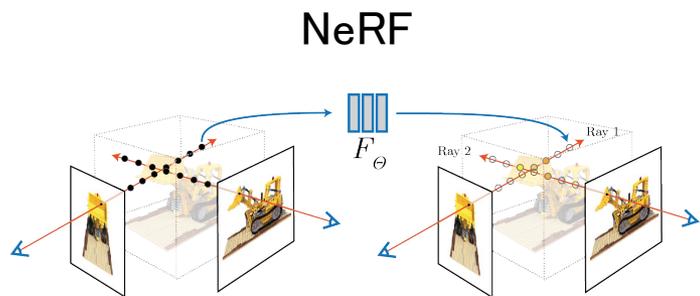
「ガウス同士の重複が無視できる」仮定はあまり成り立たない

→Sortの妥当性や、視点間での一貫性が保証されない

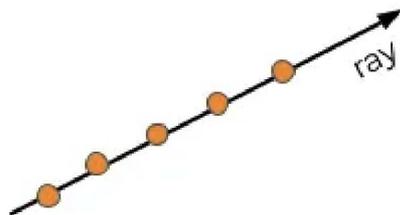
対策研究→[StopThePop](#)

# 形状表現: Gaussian SplattingとNeRFの違い

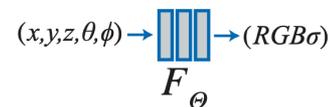
## 論文の背景にある知識等に着目



MLPを用いたImplicitな表現



Rayに沿って連続的なMLPのクエリを行った結果を描画

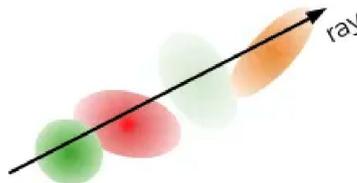


MLPのweightが必要

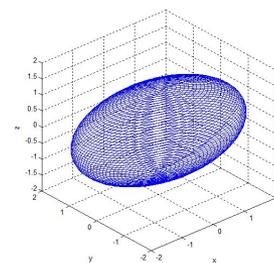
## Gaussian Splatting



3D Gaussianを用いたExplicitな表現



与えられたレイに関連する離散的な3D Gaussianの集合をブレンドした結果を描画



各3D Gaussianのパラメタ  
i.e 位置,共分散行列,色,透過度  
が必要

# 3DGS関連研究の動向

---

研究動向:

- NeRFの研究を3DGSへの適用
- インクリメンタルな改善

が多い/出るのが早い

NeRF to 3DGS

- 動機: 3DGSの高速な学習速度※・レンダリング速度を活かしてSOTA達成
  - LERF→LangSplat,
  - 4D NeRF→4DGS
  - INGP→FMGS

インクリメンタルな改善

- 焦点: 3DGSの特定の側面を強化

例: ストレージ圧縮アルゴリズムの高度化 , レンダリング技術の最適化 ...

※収束性はNeRFの方が優れている

# 3DGS vs 他の3D表現手法

3DGSでは数十万～数百万のGaussianを用いてシーンを表現



各Gaussianは約250バイト

他の3D表現手法と比較して必要とするモデルサイズが大きい。  
この問題は実世界応用の障壁となる(特にエッジデバイスなど)



シーンをできる限り圧縮したい！

	メッシュ	点群	ボクセル	NeRF	3DGS
品質	◎	※	△	◎	◎
レンダリング速度	◎	◎	◎	△	◎
メモリ(ストレージ)効率	◎	△	△	○	△
3D再構築	△	◎	◎	◎	◎

◎:良い/早い/容易

○:やや良い

△:悪い/遅い/難しい

※:不可能

# 3DGS × モデルサイズ圧縮のトレンド

## ➤ ベクトル量子化

[C3DGS\(CVPR2024\)](#):3DGSの各パラメータを量子化とファインチューニングにより圧縮し、最大約31倍の圧縮率

## ➤ プルーニング

[Compact-3DGS\(CVPR2024\)](#):Gaussianの体積と透明度に基づくマスキングを行い冗長なGaussianを取り除く。さらにニューラルフィールドによる色パラメータの表現、ベクトル量子化。最大約25倍の圧縮率

[Trimming the Fat\(arXiv 2024/6/26\)](#):勾配情報を活用した反復的プルーニングとFine-tuningのアプローチ。ベクトル量子化と併用すると約50倍の圧縮率。

