NERF

Representing Scenes as Neural Radiance Fields for View Synthesis

Ben Mildenhall[1⋆] Pratul P. Srinivasan[1⋆] Matthew Tancik[1⋆]
Jonathan T. Barron[2] Ravi Ramamoorthi[3] Ren Ng[1]

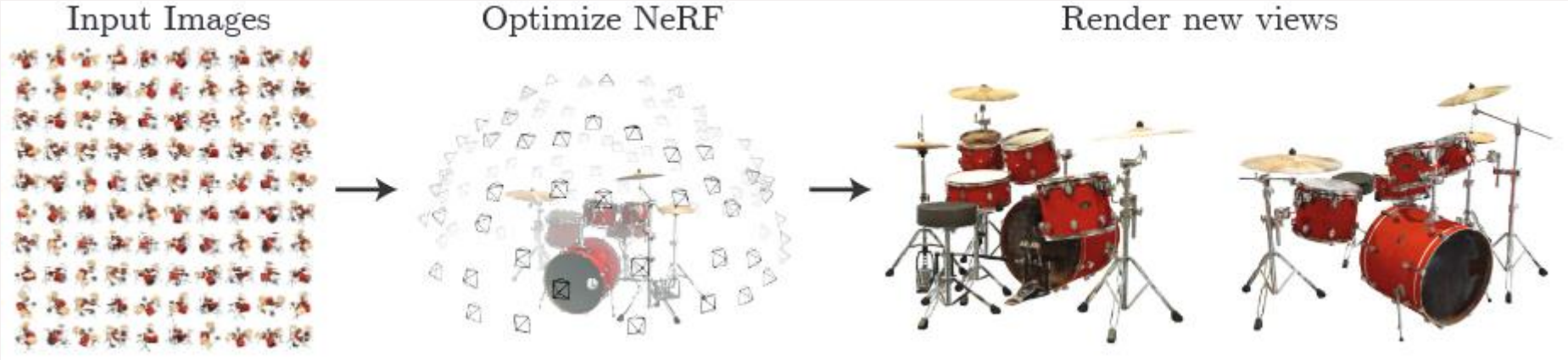
[1]UC Berkeley [2]Google Research [3]UC San Diego

PLAN

1. Introduction
2. Modèle NeRF
   a) Champs de Radiance
   b) Encodage de position
   c) Echantillonnage hiérarchique de volume
3. Les jeux de donnée
4. Résultats
5. Discussion
6. Conclusion

INTRODUCTION

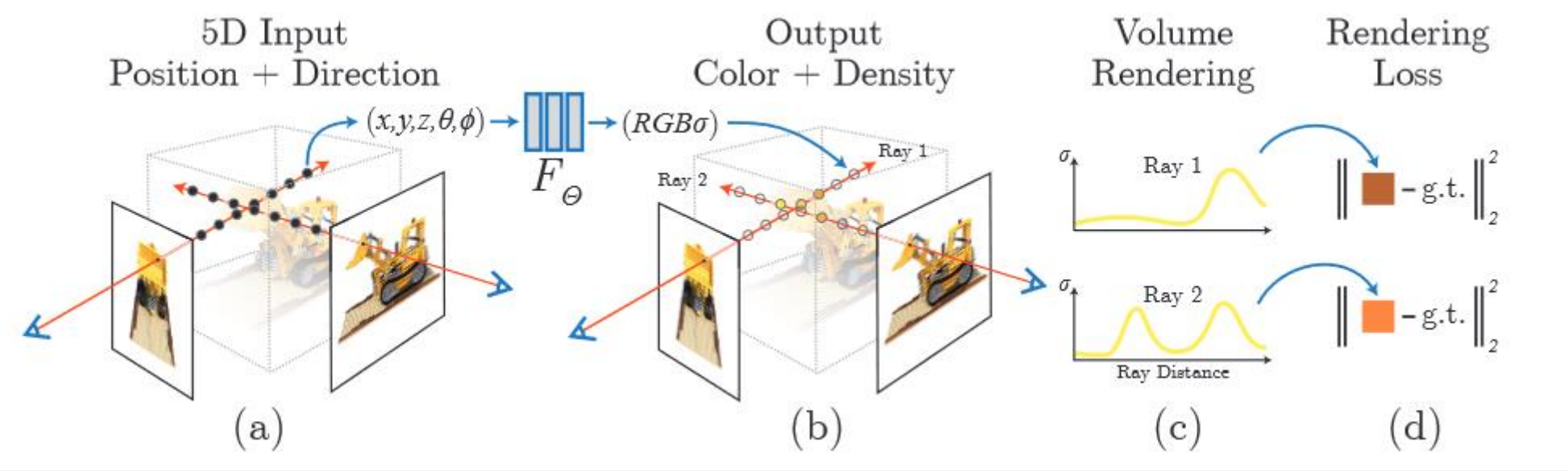
- Problème de la synthèse de vue
  - Créer de nouvelles vues d'une scène 3D à partir de données existantes
- Nouvelle approche utilisant un perceptron multi couche (MLP)
- Obtiens un rendu photoréaliste de scène 3D



