Introduction aux NeRF, SDF et 3D GS

Séminaire LASTIG

Camille Billouard (CNES, IGN)

Historique

Photosculpture

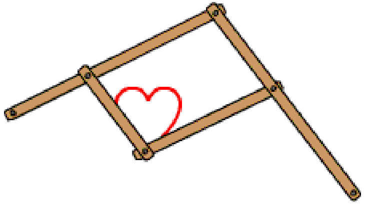
1859 …

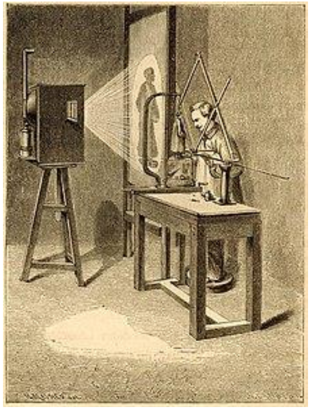
François Willème

Historique

Photosculpture

Pantographe !





1859 …

François Willème

Historique

Historique

Photosculpture

1859 …… 2020

NeRF

François Willème

Historique

SDF

Photosculpture

Qualité du rendu

1859 …… 2020 2021

NeRF

François Willème

Historique

SDF

Photosculpture

Qualité du rendu

1859 …… 2020 2021 2022

NeRF

Accélération de l'entraînement

François Willème

Historique

SDF

Photosculpture

Qualité du rendu

Dynamique, éditable, segmentation, …

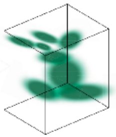
1859 …… 2020 2021 2022 2023

NeRF

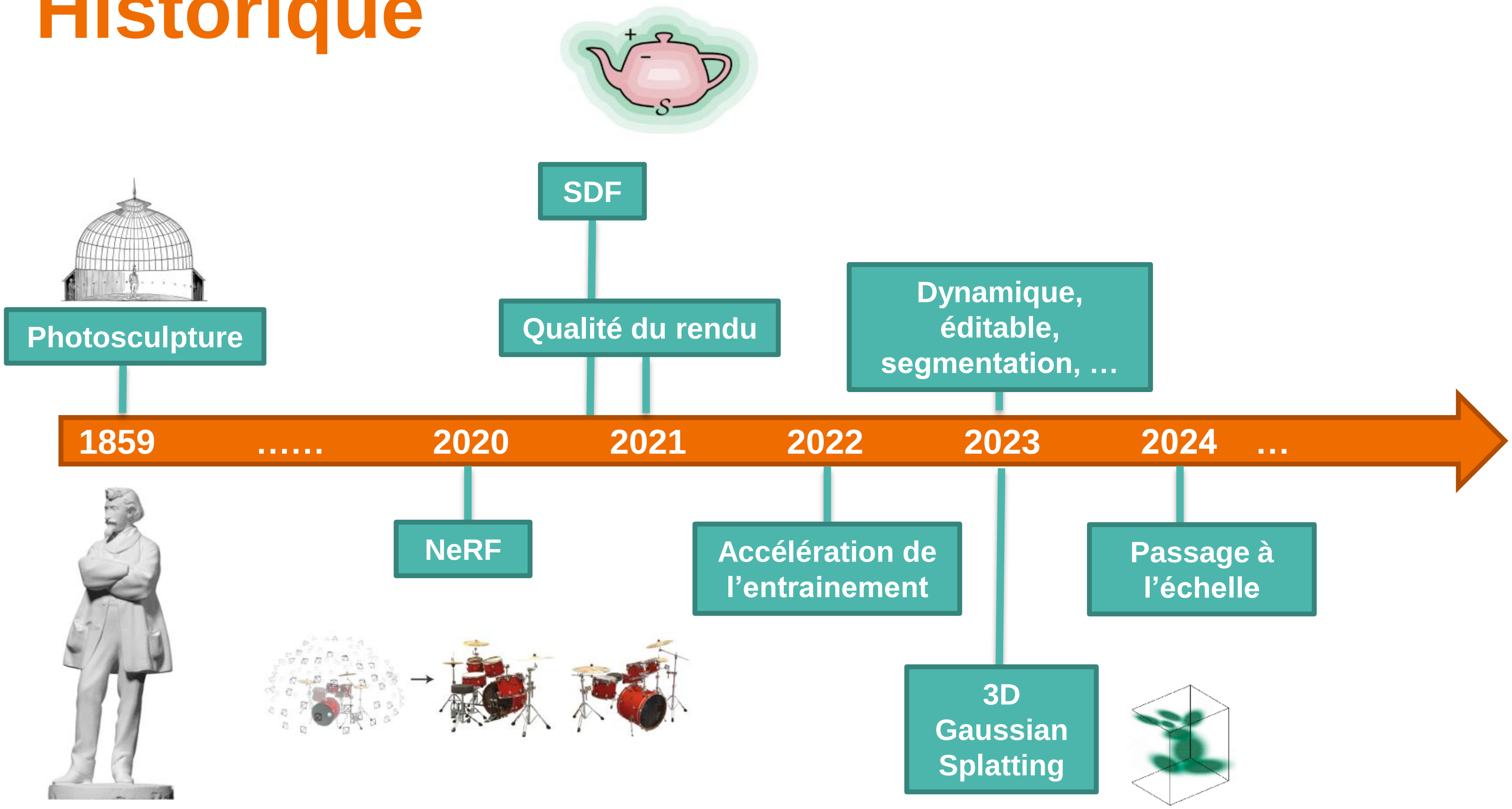
Accélération de l'entrainement

François Willème

3D Gaussian Splatting



Historique



François Willème

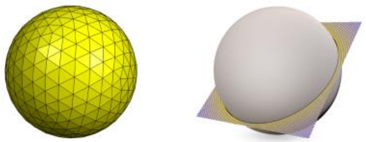
Terminologies

- Field : quantité définie pour des coordonnées spatiales et/ou temporelles
- Neural Net (MLP) : Théorème d'approximation universelle
- Neural Field : Champ paramétré par un MLP

Représentations de scènes

- Explicite : décrit directement la géométrie et la surface des objets à l'aide d'éléments numérotés

- Implicite : décrit une scène avec une fonction ou un *field,* en n'importe quel point de l'espace
  - SDF



Polygon Mesh $f(x, y, z) = \sqrt{x^2 + y^2 + z^2} - 1$

  - NeRF : radiance en tout point de la scène

Qu’est ce qu’un NERF ?

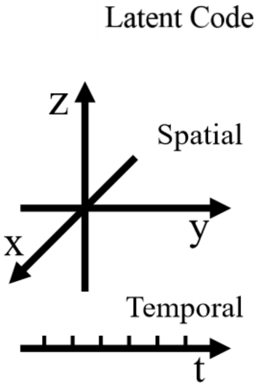
- Un Neural Radiance Field (NeRF) est une méthode basée sur l'apprentissage profond pour reconstruire une représentation 3D d'une scène à partir d'images orientées

- Un NeRF apprend à synthétiser des nouvelles vues de la scène

- L’apprentissage est spécifique (il doit être ré-appris) à chaque scène

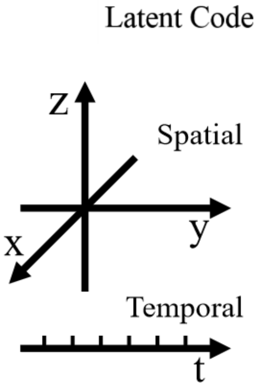
- Un NeRF contient la géométrie et les propriétés de réflectance de la scène.

