A Compact Dynamic 3D Gaussian Representation for Real-Time Dynamic View Synthesis

Kai Katsumata *Duc Minh Vo* *Hideki Nakayama*

THE UNIVERSITY OF TOKYO

Project *Presenter

Paper ID: 11720

Contact: katsumata@nlab.ci.i.u-tokyo.ac.jp

A Compact Dynamic 3D Gaussian Representation for Real-Time Dynamic View Synthesis

Kai Katsumata   Duc Minh Vo   Hideki Nakayama

Highlight

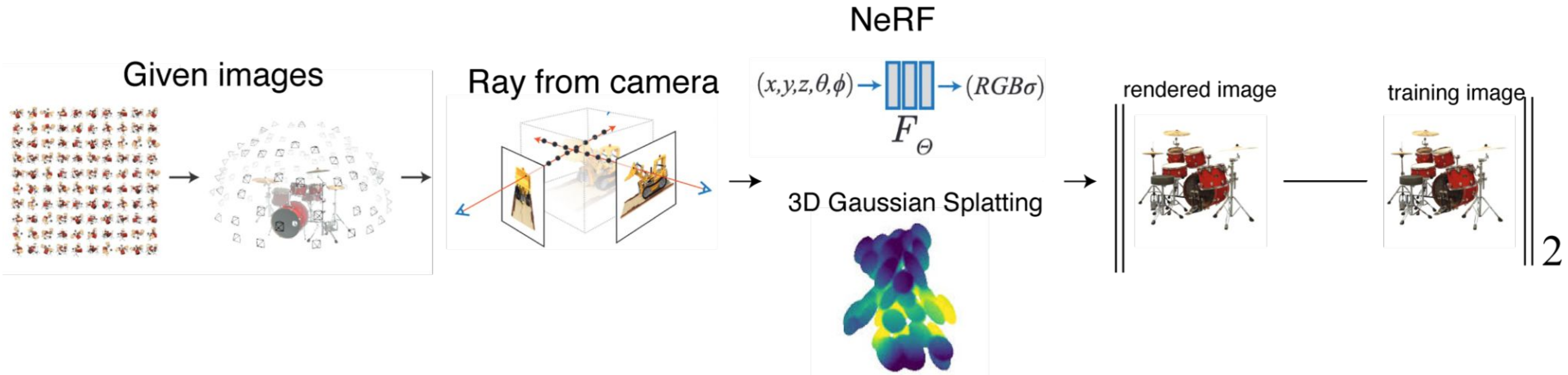
PSNR: 33.56	PSNR: 36.27	PSNR: 33.61	PSNR: 21.52	PSNR: **37.45**	
FPS: 0.54	FPS: 1.23	FPS: 0.25	FPS: 169	FPS: **146**	
Mem: 471MB	Mem: 1.2GB	Mem: **48MB**	Mem: 21MB	Mem: 89MB	
K-Planes	V4D	TiNeuVox-B	3DGS	**Our method**	Ground Truth

Our method achieves high quality, high speed, and compact dynamic view synthesis

Kai Katsumata Duc Minh Vo Hideki Nakayama

Novel View Synthesis

Given multiple images from different viewpoints, novel view synthesis aims to generate images from new viewpoints. Two popular approach: neural radiance fields (NeRF) and 3D Gaussian Splatting.

