Complementarity between 2D and 3D tasks

陈伟德

2025.3.2

论文分享

1. Latent Radiance Fields with 3D-aware 2D Representations

Chaoyi Zhou∗, Xi Liu,∗ Feng Luo, Siyu Huang†
Visual Computing Division School of Computing Clemson University

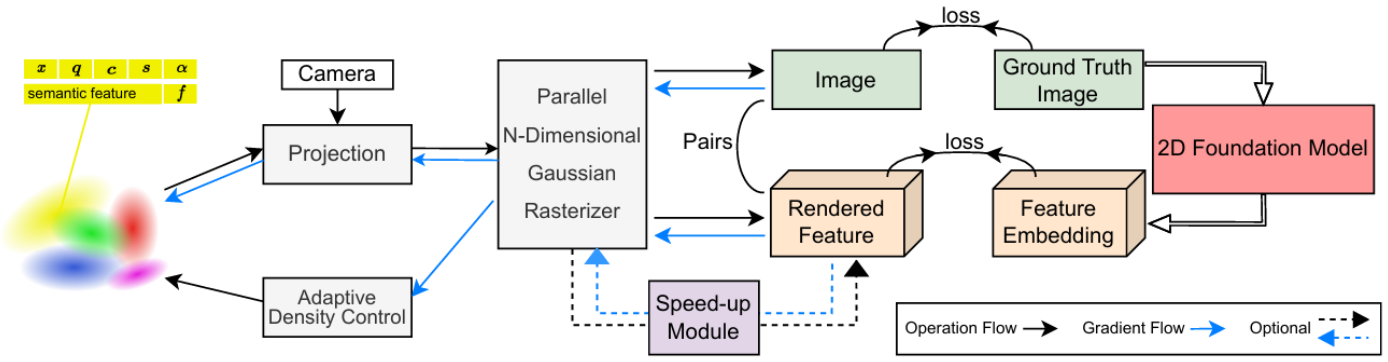
2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting

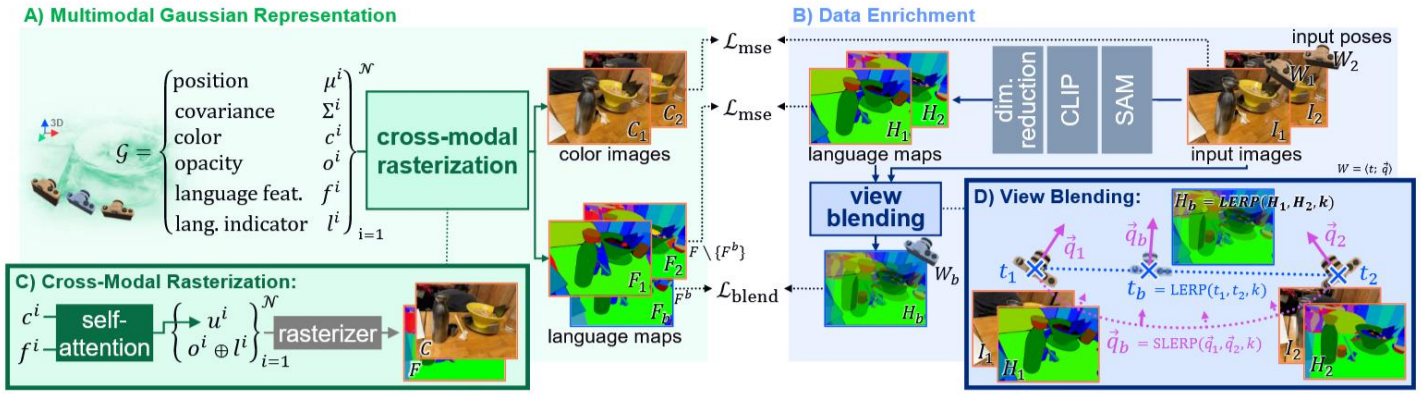
Yue Chen, Xingyu Chen, Anpei Chen, Gerard Pons-Moll, Yuliang Xiu
1Westlake University 2Max Planck Institute for Intelligent Systems 3University of T¨ ubingen, T¨ ubingen AI Center
4Max Planck Institute for Informatics, Saarland Informatics Campus

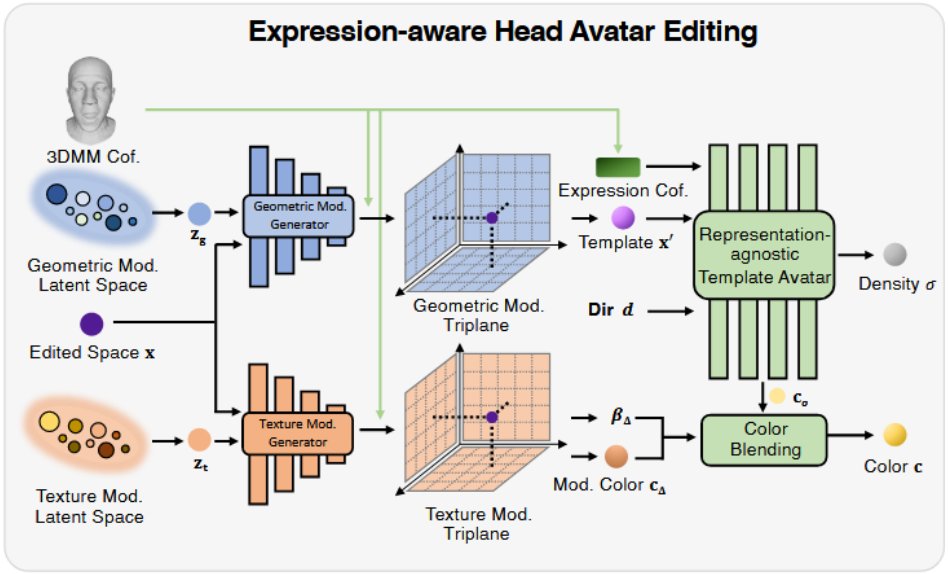
1. Latent Radiance Fields with 3D-aware 2D Representations，ICLR 2025

背景

- 将 2D 特征蒸馏到 3D 空间，有助于增强 3D 语义理解和 3D 生成
- 现有方法没有解决 2D 特征空间和 3D 表示之间的 domain gap，导致渲染性能下降



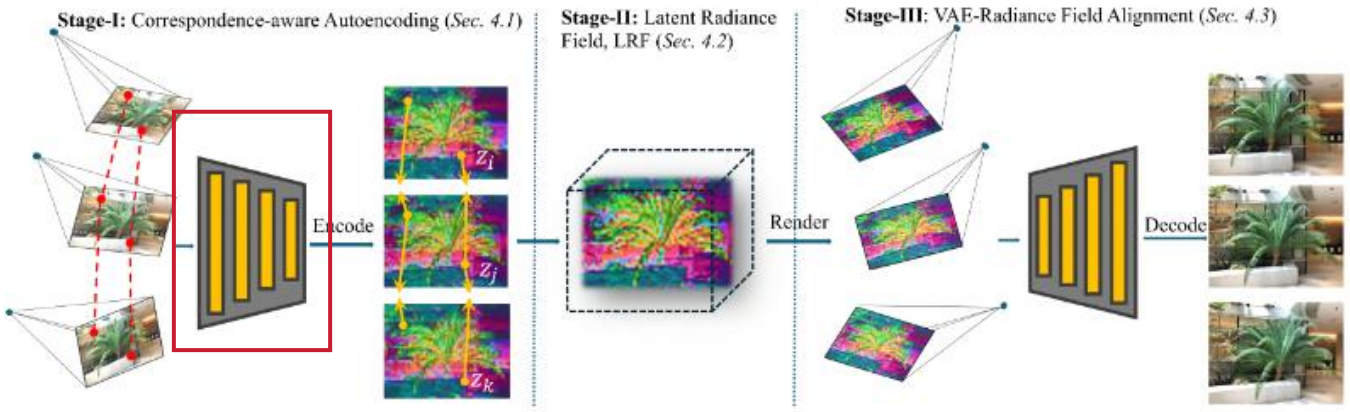




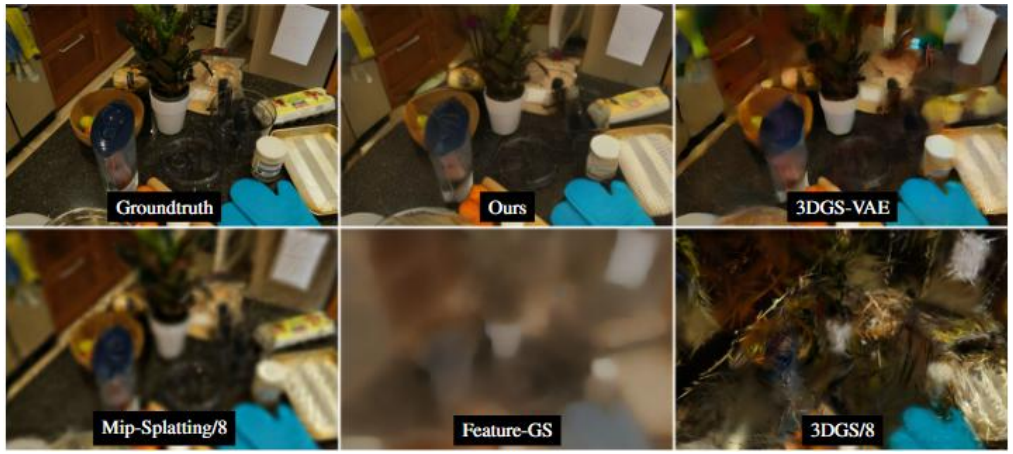
1. Latent Radiance Fields with 3D-aware 2D Representations，ICLR 2025