## ➤ Neural Gaussian

[Scaffold-GS\(CVPR2024\)](#):SfMから得られたanchor pointsを中心に、局所的なNeural Gaussiansを生成。各Neural GaussianのパラメータはMLPを用いて推論。マルチスケールのシーンや新しい視点での優れた性能。最大約10倍の圧縮率。

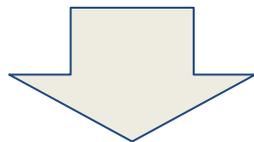
# 3DGS × モデルサイズ圧縮: ベクトル量子化

## ➤ ベクトル量子化

3DGSでは大量のGaussianを用いてシーンを表現

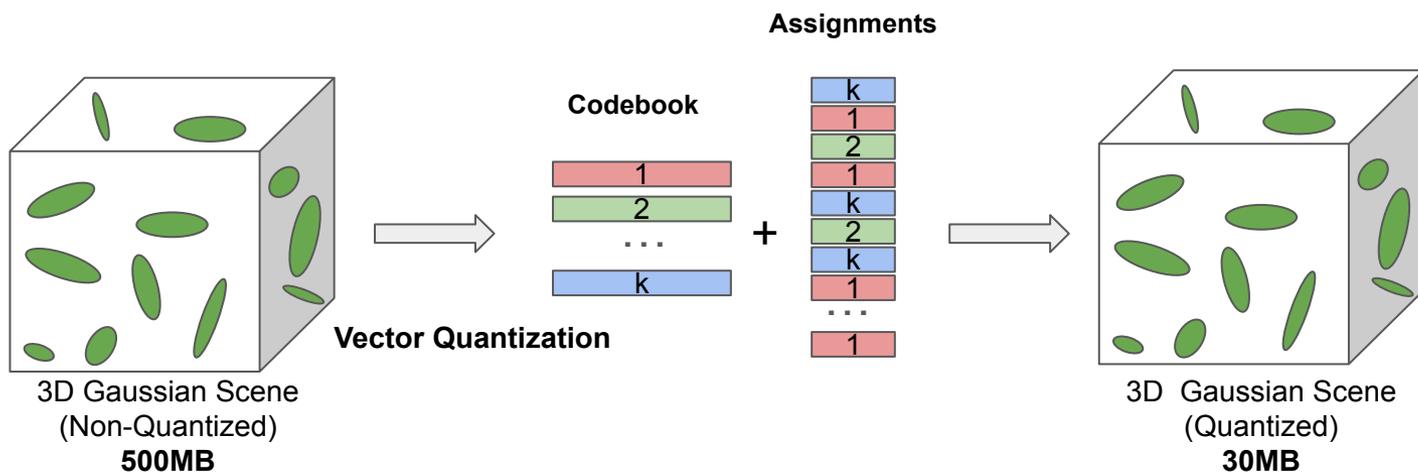
Gaussianの数が多いので似たような(パラメータを持つ)Gaussianがほぼ必ず存在する

→これらの似たようなGaussianのパラメータを、代表的なパラメータ(クラスタ中心など)で置き換えればいいんじゃないか？



結果的に

- **代表的なパラメータのみ** を保存するコードブック
- どのGaussianにどのパラメータを割り当てられるかという情報のみで、シーンの表現が可能となる。



# Text to 3Dに関する気づき

- [Dream Fusion](#)をはじめ多くのText to 3DモデルではNeRFやGaussian Splattingを使用したものがほとんど
  - Text to 3D Point CloudやText to 3D Meshは少数？
- Text to 3Dモデルの流れ
  1. Text Promptを入力
  2. Stable DiffusionなどのText image modelにより画像を生成
  3. NeRF/Gaussian Splattingに生成画像を入力
  4. 3Dモデルを生成



# Text to 3DのLimitation

- [Dream Fusion](#)や[GSGEN](#)など多くのText to 3DモデルでStable Diffusionを使用
  - Stable Diffusionの性能に大きく依存
  - ヤヌス問題
    - 3Dモデルが複数の面を持ち、ジオメトリが崩れてしまうという重大な問題
    - 単一視点画像がStable Diffusionにより生成されることが問題?



