Accurate Geometric Reconstruction and High-Fidelity Photorealistic Rendering of Reflective Glass Façades using Gaussian Splatting with a Novel Reflection MLP

Zhenyu Liang[1 2], Xiao Zhang[1], Boyu Wang[3], Ang Li[1], Zhaolun Liang[1], Jeff Chak Fu Chan[1], Mingzhu Wang[4], Jack Chin Pang Cheng[1]

[1]Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong S.A.R. (China)
[2]Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, USA
[3]S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[4]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong S.A.R. (China)
zhenyu.liang@connect.ust.hk, xiao.zhang@connect.ust.hk, bw3005@nyu.edu, lionel_a.li@connect.ust.hk, zliangaq@connect.ust.hk, cfchanay@connect.ust.hk, m.wang@cityu.edu.hk, cejcheng@ust.hk.

Abstract –

Generating building digital twins from UAV imagery via 3D reconstruction is significant. However, reflective glass façades remain particularly challenging: their high-frequency, view-dependent appearance often causes photogrammetry to yield distorted geometry and corrupted textures. While Gaussian Splatting supports view-dependent rendering, its spherical harmonics color model mainly captures low-frequency components and fails to reconstruct high-frequency specular reflections. In addition, vanilla Gaussian Splatting tends to "explain" reflections by reconstructing ghost geometry behind the glass surface and producing a visually plausible view-dependent appearance, rather than recovering the true façade geometry. To address these issues, this paper proposes a novel Gaussian Splatting–based reconstruction and rendering framework for reflective glass façades, enabling both accurate geometry recovery and high-fidelity photorealistic rendering. First, UAV images are processed using photogrammetric reconstruction to obtain a coarse façade mesh with distorted geometry. Second, individual glass panels are segmented in 2D images using structural regularities and then projected onto the coarse mesh to correct geometric distortions in the façade. Third, using the corrected façade mesh as a geometric prior, a Gaussian Splatting model is trained with a rendering formulation that considers both diffuse color and specular reflection. For specular reflection, a novel reflection Multilayer Perceptron (MLP) is developed that takes the 3D surface position and the reflection direction as input to efficiently model near-field illumination with low computational and memory costs. Experiments on two buildings in Hong Kong show that our method reliably reconstructs reflective façade geometry and outperforms state-of-the-art approaches in rendering high-frequency, view-dependent appearances.

Keywords –

Gaussian Splatting; Reflective Glass Façades; 3D Reconstruction; Photorealistic Rendering; Reflection MLP; Digital Twin Modelling

1 Introduction

Using UAV-acquired imagery to reconstruct a building's 3D model can serve as a foundation for a digital twin [1-4]. Such a model can be further converted into a BIM representation [5-7], or used to inspect and annotate façade-level defects. This research direction is therefore highly valuable. However, for buildings with reflective glass façades, this pipeline faces challenges.

Optics-based passive 3D reconstruction, commonly referred to as photogrammetry, fundamentally relies on extracting and matching image keypoints across views and leveraging epipolar geometry to recover both the 3D coordinates of scene points and the camera poses [1]. A core assumption is multi-view consistency, i.e., that the same physical point exhibits similar appearance in overlapping images and can thus be reliably matched. For reflective surfaces, this assumption is violated: the observed color at the same surface point can differ significantly between adjacent viewpoints because the reflection directions change significantly. As a result, photogrammetry-based reconstructions often suffer from severe geometric and textural distortions when applied to reflective façades.

For such view-dependent textures on reflective glass, conventional point cloud or mesh representations are inadequate, as they can only represent static textures. In recent years, learning-based 3D reconstruction and rendering methods such as Neural Radiance Fields (NeRF) [8] and Gaussian Splatting (GS) [9] have emerged. NeRF trains an implicit MLP that takes a 5D input (3D position and 2D viewing direction) and outputs color together with a density value representing geometry. In contrast, GS represents scenes as an explicit set of Gaussian geometries and renders novel views via rasterization. Compared with NeRF, GS enables real-time rendering and faster training, making it more suitable as a base representation for building digital twins.

GS typically models view-dependent color using spherical harmonics (SH). While SH can capture view dependence to some extent, it is primarily effective for low-frequency effects (i.e., appearance changes that vary smoothly with camera pose.) Specular reflections, however, are high-frequency view-dependent effect and can change dramatically with viewpoint. Consequently, SH provides an inadequate representation for reflections on glass façades.

