MEGA: Memory-Efficient 4D Gaussian Splatting for Dynamic Scenes

Supplementary Material

A. Experimental Results

We provide the complete results on the Technicolor and Neural 3D Video datasets in Table 4 and Table 5. More visualizations are available in Fig. 6 and Fig. 7.

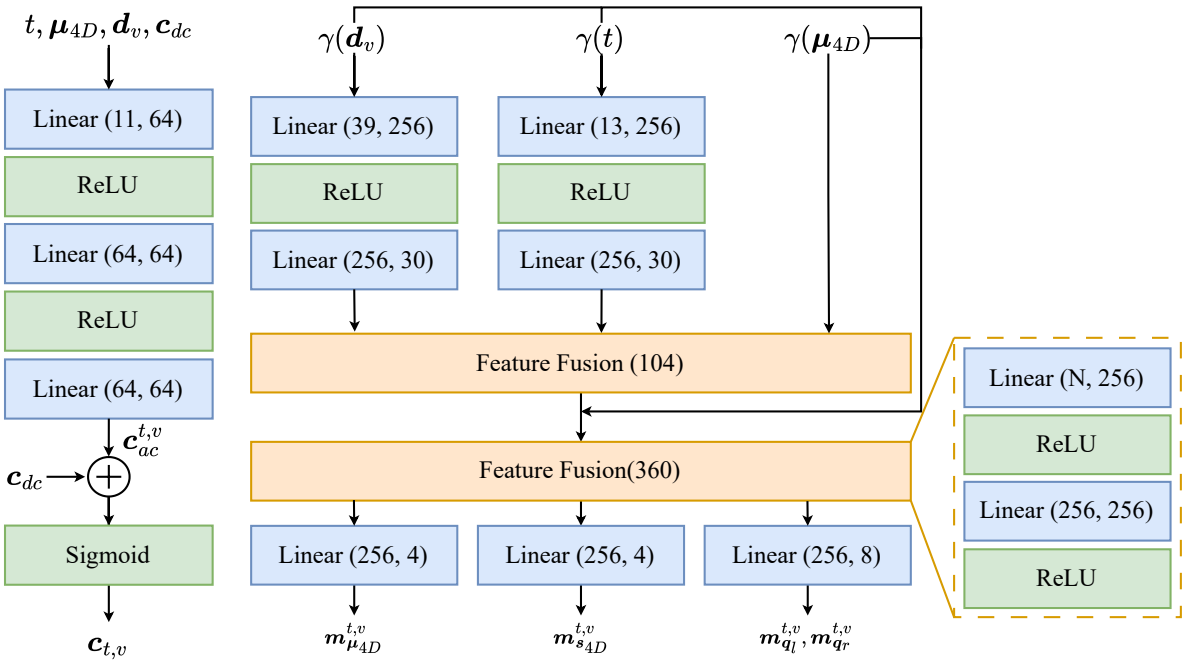
B. Network Structure

AC Color Predictor. Fig. 8 (a) shows the details of the AC color predictor. After generating the AC color component $\boldsymbol{c}_{ac}^{t,v}$, we combine the DC component $\boldsymbol{c}_{dc}$ to produce the final color $\boldsymbol{c}_{t,v}$.

Deformation Predictor. Fig. 8 (b) provides the details of the deformation predictor. For the feature fusion module, we apply two linear layers with ReLU activation function.

Figure 6. Subjective comparison of various methods on *Cut Roasted Beef* scene (Top) and *Sear Steak* scene (Bottom) from the Neural 3D Video Dataset.

Figure 7. Subjective comparison of various methods on *Birthday* scene (Top), *Trains* scene (Medium) and *Painter* scene (Bottom) from the Technicolor Dataset.



(a) AC Color Predictor (b) Deformation Predictor

Figure 8. The network structures of (a) AC color predictor, (b) Deformation predictor.

Table 4. Quantitative comparisons with various competitive baselines on the Technicolor Dataset.