Stable DreamFusion [50, 69]

threestudio ProlificDreamer [23, 77]  
A DSLR photo of a panda



Fantasia3D [12]

threestudio ProlificDreamer [23, 77]  
A high quality photo of a furry corgi

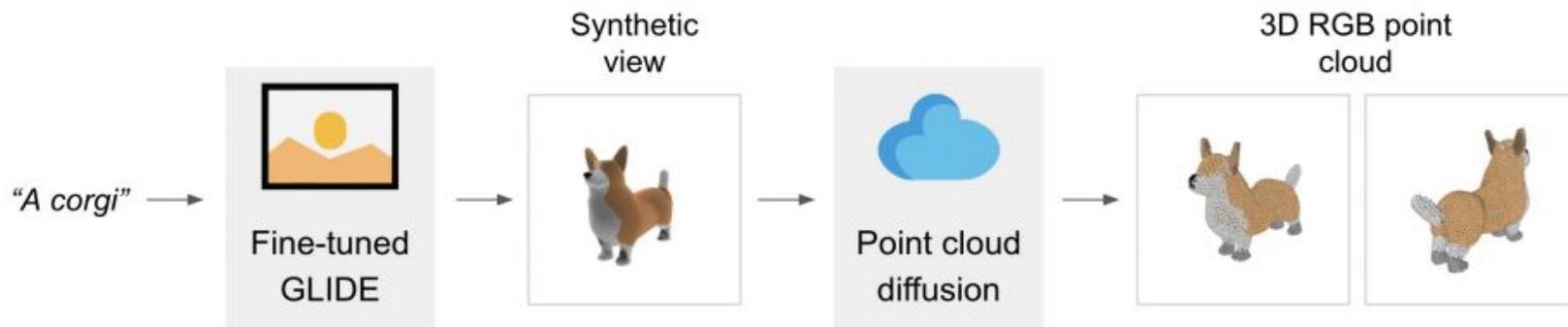
さらに、実用性を考慮すると生成されるデータは

- Point cloud
- Mesh

が好ましい

# Text to 3Dの実用性

- Text to Point cloud/Meshモデルはないのか？
  - [Point-E](#)
    - Text to Point cloudが可能
  - [Shap-E](#)
    - Text to Meshが可能
- NeRFやGaussian Splattingを使用したモデルとの違い
  - Text to Point cloud/Meshモデルでは3Dデータセットが必要
    - Point-Eの概要におけるPoint cloud diffusionの段階で3Dデータセットを学習
    - Point-EやShap-Eで使用している3Dデータセットは非公開
      - 学習データがクリアなモデルの必要性

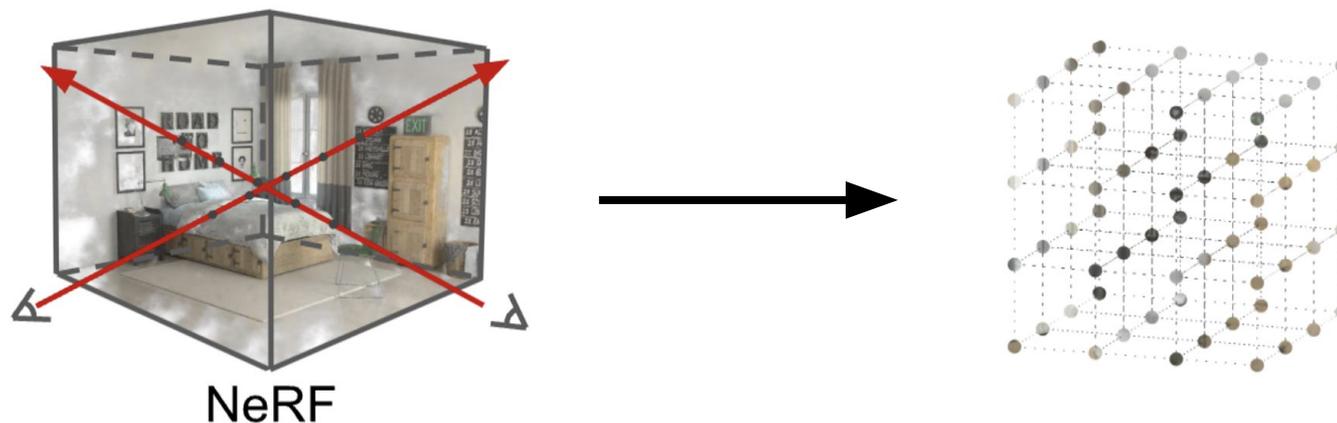


Point-Eの概要

# 3D空間認識への応用:NeRF由来のボクセル表現

NeRF空間をボクセル表現(離散化した各座標における色の平均と密度を抽出したもの)に変換

→ 3D CNNなどの一般的な3D 空間認識手法へ入力可能な形式に

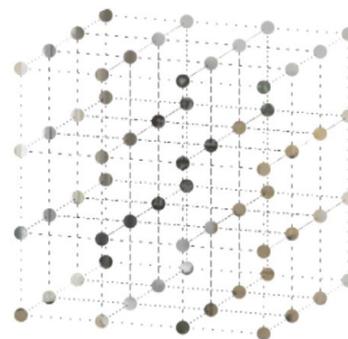
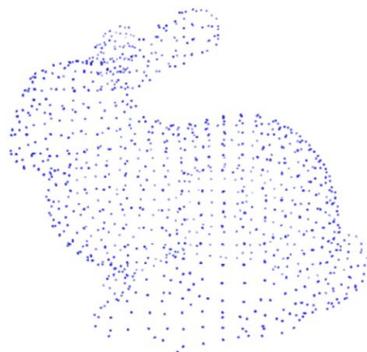