LE MODÈLE NERF : CHAMP DE RADIANCE

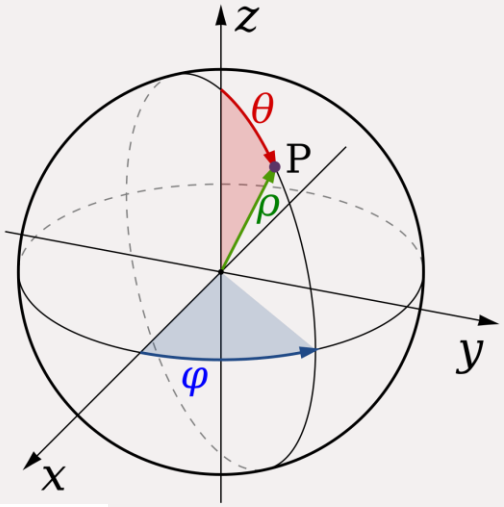
- Changement de paradigme : on apprend une scène.
- La scène est représentée par une fonction à 5 variables et 4 sorties

$$F_\Theta: \ (x, y, z, \theta, \phi) \rightarrow (r, g, b, \sigma)$$

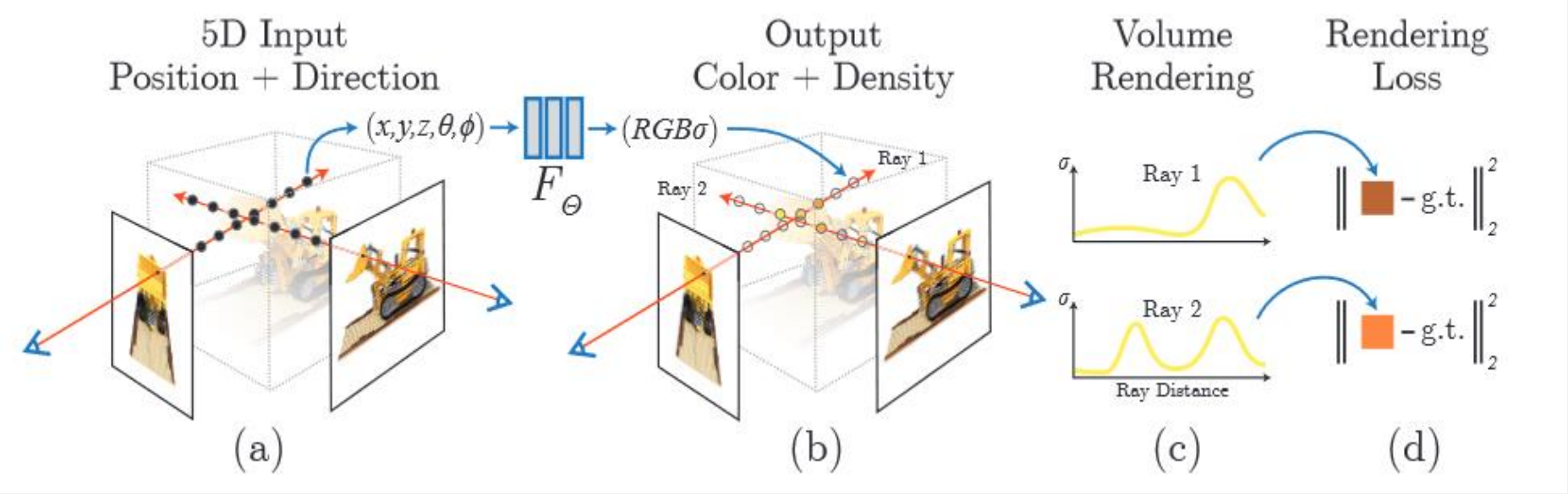


LE MODÈLE NERF : CHAMP DE RADIANCE

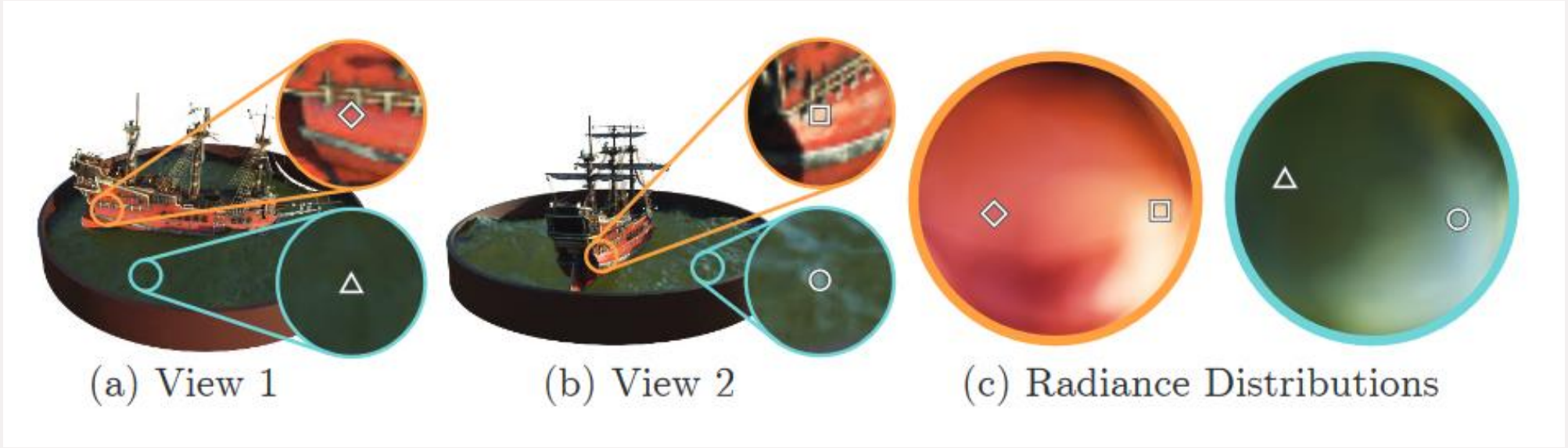
$$F_\Theta:\ (x, y, z, \theta, \phi) \rightarrow (r, g, b, \sigma)$$



- (x, y, z) : la position que l'on veut observer
- (θ, ϕ) : l'angle depuis lequel on observe
- (r, g, b) : la couleur prédite
- σ : la densité du volume (~Transparence)

