ClaraVid: A Holistic Scene Reconstruction Benchmark From Aerial Perspective With Delentropy-Based Complexity Profiling

Radu Beche *Sergiu Nedevschi*

TECHNICAL UNIVERSITY OF CLUJ-NAPOCA ROMANIA · hria HUB ROMÂN DE INTELIGENȚĂ ARTIFICIALĂ

Project page

ICCV OCT 19-23, 2025 HONOLULU HAWAII

1. Introduction

Context. Neural reconstruction has advanced rapidly, yet UAV scene understanding lacks datasets unifying *semantics and geometry*. Existing resources remain *low-resolution or task-specific*, limiting holistic evaluation.

Goal. A dataset for evaluating *holistic semantic and geometric reconstruction* from UAV imagery that spans multiple *levels of visual complexity*

Contributions.

- 📊 *Delentropic Scene Profile* -- entropy-based measure of scene complexity
- 🌐 *ClaraVid* -- synthetic 4K aerial benchmark with multi-view, multimodal labels
- 📝 Study showing strong correlation shown between *complexity and neural reconstruction error*

2. Scene Complexity Estimation

- In neural scene reconstruction…

  scene complexity can be estimated directly from 2D imagery*

$$p_{ij} = \frac{1}{4XY} \sum_{n=-N}^{N-1} \sum_{m=-M}^{M-1} \delta_{i,f_x(m,n)} \delta_{j,f_y(m,n)} \qquad H_{\text{del}} = -\frac{1}{2} \sum_{i}^{I} \sum_{j}^{J} p_{ij} \log_2 p_{ij}$$

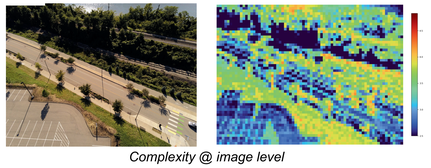
- We propose a framework for quantifying scene complexity

  Delentropic Scene Profile (DSP)

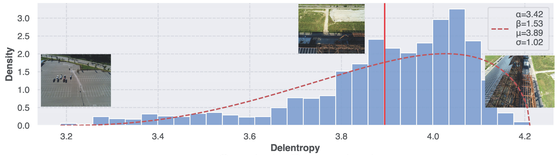
$$\begin{aligned} \text{DSP}_S &= Beta(H_{\text{del}} \mid \alpha, \beta, a, b) \\ &= \frac{(H_{\text{del}} - a)^{\alpha-1} (b - H_{\text{del}})^{\beta-1}}{(b-a)^{\alpha+\beta-1} B(\alpha, \beta)} \end{aligned}$$

  - $\mathbf{H_{del}}$ **-** image delentropy (complexity)
  - **α, β -** skew toward easy or hard regions
  - **a, b -** normalizing factors
  - **μ -** average scene complexity

- **A higher delentropy means a more complex image**



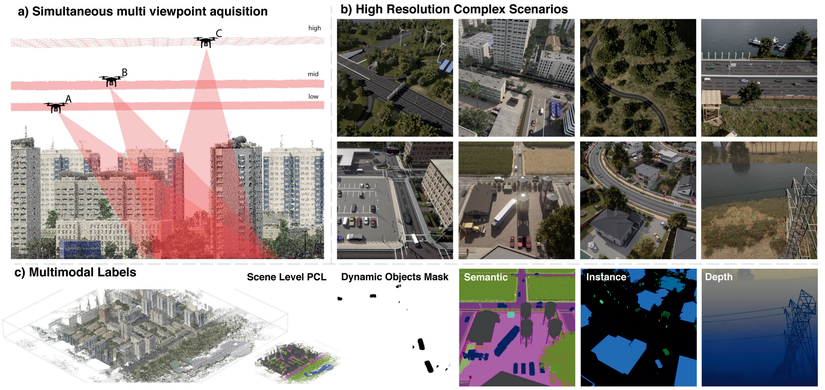
Complexity @ image level



Complexity @ scene level

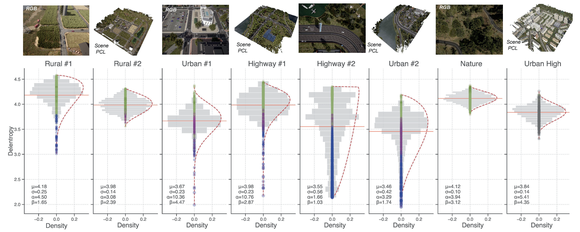
* given a structured collection policy

3. ClaraVid Dataset



16 917 multimodal samples across **5 environments** and **8 missions**

- Targeted toward neural reconstruction and holistic scene understanding
- Multimodal samples captured from **3 viewpoints** in a **mapping-oriented mode**