	Birthday						*Fabien*					
Method	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓
DyNeRF [22]	29.20	-	0.0240	0.0668	-	-	32.76	-	0.0175	0.2417	-	-
HyperReel [1]	29.99	0.0390	-	0.0531	-	-	34.70	0.0525	-	0.1864	-	-
Deformable 3DGS [44]	30.68	0.0440	0.0237	0.0775	52.83	90.61MB	33.33	0.0673	0.0273	0.1851	95.52	42.81MB
E-D3DGS [2]	31.88	0.0328	0.0172	0.0506	62.41	66.50MB	34.69	0.0612	0.0236	0.1689	124.71	20.02MB
STG [23]	31.65	0.0293	0.0156	0.0413	128.43	51.81MB	35.61	0.0468	0.0177	0.1140	138.03	40.23MB
4DGS [46]	31.00	0.0383	0.0211	0.0629	39.61	7986.31MB	33.57	0.0582	0.0226	0.1555	87.54	3334.57MB
Ours	32.02	0.0309	0.0163	0.0460	61.26	31.43MB	34.89	0.0597	0.0233	0.1760	147.58	10.26MB
	Painter						*Theater*					
Method	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓
DyNeRF [22]	35.95	-	0.0140	0.1464	-	-	29.53	-	0.0305	0.1881	-	-
HyperReel [1]	35.91	0.0385	-	0.1173	-	-	33.32	0.0525	-	0.1154	-	-
Deformable 3DGS [44]	34.71	0.0497	0.0211	0.1302	84.37	51.56MB	29.65	0.0768	0.0382	0.1795	80.40	54.75MB
E-D3DGS [2]	35.97	0.0360	0.0149	0.0903	94.91	38.00MB	31.04	0.0643	0.0307	0.1493	56.88	77.61MB
STG [23]	35.73	0.0369	0.0148	0.0963	157.01	54.84MB	31.16	0.0595	0.0286	0.1332	137.48	48.52MB
4DGS [46]	35.73	0.0423	0.0176	0.1125	54.73	5667.79MB	31.29	0.0696	0.0341	0.1653	54.05	5770.69MB
Ours	36.73	0.0380	0.0154	0.1014	121.72	14.03MB	31.54	0.0622	0.0297	0.1475	56.91	34.31MB
	Trains						Average					
Method	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓
DyNeRF [22]	31.58	-	0.0190	0.0670	-	-	31.80	-	0.0210	0.1400	0.02	30.00MB
HyperReel [1]	29.74	0.0525	-	0.0723	-	-	32.70	0.0470	-	0.1090	4.00	60.00MB
Deformable 3DGS [44]	26.39	0.1104	0.0663	0.2040	67.32	67.08MB	30.95	0.0696	0.0353	0.1553	76.09	61.36MB
E-D3DGS [2]	30.87	0.0525	0.0289	0.0976	56.81	78.23MB	32.89	0.0494	0.0231	0.1114	79.14	56.07MB
STG [23]	32.61	0.0296	0.0169	0.0380	147.70	61.34MB	33.35	0.0404	0.0187	0.0846	141.73	51.35MB
4DGS [46]	28.79	0.0590	0.0362	0.0985	40.36	7775.97MB	32.07	0.0535	0.0263	0.1189	55.26	6107.07MB
Ours	32.69	0.0301	0.0172	0.0362	28.25	72.21MB	33.57	0.0442	0.0204	0.1014	83.14	32.45MB

C. Ablation Study

MLP. As shown in Table 6, MLPs of various sizes exhibit similar results, because $\mathcal{F}_\phi$ and $\mathcal{F}_\theta$ only provide temporal and viewpoint varying information, which can be effectively captured by lightweight MLPs.

Trade-off coefficients. Table 6 shows the results of various trade-off coefficients λ and κ. Our default trade-off coefficients are chosen empirically.

