HyperGS: Hyperspectral 3D Gaussian Splatting

Supplementary Material

In this supplementary materials document, we present the complete results for all scenes in the Bayspec and SOP datasets in Section 8. Additionally in Section 9 we evaluate our performance on third dataset: a simulated ScanNetv200 dataset [7], where spectral replacements are applied to the target images. Furthermore, we provide ablation studies on various aspects of our approach, including the pruning strategy scoring function (Section 10.1), the performance of the system's autoencoder with varying latent space dimensionality and the global pruning call frequency to the system (Section 4.3).

We chose to ablate on the Bayspec dataset scenes in development, and we found larger improvements in optimising from the Bayspec dataset that translated into the others. This can be attributed to Bayspec's high-frequency information in most scenes, more noisy spectra, and general significance of images in the scenes.

8. Per-scene quantitative results

Tables 4 and 5 break down the results of the SOP and Bayspec datasets results of Tables 2 and 1 from the main paper, into metrics for each scene. Our method consistently enhances scene modeling performance on almost every combination of metric and scenarios. This is especially true in terms of improving the PSNR index for scene modeling. PSNR measures the extremism of outliers in the spectra of the predicted image which highlights the overall benfits of HyperGS over other methods.

9. Simulated Scannet dataset

Since no room-scale multi-view hyperspectral dataset currently exists, we propose simulating such a dataset using the ScanNetV2 dataset. To enhance the diversity and spectral richness, we utilize longer channel depths by incorporating downsampled Raman spectra from the open-source RRUFF dataset. Specifically, we select 200 suitable spectra (from their material description) from the RRuFF dataset and downsample them by a factor of 16, resulting in spectra with 228 channels. These spectra replace the segmentation labels in ScanNetV2 to generate a large number of simple simulated hyperspectral images. This approach provides valuable insights into the performance degradation of systems at high channel depths, even when a dataset includes a large number of viewpoints. Furthermore, to perform camera pose estimation from the colmap scenes, we downsampled the number of images and downscaled the resolutions of the image by 2 in order to get a working camera intrinsic estimation from the scene with COLMAP.

For fairness, we evaluate against the same baselines used throughout the mains paper, excluding MipNeRF360, as it was explicitly designed for "turntable style" 360-degree datasets. Table 6 summarizes the average results for four randomly selected scenes from the ScanNetV2 dataset.

The results consistently demonstrate that HyperGS is more robust and accurate than all other methods evaluated again. While 3DGS also performed well compared to other baselines, its success can be attributed to the use of cleaner and less varied spectra, which was found to particularly benefit 3DGS in the SOP dataset. NeRF-based methods, as expected, were to perform better on ScanNet than on the SOP dataset due to the larger number of viewpoints available. However, HyperGS still provides greater results. We attribute this to other baselines performance degrading substantially when tasked with handling high channel counts, highlighting their limitations in hyperspectral scenarios. HyperGS overcomes these issues with the use of it's learned latent space.

We also provide the full set of quantitative results for the scannet dataset performance in Table 7.

10. Additional ablation studies

In this section we aim to provide greater ablation studies on the bayspec dataset for HyperGS. We found that refining the technique on this dataset yielded greater results in the SOP dataset that we tested on as part of the main paper.

10.1. Pruning Strategy

Choosing the appropriate pruning score function is crucial for maintaining spectral fidelity after the pruning is performed. We tested several functions—including L1, L2, Huber, SAM, and mean average error (MAE) as shown in Table 11. We also provide the results without the pruning to highlight the positive effect it has on the system (called 'None' in the table). The L1-Norm loss performed best, balancing detail preservation and model simplicity by penalizing large deviations while keeping the structure intact. SAM preserved angular relationships but resulted in worse channel intensity preservation. L2 led to over-pruning, degrading reconstruction quality in regions with complex spectral features due to smoothing of the latent spectra when global pruning was activated. Hence in the final model we used an L1 loss in equation 17.