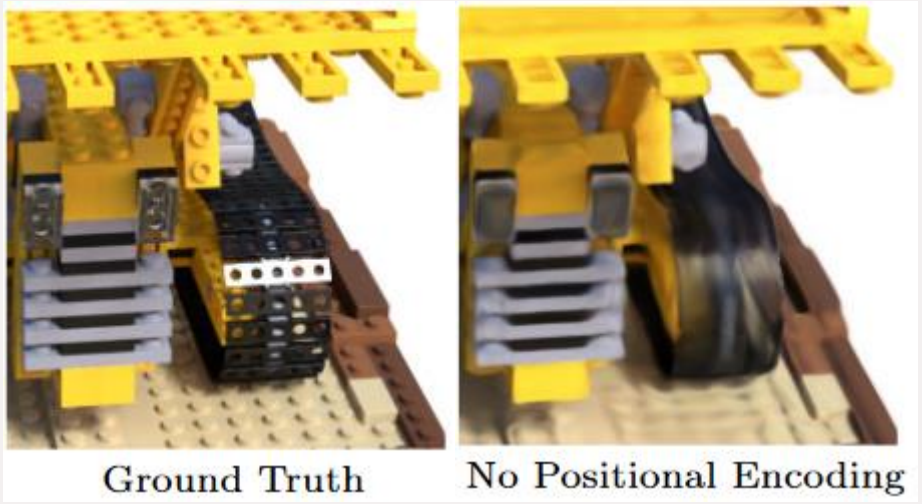


LE MODÈLE NERF : CHAMP DE RADIANCE



LE MODÈLE : ENCODAGE DE POSITION

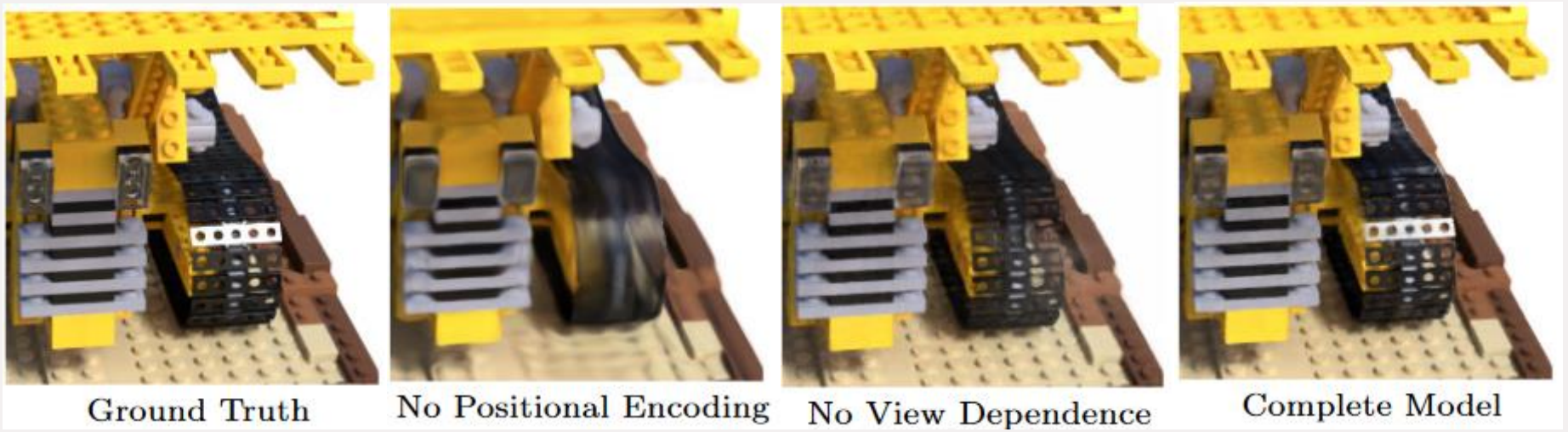
- Les MLP ont tendances à apprendre des fonctions de basses fréquences
- Les scènes présentent des variations très fréquentes de couleur et de géométrie
- Augmenter la dimensionalité des données avec une fonction bien choisie



LE MODÈLE : ENCODAGE DE POSITION

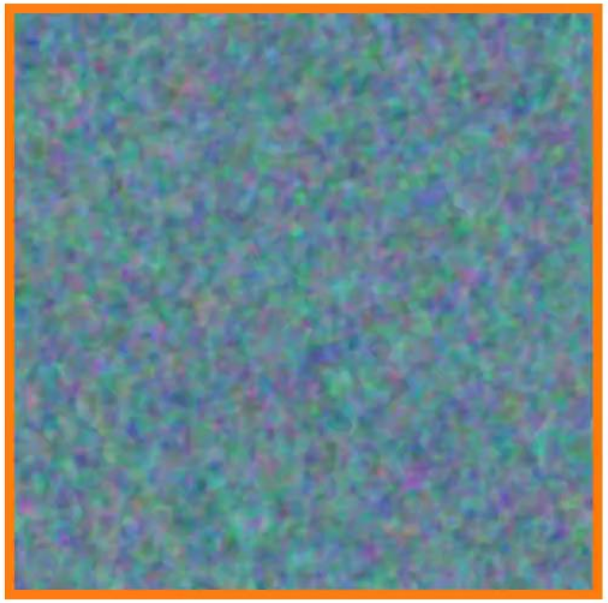
- Augmenter la dimensionalité des données avec une fonction bien choisie