Recent studies have attempted to enhance GS for reflection modeling. One line of work introduces a 2D environment map to represent global illumination and queries it using reflection directions to reproduce high-frequency, view-dependent appearance [10-11]. However, such queries inherently assume illumination originates from infinity. In real-world 3D environments, reflections on glass façades are induced by a near-field light field, where contributing sources lie at varying distances. Representing this with a 2D environment map leads to parallax errors. Another line of work attempts to reconstruct the 3D environmental light field [12], e.g., by fitting millions of "environment Gaussians" to approximate where in 3D space light of different colors is emitted and how it reaches and reflects off the façade. Although this explicitly approximates the reflection formation process, inferring a complete 3D environment solely from reflections is highly complex, and representing the entire surrounding scene with Gaussians leads to substantial computation and storage costs.

To address these limitations, this paper proposes a new approach: we train a reflection Multilayer Perceptron (MLP) that takes as input the 3D position on the reflective surface and the reflection direction computed from the viewing direction and the surface normal, and outputs the corresponding color. In effect, the MLP learns a surface light field, enabling efficient modeling of near-field illumination effects without reconstructing the entire environment, thereby providing an effective and computationally efficient solution.

Effective training of the reflection MLP requires accurate surface geometry and normal directions of the reflective façade. Photogrammetry often produces distorted and incomplete geometry for reflective surfaces. Notably, vanilla GS also tends to place Gaussians behind reflective surfaces to reproduce the illusion of reflection in rendering, rather than reconstructing the true reflective surface itself. Therefore, we further propose a new geometric correction method for reflective glass façades by combining (i) the distorted coarse façade mesh obtained from photogrammetry and (ii) individual 3D panel models derived from façade panel segmentation, resulting in an accurate reconstruction of the reflective glass façade surfaces.

In summary, our main contributions are as follows: (1) proposing a novel model that integrates GS with a reflection MLP to efficiently model near-field specular reflections. By decomposing diffuse color and reflection via a shading formulation, our method enables high-fidelity photorealistic rendering of reflective glass façades; (2) developing a mesh correction approach that combines photogrammetry with a façade panel segmentation model, enabling accurate reconstruction of reflective glass façade geometric surfaces; (3) validating the proposed method on two real-world buildings with reflective glass façades in Hong Kong, and the experimental results demonstrate that our method outperforms state-of-the-art methods in rendering high-frequency, view-dependent facades' appearance.

2 Preliminary of Gaussian Splatting

This paper adopts 2D Gaussian Splatting (2DGS) [13] as the basic primitive for reconstructing and rendering reflective glass façades. Unlike 3D Gaussian Splatting (3DGS), which represents scenes using volumetric Gaussian ellipsoids, 2DGS represents objects using Gaussian elliptical disks. This representation improves multi-view consistency in both surface normals and depth, facilitating accurate querying of façade surface geometry from different camera viewpoints.

Each 2D Gaussian is parameterized by a central point p_k, two principal tangential vectors t_u and t_v, two scaling vectors s_u and s_v, an opacity α, a color coefficient vector c represented using spherical harmonics. In the local tangent plane defined in world space, the 2D Gaussian can be formulated by Eq. (1) and Eq. (2). Given local coordinates $\mathrm{u} = (u, v)$ on the tangent plane, the corresponding Gaussian value is computed by Eq. (3).

$$P(u,v) = p_k + s_u t_u u + s_v t_v v = H(u,v,1,1)^\top \quad (1)$$

$$H = \begin{bmatrix} s_u t_u & s_v t_v & 0 & p_k \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} RS & p_k \\ 0 & 1 \end{bmatrix} \quad (2)$$

$$\mathcal{G}(\mathrm{u}) = \exp\left(-\frac{u^2+v^2}{2}\right) \quad (3)$$

Based on these definitions, rendering in 2DGS amounts to establishing the correspondence between a camera pixel $\mathrm{x} = (x, y)$ and the local tangent-plane coordinates $\mathrm{u} = (u, v)$ of each 2D Gaussian. This can be computed via a ray–splat intersection, where $W \in \mathbb{R}^{4\times 4}$ denotes the transformation matrix from world space to screen space:

$$\mathrm{x} = (xz, yz, z, 1)^{\top} = WH(u, v, 1, 1)^{\top} \tag{4}$$

After obtaining the mapping from pixel $(x, y)$to the local coordinates (u, v), the Gaussian contribution to that pixel is evaluated using Eq. (3), and the intersection depth z is obtained from Eq. (4). The final pixel color is then computed by rasterization using volumetric alpha blending:

$$\mathrm{c}(\mathrm{x}) = \sum_{i=1} c_i \alpha_i \hat{\mathcal{G}}_i(\mathrm{u}(\mathrm{x})) \prod_{j=1}^{i-1} (1 - \alpha_j \hat{\mathcal{G}}_j(\mathrm{u}(\mathrm{x}))) \tag{5}$$

where $\hat{\mathcal{G}}$ denotes a low-pass filtered version of the Gaussian kernel, introduced to mitigate unstable optimization caused by strongly slanted viewpoints.

