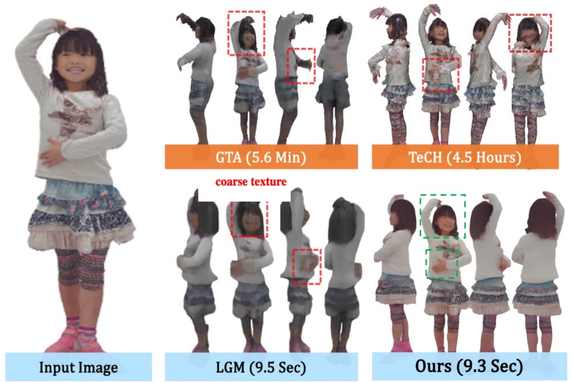
HumanSplat: Generalizable Single-Image Human Gaussian Splatting with Structure Priors

Panwang Pan*[1], Zhuo Su*†[1], Chenguo Lin*[1,2], Zhen Fan[1], Yongjie Zhang[1], Zeming Li[1], Tingting Shen[3], Yadong Mu[2], Yebin Liu‡[4]

[1]ByteDance, [2]Peking University, [3]Xiamen University, [4]Tsinghua University

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation



The Drawbacks of Related Works:

❖ **Lack human priors as inductive biases.**

❖ **The requirements for dense input images.**

❖ **Time-consuming per-instance optimization.**

Contribution

❖ **It is the first work to propose a novel generalizable human Gaussian Splatting network for high-fidelity human reconstruction from a single image.**

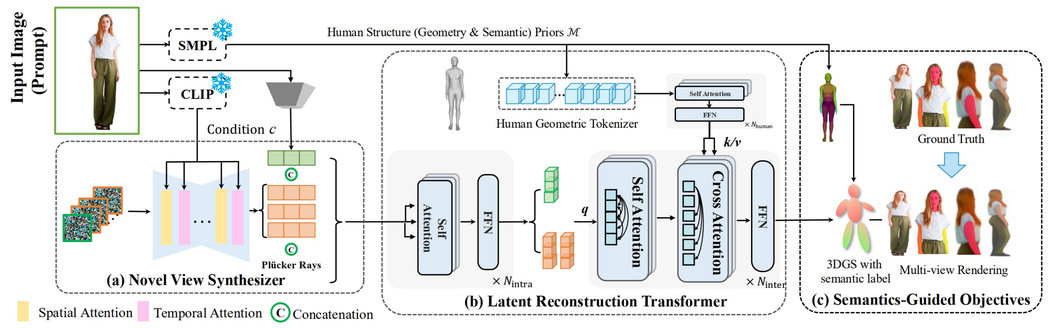
❖ **We integrate a universal Transformer framework by leveraging human geometry priors and appearance priors from the 2D generative diffusion model.**

❖ **We enhance the fidelity of reconstructed human models by introducing semantic cues, hierarchical supervision, and tailored loss functions.**

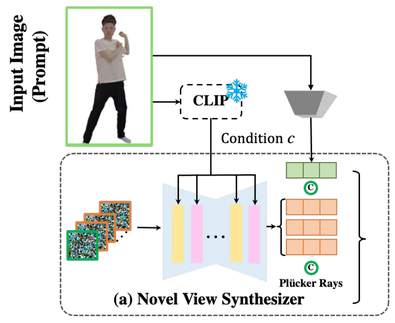
❖ **Extensive experiments demonstrate that our method achieves SOTA performance.**

Methodology

Overview of HumanSplat



Novel-view Synthesizer

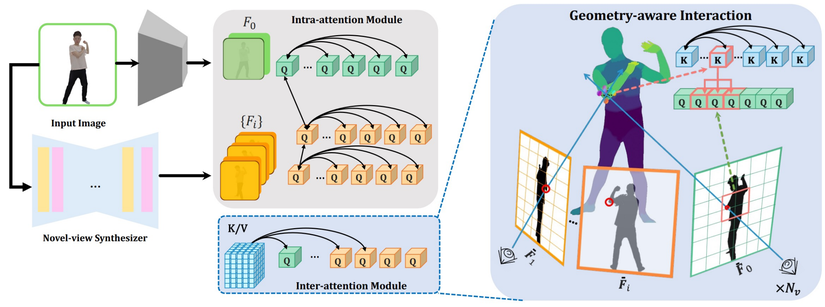


$$\mathbb{E}_{\epsilon\sim p(\epsilon)}\left[\lambda(\epsilon)\left|\left|(\mathrm{D}_\theta\{F_i^\epsilon\}_{i=1}^N;\{e_i,a_i\}_{i=1}^N,c,F_0,\epsilon)-\{\mathrm{F_i}\}_{\mathrm{i=1}}^{\mathrm{N}}\right|\right|\right]$$

Semantics-guided Objectives

$$\mathcal{L}_H=\frac{1}{n}\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_i\lambda_j(\mathcal{L}_{\text{MSE}}(\mathbf{I}_{i,j},\hat{\mathbf{I}}_{i,j})+\lambda_p\mathcal{L}_{\text{p}}(\mathbf{I}_{i,j},\hat{\mathbf{I}}_{i,j})),$$

Latent Reconstruction Transformer



Experiments

❖ Quantitative Comparison

Quantitative comparison against SOTA methods.

Method	Category		THuman2.0 [105]			Twindom [106]			Time
	Diffusion	Human Prior	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
PIFu [107]	✗	✗	18.093	0.911	0.137	-	-	-	30.0s
LGM* [29]	✓	✗	20.013	0.893	0.116	19.840	0.851	0.292	9.5s
GTA [30]	✗	✓	18.050	-	-	17.669	0.741	0.418	43s
SIFU [102]	✗	✓	22.025	0.921	0.084	19.714	0.832	0.312	44s
SIFU† [102]	✓	✓	22.102	0.923	0.079	-	-	-	6min
Magic123† [108]	✓	✗	14.501	0.874	0.145	-	-	-	1h
HumanSGD† [22]	✓	✓	17.365	0.895	0.130	-	-	-	7min
TeCH† [21]	✓	✓	**25.211**	**0.936**	0.083	21.192	0.884	0.188	4.5h
HumanSplat (Ours)	✓	✓	24.033	0.918	**0.055**	**23.346**	**0.913**	**0.125**	**9.3s**

❖ Qualitative Comparison

Qualitative comparison against SOTA methods.