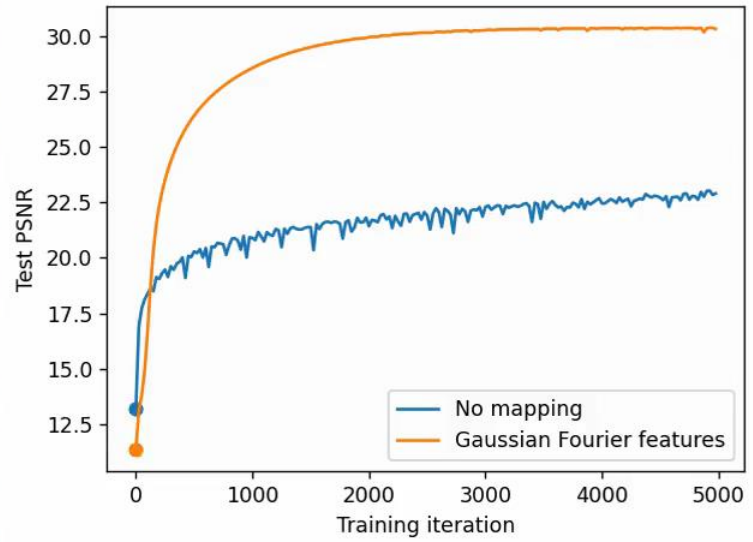
$$\gamma(p) = \left(\sin\left(2^0\pi p\right), \cos\left(2^0\pi p\right), \cdots, \sin\left(2^{L-1}\pi p\right), \cos\left(2^{L-1}\pi p\right)\right)$$



LE MODÈLE : ENCODAGE DE POSITION

- Augmenter la dimensionalité des données avec une fonction bien choisie





LE MODÈLE : ÉCHANTILLONNAGE HIÉRARCHIQUE

- Deux résolutions différentes
- Grossier : Echantillonne un ensemble de points N_C
- Précis : En utilisant l'échantillonnage grossier, on échantillonne intelligemment le volume sur N_F points

$$\mathcal{L} = \sum_{\mathbf{r}\in\mathcal{R}} \left[\left\| \hat{C}_c(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 + \left\| \hat{C}_f(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 \right]$$

LES JEUX DE DONNÉE

Rendus synthétiques d'objet

- Le DeepVoxels dataset
  - 512 × 512 pixels
  - Non Lambertians

Image réelle de scène complexe

LES JEUX DE DONNÉE

Rendus synthétiques d'objet

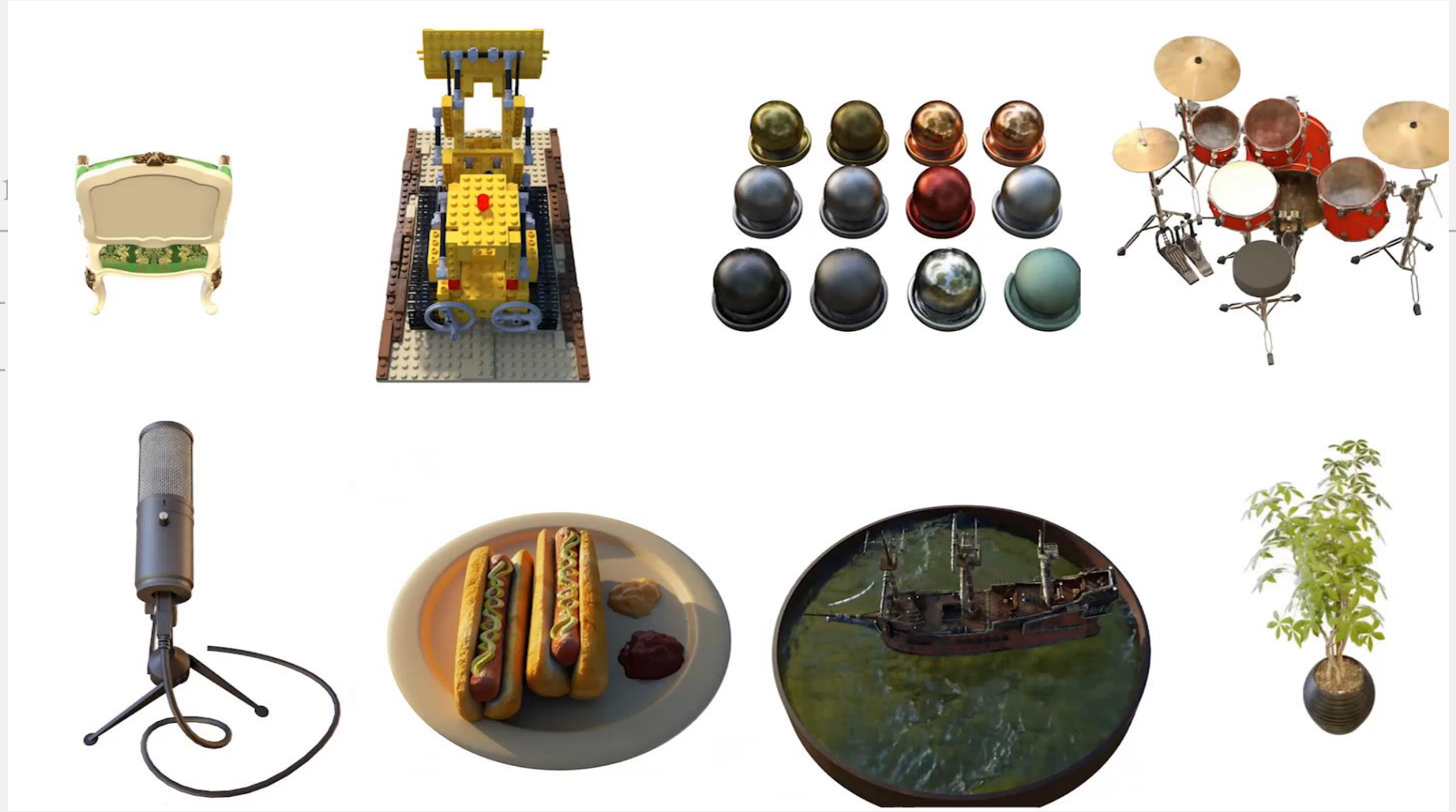
- Le DeepVoxels dataset
- Le dataset non lambertien
  - 100 vues par scènes
  - 800 × 800 pixels