10.2. Autoencoder Training Strategy

We investigated the performance differences between an autoencoder trained individually for all the scenes in Bayspec

Surface Optics Datasets

Method	Rosemary				Basil				Tools				Origami				
	PSNR ↑	SSIM↑	SAM↓	RMSE↓	PSNR ↑	SSIM↑	SAM↓	RMSE↓	PSNR ↑	SSIM↑	SAM↓	RMSE↓	PSNR ↑	SSIM↑	SAM↓	RMSE↓	FPS↑
NeRF	8.42	0.7461	0.0284	0.3560	9.91	0.5534	0.0769	0.5256	11.61	0.4962	0.0610	0.3018	13.64	0.5684	0.0835	0.2083	0.12
MipNeRF	13.64*	0.5684*	1000*	0.2083*	10.11	0.5878	0.0728	0.5334	12.78	0.5213	0.0598	0.2781	11.697	0.5149	0.0956	0.2595	0.092
TensoRF	12.1	0.73351	0.0212	0.2662	15.23	0.5811	0.0435	0.3628	11.697*	0.5149*	0.0956*	0.2595*	12.98	0.4488	0.0776	0.2314	0.195
Nerfacto	18.66	0.8836	0.0078	0.1205	16.54	0.7915	0.0176	0.1655	16.254	0.6135	0.0198	0.1549	14.02	0.5028	0.0953	0.1993	0.572
MipNerf360	8.47	0.7518	0.0876	0.3825	13.92	0.8584	0.0497	0.2035	16.80	0.7241	0.0832	0.1482	9.93	0.3951	0.3271	0.3288	0.011
HS-NeRF	*18.60	*0.887	*0.0077	*0.1187	*16.81	*0.771	*0.0172	*0.1587	*12.001	*0.355	*0.470	*0.185	10.359	0.4530	0.3197	0.3188	0.488
3DGS	*25.56*	*0.9695*	*0.0028*	*0.0534*	*21.19*	*0.9385*	*0.0101*	*0.0897*	*29.13*	*0.9596*	*0.0165*	*0.0391*	*38.46*	*0.9833*	*0.0003*	*0.0128*	**79.0**
HyperGS	**26.77**	**0.9845**	**0.0021**	**0.0445**	**25.30**	**0.9503**	**0.00514**	**0.0569**	**30.86**	**0.9773**	**0.0091**	**0.0288**	**39.12**	**0.9906**	**0.0002**	**0.0114**	*3.56*

Table 4. Quantitative results using the HS-NeRF dataset against separate hyperspectral methods and baseline NeRF and 3DGS.

BaySpec Datasets

Method	Pinecone				Caladium				Anacampseros				
	PSNR ↑	SSIM↑	SAM↓	RMSE↓	PSNR ↑	SSIM↑	SAM↓	RMSE↓	PSNR ↑	SSIM↑	SAM↓	RMSE↓	FPS↑
NeRF	22.82	0.6113	0.0446	0.0728	23.12	0.58348	0.0491	0.0709	24.12	0.6220	0.0384	0.0623	0.13
MipNeRF	21.45	0.5738	0.0410	0.0856	23.36	0.5935	0.0487	0.0685	23.43	0.6160	0.0408	0.0786	0.090
TensoRF	24.12	0.6454	0.0593	0.0625	24.79	0.6424	0.0516	0.0577	25.07	0.6569	0.0394	0.0558	0.17
Nerfacto	15.36	0.4935	0.0707	0.1709	20.67	0.6208	0.0529	0.0945	21.32	0.6423	0.0417	0.0867	0.50
MipNeRF360	*25.93*	*0.7355*	*0.0279*	*0.0507*	*26.93*	*0.7371*	*0.0332*	*0.0461*	**26.73**	**0.7601**	*0.0230*	*0.0461*	0.010
HS-NeRF	20.07	0.581	0.0725	0.1521	19.084	0.705	0.0533	0.0902	20.32	0.7260	0.0345	0.0789	0.47
3DGS	22.65	0.6039	0.0668	0.0819	23.50	0.7131	0.2889	0.0758	22.59	0.5786	0.0447	0.0853	**78.1**
HyperGS	**27.0**	**0.7509**	**0.0309**	**0.0447**	**27.70**	**0.8354**	**0.0271**	**0.0414**	*26.62*	*0.7545*	**0.0183**	**0.0460**	*2.31*

Table 5. Quantitative results using the HS-NeRF dataset against separate hyperspectral methods and baseline NeRF and 3DGS.