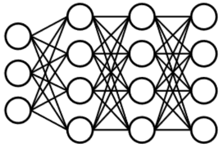
Framework



Echantillonnage des coordonnées

Framework



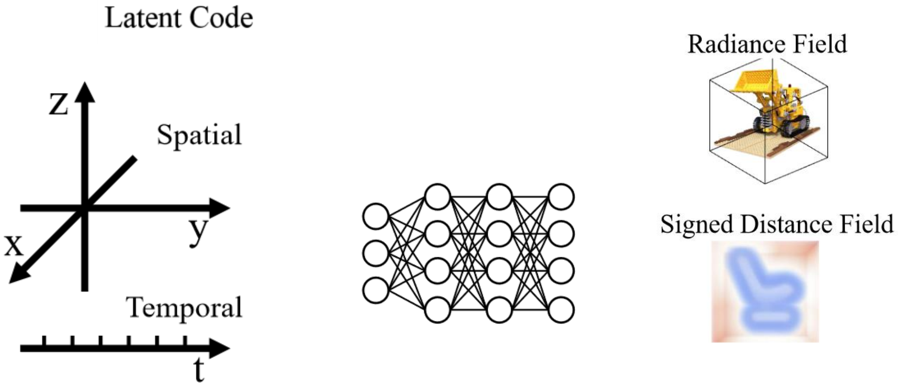


MLP

Echantillonnage des coordonnées

Framework

Ce que l'ont veut reconstruire



Echantillonnage des coordonnées

MLP

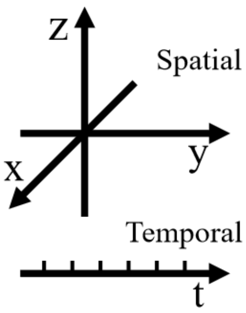
Domaine de reconstruction

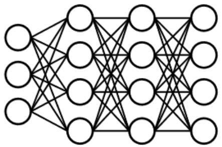
Framework

Ce que l’ont veut reconstruire

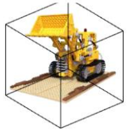
Comment on le reconstruit

Latent Code



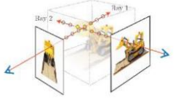


Radiance Field

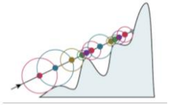


Signed Distance Field

Volume Rendering



Sphere Tracing



Echantillonnage des coordonnées

MLP

Domaine de reconstruction

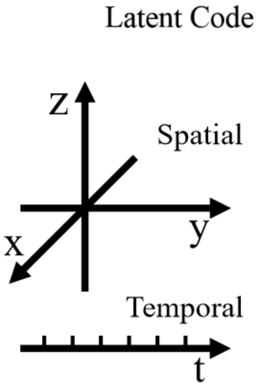
Rendering différentiable

Framework

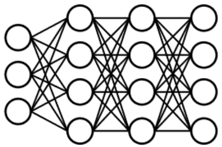
Ce que l’ont veut reconstruire

Comment on le reconstruit

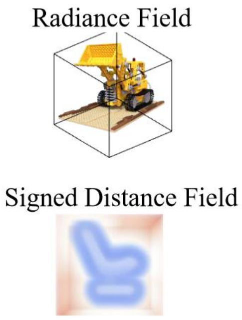
Ce qu’on observe/mesure



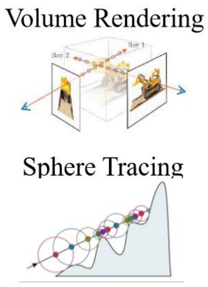
Echantillonnage des coordonnées



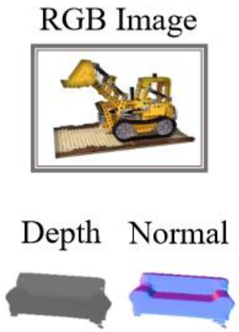
MLP



Domaine de reconstruction



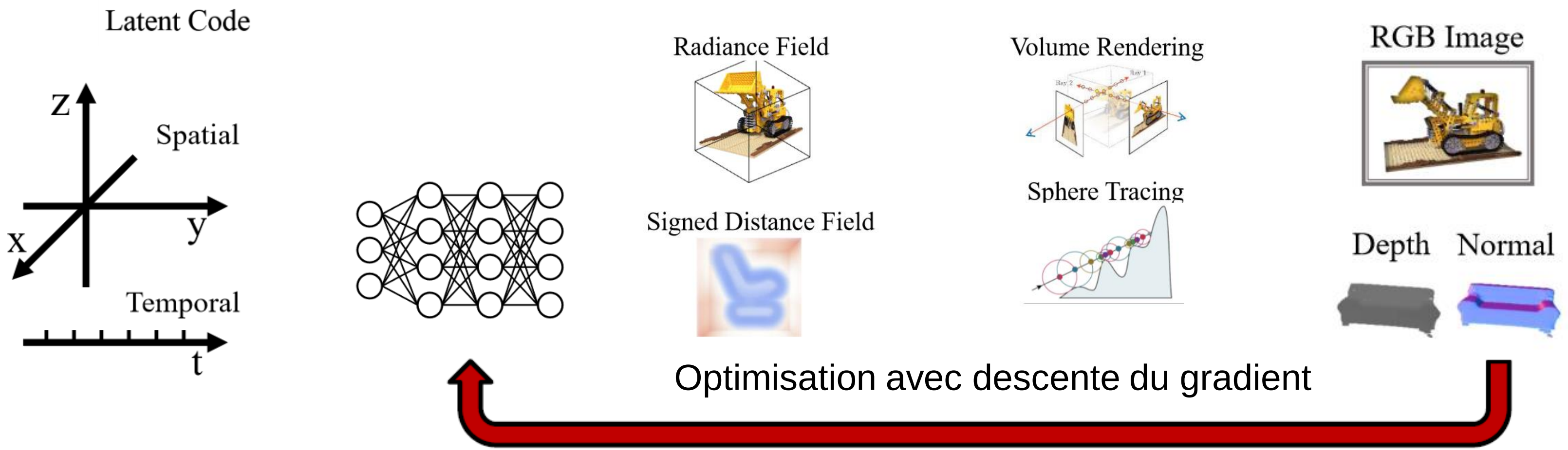
Rendering différentiable



Domaine du capteur

Framework

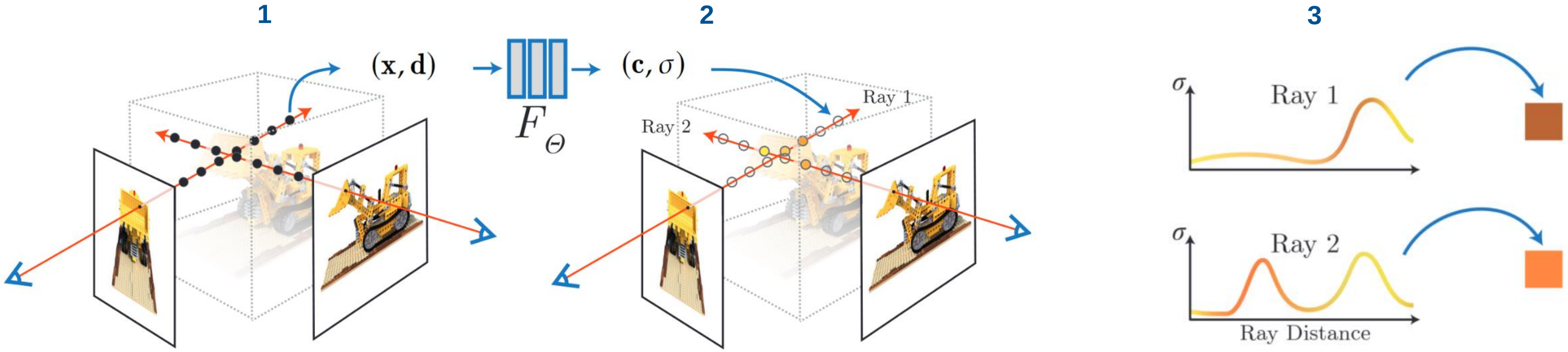
Ce que l'ont veut reconstruire | **Comment on le reconstruit** | **Ce qu'on observe/mesure**



MLP

Echantillonnage des coordonnées | **Domaine de reconstruction** | **Rendering différentiable** | **Domaine du capteur**

NeRF



1. Tracé de rayons avec échantillonnage de points (coordonnées et directions)

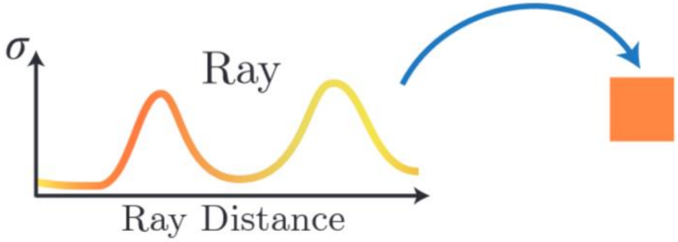
2. Estimation de la couleur et de la densité pour chaque point

3. Rendu de volume pour estimer la couleur des pixels

*<u>Illustration</u> :
*Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng, **NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,** Mar 2020*

Volume rendering

Sorties MLP

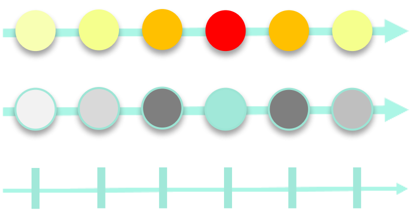


$$\mathbf{c}(r) = \sum_{i=1}^{N} w_i \boxed{c_i}$$

c_i

σ_i

δ_i *



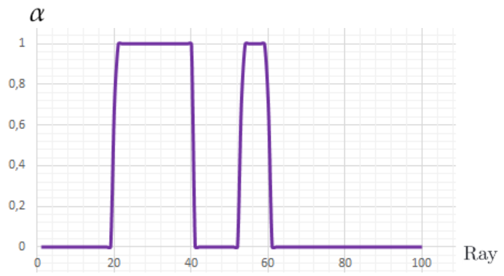
$$\alpha_i = 1 - \exp\left(-\boxed{\sigma_i}\delta_i\right)$$