Stage Ⅰ

- 在 latent space 做几何一致性约束
- 建立一个 latent Radiance Field 来表示 3D 场景
- 对齐 VAE 和 Radiance Field



$$x_j^\top F_{ij} x_i = 0.$$



latent space 破坏一致性 引入大量高频噪声

$$\mathcal{L}_{\text{VAE}}(\theta, \phi; \mathbf{X}) = \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{X})}[\log p_\theta(\mathbf{X}|\mathbf{Z})] - \mathrm{KL}(q_\phi(\mathbf{Z}|\mathbf{X}) \| p(\mathbf{Z})).$$

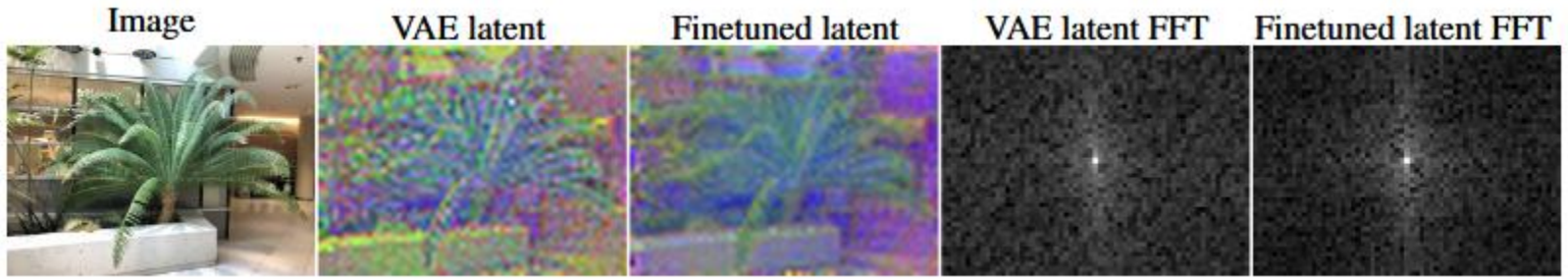
$$\mathcal{L}_{\text{corres}} = \sum_{i=1}^{K} \sum_{j \in \mathcal{K}(i)} \lambda_{ij} \left\| z_i - z_j \right\|_1 \qquad \mathcal{L}_{\text{reg}} = -\mathrm{KL}\left(q(Z|X) \parallel q_{\text{original}}(Z|X)\right)$$

$$\mathcal{L}_{\text{Stage1}} = \mathcal{L}_{\text{VAE}} + \lambda_1 \mathcal{L}_{\text{corres}} + \lambda_2 \mathcal{L}_{\text{reg}}.$$

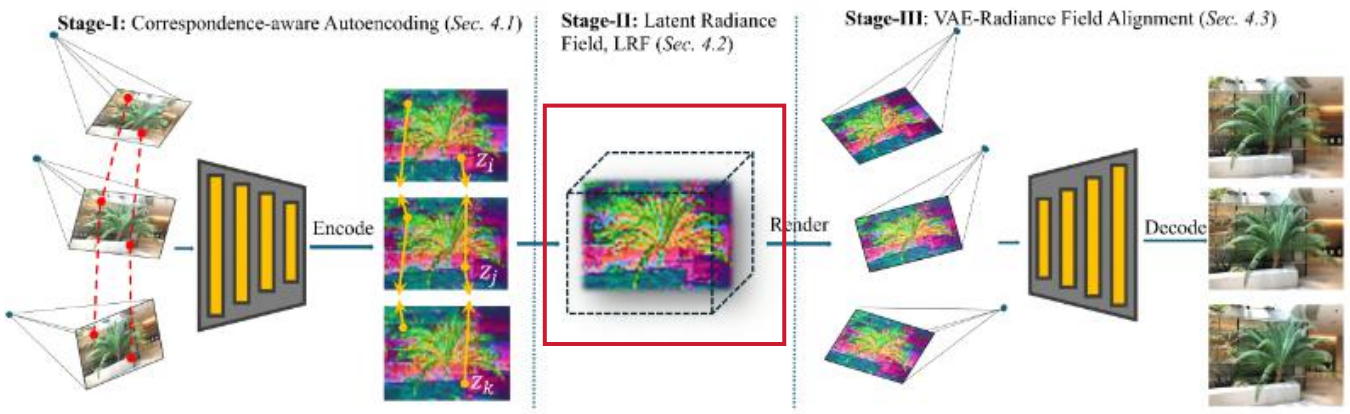
保留多视角几何一致性，能够兼容现有的 NVS 框架

1. Latent Radiance Fields with 3D-aware 2D Representations，ICLR 2025

Stage Ⅱ



高频噪声得以有效去除



在 VAE 的 2D 潜在空间中直接创建 3D 表示

可微光栅化过程

$$C = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j).$$

$$\mathbf{Z} = \sum_{i \in \mathcal{N}} \mathbf{z}_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_i),$$

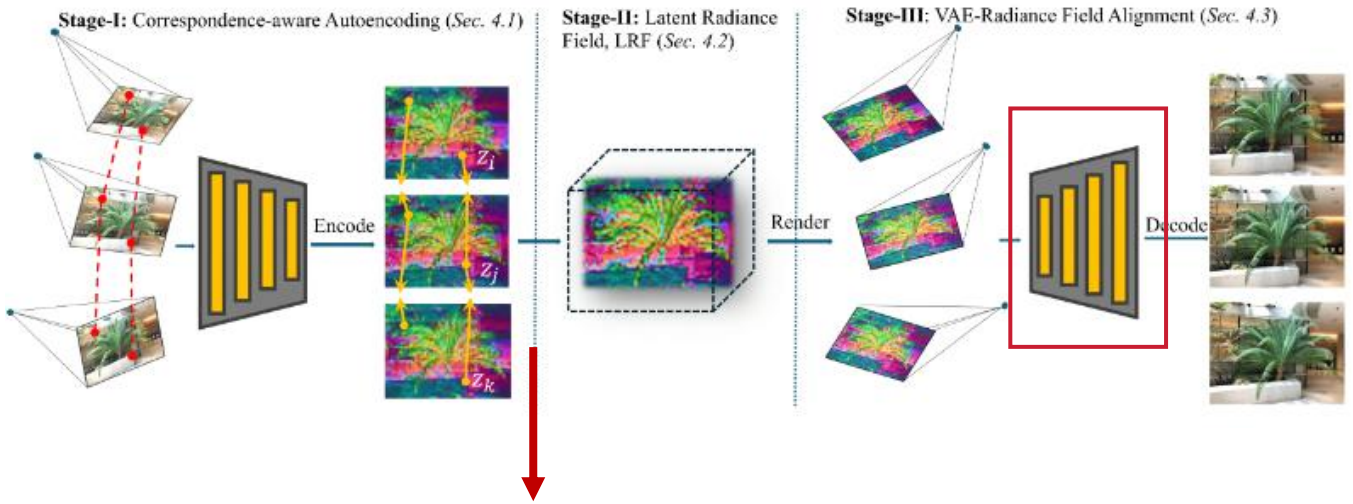
$$\hat{\mathcal{G}} = \underset{\{(\boldsymbol{\mu},\mathbf{s},\mathbf{R},\alpha,\mathbf{SH}_f\}}{\arg\min} \sum_{i=1}^{N} \mathcal{L}^f(r(\mathcal{G}, \mathbf{P}_i), \mathbf{Z}_i),$$

不再需要引入额外的几何一致性约束

1. Latent Radiance Fields with 3D-aware 2D Representations，ICLR 2025

Stage Ⅲ

- 由于神经渲染的非线性特性，$p(Z_{NVS})$ 偏离 $p(Z_{VAE})$
- 现有 NVS 方法对训练视角进行过拟合



仅用部分视角训练LRF

非线性特性的来源

- 场景的复杂几何和光照
- 渲染过程中的累积效应
- 视图依赖性

$$\mathcal{L}_{\text{StageIII}} = \lambda_{\text{train}} \left\| D(Z_{\text{train}}) - I_{\text{train}} \right\|_1 + \lambda_{\text{novel}} \left\| D(Z_{\text{novel}}) - I_{\text{novel}} \right\|_1 ,$$

平衡LRF的训练视角和新视角

1. Latent Radiance Fields with 3D-aware 2D Representations，ICLR 2025

实验

Table 1: Our method outperforms the image and latent space NVS baselines on most settings and metrics, from object-level to unbounded outdoor scenes. Latent-NeRF* denotes we adapt it to NVS.