Image réelle de scène complexe

LES JEUX DE DONNÉE

Ren

- L
- L



LES JEUX DE DONNÉE

Rendus synthétiques d'objet

- Le DeepVoxels dataset
- Le dataset non lambertien

Image réelle de scène complexe

- 8 scènes
- Capturées avec un smartphone
- 20 à 62 images
- 1008 × 756 pixels

LES JEUX DE DONNÉE

Rendus synthétiques d'objet

Fern

T-Rex

Image réelle de scène complexe

Orchid

MODÈLES DE COMPARAISON

- Neural Volumes (NV) :
  - Réseau à convolution
  - Prédit une grille 3D de voxels
  - 1 modèle par scène
- Scene Representation Network (SRN) :
  - Perceptron multi couche
  - Associe $(x, y, z) \rightarrow (RGB\alpha)$
  - 1 modèle par scène
- Local Light Field Fusion (LLFF) :
  - Réseau à convolution
  - Prédit directement une image $RGB\alpha$
  - Scène = Entrée du modèle

MESURES DE CONTRÔLE

- Learned Perceptual Image Patch Similarity (LPIPS) : Mesure la similarité dans l'activation d'un reseau prédéfini pour deux images.
- Structural Similarity (SSIM) : Mesure une similarité de « structure »
- Peak Signal to Noise Ration (PSNR) : Ratio entre un signal (l'image source) et un bruit (l'erreur de reconstruction)