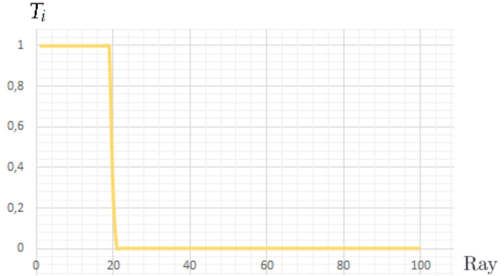
$$T_i = \prod_{j=1}^{i-1} \left(1 - \alpha_j\right)$$

* Pas de sampling

Si le point *i* a une densité <u>et</u> est visible

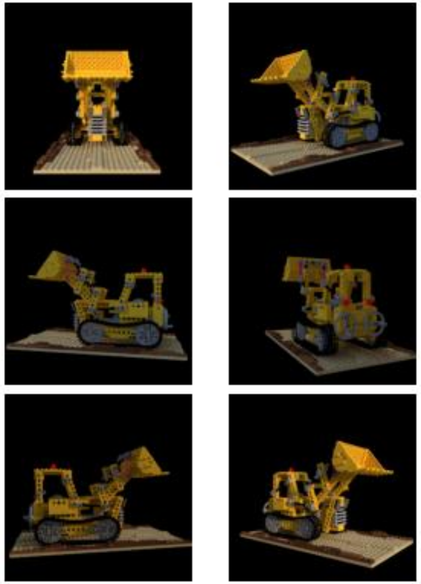
$$w_i = T_i \alpha_i$$



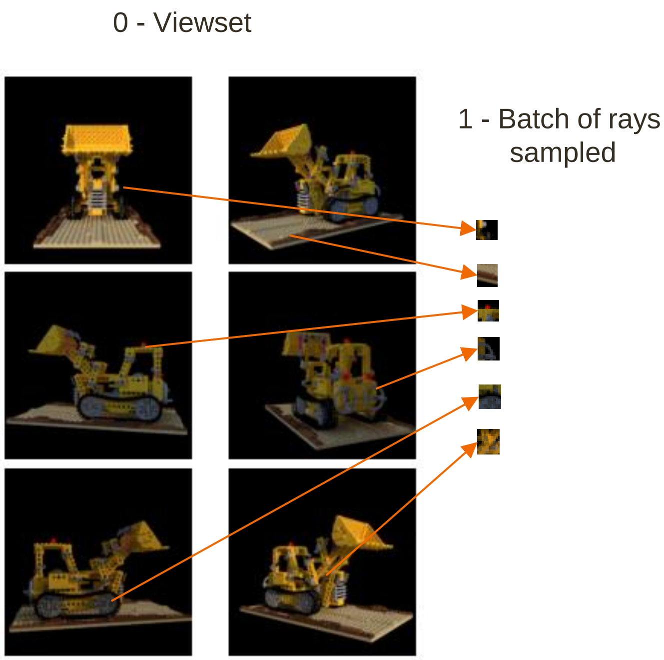


Training process

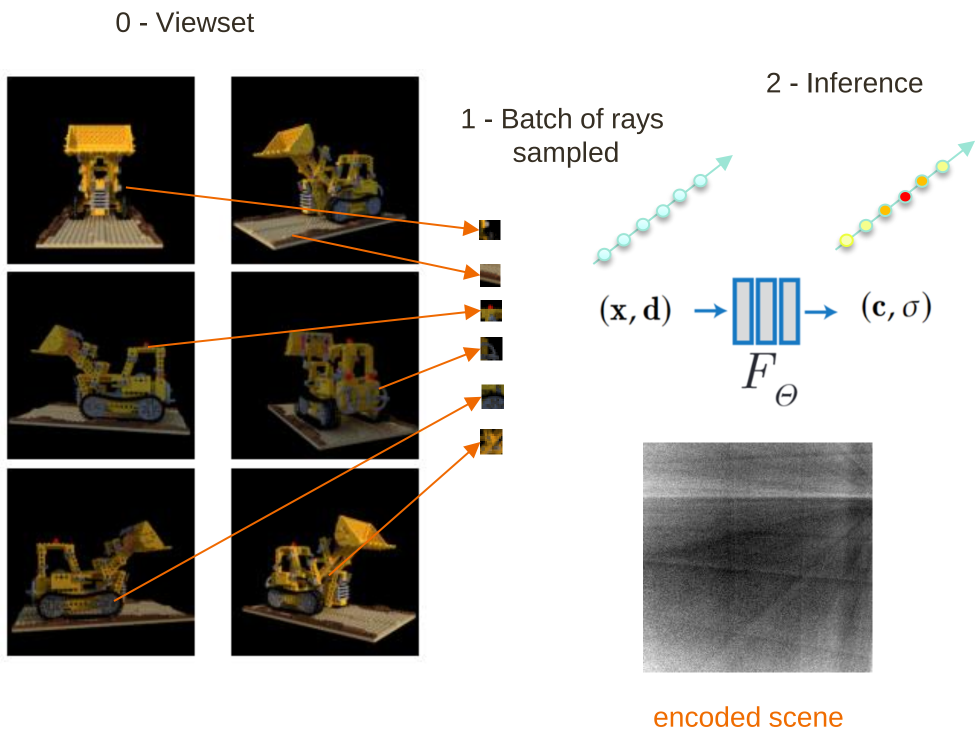
0 - Viewset



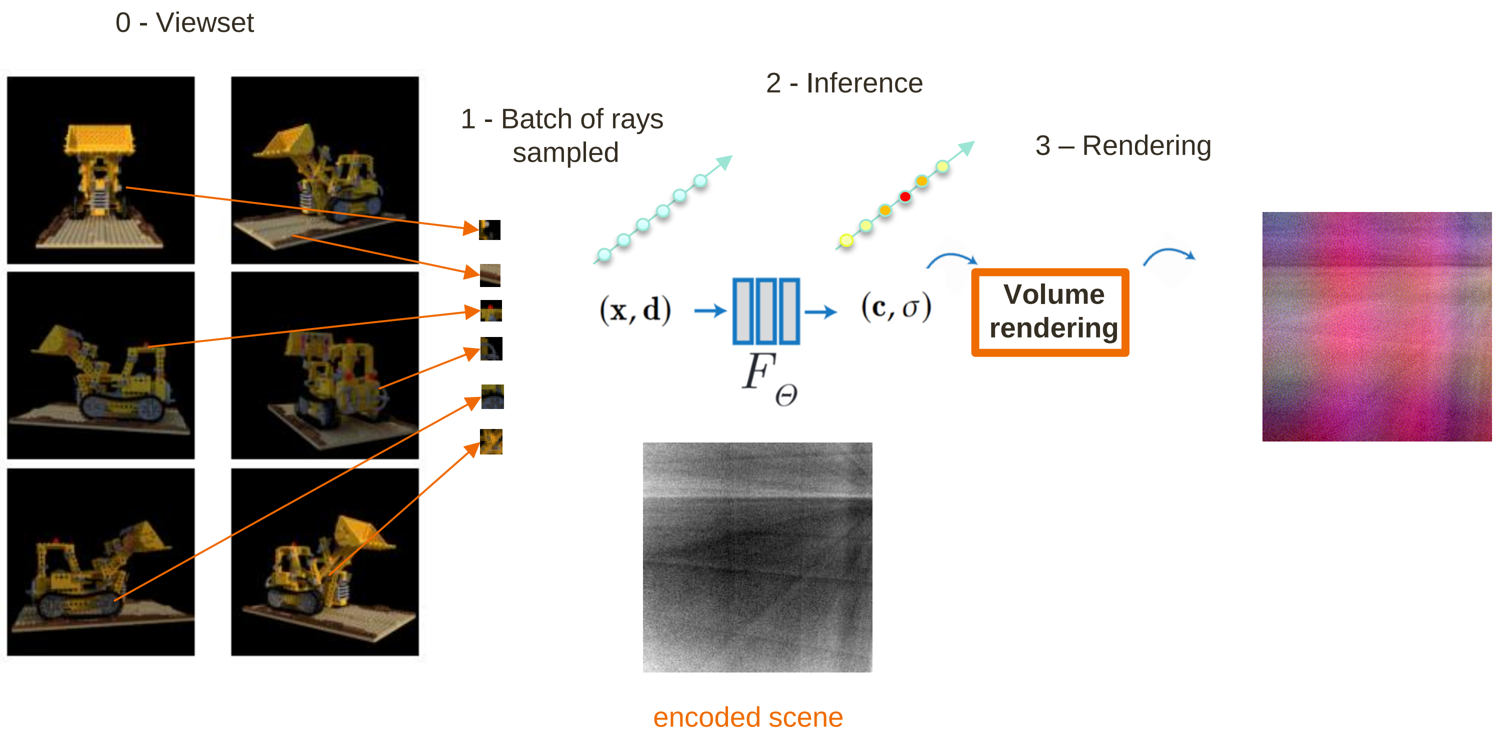
Training process



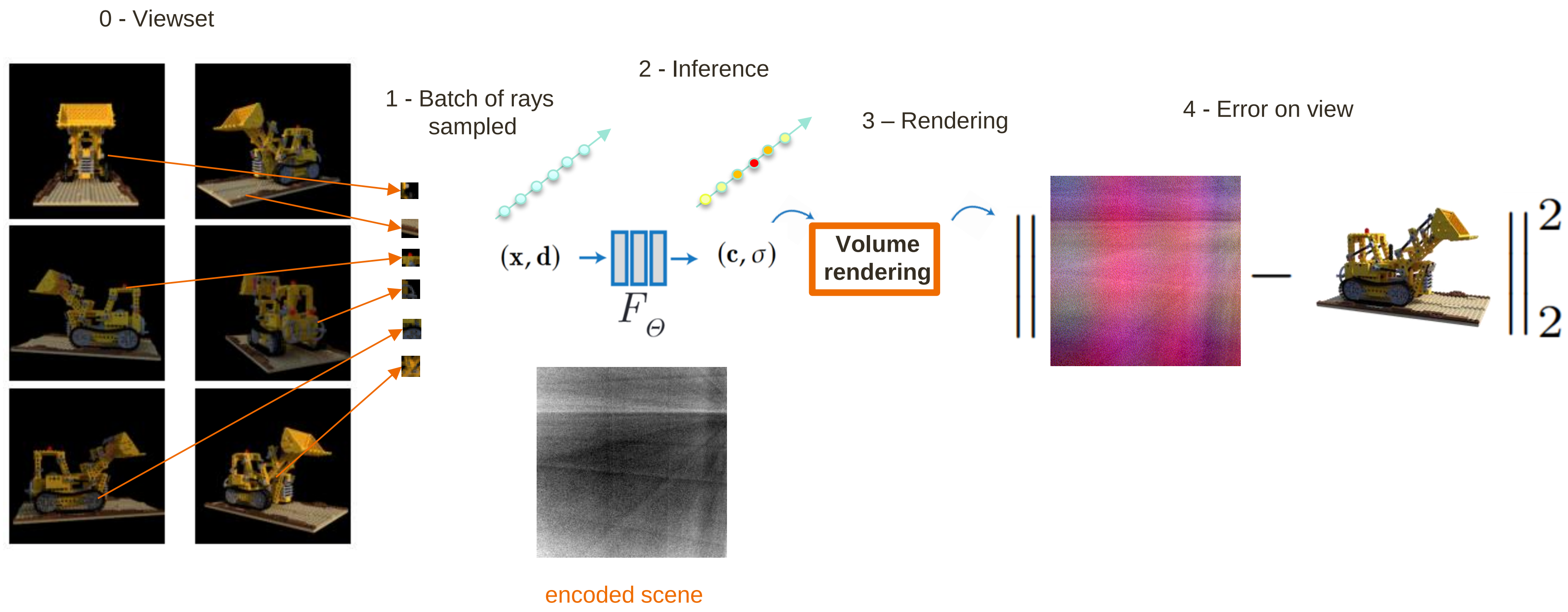
Training process



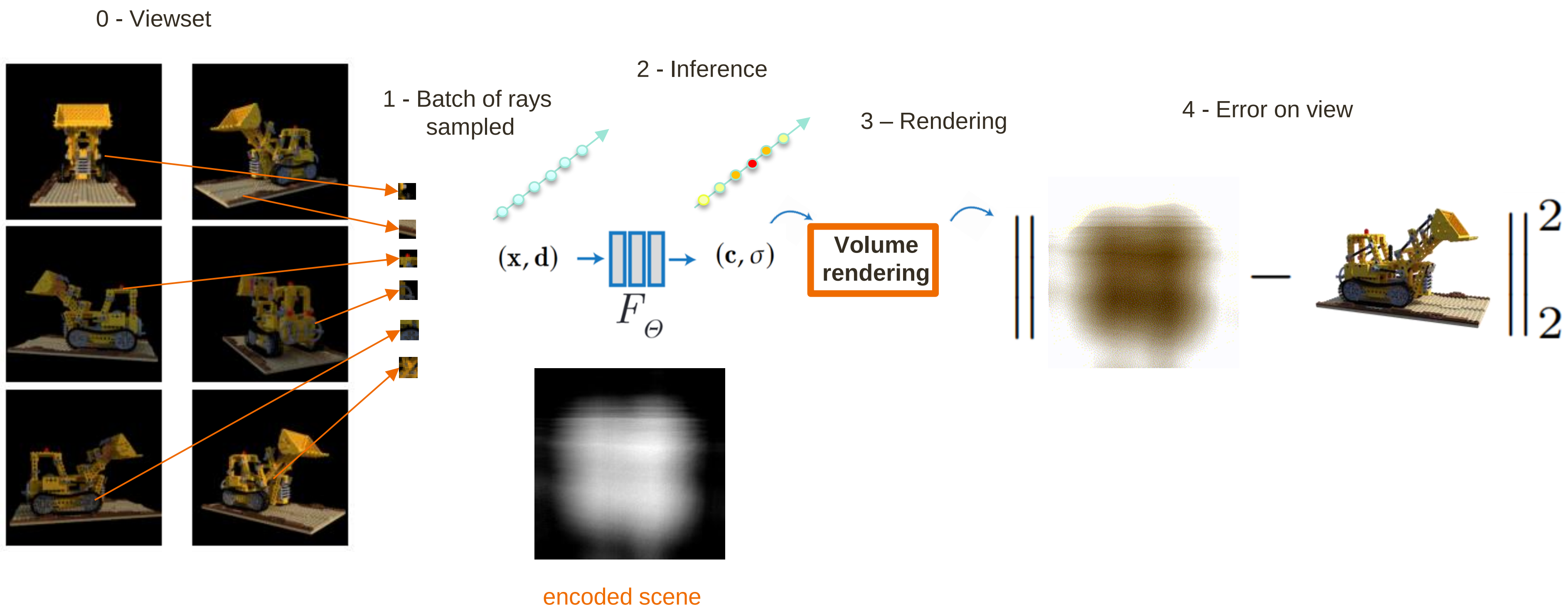