Dataset	Metric	Image Space 3DGS/8	Image Space Mip-Splatting/8	Latent Space 3DGS-VAE	Latent Space Latent-NeRF*	Latent Space Feature-GS	Latent Space 3DGS-LRF (Ours)
MVImgNet	PSNR ↑	16.93	24.89	25.04	18.50	21.09	26.26
	SSIM ↑	0.561	0.799	0.824	0.709	0.772	0.863
	LPIPS ↓	0.466	0.328	0.250	0.403	0.372	0.178
NeRF-LLFF	PSNR ↑	9.98	19.68	19.07	18.31	16.48	20.00
	SSIM ↑	0.110	0.484	0.493	0.457	0.415	0.541
	LPIPS ↓	0.631	0.513	0.364	0.387	0.539	0.289
DL3DV-10K	PSNR ↑	14.03	21.81	20.57	18.16	16.60	22.45
	SSIM ↑	0.352	0.609	0.595	0.530	0.449	0.667
	LPIPS ↓	0.541	0.451	0.346	0.432	0.602	0.197
Mip-NeRF360	PSNR ↑	14.79	22.38	19.44	15.93	17.13	20.83
	SSIM ↑	0.273	0.502	0.404	0.312	0.337	0.469
	LPIPS ↓	0.586	0.521	0.432	0.537	0.642	0.328

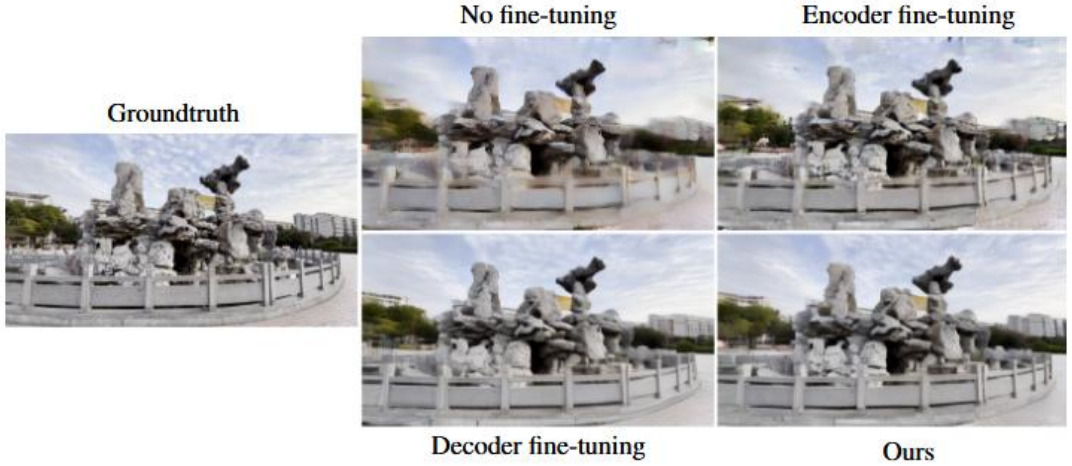
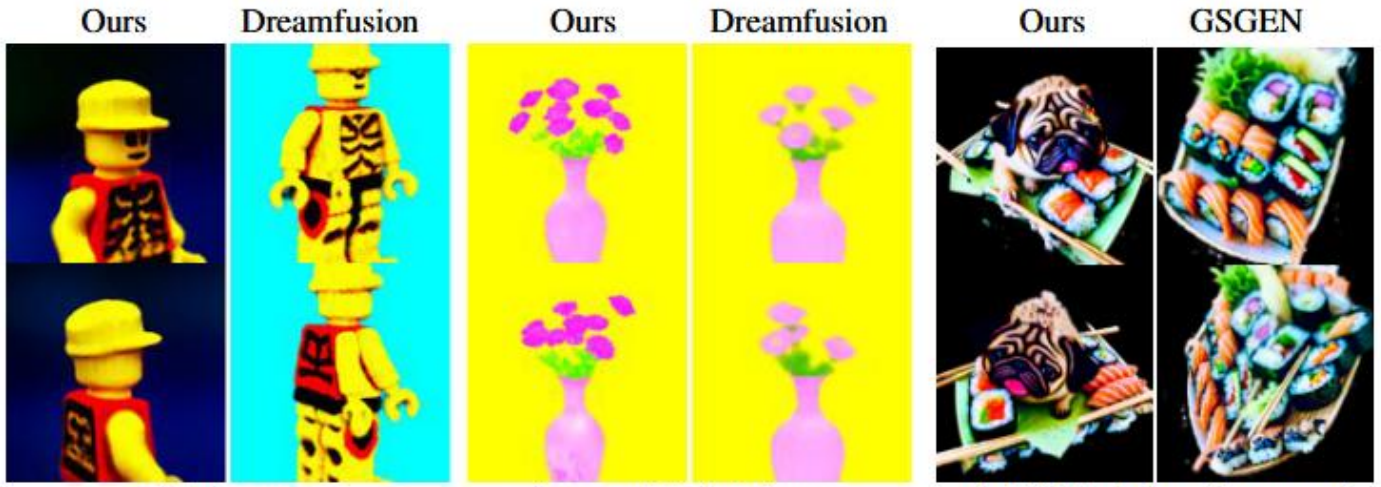


Table 3: We ablate correspondence-aware autoencoding and VAE-radiance field aligned decoder fine-tuning on DL3DV-10K dataset to reveal their necessity in latent 3D reconstruction .

VAE	Encoder fine-tuned	Decoder fine-tuned	PSNR↑	SSIM↑	LPIPS↓
✓	-	-	20.57	0.595	0.346
✓	✓	-	21.16	0.620	0.282
✓	-	✓	21.73	0.645	0.208
✓	✓	✓	**22.45**	**0.667**	**0.197**

1. Latent Radiance Fields with 3D-aware 2D Representations，ICLR 2025

实验



A lego man *A vase with pink flowers* *A DSLR photo of a tray of sushi containing pugs*

A DSLR photo of a tray of sushi containing pugs *A zoomed out DSLR photo of a cake in the shape of a train* *A zoomed out DSLR photo of a plate of fried chicken and waffles*

对下游任务影响

Method	Metric	NeRF-LLFF	DL3DV-10K	Mip-NeRF360
VAE	PSNR ↑	23.47	24.59	24.54
Our-VAE	PSNR ↑	23.59	23.25	24.24

稀疏视角下的性能对比

Table 2: A comparison of different methods on LLFF dataset using 3 views.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
3DGS-VAE	13.06	0.283	0.570
3DGS	13.79	0.331	0.468
Mip-Splatting	13.70	0.315	0.486
Ours	**15.51**	**0.379**	**0.465**

1. Latent Radiance Fields with 3D-aware 2D Representations，ICLR 2025

总结

核心思想：

将三维重建任务从传统的三维空间转化为二维潜在空间，并通过潜在辐射场（LRF）实现更高效的三维表示。

核心创新点：

- **二维潜在空间的辐射场表示**：传统辐射场方法通常在三维空间中构建，捕捉光照信息和视角变化的细节。而 LRF 将这种表示方法转移到二维潜在空间，把三维场景的几何和纹理信息压缩成二维潜在表示。
- **三维感知的二维表示学习**：引入一个对应感知的自编码方法，使得二维潜在表示在编码过程中能够捕捉到三维空间中的深度和形状信息。
- **VAE-辐射场对齐策略：**缩小二维潜在空间和三维自然空间之间的差距，使得模型能够在二维潜在空间中实现更高效的三维重建，并增强视觉质量。

