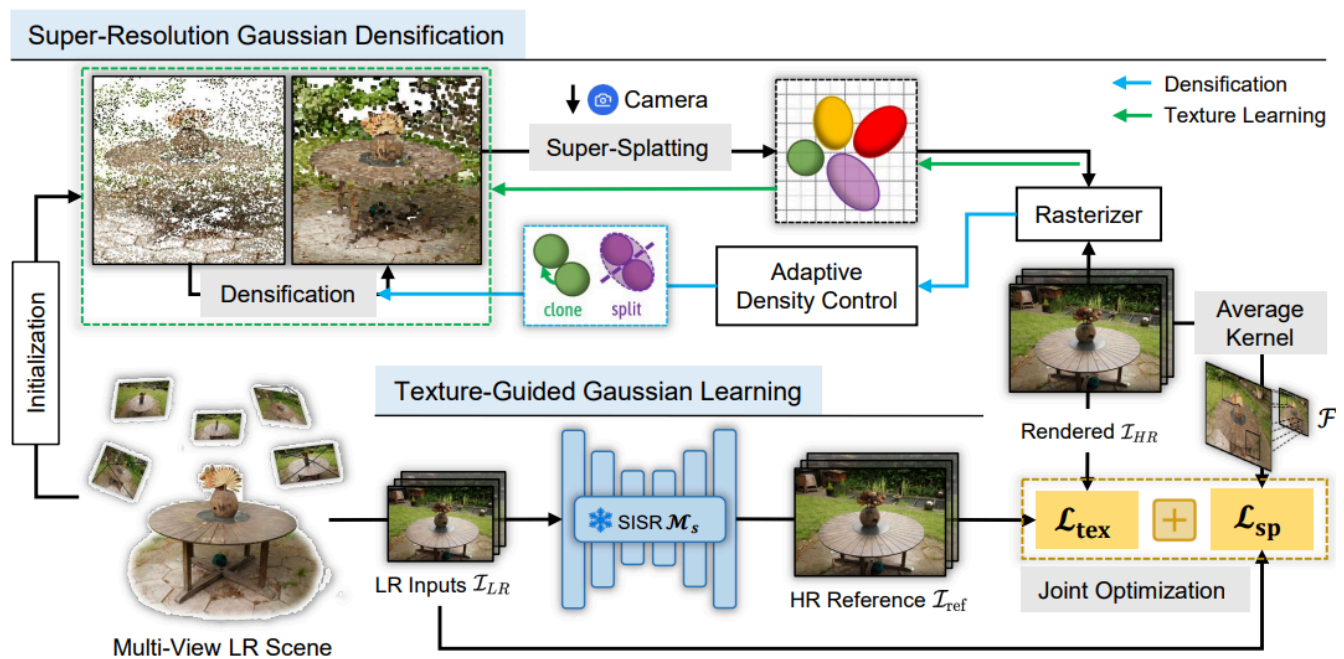


# SRGS: Super-Resolution 3D Gaussian Splatting

## 概要

- タスクはHRNVS(High Resolution Novel View Synthesis)
- 低画質画像で最適化するとsparsityとtexture deficiencyが問題
- 上記の問題を解決するためにSRGSを提案
  - Super-Resolution Gaussian Densification
  - Texture-Guided Gaussian Learning
- Mip-NeRF 360とTanks&TemplatesというデータセットにおいてSoTA達成
- HRNVSで初めて3DGSを使用
- 要約するとsplattingするときに解像度を上げて、2Dの学習済みモデルで生成したものを正解としたlossを定義しただけで、gaussianの最適化アルゴリズムは特にいじっていない



(論文より引用)

## Related work

- HRNVSに対してNeRFを使ったもの

- NeRF-SR
- CROC(sota)
- Mip-NeRF

## 3DGS復習

- 3DGSは3D colored Gaussian primitivesの集合である $\mathcal{G} = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N\}$ を学習する
- $\mathbf{g} \in \mathcal{G}$ は以下のように書ける

$$\mathbf{g}^{3D}(\mathbf{x} \mid \mu, \Sigma) = \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right)$$

- $\Sigma$ はscaling matrix  $S$ とrotation matrix  $R$ を用いて $\Sigma = RSS^\top R^\top$ と分解できる
- ここからRasterization renderingの説明
- カメラポーズの情報としてextrinsic matrix  $\mathbf{T}$ とprojection matrix  $\mathbf{K}$ が与えられたときに2D screen spaceにおけるposition  $\hat{\mu}$ と $\hat{\Sigma}$ は以下で定まる

$$\hat{\mu} = \mathbf{KT}[\mu, 1]^\top, \quad \hat{\Sigma} = \mathbf{JT\Sigma T^\top J^\top}$$

- ここで $\mathbf{J}$ はヤコビ行列
- このときピクセル $u$ でのgaussianは以下で定まる

$$\mathbf{g}^{2D}(\mathbf{u} \mid \hat{\mu}, \hat{\Sigma}) = \exp\left(-\frac{1}{2}(\mathbf{u} - \hat{\mu})^\top \hat{\Sigma}^{-1}(\mathbf{u} - \hat{\mu})\right)$$