RÉSULTATS

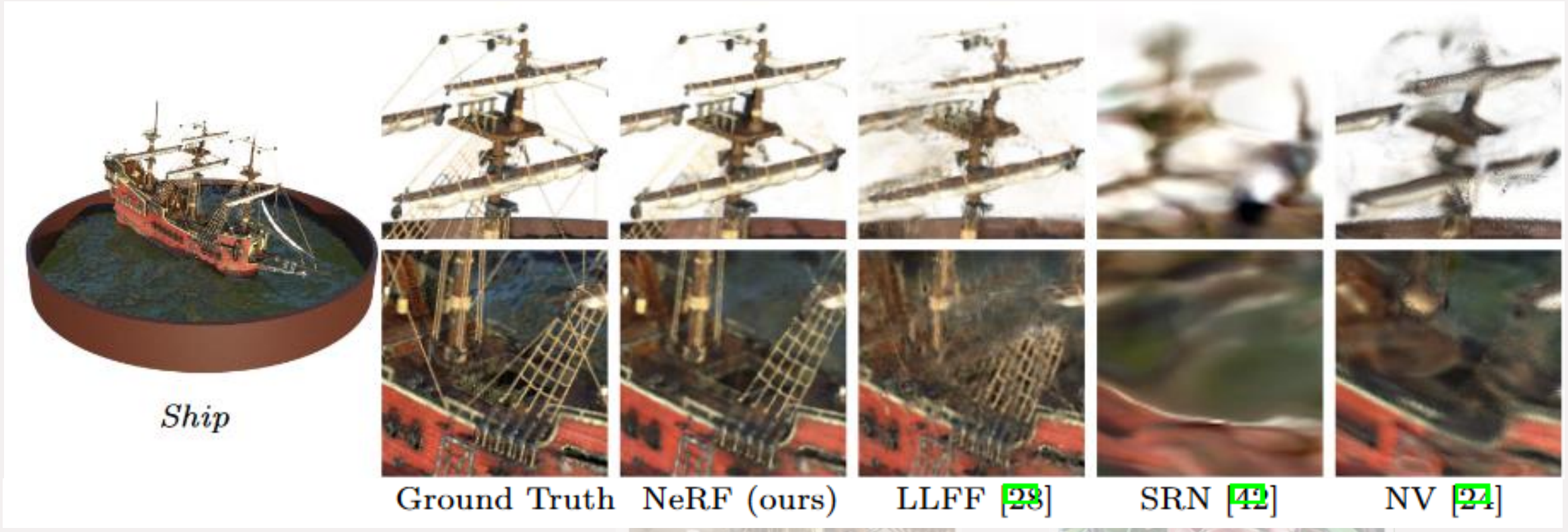
Method	Diffuse Synthetic 360° [41]			Realistic Synthetic 360°			Real Forward-Facing [28]		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SRN [42]	33.20	0.963	0.073	22.26	0.846	0.170	22.84	0.668	0.378
NV [24]	29.62	0.929	0.099	26.05	0.893	0.160	-	-	-