2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting，CVPR 2025　　背景

问题：视觉基础模型（VFMs）通常在大规模数据集上训练，但往往局限于2D图像，缺少对3D世界的理解能力的评估。

VFMs 在学习策略和代理任务等方面的多种差异使得公平且全面的基准测试变得困难。

现有的3D感知工作

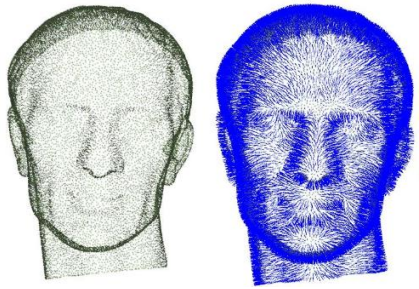
单视角2.5D估计

双视角稀疏2D对应

→

忽略了纹理感知、多视角一致性

需要 3D GT，限制评估集规模和多样性



深度估计与法线估计

图像匹配

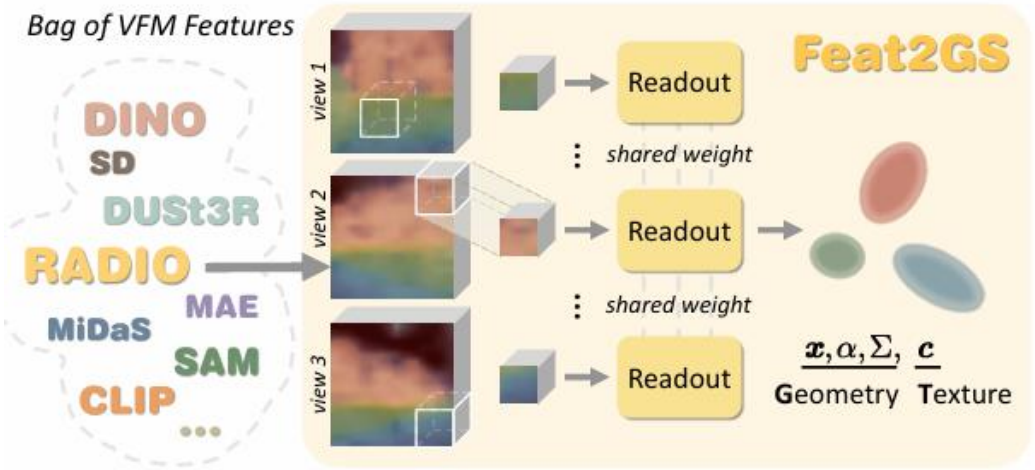
3D感知能力是否必须来自3D数据?

训练方法重要性有多大?

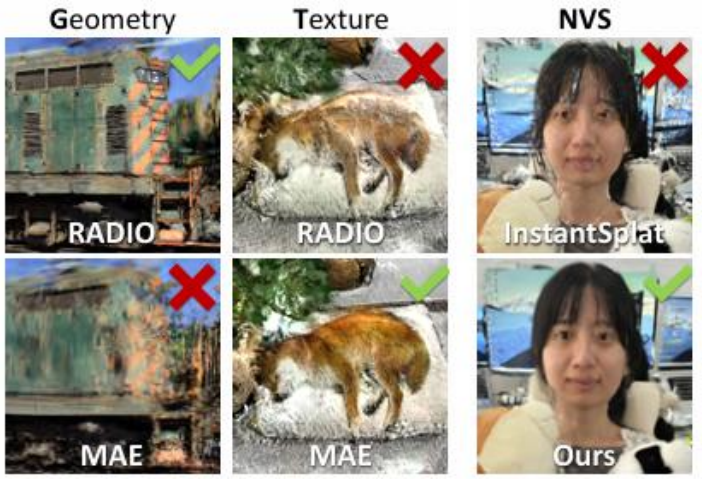
目标：探测多种 VFMs 的 3D 感知能力，并研究哪些因素有助于构建具备 3D 感知能力的 VFM。

2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　背景

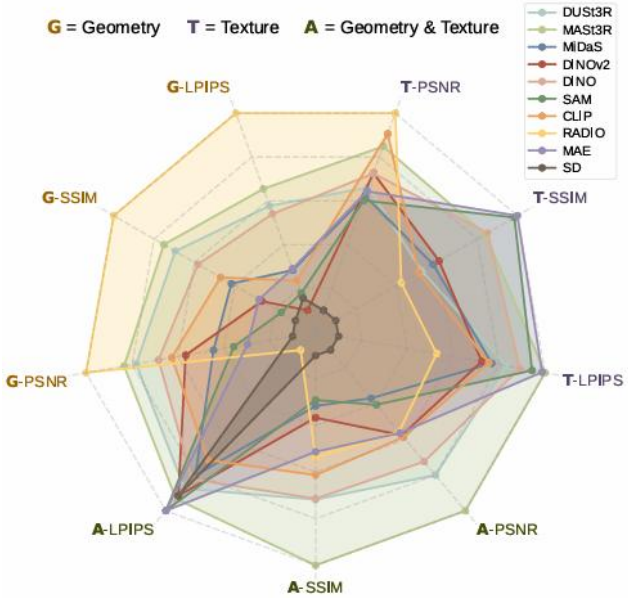
NVS：允许每个输入像素都对评估做出贡献，而不仅是稀疏匹配点。与 2D 稀疏匹配不同，NVS 仅需要图像，无需昂贵的视觉对应标注。大量公开的多视图数据集涵盖多样化场景和视角，为评估提供支撑。



(a) Feat2GS as VFM Probe



(b) Extensive Analysis of VFM　(c) Strong Baseline

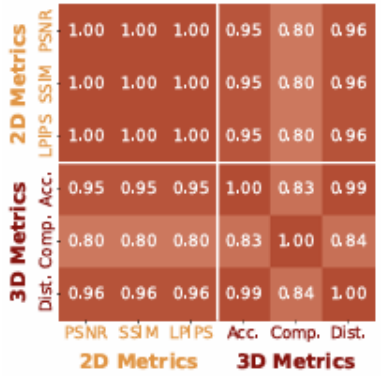


- 训练用多视图光度损失最小化渲染结果与输入之间的视觉差异
- 测试通过未见过的视图，在多样化数据集中测量视觉相似性指标
- DUSt3R 初始化相机参数，通过光度损失优化稀疏且未校准的随意图像
- GS 的参数被分为几何 (x,α,Σ) 和纹理 c，可以切换学习模式

Feat2GS	-Geometry	-Texture	-All	InstantSplat [22]
Feature-Readout	x, α, Σ	c	x, c, α, Σ	-
Free-Optimize	c	x, α, Σ	-	x, c, α, Σ

	2D Metrics			3D Metrics		
Feature	PSNR↑	SSIM↑	LPIPS↓	Acc.↓	Comp.↓	Dist.↓
DUSt3R	21.36	.7772	.2195	2.439	1.316	6.955
MASt3R	21.44	.7792	.2177	2.321	1.286	6.557
MiDaS	21.09	.7712	.2254	2.934	1.412	8.230
DINOv2	21.01	.7695	.2277	3.101	1.337	8.588
DINO	21.40	.7783	.2187	2.440	1.316	6.885
SAM	20.93	.7660	.2304	3.176	1.339	8.785
CLIP	21.26	.7752	.2215	2.357	1.209	6.739
RADIO	21.78	.7871	.2042	1.886	1.326	5.431
MAE	20.96	.7666	.2289	2.963	1.337	8.374
SD	20.76	.7638	.2343	4.334	1.603	11.594
IUVRGB	16.09	.6825	.3134	13.015	16.957	46.671