- このときピクセル $\mathbf{u}$ の色は以下で定まる

$$\begin{aligned} \mathbf{C}(\mathbf{u}) &= \sum_{i \in N} T_i \mathbf{g}_i^{2D}(\mathbf{u} \mid \hat{\mu}_i, \hat{\Sigma}_i) \alpha_i \mathbf{c}_i \\ T_i &= \prod_{j=1}^{i-1} \left(1 - \mathbf{g}_j^{2D}(\mathbf{u} \mid \hat{\mu}_j, \hat{\Sigma}_j) \alpha_j\right) \end{aligned} \tag{1}$$

- lossは以下

$$\mathcal{L}_{gs} = (1 - \lambda) \mathcal{L}_1 + \lambda \mathcal{L}_{D-SSIM}$$

- また、以下の勾配 $\nabla g$ が閾値を超えるGaussianをsplitting, cloningの対象とする

$$\nabla g = \frac{1}{M^i} \sum_{k=1}^{M^i} \sqrt{\left(\frac{\partial L_k}{\partial u_{ndc,x}^{i,k}}\right)^2 + \left(\frac{\partial L_k}{\partial u_{ndc,y}^{i,k}}\right)^2} > \tau_{pos}$$

- $M^i$ は*i*番目のgaussianが関与している視点の個数

## 提案手法1. Super-Resolution Gaussian Densification

- Screen spaceに射影された2D Gaussian  $\{\mathcal{G}_k^{2D} \mid k = 1, \dots, K\}$ を超解像してHR novel views  $\mathcal{I}_{HR}$ を生成
- lossについては生成された画像をaverage kernelでダウンサンプリングしてgtと比較する
- あるピクセル*p*についてダウンサンプリング後の値は以下のように定義される

$$\mathcal{F}^p = \frac{1}{s^2} \sum_{r \in \mathcal{R}_s(p)} I_{HR}^r$$

- $s$ は超解像の倍率で $\mathcal{R}_s$ は*p*の近傍ピクセルの集合(これをsub pixelという)
- このときlossは以下のように定義される

$$\begin{aligned} \mathcal{L}_1^{sp} &= \frac{1}{|p|} \sum_{p \in p} \|\mathcal{F}^p - \mathcal{I}_{LR}^p\| \\ \mathcal{L}_{D-SSIM}^{sp} &= 1 - \text{SSIM}(\mathcal{F}, \mathcal{I}_{LR}) \\ \mathcal{L}_{sp} &= (1 - \lambda) \mathcal{L}_1^{sp} + \lambda \mathcal{L}_{D-SSIM}^{sp} \end{aligned} \tag{2}$$

## 提案手法2. Texture-Guided Gaussian Learning

- 学習済み2D SRモデルの出力を正解とするアイデア
- 学習済み2D SRモデルを $\mathcal{M}_s$ として、その出力を $\mathcal{I}_{ref} = \mathcal{M}_s(\mathcal{I}_{LR})$ とする
- このときlossを以下で定義

$$\mathcal{L}_{tex} = (1 - \lambda) \mathcal{L}_1(\mathcal{I}_{ref}, \mathcal{I}_{HR}) + \lambda \mathcal{L}_{D-SSIM}(\mathcal{I}_{ref}, \mathcal{I}_{HR})$$

## 損失関数

- 以上まとめて損失関数は以下

$$\mathcal{L} = (1 - \lambda_e) \mathcal{L}_{sp} + \lambda_e \mathcal{L}_{tex}$$

## experiments

- datasetはSYnthetic NeRF, Tanks & Temples, Mip-NeRF 360

- 比較対象は以下

Method	Synthetic NeRF $\times 2$			Synthetic NeRF $\times 4$			Tanks & Temples $\times 2$			Tanks & Temples $\times 4$		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
TensoRF [22]	31.45	0.952	0.049	28.01	0.910	0.113	27.14	0.907	0.148	26.82	0.896	0.174
NeRF-SR [48]	30.08	0.939	0.050	28.90	0.927	0.099	26.78	0.897	0.152	26.36	0.887	0.182
CROC [3]	—	—	—	30.71	0.945	0.067	—	—	—	—	—	—
3DGS [2]	24.64	0.923	0.064	20.31	0.852	0.121	25.84	0.937	0.093	24.02	0.905	0.125
Mip-splatting [56]	31.37	0.957	0.048	28.44	0.930	0.087	27.31	0.947	0.085	26.40	0.929	0.108
SRGS (Ours)	<b>32.67</b>	<b>0.965</b>	<b>0.036</b>	<b>30.83</b>	<b>0.948</b>	<b>0.056</b>	<b>28.44</b>	<b>0.950</b>	<b>0.085</b>	<b>28.18</b>	<b>0.938</b>	<b>0.102</b>
HR-3DGS	33.32	0.974	0.023	33.32	0.974	0.023	28.90	0.953	0.082	28.90	0.953	0.082

Method	Mip-NeRF 360 $\times 4$			Mip-NeRF 360 $\times 8$		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Mip-NeRF 360 [51]	24.16	0.670	0.370	24.10	0.706	0.428
Zip-NeRF [50]	20.87	0.565	0.421	20.27	0.559	0.494
3DGS [2]	20.71	0.619	0.394	19.59	0.619	0.476
Mip-splatting [56]	26.43	0.754	0.305	26.22	0.765	0.392
SRGS (Ours)	<b>26.88</b>	<b>0.767</b>	<b>0.286</b>	<b>26.57</b>	<b>0.775</b>	<b>0.377</b>

(論文より引用)

## 英単語

- facilitate : 促進する
- densification : 高密度化
- outperform : 上回る
- long-standing : 長く続いている
- intricate : 複雑な