3 Methodology

The proposed Gaussian Splatting–based reconstruction and rendering framework is illustrated in Fig. 1. It consists of two main stages: (1) Geometric Reconstruction and Correction, and (2) Gaussian Splatting Reconstruction and Rendering.

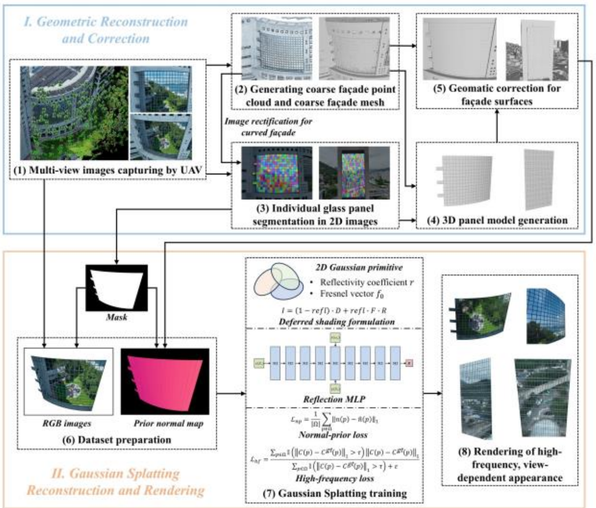


Fig. 1 Overall framework for the proposed method

3.1 Geometric Reconstruction and Correction for Reflective Glass Façade

The UAV-captured images were first used to generate a geometrically distorted coarse façade mesh and the corresponding coarse façade point cloud via photogrammetric 3D reconstruction. Notably, the coarse façade mesh exhibits uneven and broken glass panel surfaces, while the coarse façade point cloud contains missing points on the glass panels, as illustrated in Fig. 2.

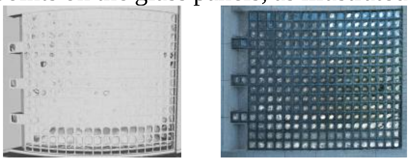


Fig. 2 Coarse façade mesh (left) and coarse façade point cloud (right) from photogrammetric 3D reconstruction

3.1.1 Image Rectification for Curved Façades

In the 2D images, the glass-panel segmentation method proposed in the authors' prior work is adopted. Leveraging the structural regularities of glass panels, specifically linear features and repetitive patterns, this method formulates a maximum a posteriori (MAP) estimation problem to extract panel corner points for the segmentation of the individual glass panels (see [14] for details). This method is effective in mitigating the adverse impact of reflections and background clutter on panel segmentation, and it performs well on frontal-view façade images. However, because it relies on line-based features, it is directly applicable to planar (flat) façade surfaces. For curved façades, an image rectification step is required to warp the curved surface into a planar surface in the image before segmentation.

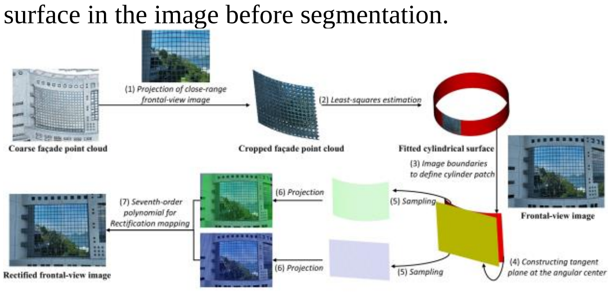


Fig. 3 Image rectification for curved façade surfaces

Fig. 3 illustrates the image rectification process for curved façades. First, a close-range frontal-view image is captured near the façade surface. This image contains only the façade, without any background clutter. Using the 3D-to-2D projection relationship between the coarse façade point cloud (3D points (X, Y, Z)) and the image pixels (u, v), as given in Eq. (6), the subset of 3D points that project into the image is extracted to form a cropped point cloud. A cylindrical surface is then fitted to the cropped point cloud via least-squares estimation, yielding the cylinder equation that represents the curved façade geometry.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \cdot \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{6}$$

where K is the camera intrinsic matrix; R and T are extrinsic rotation matrix and translation vector, respectively.

For image that requires panel segmentation, the image boundaries are used to determine the corresponding extent of the fitted cylindrical surface, thereby defining a patch on the cylinder (highlighted in red in the figure). This cylindrical patch is first sampled to obtain a set of surface points (green points). A tangent plane is then constructed at the angular center of the patch (yellow plane), and the sampled surface points are orthogonally projected onto this plane to form a planar point set (blue points).