# 3D空間認識への応用: ボクセル表現の利点

## 点群由来のボクセル表現とNeRF由来のボクセル表現の比較

	表現の対象
点群由来のボクセル表現	物体の表面のみ
NeRF由来のボクセル表現	空間全体(物体の内部などを含む)

→ 物体の内部情報は主に物体認識タスクで効果的



# 3DGSにおける注目研究者

## 3DGS 著者ランキング Top3 (2024/07/11時点)

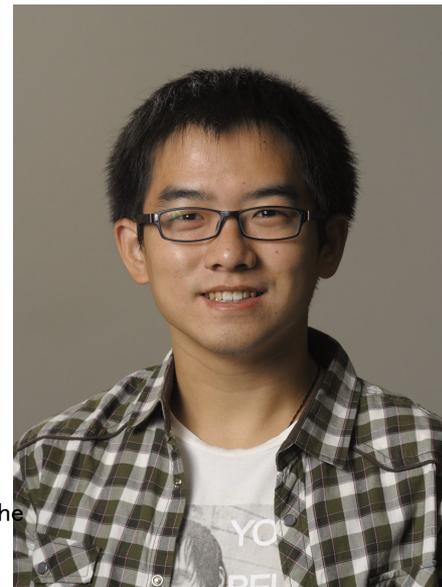
参照先: <https://github.com/MrNeRF/awesome-3D-gaussian-splatting>

Rank	Name	Papers
1	Zhangyang Wang	9
2	Yebin Liu	6
	Zhiwen Fan	
	Dejia Xu	
	Jiemin Fang	
	Lingxi Xie	
	Xiaopeng Zhang	
3	Andreas Geiger	5
	Chen Yang	

# 3DGSにおける注目研究者

- Zhangyang Wang (VITA, UT Austin)
  - 3DGS元論文が公開されて約1年弱で9本論文を公開済
  - 代表論文
    - 3D再構成: InstantSplat, NeuralLift-360
    - NVS: GNT, SinNeRF
  - VITAの特色
    - 教員と学生が密に議論している? [以下WEBより引用](#)

We have a very flat management structure with minimal communication overhead (“Talk is cheap. Show me the code/math”). Every student works directly and closely with Dr. Wang. The group also benefits a lot from its highly interactive, intimate, and helpful culture.



<https://vita-group.github.io/group.html>

## ECCV2024にも3DGSで3本採択!!

- Z. Zhu\*, Z. Fan\*, Y. Jiang\*, and Z. Wang  
“FSGS: Real-Time Few-shot View Synthesis using Gaussian Splatting”  
European Conference on Computer Vision (ECCV), 2024. [\[Paper\]](#) [\[Code\]](#)
- S. Zhou, Z. Fan\*, D. Xu\*, H. Chang, P. Chari, T. Bharadwaj, S. You, Z. Wang, and A. Kadambi  
“DreamScene360: Unconstrained Text-to-3D Scene Generation with Panoramic Gaussian Splatting”  
European Conference on Computer Vision (ECCV), 2024. [\[Paper\]](#) [\[Code\]](#)
- R. Li, Z. Fan\*, B. Wang, P. Wang\*, Z. Wang, and X. Wu  
“VersatileGaussian: Real-time Neural Rendering for Versatile Tasks using Gaussian Splatting”  
European Conference on Computer Vision (ECCV), 2024. [\[Paper\]](#) [\[Code\]](#)



# まとめ

---

- GaussianSplattingの影響が大きい
  - 微分可能レンダリングの速度面が実質解決
  - Explicitな表現になったことでNVSだけでなく識別やFusionへ応用が波及
  - 一方で収束の安定性等やモデルサイズ圧縮がNeRFに比べ未成熟
- 生成タスクへの応用が加熱
  - 2D教師のみで(3Dデータセットなしで)学習できる性質が有用
  - 今後3Dデータが増えた時にトレンドがどうなるかに注目