Opacity loss. As shown in Table 6, simply applying $\mathcal{L}_{opa}$ to existing baselines does not improve performance. For 4DGS and STG, the reason may arise from their explicit

Table 5. Quantitative comparisons with various competitive baselines on the Neural 3D Video Dataset. [1]: Only report the result on the *Flame Salmon* scene. [2]: Exclude the *Coffee Martini* scene. [3]: These methods train each model with a 50-frame video sequence to prevent memory overflow, requiring six models to complete the overall evaluation. [4]: Only report the overall results.

	Coffee Martini						*Cook Spinach*					
Method	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓
HexPlane[2,3] [3]	-	-	-	-	-	-	32.04	-	0.0150	0.0820	-	-
NeRFPlayer[3] [39]	31.53	0.0245	-	0.085	-	-	30.56	0.0355	-	0.1130	-	-
HyperReel [1]	28.37	0.0540	-	0.1270	-	-	32.30	0.0295	-	0.0890	-	-
K-Planes [13]	29.99	-	0.0170	-	-	-	31.82	-	0.0170	-	-	-
MixVoxels-L [41]	29.63	-	0.0162	0.099	-	-	32.40	-	0.0157	0.088	-	-
MixVoxels-X [41]	30.39	-	0.0160	0.062	-	-	32.63	-	0.0146	0.057	-	-
Dynamic 3DGS [29]	26.49	0.0263	0.0129	0.087	-	-	30.72	0.0295	0.0161	0.090	-	-
Deformable 3DGS [44]	27.88	0.0470	0.0284	0.0855	26.89	33.84MB	33.06	0.0267	0.0142	0.0519	31.06	33.21MB
E-D3DGS [2]	29.56	0.0319	0.0193	0.0300	51.94	57.97MB	32.71	0.0219	0.0123	0.0255	74.11	36.82MB
STG[3] [23]	28.55	0.0418	0.0253	0.0692	221.76	214.52MB	33.18	0.0215	0.0113	0.0367	290.03	151.52MB
4DGS [46]	27.98	0.0435	0.0265	0.0847	78.79	3704.58MB	32.73	0.0245	0.0133	0.0489	111.77	2474.94MB
Ours	27.84	0.0440	0.0270	0.0770	75.66	24.90MB	33.08	0.0230	0.0125	0.0471	92.51	19.83MB
	Cut Roasted Beef						*Flame Salmon*					
Method	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓
Neural Volume[1] [25]	-	-	-	-	-	-	22.80	-	0.0620	0.2950	-	-
LLFF[1] [30]	-	-	-	-	-	-	23.24	-	0.0200	0.2350	-	-
DyNeRF[1] [22]	-	-	-	-	-	-	29.58	-	0.0200	0.0830	0.015	28.00MB
HexPlane[2,3] [3]	32.55	-	0.0130	0.0800	-	-	29.47	-	0.0180	0.0780	-	-
NeRFPlayer[3] [39]	29.35	0.0460	-	0.1440	-	-	31.65	0.0300	-	0.098	-	-
HyperReel [1]	32.92	0.0275	-	0.084	-	-	28.26	0.0590	-	0.136	-	-
K-Planes [13]	31.82	-	0.0170	-	-	-	30.44	-	0.0235	-	-	-
MixVoxels-L [41]	32.40	-	0.0157	0.088	-	-	29.81	-	0.0255	0.116	-	-
MixVoxels-X [41]	32.63	-	0.0146	0.057	-	-	30.60	-	0.0233	0.078	-	-
Dynamic 3DGS [29]	30.72	0.0295	0.0161	0.0900	-	-	26.92	0.0512	0.0302	0.1220	-	-
Deformable 3DGS [44]	31.43	0.0333	0.0204	0.0551	28.43	33.14MB	28.70	0.0432	0.0255	0.0804	28.72	34.17MB
E-D3DGS [2]	33.02	0.0213	0.0116	0.0258	74.33	36.63MB	29.79	0.0363	0.0216	0.0535	61.03	45.08MB
