Cross-Resolution Gaussian Splatting (XR-GS) for Hybrid Resolution Photography

Shreyans Sethi, Vaishnavi Khindkar – 16-825: Learning for 3D Vision

Carnegie Mellon University

Motivation:

Modern 3D Gaussian Splatting (3D-GS) achieves high-quality novel-view synthesis, but it assumes **all training images share a uniform resolution**. In real photography, especially handheld phone capture, this rarely holds:

- Some images are **high-resolution (HR)** → Optical zoom, moved-in close-ups, better cameras
- Many are **low-resolution (LR)** → Digital zoom, compression, wide shots

When forced into a single-scale optimization, GS suffers from:

- Hallucinated high-frequency details
- Over-sharpened edges / inconsistent geometry
- Overfitting to noise in LR images
- Unnecessary Gaussian growth (Leading to slow rendering)
- Loss of genuine fine structure from HR views

Goal:

Enable 3D-GS to *intelligently utilize* mixed-resolution supervision so the model learns **global structure** from LR images and **true detail** from HR images.

By being *resolution-aware*, XR-GS can treat views differently

Dataset:

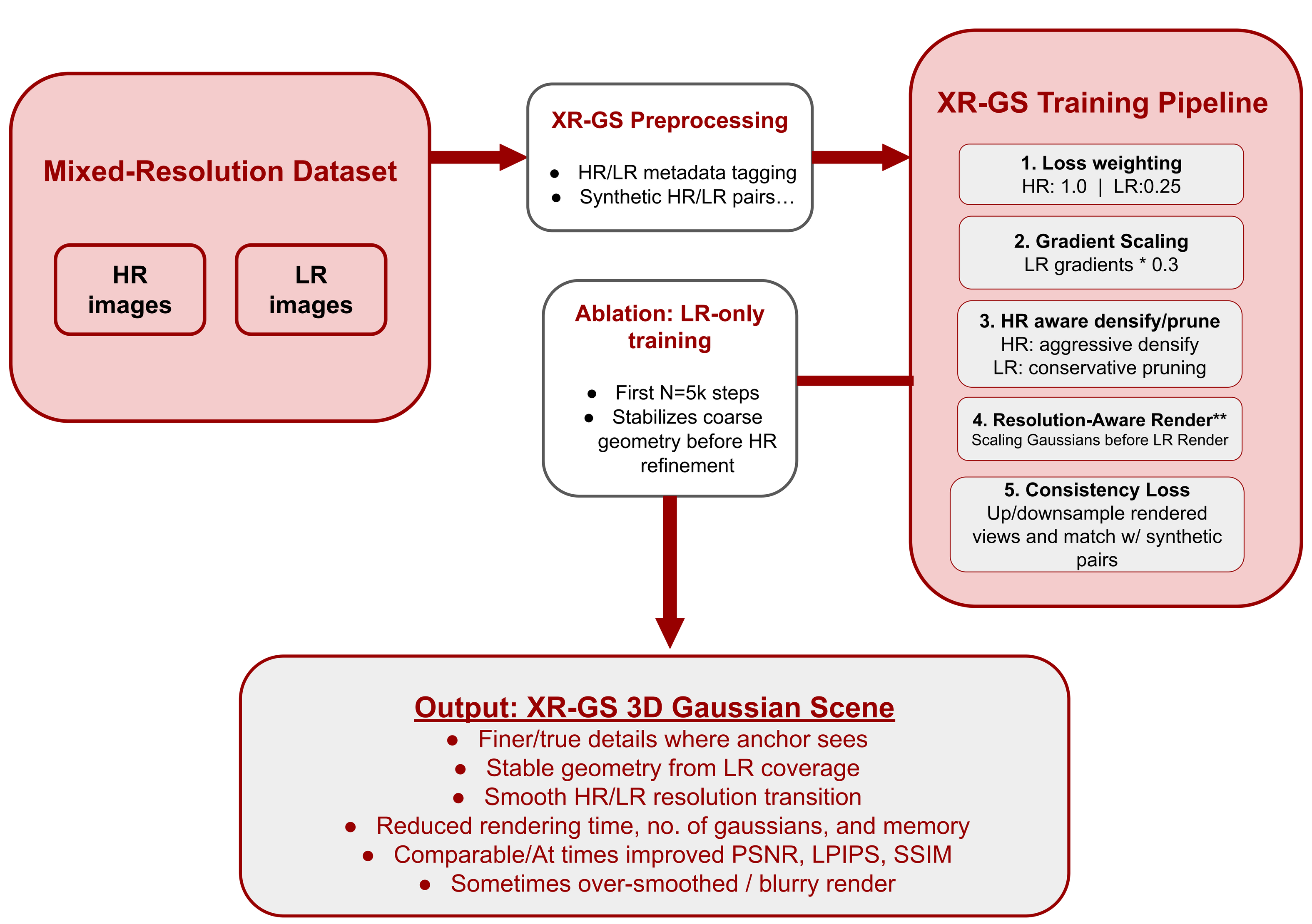
Created a synthetic hybrid-resolution benchmark using **LLFF (*7 Scenes)**