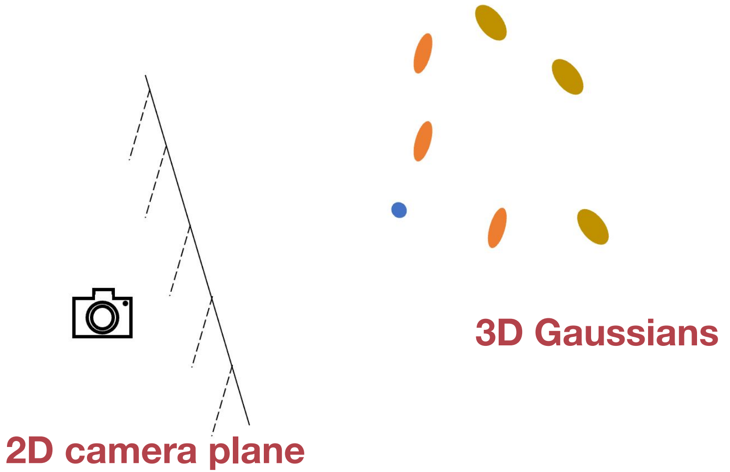


Fast, data-efficient, compact representation is essential for synthesis

3D Gaussian Splatting

1) a 3D position: $[x, y, z]^\top \in \mathbb{R}^3$,
2) a 3D rotation represented by a quaternion: $[q_x, q_y, q_z, q_w]^\top \in \mathbb{R}^4$
3) a scaling factor: $[s_x, s_y, s_z]^\top \in \mathbb{R}^3$
4) spherical harmonics coefficients representing color with the degrees of freedom k: $h \in \mathbb{R}^{3\times(k+1)^2}$
5) an opacity: $o \in \mathbb{R}$

3D Gaussian parameters



project and blend 3D Gaussians

Kerbl et al., "3D Gaussian Splatting for Real-Time Radiance Field Rendering," SIGGRAPH 2023

Extending to dynamic scenes

Static 3D Gaussian parameters

1) a 3D position: $[x, y, z]^\top \in \mathbb{R}^3$,
2) a 3D rotation represented by a quaternion: $[q_x, q_y, q_z, q_w]^\top \in \mathbb{R}^4$
3) a scaling factor: $[s_x, s_y, s_z]^\top \in \mathbb{R}^3$
4) spherical harmonics coefficients representing color with the degrees of freedom k: $h \in \mathbb{R}^{3\times(k+1)^2}$
5) an opacity: $o \in \mathbb{R}$

Dynamic 3D Gaussian parameters

1) a 3D position at time t: $[x_t, y_t, z_t]^\top \in \mathbb{R}^3$,
2) a 3D rotation at time t represented by a quaternion: $[q_t^x, q_t^y, q_t^z, q_t^w]^\top \in \mathbb{R}^4$,
3) a scaling factor: $[s_x, s_y, s_z]^\top \in \mathbb{R}^3$,
4) spherical harmonics coefficients representing color with the degrees of freedom k: $h \in \mathbb{R}^{3\times(k+1)^2}$,
5) an opacity: $o \in \mathbb{R}$.

O(TN) parameters

Challenges: multi-view assumption and huge memory consumption.