Training process



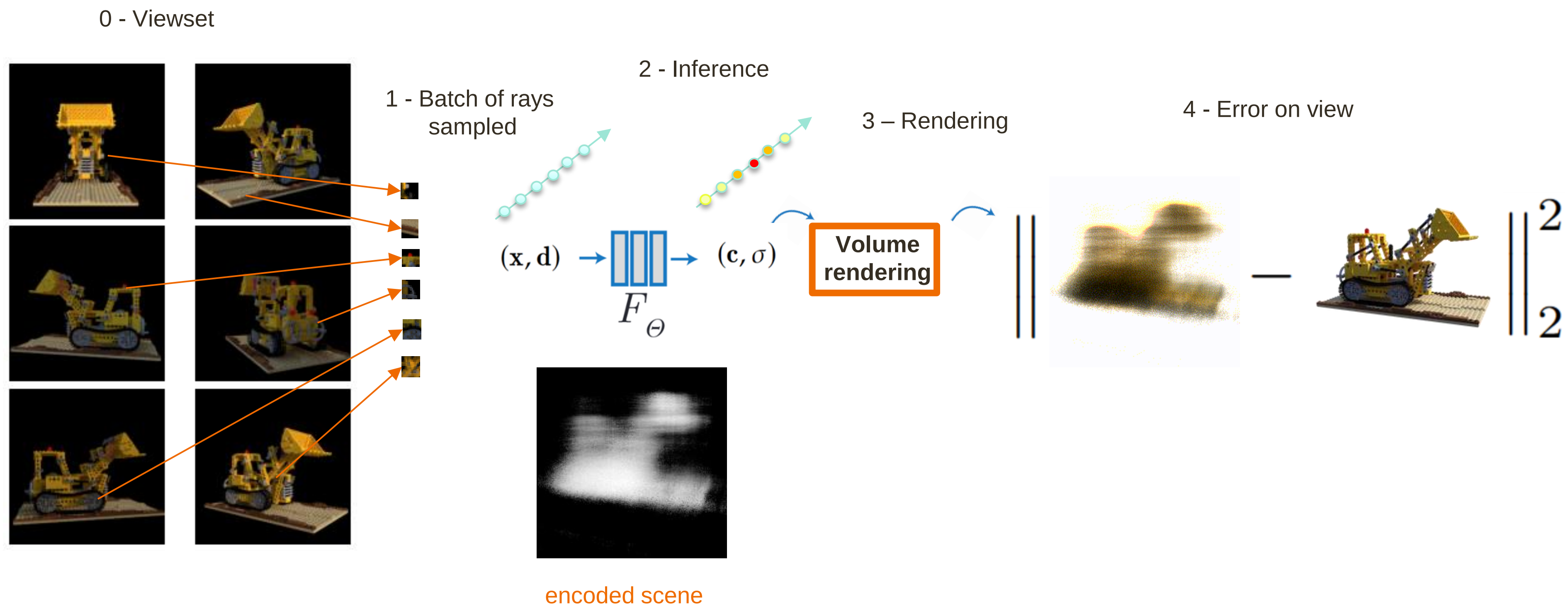
Training process



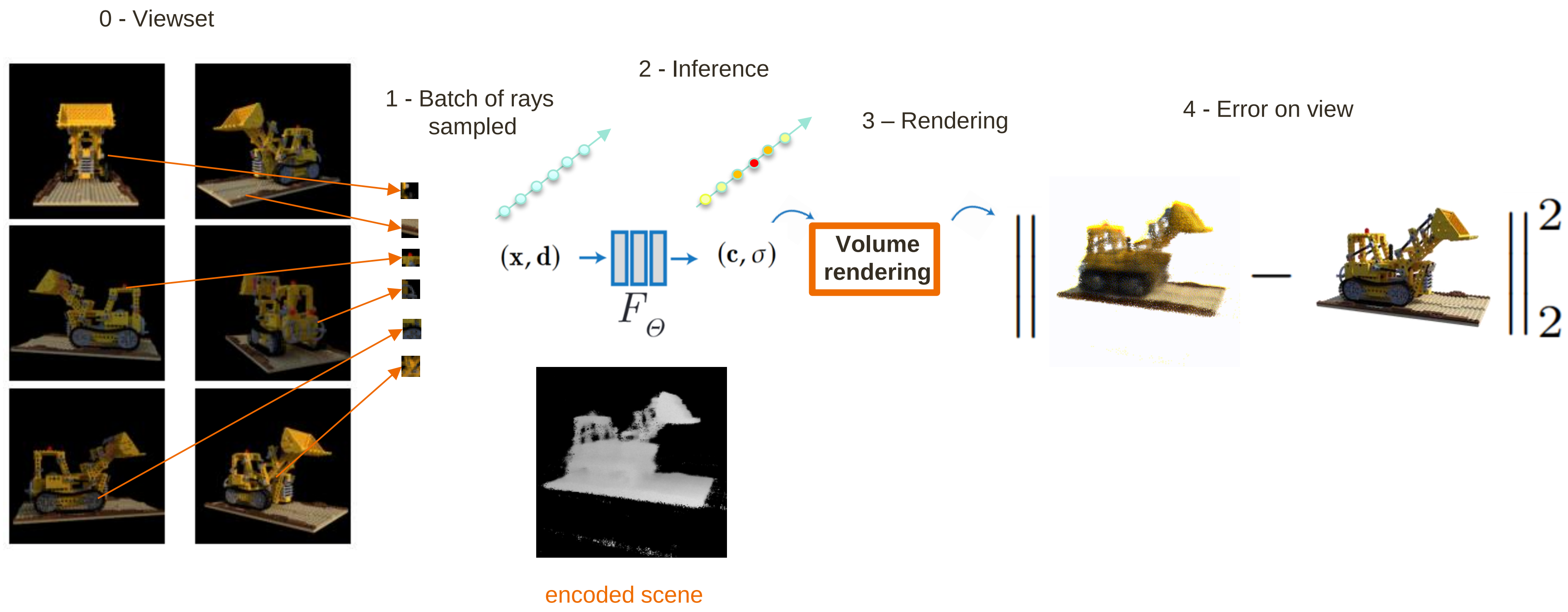
Training process



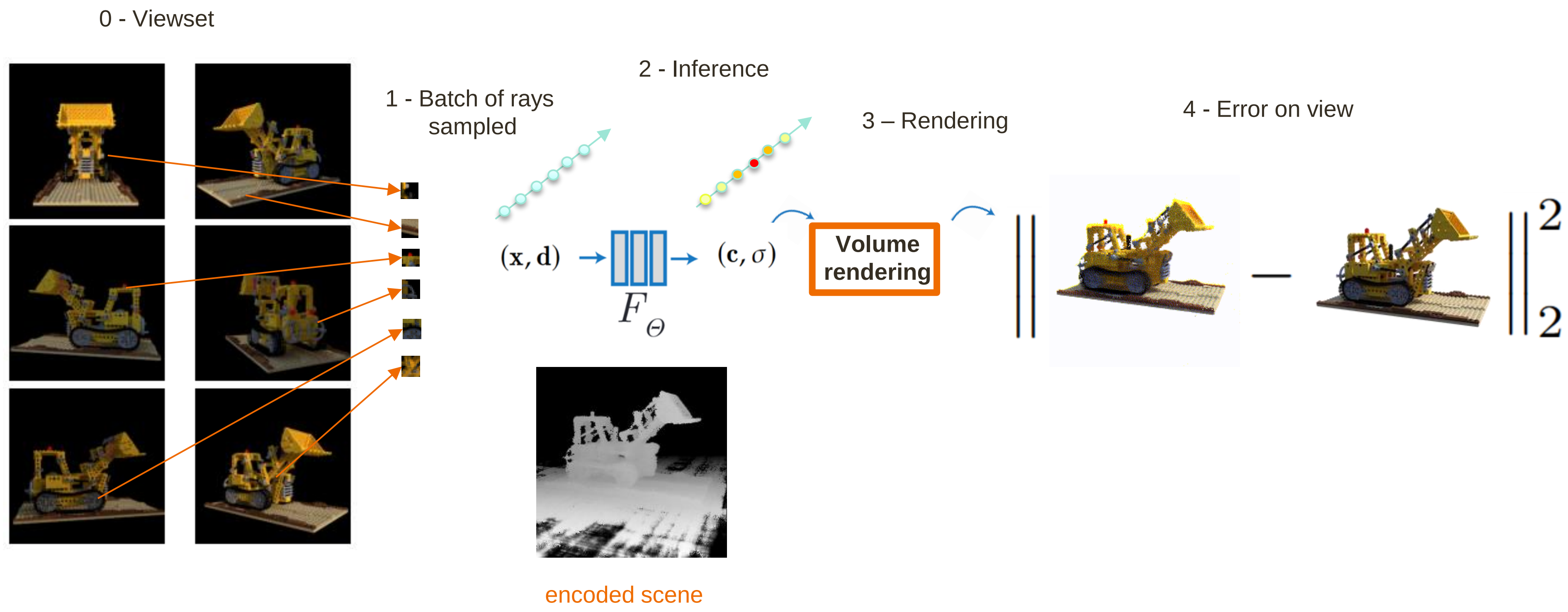
Training process



Training process



Training process



SDF : Signed Distance Field

- En apprenant un champs de distance signée à la surface apparente
- NeRF + évaluation de la distance xyz + background à traiter :
  - Mask
  - Modélisation du background avec un NeRF

  … donc plus couteux !

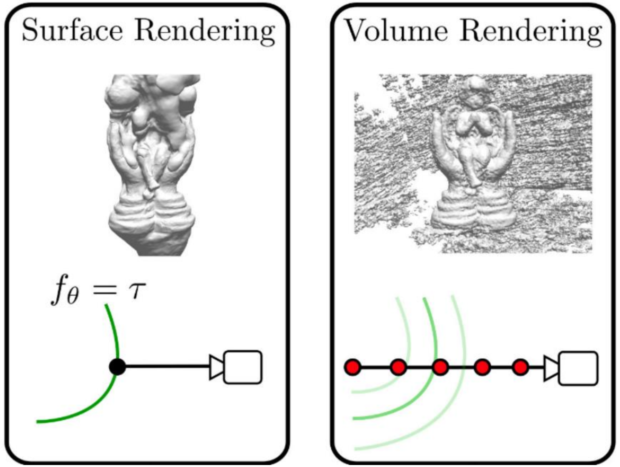
*<u>*Ilustration*</u> :*
Peng Wang and Lingjie Liu and Yuan Liu and Christian Theobalt and Taku Komura and Wenping Wang
NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, 2021

SDF : Signed Distance Field

$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^3 | f(\mathbf{x}) = 0\}$$

- Fonction en « cloche » centrée en zéro :

$$\phi_s(f(\mathbf{x}))$$



*Ilustration :
Oechsle, Michael and Peng, Songyou and Geiger, Andreas. UNISURF: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction, ICCV 2021