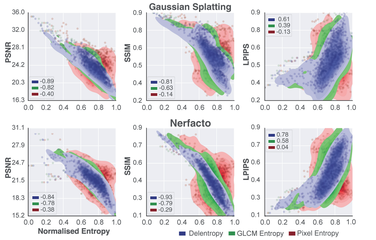


ClaraVid** has a broad spectrum of **scene complexity profiles

4. Experimental results

Delentropic Scene Profile

- **Delentropy** correlates strongly with **reconstruction error** across NeRF and Gaussian Splatting representations
- **DSP** captures scene-level trends, serving as a **quantitative measure** for evaluating neural reconstruction



Scene	DSP			Instant-NGP[40]			TensoRF [61]			Nerfacto[62]			ZipNeRF[63]			Gaussian Splatting[7]		
	μ	α	β	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Rural #1	*4.18*	*4.50*	*1.65*	18.27	0.3171	0.8492	18.11	0.2898	0.9204	18.36	0.2981	0.9022	18.56	0.4462	0.8818	19.97	0.5351	0.6019
Nature	*4.12*	*3.94*	*3.12*	21.40	0.3993	0.7057	21.02	0.3338	0.8213	21.60	0.3772	0.7705	24.55	0.5184	0.5829	22.94	0.5586	0.5515
Rural #2	*3.98*	*3.08*	*2.39*	20.75	0.3769	0.7192	20.61	0.3525	0.7890	21.29	0.3721	0.8145	24.54	0.5010	0.6645	22.59	0.5402	0.5727
Highway #1	*3.98*	*10.76*	*2.87*	20.32	0.3165	0.9148	20.41	0.3240	0.9043	20.96	0.3548	0.8862	21.73	0.4792	0.7960	22.03	0.4769	0.6814
Urban High	*3.84*	*5.41*	*4.35*	21.26	0.4878	0.5150	20.75	0.4465	0.5564	22.39	0.4797	0.5893	24.30	0.6920	0.4049	24.19	0.6211	0.4324
Urban #1	*3.67*	*10.36*	*4.47*	22.25	0.5507	0.4829	21.96	0.5230	0.5271	23.24	0.5680	0.5321	26.39	0.7451	0.3614	26.37	0.7507	0.2981
Highway #2	*3.55*	*1.66*	*1.03*	22.97	0.5174	0.6389	22.78	0.4944	0.6887	22.66	0.5238	0.6668	24.51	0.6964	0.4609	24.61	0.6504	0.4858
Urban #2	*3.29*	*3.29*	*1.74*	23.85	0.5734	0.4925	23.66	0.5511	0.5438	24.24	0.5798	0.5485	25.42	0.6945	0.3443	26.68	0.7125	0.3658
Average	***3.97***	***5.37***	***2.70***	21.38	0.4424	0.6648	21.16	0.4144	0.7189	21.84	0.4442	0.7138	**23.75**	0.5966	0.5622	23.67	**0.6057**	**0.4987**

ClaraVid

- Enables controlled evaluation across **multiple complexity levels**.
- Provides multimodal data for **semantic + geometric reconstruction evaluation** from multiple viewpoints
- For more results, please check the paper

Train \Pred	Nerfacto[62] 45° L	55° M	90° H	Gaussian Splatting[7] 45° L	55° M	90° H
Appearance Synthesis						
PSNR ↑						
45° L	**21.36**	20.46	19.08	**22.98**	21.56	20.50
55° M	20.25	**21.43**	20.48	20.75	**22.75**	21.06
90° H	18.23	19.42	**22.45**	17.88	19.26	**23.00**
SSIM ↑						
45° L	**0.42**	0.40	0.36	**0.58**	0.50	0.42
55° M	0.39	**0.42**	0.40	0.45	**0.58**	0.41
90° H	0.34	0.40	**0.47**	0.36	**0.65**	0.59
LPIPS ↓						
45° L	0.75	**0.74**	0.76	**0.55**	0.61	0.78
55° M	0.77	**0.73**	0.74	0.59	**0.53**	0.62
90° H	0.79	0.71	**0.68**	0.81	0.73	**0.50**
Semantic Segmentation						
mIoU ↑						
45° L	**44.82**	44.14	40.78	**56.53**	55.11	53.50
55° M	43.12	**46.57**	45.74	53.34	**59.59**	56.50
90° H	36.95	41.64	**51.08**	36.68	40.82	**58.54**
Depth						
Abs Rel ↓						
45° L	0.059	**0.042**	0.066	0.025	0.017	**0.010**
55° M	0.042	0.043	**0.027**	0.023	0.016	**0.009**
90° H	0.054	0.039	**0.022**	0.028	0.019	**0.010**
RMSE ↓						
45° L	12.34	**10.64**	14.22	4.67	3.37	**2.36**
55° M	11.47	9.09	**6.55**	4.61	3.07	**1.93**
90° H	14.20	8.95	**4.68**	5.41	3.53	**2.08**
$\delta < 1.05$ ↑						
45° L	0.8193	**0.8467**	0.8317	0.881	0.92	**0.95**
55° M	0.8315	0.8584	**0.9024**	0.895	0.93	**0.96**
90° H	0.7886	0.8354	**0.8980**	0.871	0.91	**0.95**