Both the original cylindrical surface points (green) and their planar counterparts (blue) are projected onto the target image. In this projection, the green points indicate the pixel locations corresponding to the true 3D geometry of the curved façade, whereas the blue points indicate the pixel locations after unfolding the curved façade into a planar façade. Consequently, rectifying a curved-façade image into a planar-façade image amounts to estimating a mapping from the green-point locations to the blue-point locations in the image. A seventh-order polynomial is used to fit this mapping, enabling effective image rectification for curved façades.

3.1.2 Glass Panel Segmentation and Projection for 3D Panel Model Generation

For frontal-view image of planar façade and rectified frontal-view image of curved façade, glass-panel corner points are extracted using structural regularities-based MAP estimation. By pairing the four corner points belonging to the same panel, each glass panel can be segmented. For curved façades, the corner points extracted in the rectified image are further mapped back to the unrectified image using the fitted seventh-order polynomial. Fig. 4 shows the glass-panel segmentation results for both planar and curved façades.

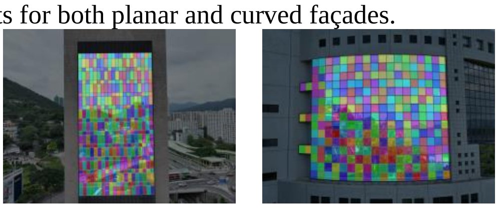


Fig. 4 Glass panel segmentation in images

The extracted glass-panel corner points are further projected onto the corresponding coarse façade mesh via ray casting, producing the 3D locations of the panel corners. Because corner extraction in the image may incur subtle pixel-level errors, which can be amplified after ray casting over tens of meters, the resulting 3D panel corners are regularized: for each row of corners, the z-coordinates are averaged, and for each column, the x- and y-coordinates are averaged, thereby reducing segmentation-induced deviations. After pairing the regularized 3D corner points, the reconstructed 3D panel models are obtained, as shown in Fig. 5.

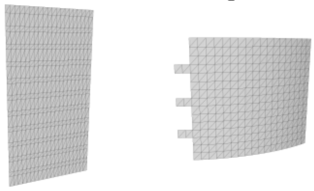
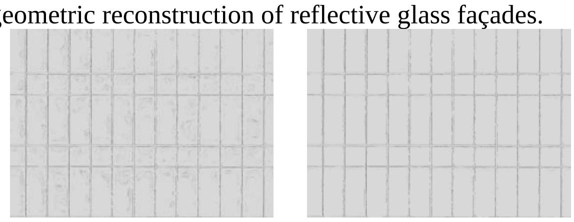


Fig. 5 3D panel models for planar façade (left) and curved façade (right)

3.1.3 Geometric Correction on Coarse Façade Mesh

Based on the obtained 3D panel models, geometric correction can be applied to the coarse façade mesh. Notably, simply overlapping the 3D panel models with the coarse mesh still results in an uneven façade surface (Fig. 6(a)), because the coarse mesh often contains local protrusions and irregularities on the panel regions. Conversely, directly replacing the corresponding mesh regions with the full 3D panel models would remove the surrounding frames.

To address this, each 3D panel is slightly shrunk in-plane about its center, and a small cuboidal region is defined along the panel normal. The portion of the coarse façade mesh within this cuboid is then replaced with the shrunk 3D panels. This procedure produces the corrected façade surface shown in Fig. 6(b), enabling accurate geometric reconstruction of reflective glass façades.



(a) (b)

Fig. 6 Geometric correction by (a) simply overlapping and (b) cuboidal region replacement

3.2 Gaussian Splatting Reconstruction and Rendering for Reflective Glass Façade

As the mesh models are unable to represent the high-frequency, view-dependent appearance of reflective glass façades, the final digital representation is constructed using the 3D model reconstructed by Gaussian Splatting. The accurate façade geometry obtained in Section 3.1 is incorporated as a geometry prior in the training of the Gaussian Splatting model.

3.2.1 Gaussian Augmentation and Deferred Shading Function

For each Gaussian, in addition to the parameters introduced in Section 2, two attributes are further included: a reflectivity coefficient r and a 3-dimensional

Fresnel vector f_0 for calculating the Fresnel term. The Fresnel term modulates specular reflectance as a function of the viewing angle, capturing the characteristic behaviour of glass surfaces, i.e., reflectance increases toward grazing angles and decreases near normal incidence. During rendering, different appearance components are first rendered into 2D images independently and then composited, following a deferred shading strategy.