Résultats NeuS

Scan ID	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	Mean
PSNR(NeRF)	24.83	25.35	26.87	27.64	30.24	29.65	28.03	28.94	26.76	29.61	32.85	31.00	29.94	34.28	33.69	29.31
PSNR(Ours)	23.98	22.79	25.21	26.03	28.32	29.80	27.45	28.89	26.03	28.93	32.47	30.78	29.37	34.23	33.95	28.55
SSIM(NeRF)	0.753	0.794	0.780	0.761	0.915	0.805	0.803	0.822	0.804	0.815	0.870	0.857	0.848	0.880	0.879	0.826
SSIM(Ours)	0.732	0.778	0.722	0.739	0.915	0.809	0.818	0.831	0.812	0.815	0.866	0.863	0.847	0.878	0.878	0.820

	w/ mask			**w/o** mask			
ScanID	IDR	NeRF	Ours	COLMAP	NeRF	UNISURF	Ours
scan24	1.63	1.83	**0.83**	**0.81**	1.90	1.32	1.00
scan37	1.87	2.39	**0.98**	2.05	1.60	**1.36**	1.37
scan40	0.63	1.79	**0.56**	**0.73**	1.85	1.72	0.93
scan55	0.48	0.66	**0.37**	1.22	0.58	0.44	**0.43**
scan63	**1.04**	1.79	1.13	1.79	2.28	1.35	**1.10**
scan65	0.79	1.44	**0.59**	1.58	1.27	0.79	**0.65**
scan69	0.77	1.50	**0.60**	1.02	1.47	0.80	**0.57**
scan83	1.33	**1.20**	1.45	3.05	1.67	1.49	**1.48**
scan97	1.16	1.96	**0.95**	1.40	2.05	1.37	**1.09**
scan105	**0.76**	1.27	0.78	2.05	1.07	0.89	**0.83**
scan106	0.67	1.44	**0.52**	1.00	0.88	0.59	**0.52**
scan110	**0.90**	2.61	1.43	1.32	2.53	1.47	**1.20**
scan114	0.42	1.04	**0.36**	0.49	1.06	0.46	**0.35**
scan118	0.51	1.13	**0.45**	0.78	1.15	0.59	**0.49**
scan122	0.53	0.99	**0.45**	1.17	0.96	0.62	**0.54**
mean	0.90	1.54	**0.77**	1.36	1.49	1.02	**0.84**

**Tab* :
Peng Wang and Lingjie Liu and Yuan Liu and Christian Theobalt and Taku Komura and Wenping Wang
NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, 2021

Résultats NeuS

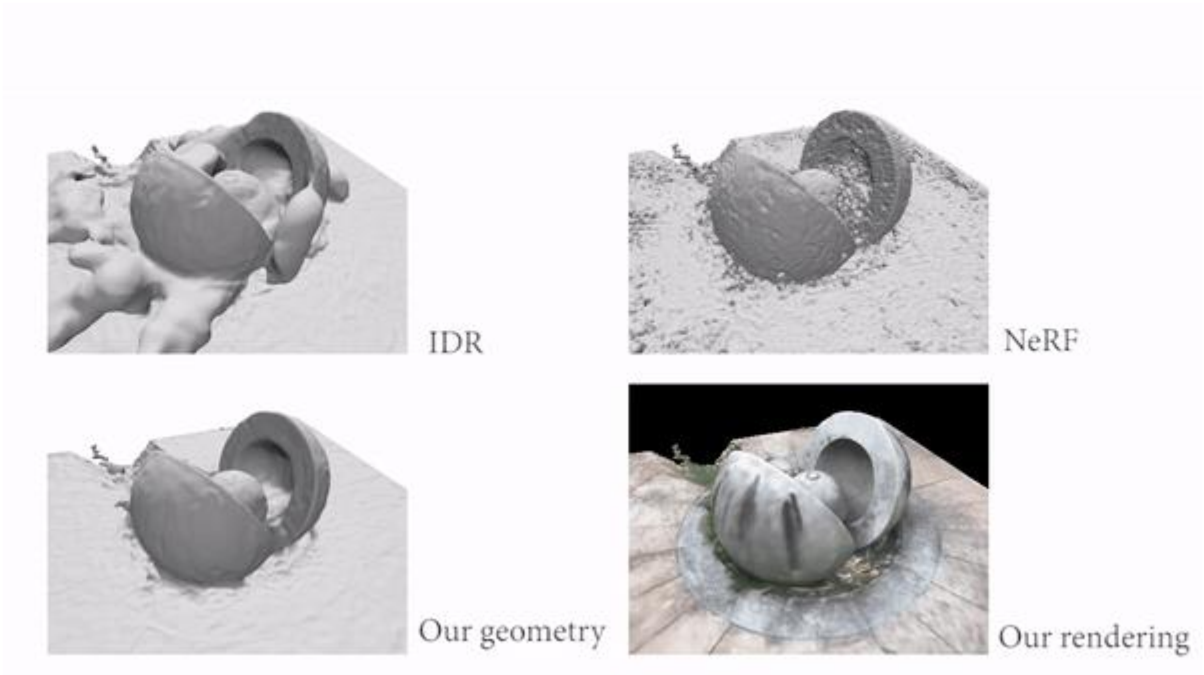


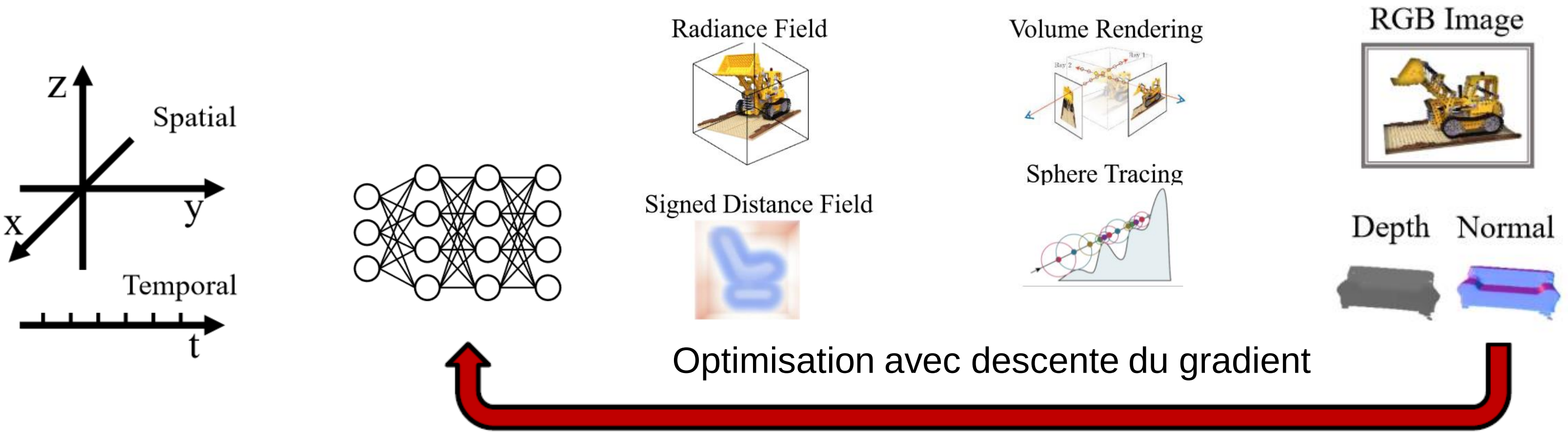
**Illustration* :
Peng Wang and Lingjie Liu and Yuan Liu and Christian Theobalt and Taku Komura and Wenping Wang
NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, 2021