(a) 2D Metrics *vs.* 3D Metrics

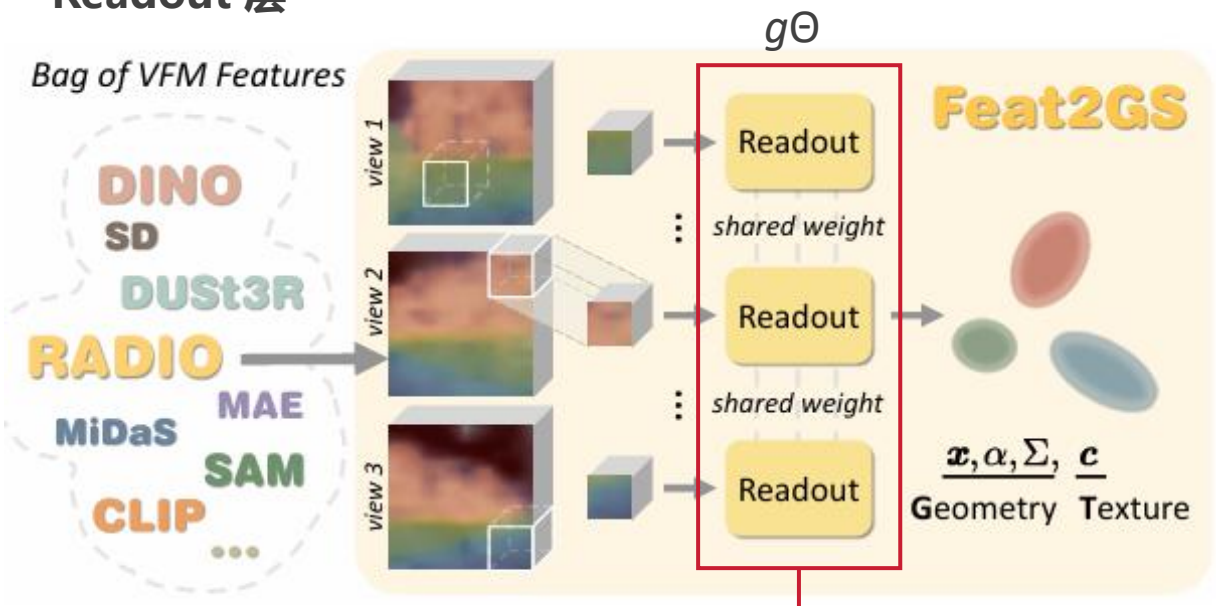


(b) Correlation Matrix

证明可以使用NVS的2d指标来评估

2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　方法

Readout 层



(a) Feat2GS as VFM Probe

$$\boldsymbol{G}_i = g_\Theta(\mathbf{f}_i),$$　　两层 256 nodes 的MLP层

- 用 DUSt3R 初始化相机姿态 T，然后通过光度损失联合优化读出层 $g\Theta$ 和相机姿态T：　　$\min_{\Theta,\boldsymbol{T}} \|\mathcal{R}(g_\Theta(\mathbf{f}), \boldsymbol{T}) - \mathcal{I}\|$.
- 为了确保对来自不同基础模型的特征进行鲁棒评估，使用点云回归进行热启动优化：　　$\min_{\Theta} \|g_\Theta(\mathbf{f}) - \boldsymbol{G}_{init}\|$,

用于评估的 VFMs

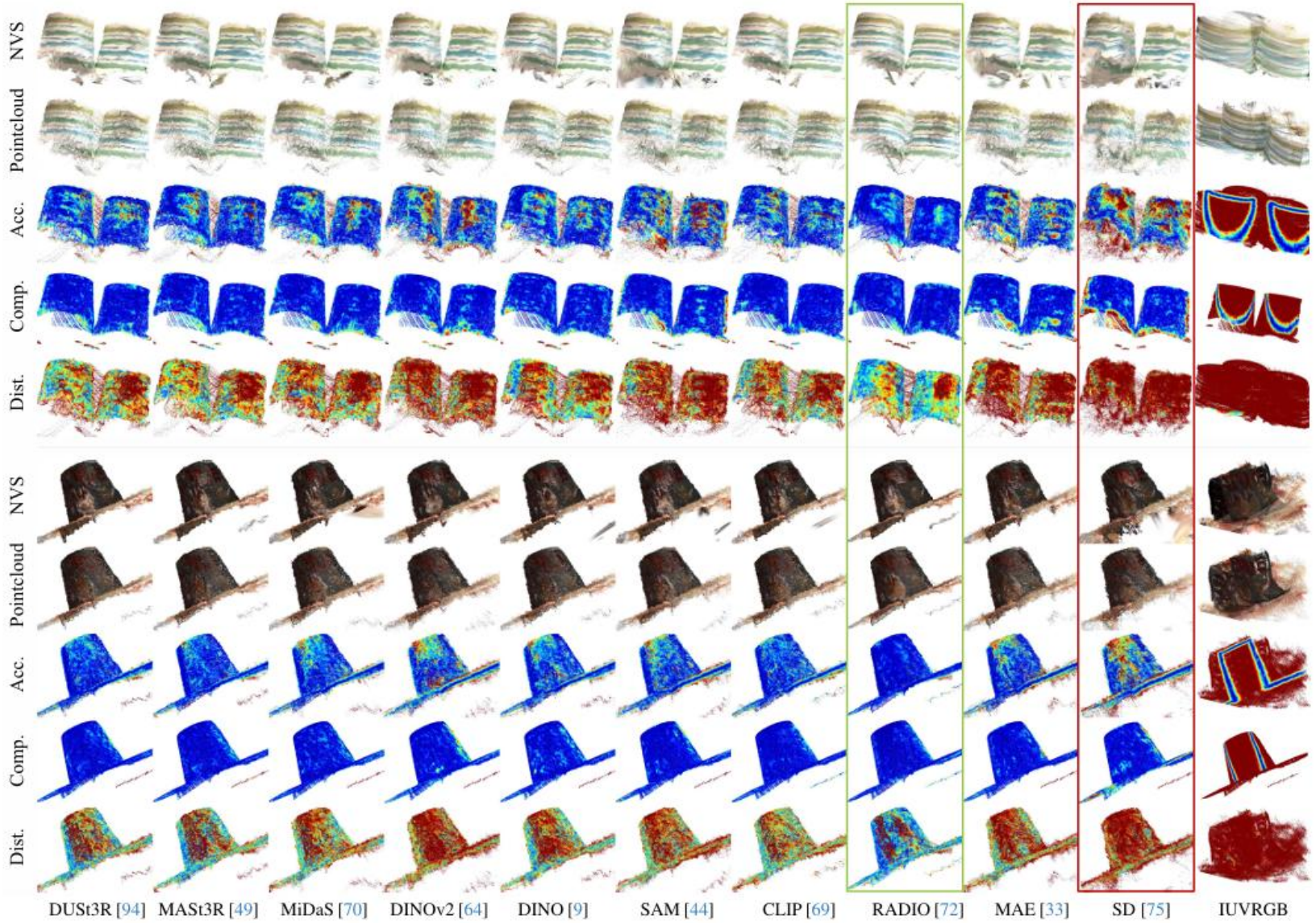
VFM	Arch.	Channel	Supervision	Dataset
DUSt3R [94]	ViT-L/16	1024	Point Regression	3D DUSt3R-Mix
MASt3R [49]	ViT-L/16	1024	Point Regression	3D MASt3R-Mix
MiDaS [70]	ViT-L/16	1024	Depth Regression	3D MiDaS-Mix
DINOv2 [64]	ViT-B/14	768	Self Distillation	2D LVD-142M
DINO [9]	ViT-B/16	768	Self Distillation	2D ImageNet-1k
SAM [44]	ViT-B/16	768	Segmentation	2D SA-1B
CLIP [69]	ViT-B/16	512	Contrastive VLM	2D WIT-400M
RADIO [72]	ViT-H/16	1280	Multi-teacher Distillation	2D DataComp-1B
MAE [33]	ViT-B/16	768	Image Reconstruction	2D ImageNet-1k
SD [75]	UNet	1280	Denoising VLM	2D LAION

用于评估的数据集

Dataset	Scene Type	Complexity	View Range	Views
LLFF [60]	Indoor	Simple	Small	2
DTU [1]	Indoor Object	Simple	Small	3
DL3DV [52]	Indoor / Outdoor	Moderate	Medium	5-6
Casual	Daily Scenario	Moderate	Medium	4-7
MipNeRF360 [4]	Unbounded	Moderate	360	6
MVimgNet [111]	Outdoor Object	Moderate	180-360	2-4
T&T [46]	Indoor / Outdoor	High	Large	6

2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　实验

可视化进一步展示 NVS 的质量与 3D 指标的高相关性



2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　实验

总体性能对比

在多个数据集上对 VFMs 的特征进行三种模式的测试

worst to best.