For an image I, the deferred shading function is defined in Eq. (7). The diffuse color D is computed using Eq. (5) and represented with spherical harmonics. The specular reflection R is estimated by the reflection MLP described in Section 3.2.2. The reflectivity $refl$ is computed using the same volumetric alpha blending formulation as in Eq. (5). Since the surface roughness of reflective glass façades is negligible and can be approximated as ideal specular reflection, the Fresnel term F can be calculated using Schlick's approximation in Eq. (8), parameterized by F_0obtained via volumetric alpha blending of the per-Gaussian $f_0$values.

$$I = (1 - refl) \cdot D + refl \cdot F \cdot R \tag{7}$$
$$F = F_0 + (1 - F_0) \cdot (1 - \cos\theta)^5 \tag{8}$$

where θ denotes the angle between the surface normal n and the outgoing radiance direction ω_o (i.e., the direction opposite to the viewing direction).

3.2.2 Reflection MLP

To model near-field specular reflections, we design a reflection MLP that predicts the specular reflected color R from the surface position X_r and the reflection direction ω_r at that position (computed using Eq. (9)).

$$\omega_r = \omega_o - 2\omega_o^{\top} n \cdot n \tag{9}$$

The architecture of the reflection MLP is illustrated in Fig. 7. It comprises eight fully connected layers with ReLU activations and 512 hidden units per layer, followed by a final sigmoid-activated output layer. Since standard MLPs exhibit a spectral bias toward low-frequency signals whereas specular reflections contain high-frequency details, the inputs X_r and ω_r are first mapped to high-dimensional features using Fourier feature encoding, as defined in Eq. (10). Before encoding, X_r is normalized to the range $[-1,1]$, and ω_r is enforced to be a unit direction vector. We set $L = 12$ for both $\gamma(X_r)$ and $\gamma(\omega_r)$ to ensure sufficient capacity for modeling high-frequency content.

$$\begin{aligned}\gamma(p) = (&\sin(2^0\pi p), \cos(2^0\pi p), \\ &\cdots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p))\end{aligned} \tag{10}$$

The encoded position $\gamma(X_r)$ is first processed by the first four layers alone, rather than being concatenated with the direction input from the beginning, so that the network can learn position-dependent material characteristics and extract relevant features. This staged fusion helps disentangle spatially varying properties from direction-dependent reflection. At the fifth layer, the learned position features are concatenated with the encoded reflection direction $\gamma(\omega_r)$, and $\gamma(X_r)$ is re-injected via a skip connection. The skip connection facilitates gradient flow and preserves positional information, mitigating over-smoothing in deep networks. Finally, a sigmoid activation is applied to produce the specular reflected RGB color R.

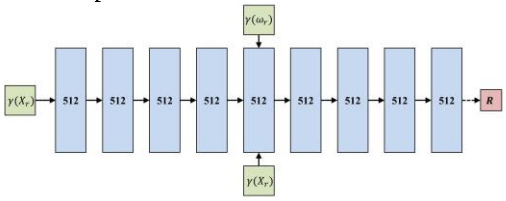


Fig. 7 Architecture of reflection MLP

3.2.3 Regularization

To effectively leverage the accurate façade geometry obtained in Section 3.1 and to ensure that high-frequency appearance details are properly learned, we introduce two additional regularization terms: a normal-prior loss and a high-frequency loss.

Normal-prior loss. As shown in Eq. (11), we compute an $\ell_1$loss between the rendered surface normals n and the reference normals $\hat{n}$ derived from the accurate façade geometry under the same viewpoint. This term enforces geometric fidelity of the reconstructed façade surface.

$$\mathcal{L}_{np} = \frac{1}{|\Omega|} \sum_{p \in \Omega} \|n(p) - \hat{n}(p)\|_1 \tag{11}$$

where Ω denotes the set of valid pixels, and p indexes image pixels.

High-frequency loss. As defined in Eq. (12), we focus the supervision on pixels whose color error exceeds a threshold τ. Specifically, we select the subset of pixels with large $\ell_1$residuals between the rendered color C and the ground truth C^{gt}, and average the residuals over this subset. Empirically, we observe that using only the standard RGB reconstruction loss $\mathcal{L}_c$ in Gaussian Splatting tends to distribute errors uniformly, yielding overly smooth reconstructions that fail to recover sharp, high-frequency reflections. The proposed high-frequency loss explicitly corrects these high-error regions.

$$\mathcal{L}_{hf} = \frac{\sum_{p \in \Omega} \mathbb{I}(\|C(p) - C^{\mathrm{gt}}(p)\|_1 > \tau)\, \|C(p) - C^{\mathrm{gt}}(p)\|_1}{\sum_{p \in \Omega} \mathbb{I}(\|C(p) - C^{\mathrm{gt}}(p)\|_1 > \tau) + \varepsilon} \tag{12}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, ε is a small constant to avoid division by zero, and τ is initialized to 0.2 and decreased in a stepwise manner over training iterations to a final value of 0.09.

Finally, the overall objective is given in Eq. (13).