Simulated Scannet Dataset

Method	Average Results			
	PSNR ↑	SSIM↑	SAM↓	RMSE↓
NeRF	15.85	0.7200	0.1509	0.1742
MipNeRF	14.45	0.7180	0.1700	0.2094
TensoRF	7.353	0.3522	0.8201	0.4599
Nerfacto	7.928	0.3896	0.6711	0.4520
HS-NeRF	7.363	0.3258	0.4107	0.4649
3DGS	*20.618*	*0.8224*	*0.06421*	*0.1140*
HyperGS	**25.12**	**0.8805**	**0.04602**	**0.05833**

Table 6. Quantitative results using the simulated hyperspectral scannet dataset against separate hyperspectral methods and baseline NeRF and 3DGS. Change colours to yellow, orange, red

versus a single autoencoder trained on all scenes within the dataset. We perform this experiment because all baselines are per-scene models and we aimed to provide a fairer test experiment. The single-scene AE approach is tailored to each scene's unique characteristics, potentially capturing finer details, while the general approach may benefit from broader exposure, improving robustness across different scenes. Our results, Table 9 show that the single AE provides better reconstruction quality in scenes with high variability, as it can specialize in scene-specific features. However, the general autoencoder, trained on all scenes, offers more consistent performance and reduced outliers across varied environments, albeit with a slight trade-off in specific scene detail and overall performance. In development, we found adding the MLP from Section 4.2 takes a stronger role in providing better spectral reflections in the scene when added to the system, providing better latent spectral understanding.

10.3. Latent space ablation

In this ablation, we tested the performance of HyperGS against the change in latent space size. Reducing latent space size can make the feature space more meaningful and compact. However, this necessitates additional AE layers, leading to a more rigid latent space, increased prediction errors and slower performance. In table 10, we test against all of the datasets presented in the main and supplementary materials with ranging latent space sizes. We chose to do divisions of 4 and 6 of the full channel depth of the hyperspectral images to highlight the changes in performance when the latent space is reduced on the differing types of camera datasets. Interestingly, latent space performance varies significantly for the noisier Bayspec dataset, likely due to reduced expressiveness in handling noisy regions. Whereas smoother hyperspectral data like that of SOP and scannet highlights the ability to comfortably transition to smaller channel sizes and in some cases outperform the division of 4 size used in the main paper. To provide a fair and controlled experiment we use the division of 4 for all results in the main paper and the scannet results, since this provides the most consistent results.

Simulated Scannet Dataset

Method	0000-00 PSNR ↑	SSIM↑	SAM↓	RMSE↓	0009-00 PSNR ↑	SSIM↑	SAM↓	RMSE↓	0645-01 PSNR ↑	SSIM↑	SAM↓	RMSE↓	0703-01 PSNR ↑	SSIM↑	SAM↓	RMSE↓	FPS↑
NeRF	14.18	0.6912	0.1113	0.2097	16.53	0.7298	0.0978	0.1657	16.28	0.6691	0.1372	0.1650	16.39	0.7889	0.2572	0.1564	0.070
MipNeRF	13.47	0.7062	0.1266	0.2280	14.47	0.7076	0.1212	0.2290	14.62	0.6754	0.1592	0.1999	15.23	0.7824	0.2729	0.1808	0.086
TensoRF	8.94	0.5894	0.5878	0.4021	5.86	0.1951	0.8788	0.5230	7.59	0.2689	0.8645	0.4564	7.02	0.3553	0.9494	0.4581	0.340
Nerfacto	8.78	0.5912	0.5578	0.4221	8.86	0.3211	0.6671	0.4866	6.95	0.2997	0.6585	0.4546	7.12	0.3462	0.8009	0.4447	0.132
HS-NeRF	6.53	0.3091	0.2237	0.4846	8.36	0.3142	0.5621	0.4987	7.16	0.3242	0.3219	0.4453	7.40	0.3558	0.5351	0.4310	0.0642
3DGS	*23.08*	*0.8842*	*0.0194*	*0.0756*	*20.96*	*0.8172*	*0.0625*	*0.1273*	*17.47*	*0.7435*	*0.0984*	*0.1539*	*20.96*	*0.8448*	*0.0765*	*0.09921*	**81.1**
HyperGS	**23.11**	**0.9096**	**0.0193**	**0.0713**	**27.20**	**0.9372**	**0.0192**	**0.0454**	**24.03**	**0.8788**	**0.0416**	**0.0653**	**26.13**	**0.8461**	**0.0750**	**0.0513**	*3.11*