Rappel : Framework

Ce que l'ont veut reconstruire

Comment on le reconstruit

Ce qu'on observe/mesure



Echantillonnage des coordonnées

MLP

Domaine de reconstruction

Rendering différentiable

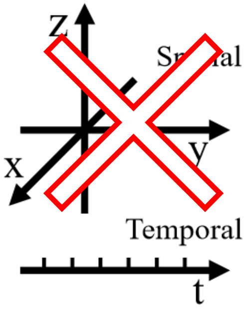
Domaine du capteur

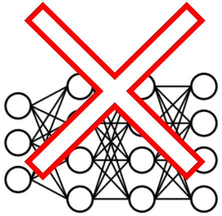
Framework sans MLP

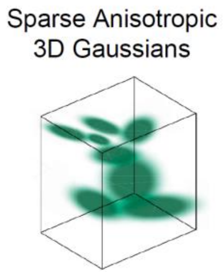
Ce que l'ont veut reconstruire

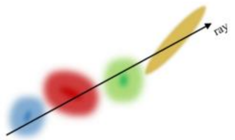
Comment on le reconstruit

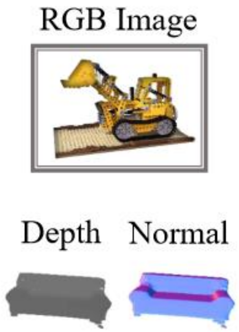
Ce qu'on observe/mesure











Optimisation avec descente du gradient

Echantillonnage des coordonnées

MLP

Domaine de reconstruction

Tile Rasterizer différentiable

Domaine du capteur

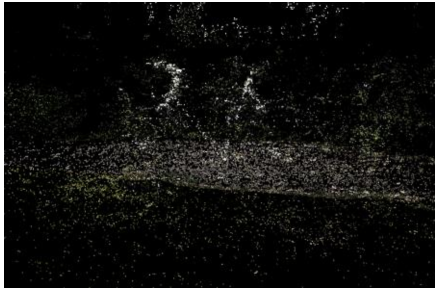
3D GS : Gaussian Splatting

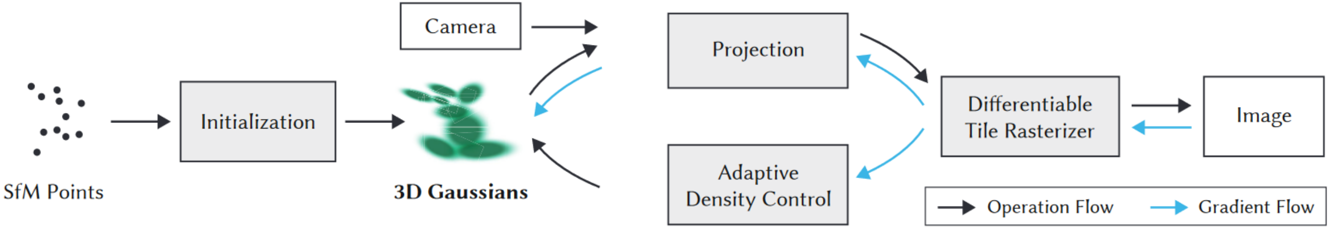
- Gaussian Splatting: en représentant la volumétrie de la scène comme un nuage d'ellipsoïdes de noyau Gaussien

- Défini par :
  - Position (moyenne μ) **:** localisation (xyz)
  - Matrice de covariance **Σ :** rotation et scaling
  - Opacité (α) : transparence
  - Couleur (RGB)
  - …

**Illustration* :
PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics
Tianyi Xie*, Zeshun Zong*, Yuxing Qiu*, Xuan Li* (equal contributions), Yutao Feng, Yin Yang, Chenfanfu Jiang.
Conference on Computer Vision and Pattern Recognition (CVPR). 2024.

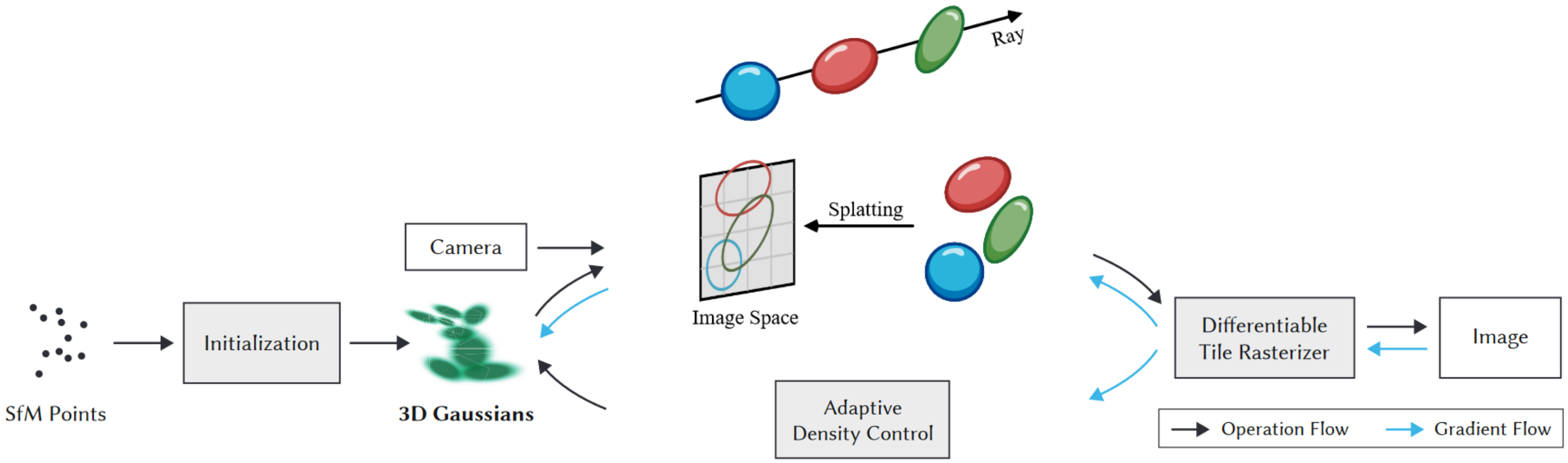
Pipeline 3D GS



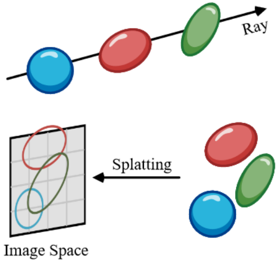


*<u>Ilustration</u> :
3D Gaussian Splatting for Real-Time Radiance Field Rendering
Kerbl, Bernhard and Kopanas, Georgios and Leimkühler, Thomas and Drettakis, George
SIGGRAPH 2023

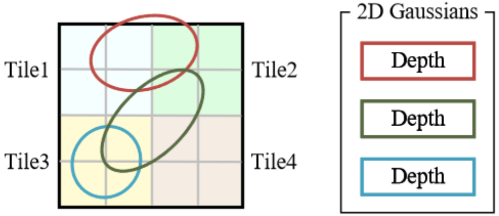
Pipeline 3D GS : Splatting



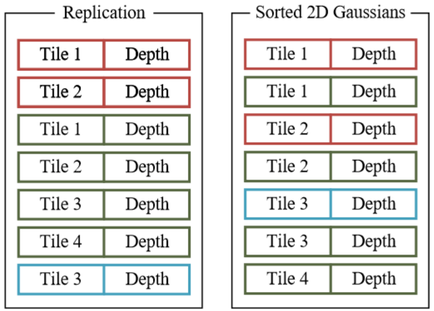
Moteur de rendu 3D GS



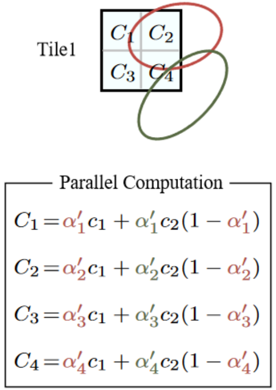
Splatting



Tuilage de l'image

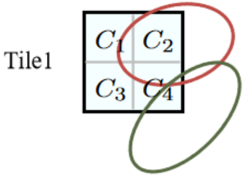


Tris des tuiles



Rendu des gaussiennes

Rendu 3D GS



Parallel Computation

$$C_1 = \alpha'_1 c_1 + \alpha'_1 c_2 (1 - \alpha'_1)$$

$$C_2 = \alpha'_2 c_1 + \alpha'_2 c_2 (1 - \alpha'_2)$$

$$C_3 = \alpha'_3 c_1 + \alpha'_3 c_2 (1 - \alpha'_3)$$

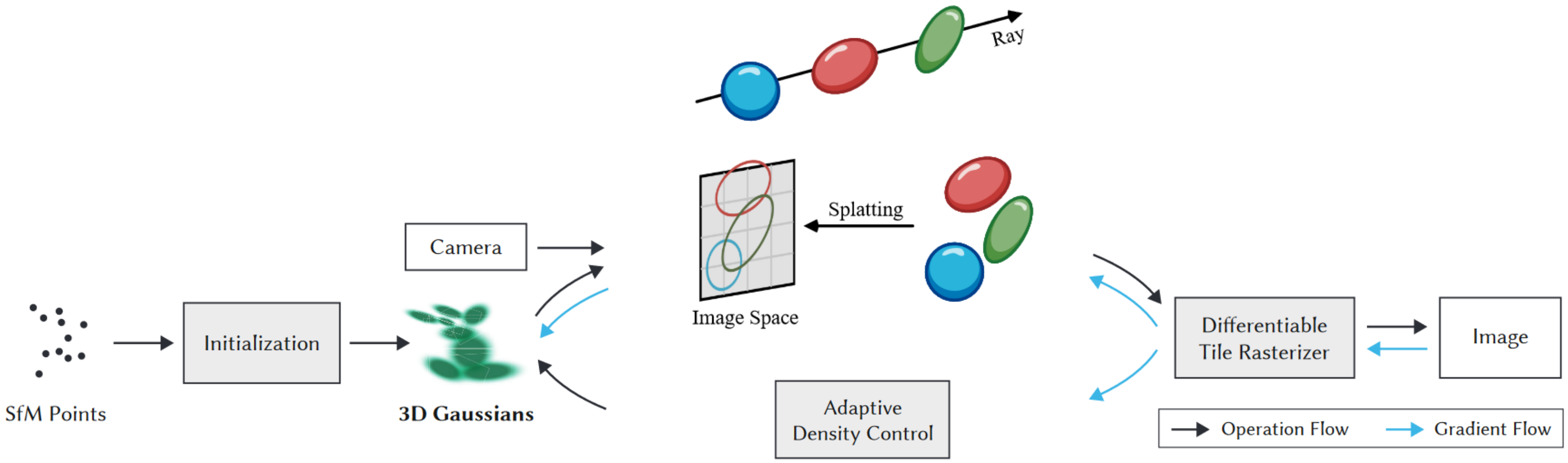
$$C_4 = \alpha'_4 c_1 + \alpha'_4 c_2 (1 - \alpha'_4)$$