Here, $\mathcal{L}_c$ combines an ℓ_1 term with the D-SSIM term, and $\mathcal{L}_n$ denotes the normal-consistency loss; both follow the original 2DGS definitions.

$$\mathcal{L} = \mathcal{L}_c + \lambda_n \mathcal{L}_n + \lambda_{np} \mathcal{L}_{np} + \lambda_{hf} \mathcal{L}_{hf} \quad (13)$$

where $\lambda_n = 0.05$, $\lambda_{\mathrm{np}} = 1$, and λ_{hf} is initialized to 1 and increased in a stepwise manner over training iterations, reaching a final value of 10.

4 Experiment

The proposed Gaussian Splatting–based reconstruction and rendering framework for reflective glass façades is validated on two buildings in Hong Kong: the HKUST Library and the Sha Tin Office Building.

4.1 Experimental Setup

For the HKUST Library, 600 multi-view images were collected, while 560 multi-view images were captured for the Sha Tin Office Building. Since this study focuses on the reconstruction and rendering of reflective glass façades, the 3D panel models of both buildings were first projected onto the images to generate façade masks, which remove background regions outside the façades. The prior normal maps used for training were likewise restricted to the masked regions.

All images were captured using a DJI M3E UAV equipped with an RTK module, providing centimeter-level GPS positioning. Due to interference from view-dependent appearance, camera poses estimated by COLMAP can exhibit substantial errors. Therefore, the captured images were first processed in DJI Terra to perform photogrammetry by incorporating GPS positions and camera orientations. This procedure supports the geometric reconstruction and correction in Section 3.1 while providing more accurate camera poses. The resulting poses were converted to the COLMAP format for Gaussian Splatting training using 100,000 epochs.

The proposed method is compared with vanilla 3DGS [9], 2DGS [13], and state-of-the-art reflection-aware Gaussian models, including GaussianShader [10] and Ref-Gaussian [11]. The proposed model is trained for a total of 2×10^5 iterations, while all baseline methods are run with their default hyperparameter settings.

In addition, different designs of the reflection MLP are explored and compared to demonstrate the rationality of our proposed architecture. Performance is evaluated using PSNR, SSIM, and LPIPS. All experiments are conducted on an NVIDIA A40 GPU.

4.2 Baseline Comparison

Figs. 8 and Table 1 illustrate the baseline comparison results on the HKUST Library dataset, while Figs. 9 and Table 2 show the results on the Sha Tin Office Building dataset. Quantitatively, the proposed method achieves the best performance across all metrics, and improves PSNR by more than 5 dB over the SOTAs on both datasets.

As illustrated by the renderings in Figs. 8 and 9, the proposed method faithfully reconstructs the view-dependent textures of reflective glass façades while preserving high-fidelity high-frequency details, such as the circular fountain in the center of the top row in Fig. 8, the reflected building in the lower-right corner of the bottom row in Fig. 8, the blue sports court in the lower-left region of the top row in Fig. 9, and the fine details of the bridge and vehicles in the bottom row of Fig. 9.

For Ref-Gaussian, it can recover view-dependent appearance to some extent; however, it represents global illumination using a 2D environment map, which causes parallax errors for real-world 3D near-field reflections, leading to noticeably blurred high-frequency details. GaussianShader also relies on a 2D environment map, but uses 3D Gaussians as primitives and defines normals using the minor-axis direction of each 3D Gaussian. This representation results in inaccurate façade-surface geometry, which in turn yields incorrect reflection estimation. Consequently, Fig. 8 shows locally chaotic appearance in the middle of the façade, and Fig. 9 exhibits unstable appearance across large façade regions, both due to poor geometric reconstruction. In addition, vanilla 3DGS and 2DGS employ spherical harmonics to model color, which captures only low-frequency view-dependent variations and fails to reproduce high-frequency specular reflections.

Overall, these comparisons demonstrate that our method reconstructs façade surfaces with 2D Gaussian primitives, estimates near-field specular reflections with a reflection MLP, and incorporates a corrected façade mesh as a geometric prior, together providing an effective solution for high-quality reconstruction and rendering of reflective glass façades.