Table 7. Quantitative results using the simulated hyperspectral scannet dataset against separate hyperspectral methods and baseline NeRF and 3DGS. Change colours to yellow, orange, red

Ablation Step	Average Results for Bayspec PSNR ↑	SSIM↑	SAM↓	RMSE↓	N.Prim(k) ↓
None	26.68	0.753	0.0340	0.0442	1301
MSE	24.11	0.712	0.0340	0.0493	**121**
Huber	*27.04*	*0.7742*	*0.0257*	0.0461	*218*
MAE	27.00	0.7753	0.0269	0.0451	532
SAM	26.89	0.7651	0.0269	*0.0447*	270
L1	**27.11**	**0.7804**	**0.0254**	**0.0440**	226

Table 8. Ablation performance difference using difference pruning functions for latent hyperspectral Gaussians in the bayspec dataset.

AE type	Average results for Bayspec dataset PSNR ↑	SSIM↑	SAM↓	RMSE↓
General	26.61	0.7722	0.02791	0.04589
Per Scene	27.11	0.7804	0.0254	0.0440

Table 9. Ablation performance difference between using a general autoencoder trained on all scenes for each camera dataset against individual autoencoders trained per scene.

	C. Depth	Average PSNR ↑	SSIM↑	SAM↓	RMSE↓	Size (GB)
B.Spec	36	27.11	0.7804	0.0254	0.0440	1.27
B.Spec	27	26.11	0.7347	0.0294	0.0481	1.15
SOP	32	30.51	0.9756	0.0415	0.0354	1.24
SOP	24	29.21	0.9701	0.0321	0.0469	1.13

Table 10. Ablation performance for differing latent space sizes for the teacher model for the average metrics over all three datasets tested.

10.4. Pruning Frequency

In this section, we aim to determine the optimal frequency of global pruning within the HyperGS approach. We evaluate three pruning strategies: **Single Pruning During Densification:** Pruning once during the densification process allows the model to recover any Gaussians lost during pruning, if needed. **Double Pruning During Densification:** Pruning twice during the densification process enables further reduction in the number of Gaussians but may impair accuracy if the model cannot achieve higher detail with fewer overall Gaussians. **Hybrid Pruning:** Pruning once during densification and once after densification concludes, to assess the stability and safety of pixel-wise pruning effects. As shown in Table 11, the best-performing method was to prune once during the densification procedure. This improvement may be attributed to the recovery of important information that can be reintroduced during densification. If global pruning is performed twice within the procedure, accuracy is further compromised, although the number of primitives decreases significantly. The downside of a single pruning call is that it leaves a larger number of primitives in the model.

Ablation Step	Average results on the Bayspec dataset PSNR ↑	SSIM↑	SAM↓	RMSE↓	N.Prim(k)↓
In Densif., 1	27.11	0.7804	0.0254	0.0440	226
In Densif, 2	26.97	0.7793	0.0268	0.0451	186
Post Densif., 1	26.11	0.7462	0.0274	0.0498	159
Hybrid, 1	26.09	0.7451	0.0281	0.0497	158

Table 11. Ablation study on performance differences using various pruning frequencies for the Bayspec dataset. 'In Densif.' refers to pruning within the densification procedure before 17.5K iterations. 'Post Densif.' refers to pruning at 17.5K iterations, and 'Hybrid' refers to pruning once before and once at 17.5K iterations.

Pruning after the densification process results in a substantial loss of accuracy, indicating that the information lost becomes unrecoverable and the overall output model is negatively affected. The hybrid method provides results similar to post-densification pruning in terms of accuracy and the number of primitives. This demonstrates that the pruning method yields consistent pruning results even when called twice. To achieve the best and most accurate results with the HyperGS system, we utilize a single global pruning call during the densification iterations for the Bayspec dataset.