$$C = \sum_{i \in \mathcal{N}} c_i \alpha'_i \prod_{j=1}^{i-1} (1 - \alpha'_j)$$

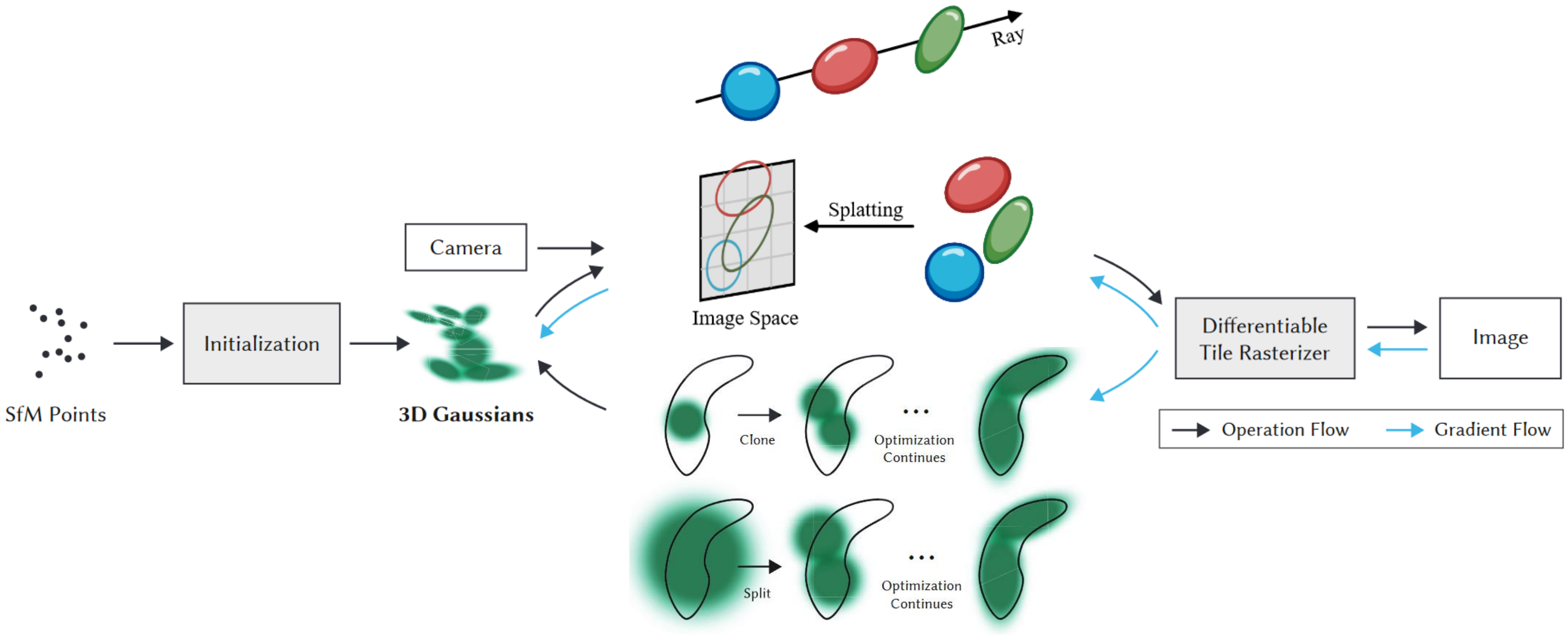
$$\alpha'_i = \alpha_i \times \exp\left(-\frac{1}{2} (\boldsymbol{x}' - \boldsymbol{\mu}'_i)^\top \boldsymbol{\Sigma}_i'^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}'_i) \right)$$

$$\boldsymbol{\Sigma} = \boldsymbol{R} \boldsymbol{S} \boldsymbol{S}^\top \boldsymbol{R}^\top$$

Pipeline 3D GS



Pipeline 3D GS : contrôle adaptatif



Résultats

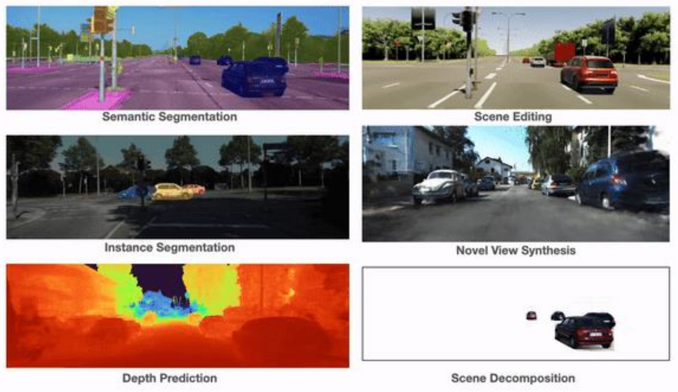
Dataset	Mip-NeRF360						Tanks&Temples						Deep Blending					
Method|Metric	$SSIM^{\uparrow}$	$PSNR^{\uparrow}$	$LPIPS^{\downarrow}$	Train	FPS	Mem	$SSIM^{\uparrow}$	$PSNR^{\uparrow}$	$LPIPS^{\downarrow}$	Train	FPS	Mem	$SSIM^{\uparrow}$	$PSNR^{\uparrow}$	$LPIPS^{\downarrow}$	Train	FPS	Mem
Plenoxels	0.626	23.08	0.463	25m49s	6.79	2.1GB	0.719	21.08	0.379	25m5s	13.0	2.3GB	0.795	23.06	0.510	27m49s	11.2	2.7GB
INGP-Base	0.671	25.30	0.371	5m37s	11.7	13MB	0.723	21.72	0.330	5m26s	17.1	13MB	0.797	23.62	0.423	6m31s	3.26	13MB
INGP-Big	0.699	25.59	0.331	7m30s	9.43	48MB	0.745	21.92	0.305	6m59s	14.4	48MB	0.817	24.96	0.390	8m	2.79	48MB
M-NeRF360	0.792†	27.69†	0.237†	48h	0.06	8.6MB	0.759	22.22	0.257	48h	0.14	8.6MB	0.901	29.40	0.245	48h	0.09	8.6MB
Ours-7K	0.770	25.60	0.279	6m25s	160	523MB	0.767	21.20	0.280	6m55s	197	270MB	0.875	27.78	0.317	4m35s	172	386MB
Ours-30K	0.815	27.21	0.214	41m33s	134	734MB	0.841	23.14	0.183	26m54s	154	411MB	0.903	29.41	0.243	36m2s	137	676MB

Ablation study :

	Truck-5K	Garden-5K	Bicycle-5K	Truck-30K	Garden-30K	Bicycle-30K	Average-5K	Average-30K
Limited-BW	14.66	22.07	20.77	13.84	22.88	20.87	19.16	19.19
Random Init	16.75	20.90	19.86	18.02	22.19	21.05	19.17	20.42
No-Split	18.31	23.98	22.21	20.59	26.11	25.02	21.50	23.90
No-SH	22.36	25.22	22.88	24.39	26.59	25.08	23.48	25.35
No-Clone	22.29	25.61	22.15	24.82	27.47	25.46	23.35	25.91
Isotropic	22.40	25.49	22.81	23.89	27.00	24.81	23.56	25.23
Full	22.71	25.82	23.18	24.81	27.70	25.65	23.90	26.05

Editable, Segmentation, ...

Panoptic Neural Fields
A Semantic Object-Aware Neural Scene Representation



PhysGaussian: Physics-Integrated 3D Gaussians for Generative Dynamics

Liens avec l’information géographique

- Geovisualisation:
  - permet le rendu de vues photoréalistes à partir de n’importe quel point de vue
  - permet (dans certaines conditions) de corriger les ombres et les objets transients
- Reconstruction: un NeRF est un objet très général dont on peut extraire:
  - MNS
  - Maillage
  - Nuage de points
  - Ortho vraie
- Les produits géométriques traditionnels (MNS, nuages de points, maillages) peuvent être utilisés pour initialiser un NeRF (accélérer l’apprentissage)

Merci de votre attention !