Table 1 Quantitative comparison of HKUST Library

Model	PSNR↑	SSIM↑	LPIPS↓
Ours	**31.66**	**0.9237**	**0.0695**
Ref-Gaussian	26.65	0.9057	0.0930
GaussianShader	22.62	0.8402	0.1599
3DGS	22.70	0.8521	0.1278
2DGS	22.40	0.8511	0.1346

Table 2 Quantitative comparison of Sha Tin Office

Model	PSNR↑	SSIM↑	LPIPS↓
Ours	**30.17**	**0.9339**	**0.0594**
Ref-Gaussian	25.02	0.9139	0.0797
GaussianShader	21.40	0.8350	0.1638
3DGS	21.98	0.8566	0.1135
2DGS	21.10	0.8474	0.1250

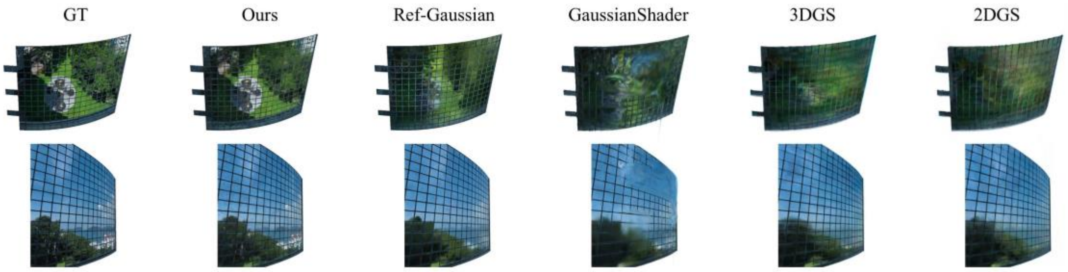


Fig. 8 Qualitative comparison for HKUST Library

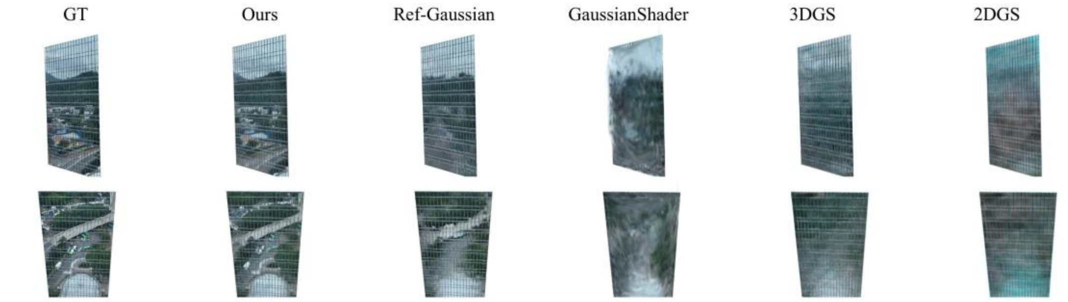


Fig. 9 Qualitative comparison for Sha Tin Office Building

4.3 Reflection MLP Architecture Evaluation

One of the core contributions of this work is the proposed reflection MLP, whose architectural design has a substantial impact on performance. The architecture is primarily governed by four factors: (1) the Fourier feature encoding order L, (2) the network depth (number of layers), (3) the hidden-layer width (hidden dimension), and (4) whether to incorporate a skip connection that re-injects the encoded position features.

In the Sha Tin Office Building, different reflection MLP architectures are evaluated, as summarized in Table 3. Comparing Cases 1, 2, and 4 indicates that increasing the hidden-layer width and network depth benefits the reconstruction of high-frequency, view-dependent textures. This is because the reflection MLP effectively serves as a high-capacity function approximator that maps position and reflected direction to reflected color; wider layers improve feature expressiveness, while deeper networks enhance the modeling of nonlinear interactions between spatial location and direction.

Comparing Cases 2 and 3 shows that introducing a skip connection that re-injects the encoded position features improves near-field reflection estimation. Additionally, Cases 3, 5, 6, and 7 suggest that moderately increasing the Fourier feature encoding order helps the MLP represent high-frequency components, whereas excessively large encoding orders degrade performance, because high-order encodings introduce very high-frequency basis functions that amplify sensitivity to small pose/calibration errors and pixel-level noise.

5 Conclusion

This paper proposes a novel Gaussian Splatting–based framework for reconstructing and rendering reflective glass façades, enabling accurate geometric reconstruction and high-fidelity photorealistic rendering. The main contributions are twofold: (1) a geometry reconstruction and correction strategy that integrates segmented 3D glass panel models with photogrammetric coarse façade mesh to refine the façade surface geometry; and (2) an enhanced 2D Gaussian representation equipped with a reflection MLP, which supports

Table 3 Comparison of reflection MLP architecture

Case	Architecture	PSNR↑	SSIM↑	LPIPS↓
1	$L = 10$, 8 layers, 256 hidden dim, w/o skip connection	29.08	0.9267	0.0636
2	$L = 10$, 8 layers, 512 hidden dim, w/o skip connection	29.68	0.9310	0.0608
3	$L = 10$, 8 layers, 512 hidden dim, w/ skip connection	29.95	0.9323	0.0603
4	$L = 10$, 6 layers, 512 hidden dim, w/ skip connection	29.77	0.9308	0.0618
5	$L = 12$, 8 layers, 512 hidden dim, w/ skip connection	30.17	0.9339	0.0594
6	$L = 14$, 8 layers, 512 hidden dim, w/ skip connection	29.84	0.9308	0.0615
7	$L = 16$, 8 layers, 512 hidden dim, w/ skip connection	29.02	0.9235	0.0672

reconstruction and rendering of high-frequency, view-dependent façade appearance.