LLFF [28]	34.38	0.985	0.048	24.88	0.911	0.114	24.13	0.798	**0.212**
Ours	**40.15**	**0.991**	**0.023**	**31.01**	**0.947**	**0.081**	**26.50**	**0.811**	0.250

Ship

T-Rex

Orchid

RÉSULTATS



RÉSULTATS

T-Rex

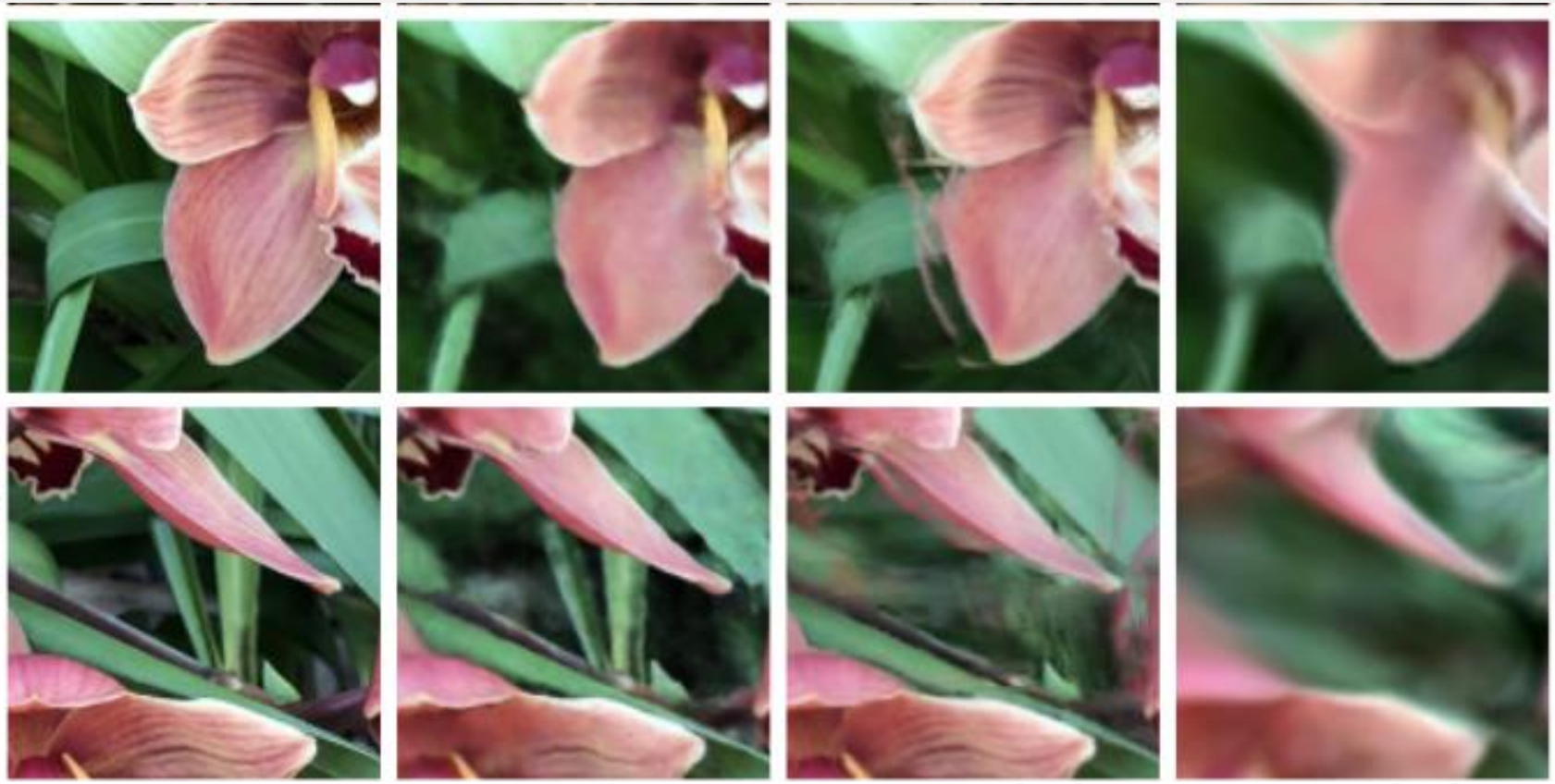
Ground Truth | NeRF (ours) | LLFF [23] | SRN [42]