	LLFF									DL3DV									Casual								
	Geometry			Texture			All			Geometry			Texture			All			Geometry			Texture			All		
Feature	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DUSt3R	19.88	.7442	.2123	19.01	.7120	.2262	19.87	.7190	.2691	19.64	.7338	.3196	18.01	.6815	.3219	19.39	.7360	.3458	19.29	.6562	.3580	17.54	.5693	.3750	19.19	.6556	.4050
MASt3R	19.89	.7447	.2123	19.01	.7115	.2261	19.99	.7250	.2657	19.64	.7334	.3188	18.07	.6813	.3211	19.41	.7373	.3464	19.30	.6550	.3576	17.59	.5708	.3722	19.37	.6588	.4027
MiDaS	19.81	.7420	.2154	19.00	.7129	.2261	19.86	.7142	.2733	19.47	.7271	.3311	17.94	.6796	.3224	19.22	.7291	.3493	19.24	.6545	.3612	17.52	.5693	.3757	18.96	.6516	.4073
DINOv2	19.77	.7345	.2226	19.04	.7133	.2254	19.91	.7163	.2637	19.47	.7293	.3288	18.00	.6805	.3223	19.27	.7317	.3479	19.42	.6524	.3698	17.64	.5701	.3754	19.21	.6535	.4023
DINO	19.81	.7423	.2140	18.98	.7121	.2260	19.97	.7212	.2744	19.60	.7324	.3209	17.97	.6790	.3219	19.41	.7359	.3476	19.24	.6513	.3614	17.50	.5683	.3756	19.10	.6566	.4056
SAM	19.72	.7354	.2181	18.98	.7133	.2260	19.76	.7144	.2629	19.48	.7297	.3271	17.97	.6822	.3218	19.20	.7272	.3459	19.32	.6469	.3704	17.52	.5725	.3736	19.19	.6569	.3981
CLIP	19.78	.7378	.2221	19.02	.7113	.2276	19.74	.7136	.2822	19.53	.7295	.3304	18.05	.6771	.3235	19.22	.7310	.3563	19.21	.6552	.3719	17.46	.5669	.3743	19.05	.6582	.4084
RADIO	19.73	.7402	.2207	19.06	.7101	.2301	19.56	.6999	.3252	19.48	.7313	.3139	18.03	.6748	.3254	19.20	.7316	.3654	19.54	.6545	.3465	17.52	.5666	.3748	18.67	.6533	.4216
MAE	19.75	.7363	.2183	19.00	.7128	.2249	19.92	.7209	.2612	19.54	.7288	.3248	17.98	.6821	.3207	19.34	.7310	.3448	19.03	.6502	.3690	17.51	.5691	.3758	19.18	.6547	.3974
SD	19.62	.7293	.2234	18.85	.7100	.2297	19.78	.7121	.2656	19.31	.7251	.3276	17.79	.6784	.3260	19.10	.7282	.3500	19.24	.6483	.3649	17.38	.5698	.3789	18.86	.6505	.4053
IUVRGB	15.55	.5765	.3986	19.75	.7303	.2262	15.38	.6175	.4308	14.78	.6326	.4541	18.75	.7023	.3250	14.05	.6431	.4386	13.17	.5454	.5248	17.88	.5927	.3846	13.71	.5917	.4955

	MipNeRF 360									MVImgNet									Tanks and Temples (T&T)								
	Geometry			Texture			All			Geometry			Texture			All			Geometry			Texture			All		
Feature	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DUSt3R	20.82	.5008	.3795	19.10	.4489	.3816	21.02	.5048	.4752	19.47	.6004	.3073	16.88	.5348	.3334	19.43	.5937	.3674	18.85	.6458	.3715	17.53	.6222	.3328	18.61	.6477	.4023
MASt3R	20.92	.5093	.3745	19.21	.4540	.3803	20.92	.5054	.4749	19.49	.6008	.3032	16.91	.5350	.3337	19.49	.5983	.3637	18.80	.6428	.3703	17.68	.6238	.3319	18.76	.6512	.3991
MiDaS	20.89	.5059	.3815	19.05	.4509	.3813	20.84	.5004	.4795	19.35	.5900	.3222	16.82	.5336	.3343	19.34	.5910	.3672	18.53	.6374	.3798	17.64	.6238	.3333	18.32	.6428	.4039
DINOv2	20.81	.4946	.3953	19.05	.4495	.3821	20.75	.4924	.4684	19.35	.5896	.3246	16.88	.5359	.3344	19.43	.5943	.3674	18.71	.6432	.3772	17.58	.6214	.3348	18.43	.6443	.4064
DINO	20.91	.5054	.3769	19.18	.4545	.3795	20.83	.5010	.4772	19.44	.5982	.3071	16.90	.5394	.3329	19.41	.5952	.3683	18.75	.6416	.3733	17.66	.6233	.3330	18.61	.6467	.4030
SAM	20.73	.4913	.3945	19.14	.4556	.3775	20.75	.4949	.4639	19.23	.5899	.3188	16.84	.5346	.3346	19.29	.5915	.3649	18.65	.6421	.3780	17.49	.6217	.3338	18.43	.6425	.4029
CLIP	20.80	.4982	.3913	19.28	.4543	.3807	20.88	.4984	.4773	19.41	.5945	.3098	16.96	.5362	.3358	19.37	.5969	.3695	18.92	.6463	.3729	17.81	.6226	.3316	18.75	.6515	.4052
RADIO	20.87	.5100	.3620	19.35	.4550	.3819	20.91	.5067	.5127	19.54	.6105	.2949	16.99	.5373	.3366	19.60	.5955	.3946	19.19	.6612	.3480	17.84	.6225	.3321	19.01	.6574	.4109
MAE	20.82	.4992	.3884	19.14	.4572	.3781	20.79	.4995	.4668	19.23	.5909	.3142	16.84	.5355	.3328	19.25	.5914	.3680	18.65	.6395	.3758	17.55	.6234	.3333	18.49	.6451	.4000
SD	20.71	.4962	.3985	18.89	.4472	.3839	20.59	.4929	.4672	19.08	.5881	.3185	16.63	.5313	.3389	19.06	.5838	.3660	18.69	.6422	.3772	17.32	.6217	.3374	18.55	.6467	.4020
IUVRGB	16.45	.4075	.5910	19.96	.4797	.3911	16.41	.4187	.5929	14.83	.5069	.4648	17.84	.5568	.3431	15.38	.5362	.4699	15.29	.5846	.4736	18.60	.6526	.3396	15.17	.5948	.4718