STG[3] [23]	33.55	0.0207	0.0106	0.0367	299.98	135.28MB	29.48	0.0375	0.0224	0.0630	215.69	268.39MB
4DGS [46]	33.23	0.0226	0.0119	0.0470	109.11	2555.56MB	28.86	0.0425	0.0257	0.0832	64.31	4695.46MB
Ours	33.58	0.0217	0.0113	0.0489	75.22	25.20MB	28.48	0.0412	0.0251	0.0736	64.07	30.26MB
	Flame Steak						*Sear Steak*					
Method	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓
HexPlane[2,3] [3]	32.08	-	0.0110	0.0660	-	-	32.39	-	0.0110	0.0700	-	-
NeRFPlayer[3] [39]	31.93	0.0250	-	0.0880	-	-	29.13	0.0460	-	0.138	-	-
HyperReel [1]	32.20	0.0255	-	0.078	-	-	32.57	0.0240	-	0.077	-	-
K-Planes [13]	32.38	-	0.0150	-	-	-	32.52	-	0.0130	-	-	-
MixVoxels-L [41]	31.83	-	0.0144	0.088	-	-	32.10	-	0.0122	0.080	-	-
MixVoxels-X [41]	32.10	-	0.0137	0.051	-	-	32.33	-	0.0121	0.053	-	-
Dynamic 3DGS [29]	33.24	0.0233	0.0113	0.0790	-	-	33.68	0.0224	0.0105	0.079	-	-
Deformable 3DGS [44]	31.83	0.0248	0.0137	0.0418	30.91	30.72MB	33.01	0.0237	0.0125	0.0416	31.73	30.74MB
E-D3DGS [2]	30.23	0.0241	0.0149	0.0243	76.92	32.244MB	31.91	0.0200	0.0110	0.0233	79.89	32.426MB
STG[3] [23]	33.59	0.0178	0.0088	0.0290	305.22	141.25MB	33.89	0.0174	0.0085	0.0295	308.15	141.16MB
4DGS [46]	33.19	0.0204	0.0106	0.0389	91.52	3173.37MB	33.44	0.0204	0.0105	0.0411	124.66	2164.07MB
Ours	32.27	0.0242	0.0129	0.0538	63.84	30.48MB	33.67	0.0200	0.0103	0.0403	93.21	19.62MB
	Average											
Method	PSNR↑	$DSSIM_1$↓	$DSSIM_2$↓	LPIPS↓	FPS↑	Storage↓						
Neural Volume[1] [25]	22.80	-	0.0620	0.2950	-	-						
LLFF[1] [30]	23.24	-	0.0200	0.2350	-	-						
DyNeRF[1] [22]	29.58	-	0.0200	0.0830	0.015	28.00MB						
HexPlane[2,3] [3]	31.71	-	0.0140	0.0750	0.56	200.00MB						
StreamRF[4] [21]	28.26	-	-	-	10.90	5310.00MB						
NeRFPlayer[3] [39]	30.69	0.0340	-	0.1110	0.05	5130.00MB						
HyperReel [1]	31.10	0.0360	-	0.0960	2.00	360.00MB						
K-Planes [13]	31.63	-	0.0180	-	0.30	311.00MB						
MixVoxels-L [41]	31.34	-	0.0170	0.0960	37.70	500.00MB						
MixVoxels-X [41]	31.73	-	0.0150	0.0640	4.60	500.00MB						
Dynamic 3DGS [29]	30.46	0.0350	0.0190	0.0990	460.00	2772.00MB						
C-D3DGS[4] [16]	30.46	-	-	0.1500	118.00	338.00MB						
Deformable 3DGS [44]	30.98	0.0331	0.0191	0.0594	29.62	32.64MB						
E-D3DGS [2]	31.20	0.0259	0.0151	0.0304	69.70	40.20MB						
STG[3] [23]	32.04	0.0261	0.0145	0.0440	273.47	175.35MB						
4DGS [46]	31.57	0.0290	0.0164	0.0573	96.69	3128.00MB						
Ours	31.49	0.0290	0.0165	0.0568	77.42	25.05MB						

modeling of motion as fixed low-order polynomials, e.g., linear or quadratic. Thus, enforcing sparsity in opacity may