Kai Katsumata Duc Minh Vo Hideki Nakayama

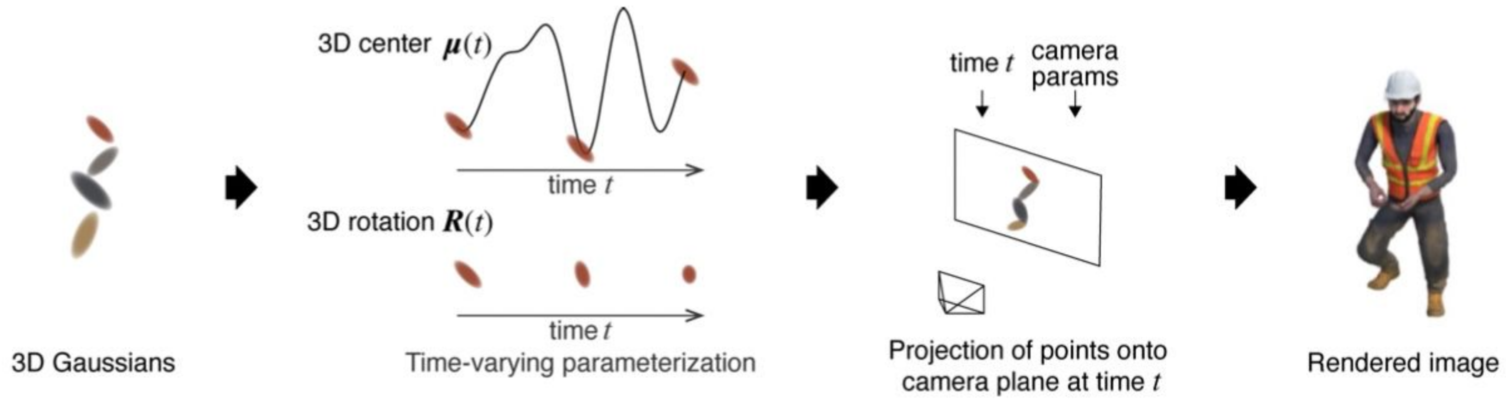
Our approach

Key idea

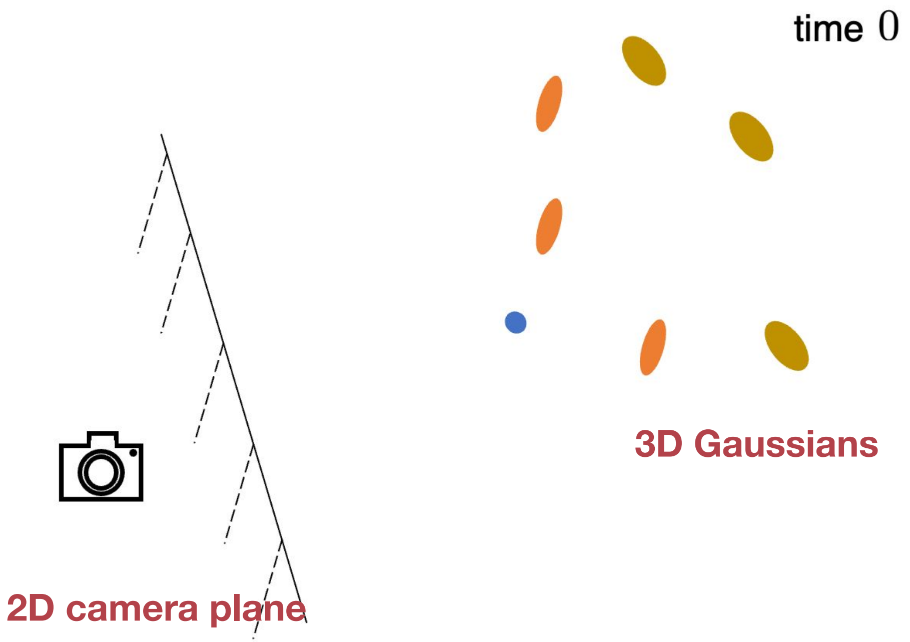
Representing time-varying parameters as functions of time with a few parameters

Our dynamic 3D Gaussian parameters

1) a 3D position at time t: $[x(t), y(t), z(t)]^\top \in \mathbb{R}^3$,
2) a 3D rotation at time t represented by a quaternion: $[q_x(t), q_y(t), q_z(t), q_w(t)]^\top \in \mathbb{R}^4$
3) a scaling factor: $[s_x, s_y, s_z]^\top \in \mathbb{R}^3$
4) spherical harmonics coefficients representing color with the degrees of freedom k: $h \in \mathbb{R}^{3\times(k+1)^2}$
5) an opacity: $o \in \mathbb{R}$



Intuition of our approach



Approximation of position and rotation

$$x(t) = w_{x,0} + \sum_{i=1}^{L} w_{x,2i-1} \sin(2i\pi t) + w_{x,2i} \cos(2i\pi t),$$

$$y(t) = w_{y,0} + \sum_{i=1}^{L} w_{y,2i-1} \sin(2i\pi t) + w_{y,2i} \cos(2i\pi t),$$

$$z(t) = w_{z,0} + \sum_{i=1}^{L} w_{z,2i-1} \sin(2i\pi t) + w_{z,2i} \cos(2i\pi t),$$

$$q_x(t) = w_{qx,0} + w_{qx,1} t,$$

$$q_y(t) = w_{qy,0} + w_{qy,1} t,$$

$$q_z(t) = w_{qz,0} + w_{qz,1} t,$$

$$q_w(t) = w_{qw,0} + w_{qw,1} t,$$

Fourier approximation for position, linear approximation for rotation

A Compact Dynamic 3D Gaussian Representation for Real-Time Dynamic View Synthesis