几何模式：RADIO > MASt3R > DUSt3R

纹理模式，表现差异显著：MAE > SAM > MASt3R

全部模式，MASt3R 和 DUSt3R最高分，其次是 DINO

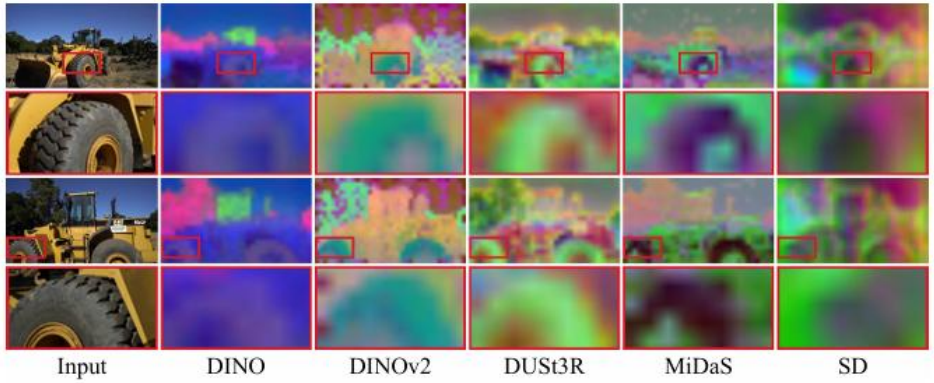
Stable Diffusion (SD) 在大多数指标中表现最差

2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　实验

纹理感知能力分析

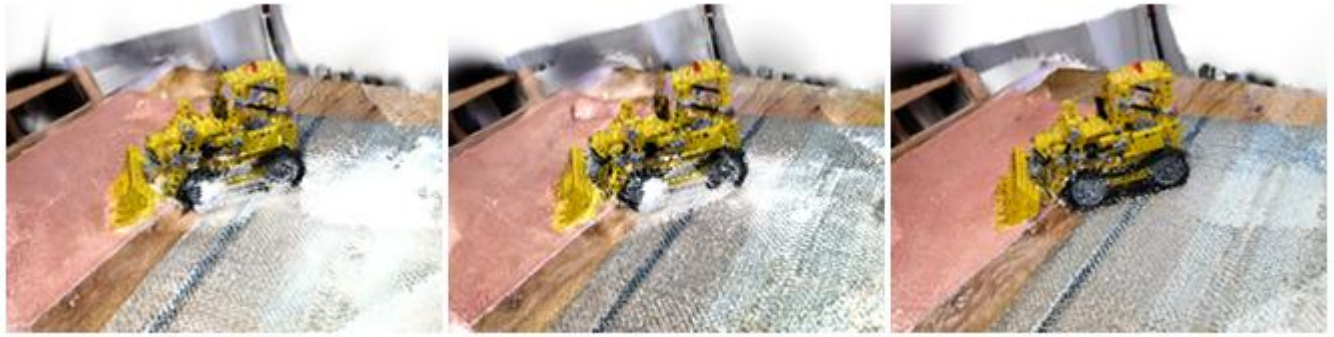
NVS 中特征一致性的对比

(a) NVS Comparison on Geometry Awareness



(b) VFM Features from Training Views

VFM 特征在纹理模式下表现较差

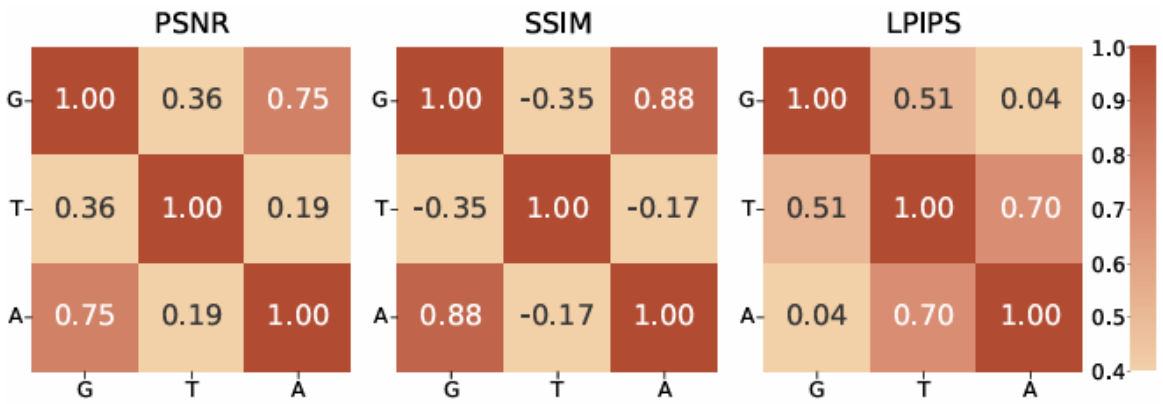


VFMs 的纹理不变性是可能的解释，对纹理细节不敏感

2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　实验

纹理感知能力分析

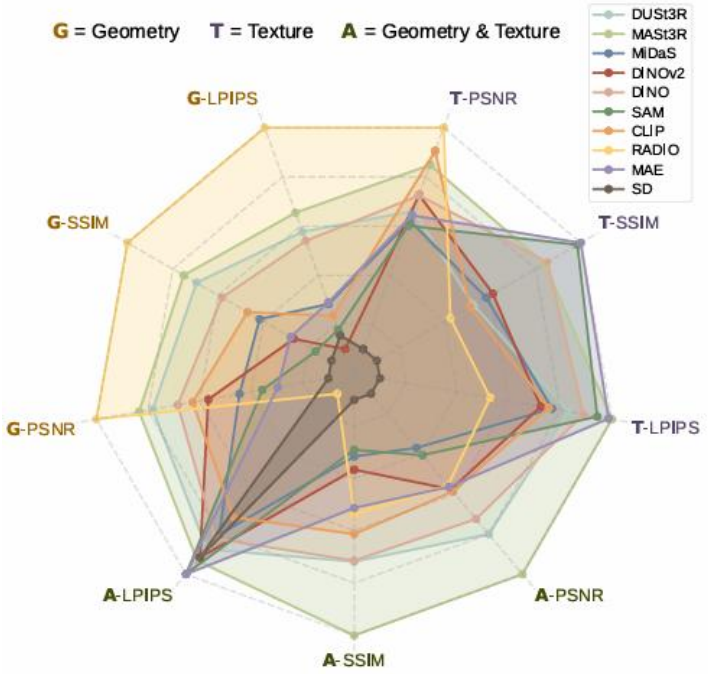
全部数据集上三种模式的性能关联性



- PSNR 和 SSIM 指标上，all 模式与 geometry 模式相关性更强
- LPIPS 指标上，all 模式与 texture 模式的相关性更高

进一步支持观点：all 模式的模糊现象源于 VFMs 缺乏纹理感知能力

什么模型的纹理感知能力强?