- **30%** of Images left as "High-Resolution", remaining (**70%**) are downsampled (5x) to be "Low-Resolution"
- COLMAP is re-run on mixed-res images, HR/LR metadata is passed into the dataset for use in XR-GS changes

Captured **2 scenes** ourselves using mixed-resolution situations:

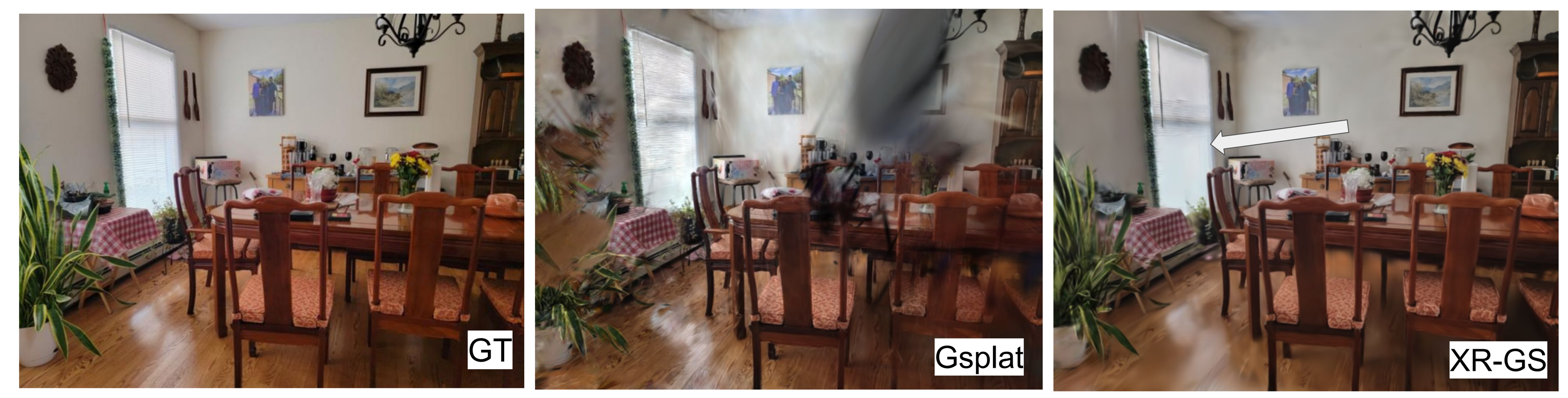
1. Dining room with zoomed out views for HR, close-ups of objects for LR
2. Keychain with 2 different cameras (1 having a higher megapixel count)

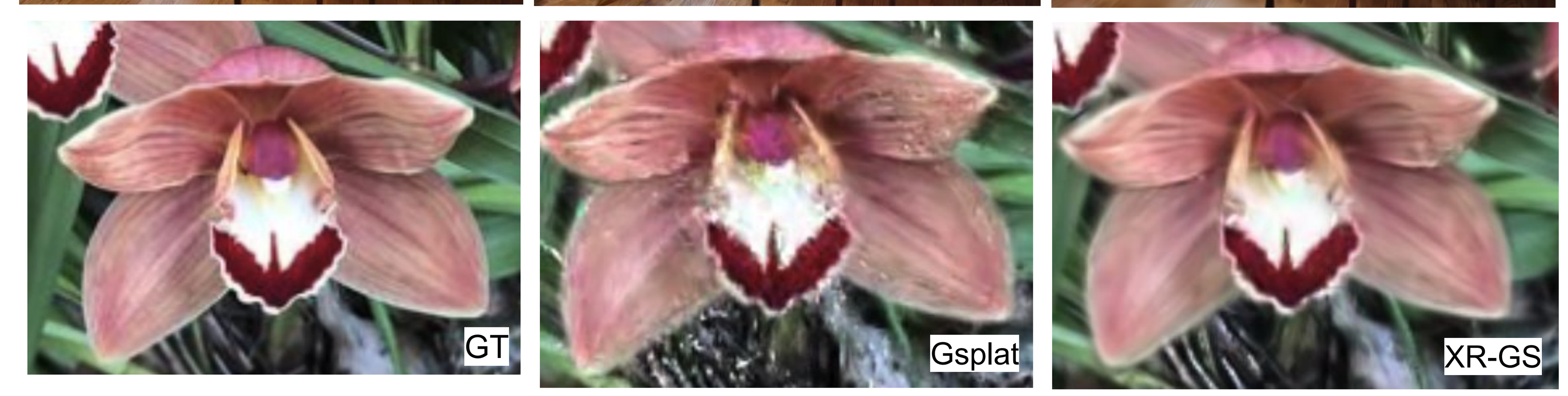
Approach:



Analysis:

Resolution-Aware Densify Prune Strategy → HR views focus Gaussian capacity on fine structure; LR views avoid over-densifying noise. Leads to structure preservation, less graininess but possible over smoothing





Loss Weighting + Gradient Scaling → Prevent LR views from overwhelming HR supervision; Helps "oversharpened/noisy look"; True to real detail in the HR view

Unsuccessful Approach: Resolution-Aware Rendering

Intuition: Splats are shared, low-res images may send a conflicting signal to "blur" fine detail
Idea: Scale radii before rendering low-res views (then scale back), mimic optical blur

Issue: Larger splats spread out the gradients (smoothing everywhere) and oversensitive to scaling values

Better Approach: Consistency Loss → Persists structure across scale by having synthetic LR/HR pairs supervise; Suppresses noise that only appears in HR renders but is not supported by the LR; Smoother geometry when visualized from novel view

$$I_{\text{proj}} = \phi(I)$$

$$\mathcal{L}_{\text{cons}} = w \begin{pmatrix} \lambda_1 \, \|I_{\text{proj}} - \tilde{I}\|_1 \\ + \; \lambda_2 \left(1 - \text{SSIM}(I_{\text{proj}}, \tilde{I})\right) \end{pmatrix}$$

Hallucinated LR Noise is Punished by Loss:

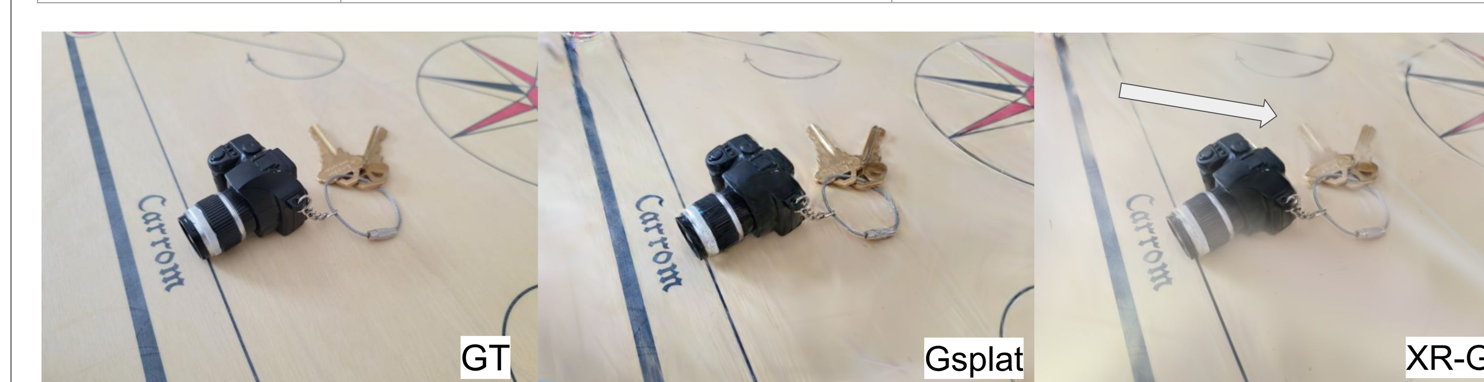
Gsplat XR-GS

Metric Results on Modified LLFF Dataset (*7 Scenes)

Metric	Average % Change from Base Gsplat to XR-GS (w/ *Scaled Losses, Gradients, Densify/Prune, and Consistency Loss)*
Training Time	-44.57
Num Gaussians	-69.68
Memory Used	-46.50
LPIPS	23.15
SSIM	0.45
PSNR	-0.80

XR-GS helps in cluttered/heavily textured scenes (plants/trees, busy rooms, etc.) but can struggle in singular object scenes. Resource usage gains remain:

Metric	Dining Room (% Change)	Keychain (% Change)
Training Time	-24.87	-16.32
Num Gaussians	-70.99	-66.05
Memory Used	-33.08	-2.59
LPIPS	-8.83	33.23
SSIM	3.60	-1.56
PSNR	3.92	-15.81



- Learning signal for small high-frequency objects (keychain, camera, etc.) being lost with down-weighted LR gradients
- Consistency Loss allows the model to create blur and "get away with it"
(In synthetic LR views, the blurriness is less noticeable important)
- Same features that lead to improvements in highly textured scenes cause issues here

Conclusion / Extensions:

- By adapting Gsplat with XR-GS, we reduce hallucination artifacts, "over-sharpened" renders and improve *true* detail reconstruction, especially in cluttered/textured scenes
- Significantly reduce the rendering time, number of gaussians rendered, memory required, while maintaining comparable improved PSNR and SSIM
- Future Work: Network for learning scaling parameters; Supporting arbitrary resolution differences (not binary HR vs LR); Cross-resolution teacher–student distillation

XR-GS shows that resolution diversity is a useful supervision signal for 3DGS