Kai Katsumata Duc Minh Vo Hideki Nakayama

Experiment

- Metrics:
  - PSNR
  - MS-SSIM

  Visual quality
  - FPS: rendering speed
  - Train time
  - Memory size: memory used to store optimized parameters

  Challenges in neural radiance field methods

DyNeRF dataset

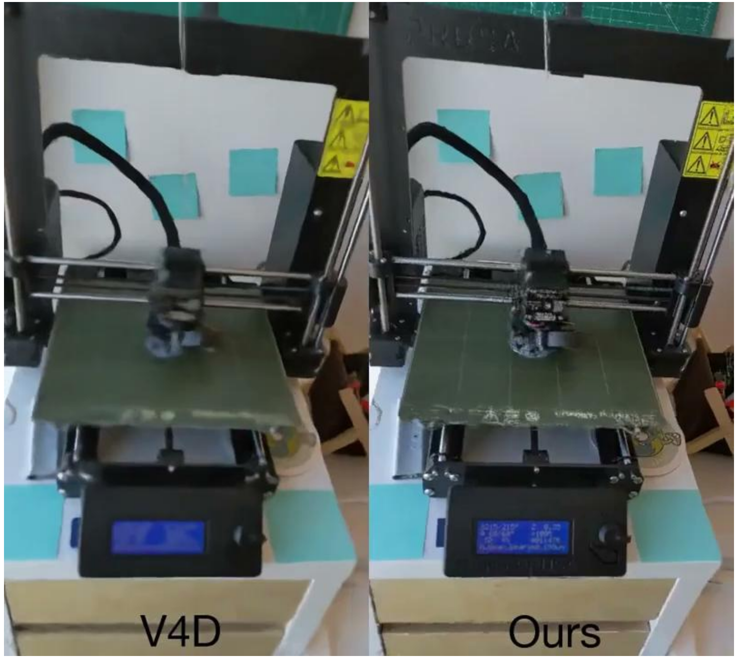
	PSNR↑	MS-SSIM↑	LPIPS↓	FPS↑	Train Time↓	Mem↓
K-Planes [Fridovich-Keil+, 2023]	**31.63**	**0.964**	-	0.31	1.8 hrs	~**309MB**
V4D [Gan+, 2024]	28.96	0.937	0.17	0.11	4 hrs	1.2GB
3DGS [Kerbl+, 2023]	20.94	0.800	0.29	109	20 mins	~198MB
D-3DGS	24.36	0.834	0.25	**119**	**51 mins**	~2.3GB
Dynamic3DGaussians [Luiten+, 2023]	27.79	0.869	0.23	51	2.1 hrs	~6.6GB
Ours	30.46	0.955	**0.15**	118	1 hrs	~338MB

Comparable **300x faster** **10x smaller**

HyperNeRF dataset

	FPS↑	Train Time↓	Mem↓	Broom		3D Printer		Chicken		Peel Banana		**Mean**	
				PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
HyperNeRF [Park+, 2021]	0.36	48 hrs	**15MB**	19.3	0.591	20.0	0.821	26.9	**0.948**	23.3	0.896	22.2	0.811
TiNeuVox-B [Fang+, 2022]	0.14	**30 mins**	48MB	21.5	0.686	22.8	0.841	28.3	0.947	24.4	0.873	24.3	0.837
V4D [GAN+, 2024]	0.15	7 hrs	1.2GB	**22.1**	0.669	23.2	0.835	**28.4**	0.929	25.2	0.873	24.7	0.827
Ours	**188**	1 hrs	~720MB	**22.1**	**0.789**	**25.5**	**0.919**	28.3	0.934	**26.6**	**0.920**	**25.6**	**0.890**

Qualitative results



Kai Katsumata Duc Minh Vo Hideki Nakayama

Application of Adding 3D object into scene

Conclusion

Goal: Compact, fast, and high-quality dynamic novel view synthesis

Challenges: Multi-view assumption and huge memory consumption

Approach: Approximation of motion of 3D Gaussians by a function with a few parameters

Result: Cutting-edge visual quality, 3DGS-level inference speed, reasonable memory footprint

Thank you for your attention.

Join us at poster session 2 Tue Oct 1st, 16:30~18:30

Contact