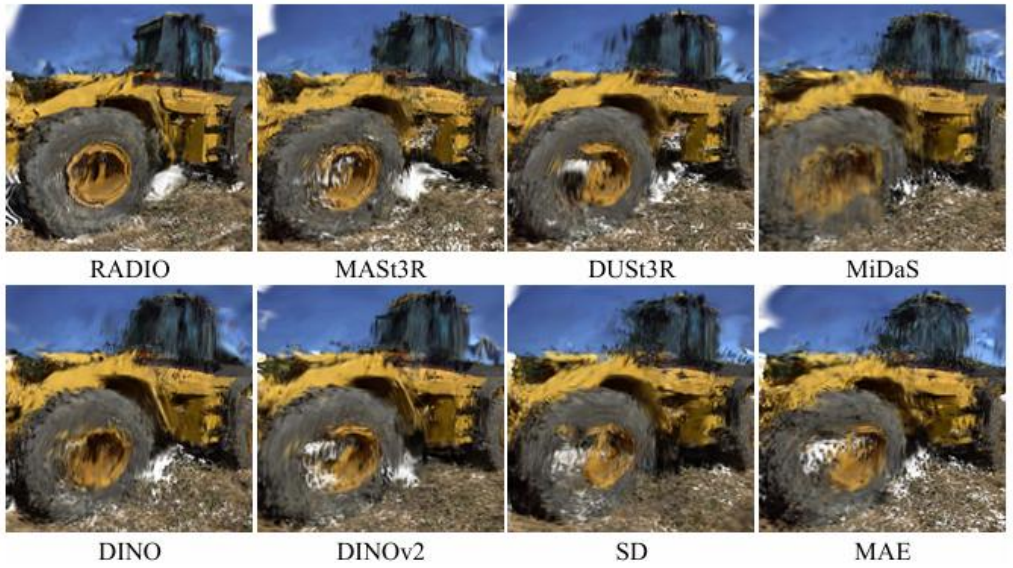


- 掩码图像重建预训练的 VFM 在恢复纹理能力上表现更加
- 基于去噪的图像重建的模型，可能导致颜色偏移

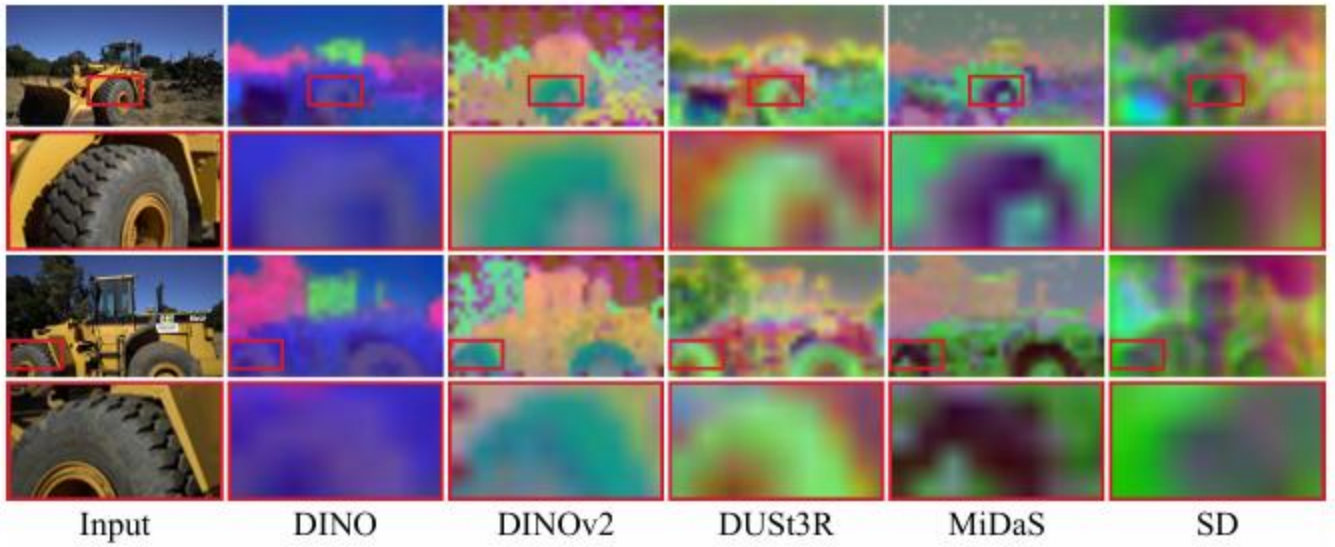
GT　　SD　　MAE

2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　实验

几何感知能力分析



(a) NVS Comparison on Geometry Awareness



(b) VFM Features from Training Views

RADIO 、**MASt3R** 、**DUSt3R** 和 **DINO** 在几何感知指标中排名前四，也意味着更强的跨视角一致性

3D 数据确实显著提高了几何感知能力 →

MASt3R 和 **DUSt3R**： 使用点云图训练（3D）

MiDaS： 使用深度图训练（2.5D）

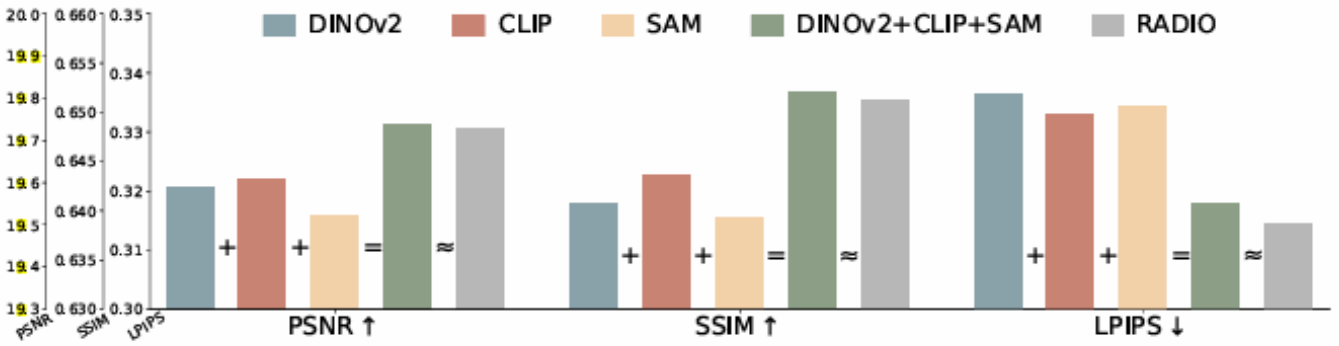
2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　实验

模型集成分析

RADIO 通过蒸馏 **DINOv2** 、**CLIP** 和 **SAM** 到一个单一模型中，实现了最佳的几何感知能力

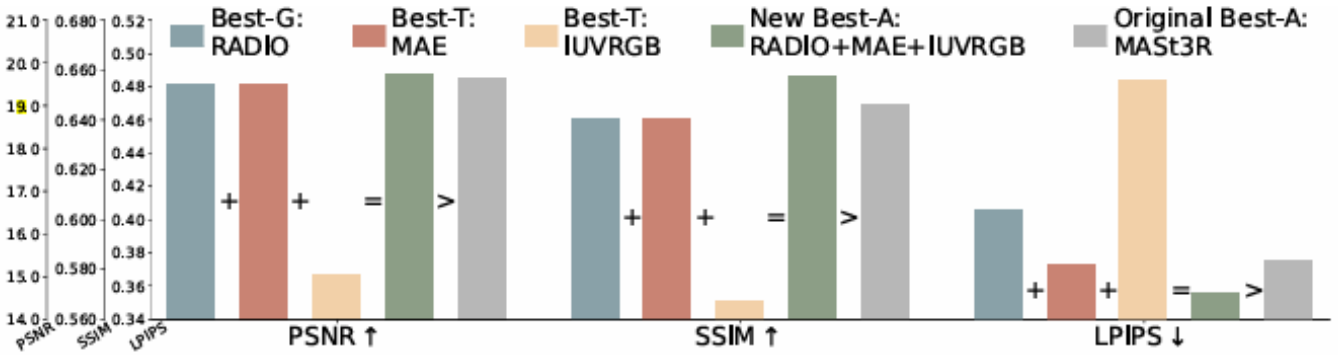
将 **DINOv2** 、**CLIP** 和 **SAM** 的特征进行拼接，使用 PCA 将特征通道数减少到 256，保持网络规模不变进行比较

Geo 模式



(a) Concatenated Features *vs.* RADIO

All 模式



(b) Concatenated Features *vs.* MASt3R

直接利用多个模型的特征可能比通过蒸馏整合它们更有效

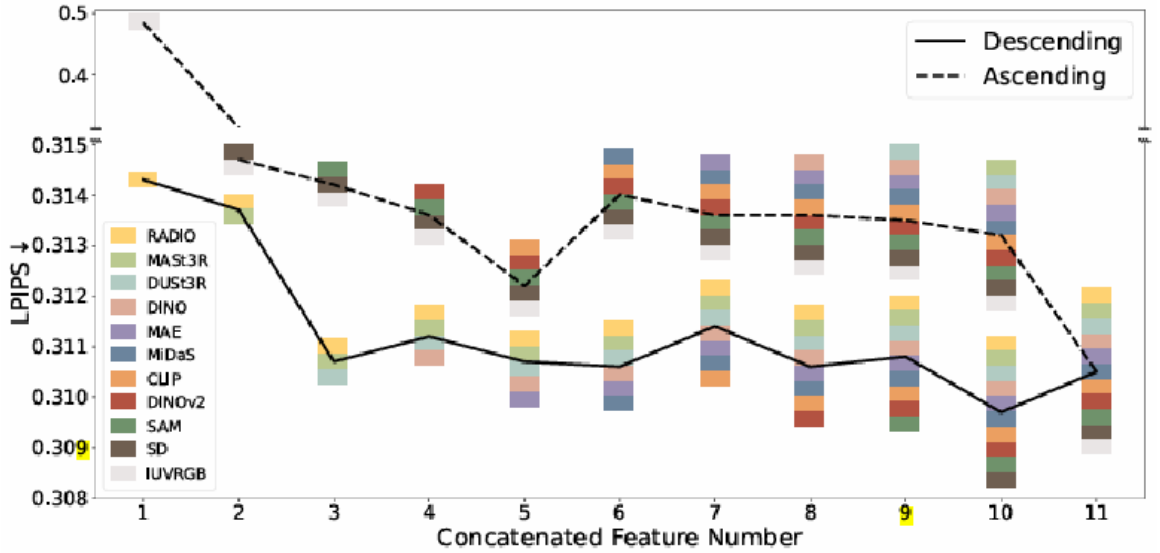
2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　实验

模型集成分析

特征微调的作用

	All Datasets								
	Geometry			Texture			All		
Feature	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DINOv2	19.59	.6406	.3364	18.03	.5951	.3291	19.50	.6388	.3760
DINOv2^{+}	19.67	.6480	.3202	18.10	.5950	.3291	19.58	.6443	.3894
DINOv2*	19.78	.6552	.2962	18.18	.5968	.3232	19.80	.6614	.3247
DINO	19.63	.6452	.3256	18.03	.5961	.3282	19.55	.6427	.3793
DINO^{+}	19.72	.6485	.3207	18.03	.5941	.3291	19.64	.6465	.3839
DINO*	19.74	.6557	.2918	18.09	.5949	.3235	19.69	.6630	.3154
CLIP	19.61	.6436	.3331	18.10	.5947	.3289	19.50	.6416	.3832
CLIP^{+}	19.68	.6466	.3222	18.09	.5941	.3286	19.63	.6468	.3842
CLIP*	19.70	.6540	.2959	18.19	.5962	.3242	19.67	.6599	.3199

Geo 模式下，按性能升降序拼接特征



Geo 模式下，几种变体与 SOTA 比较

	All Datasets		
Method	PSNR ↑	SSIM ↑	LPIPS ↓
InstantSplat [22]	18.87	0.6044	0.3039
Feat2GS w/ RADIO	19.73	0.6513	0.3143
Feat2GS w/ concat all	**19.80**	0.6545	0.3105
Feat2GS w/ DUSt3R	19.66	0.6469	0.3247
Feat2GS w/ DUSt3R*	19.75	**0.6561**	**0.2928**

2. Feat2GS: Probing Visual Foundation Models with Gaussian Splatting， CVPR 2025　　总结

核心思想：

提出了 **Feat2GS** （一种将 VFMs 的特征映射到 3D 高斯点云的方法），能够在不需要 3D GT的情况下，通过 2D 图像探索 VFMs 几何和纹理的感知能力。

关键启发：

- VFMs 在捕捉几何信息方面表现良好，但在处理纹理信息时存在困难。3D 点云图对于学习多视角一致、几何感知的模型很重要；而纹理感知能力从掩码图像重建预训练中明显受益。
- 能够有效利用 VFMs 来完成随意拍摄、稀疏视角的新视角合成（NVS）任务。
- VFMs 的特征集成是一个值得探索的方向。

Limitations：

- Feat2GS 依赖无约束立体重建器初始化相机姿态和点云。
- 基于 3D Gaussian splatting，当前仅适用于静态场景。