Ship

Orchid

RÉSULTATS

Orchid



Ground Truth | NeRF (ours) | LLFF [23] | SRN [42]

Ship | *T-Rex*

DISCUSSION

- Dépasse les modèles NV et SRN sur toutes les scènes
- Dépasse LLFF pour tout sauf une métrique
- Compromis temps vs stockage

Modèle	Temps	Stockage
NeRF	100k-300k itérations 1 à 2 jours	5 MB
LLFF	10 min	15 GB

- 5MB c'est moins de mémoire que les images d'entrée seule !

ÉTUDE D'ABLATION

	Input	#Im.	L	(N_c , N_f)	PSNR↑	SSIM↑	LPIPS↓
1) No PE, VD, H	xyz	100	-	(256, -)	26.67	0.906	0.136
2) No Pos. Encoding	$xyz\theta\phi$	100	-	(64, 128)	28.77	0.924	0.108
3) No View Dependence	xyz	100	10	(64, 128)	27.66	0.925	0.117
4) No Hierarchical	$xyz\theta\phi$	100	10	(256, -)	30.06	0.938	0.109
5) Far Fewer Images	$xyz\theta\phi$	25	10	(64, 128)	27.78	0.925	0.107
6) Fewer Images	$xyz\theta\phi$	50	10	(64, 128)	29.79	0.940	0.096
7) Fewer Frequencies	$xyz\theta\phi$	100	5	(64, 128)	30.59	0.944	0.088
8) More Frequencies	$xyz\theta\phi$	100	15	(64, 128)	30.81	0.946	0.096
9) Complete Model	$xyz\theta\phi$	100	10	(64, 128)	**31.01**	**0.947**	**0.081**

Table 2: An ablation study of our model. Metrics are averaged over the 8 scenes from our realistic synthetic dataset. See Sec. 6.4 for detailed descriptions.

CONCLUSION

- Meilleurs résultats que l'état de l'art
- Il faut chercher des méthodes plus efficaces pour faire des rendus à partir des Champs de radiance neuronaux.
- Il y a un travail d'interprétabilité à mener pour pouvoir analyser les NeRFs
- Possibilité de stocker des scènes 3D plus efficacement

RÉFÉRENCES

- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). *Nerf : Representing scenes as neural radiance fields for view synthesis*. arXiv. https://doi.org/10.48550/arXiv.2003.08934
- *Nerf : Neural radiance fields*. (s. d.-a). Consulté 9 novembre 2023, à l'adresse https://www.matthewtancik.com/nerf
- *Nerf : Representing scenes as neural radiance fields for view synthesis(MI research paper explained)*. (s. d.-b). Consulté 9 novembre 2023, à l'adresse https://www.youtube.com/watch?v=CRIN-cYFxTk