Table 6. Ablation study on the *Fabien* scene. $\mathcal{F}_{\phi}$ denotes the color MLP network, $\mathcal{F}_{\boldsymbol{\theta}}$ denotes the deformation MLP network. The first row denotes our final solution with $\lambda = 0.2$ and $\kappa = 5e^{-4}$.

Variant	PSNR↑	DSSIM$_1$↓	N ↓	Params↓
Ours	34.89	0.0597	0.31M	6.43M
Large $\mathcal{F}_{\boldsymbol{\phi}}$	33.59	0.0653	0.35M	7.30M
Large $\mathcal{F}_{\boldsymbol{\theta}}$	34.71	0.0604	0.33M	8.24M
Large $\mathcal{F}_{\boldsymbol{\phi}}$+$\mathcal{F}_{\boldsymbol{\theta}}$	34.27	0.0627	0.30M	7.85M
λ=0.1, κ=$5e^{-4}$	34.09	0.0643	0.27M	5.74M
λ=0.3, κ=$5e^{-4}$	33.99	0.0627	0.47M	9.77M
λ=0.2, κ=$1e^{-4}$	33.99	0.0639	0.48M	10.00M
λ=0.2, κ=$1e^{-3}$	34.01	0.0633	0.31M	6.41M
4DGS [46]	33.57	0.0582	5.43M	874.14M
4DGS+$\mathcal{L}_{opa}$	23.23	0.1037	8.41M	1353.13M
STG [23]	35.61	0.0468	0.30M	10.54M
STG+$\mathcal{L}_{opa}$	33.78	0.0610	0.28M	26.03M
E-D3DGS [2]	34.69	0.0612	0.06M	5.25M
E-D3DGS+$\mathcal{L}_{opa}$	34.52	0.0623	0.11M	10.21M

Table 7. Effect of view direction on the *Birthday* scene.

Variant	PSNR↑	DSSIM$_1$ ↓	N ↓	Params↓
Ours	32.02	0.0309	0.91M	18.48M
w/o view	27.35	0.0697	1.54M	31.44M

conflict with the motion priors, resulting in either over-pruning or insufficient flexibility to model more complex nonlinear temporal dynamics. For E-D3DGS, while it supports more flexible motion via multi-granularity embeddings, it induces locally similar deformation by regularizing nearby per-Gaussian embeddings. Therefore, adding $\mathcal{L}_{opa}$ may disrupt this smooth deformation by encouraging abrupt temporal activation, degrading local coherence. In contrast, our method jointly learns deformation and opacity in a unified way, enabling both localized adaptation and temporal sparsity without conflict.

View direction. Since the same time frames may reveal different visible content under different viewpoints, the view direction input is critical for disambiguating view-dependent geometry and motion, especially in sparse or occluded camera settings. As shown in Table 7, removing the view input significantly degrades rendering quality and requires more Gaussians, indicating less efficient and less accurate scene modeling.

D. Limitations

First, MEGA lacks robustness to real-world noise such as motion blur and color inconsistency. Integrating techniques from Robust GS [8] is a promising future direction. Second, while MEGA achieves significant compression, the increased flexibility of 4D Gaussians can lead to artifacts in cases of ultra-fast motion. A potential solution is to decompose the scene into static and dynamic regions, and then use 3D Gaussians for static regions and 4D Gaussians for dynamic regions. Finally, MEGA struggles with novel view extrapolation beyond the bound of training views, which can introduce artifacts. Incorporating strong scene priors could help improve generalization.