Future work will focus on leveraging multimodal large language models (MLLMs) to further optimize the overall pipeline and to enable more natural interaction with the generated façade digital models [15–17], as well as extending Gaussian-based models to downstream simulation and analysis tasks [18–20].

References

[1] Z. Liang, L. Yang, Z. Liang, J. C. F. Chan, Z. Zhang, M. Wang, and J. C. Cheng. Optimized UAV view planning for high-quality 3D reconstruction of buildings using a modified sparrow search algorithm. *Adv. Eng. Inform.*, *65*, 103344, 2025.

[2] Z. Liang, J. C. F. Chan, J. Zhang, Z. Liang, B. Wang, M. Wang, and J. C. Cheng. Multimodal LLM-driven language-embedded 3D gaussian splatting for semantic and realistic digitization of historical buildings. *Automat. Constr.*, *181*, 106628, 2026.

[3] Z. Zhao, L. Ju, L. Zhang, H. Zou, J. He, Y. Lv, T. Xiao and Cai, G. Three-dimensional geological modeling with multisource data fusion. *J. Geotech. Geoenviron.*, *152*(1), 04025164, 2026.

[4] L. Y. Ju, T. Xiao, J. He, W. F. Xu, S. H. Xiao, and L. M. Zhang. A simulation-enabled slope digital twin for real-time assessment of rain-induced landslides. *Eng. Geol.*, 108116, 2025.

[5] B. Wang, F. Lin, M. Li, Z. Liang, Z. Chen, M. Wang, and J. C. Cheng. Informative As-Built Modeling as a Foundation for Digital Twins Based on Fine-Grained Object Recognition and Object-Aware Scan-vs-BIM for MEP Scenes. *Adv. Eng. Inform.*, *65*, 103382, 2025.

[6] B. Wang, F. Lin, M. Li, Z. Liang, H. Yue, Q. Wang, and J. C. Cheng. Aligning as-built and as-designed: Local point cloud to BIM registration via hybrid visibility map encoding for construction digital twins. *Automat. Constr.*, *180*, 106551, 2025.

[7] Y. Lv, Y. Pang, L. Ju, W. Xu, Z. Liang, J. He, and L. Zhang. Automating digitalization of engineered slopes to support interactive digital twin visualization. *Automat. Constr.*, *181*, 106564, 2026.

[8] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision* (pp. 405-421), Glasgow, UK, 2020.

[9] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, *42*(4), 139-1, 2023.

[10] Y. Jiang, J. Tu, Y. Liu, X. Gao, X. Long, W. Wang, and Y. Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5322-5332), Seattle, WA, USA, 2024.

[11] Y. Yao, Z. Zeng, C. Gu, X. Zhu, and L. Zhang. Reflective Gaussian Splatting. In *The Thirteenth International Conference on Learning Representations*. Singapore, 2025

[12] T. Xie, X. Chen, Z. Xu, Y. Xie, Y. Jin, Y. Shen, S. Peng, H. Bao, and X. Zhou. Envgs: Modeling view-dependent appearance with environment gaussian. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 5742-5751). Nashville, TN, USA, 2025.

[13] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers* (pp. 1-11). Denver, CO, USA, 2024.

[14] Z. Liang, X. Zhang, Z. Liang, B. Wang, M. Wang, and J. C. Cheng. A Novel Framework for Generating Realistic and Semantic Digital Model of Reflective Glass Façades Using RefGlass-3DGS. In *Computing in Civil Engineering 2025: Computational and Intelligent Technologies* (pp. 914-923), New Orleans, LA, USA, 2025.

[15] A. Li, P. K. Y. Wong, X. Tao, J. Ma, and J. C. Cheng. An interactive system for 3D spatial relationship query by integrating tree-based element indexing and LLM-based agent. *Adv. Eng. Inform.*, *66*, 103375, 2025.

[16] C. F. Chan, P. K. Y. Wong, X. Guo, J. C. Cheng, J. P. C. Chan, P. H. Leung, X. Tao. Context-aware vision-language model agent enriched with domain-specific ontology for construction site safety monitoring. *Automat. Constr.*, *177*, 106305, 2025.

[17] K. X. Guo, P. K. Y. Wong, J. C. Cheng, C. F. Chan, P. H. Leung, and X. Tao. Enhancing visual-LLM for construction site safety compliance via prompt engineering and Bi-stage retrieval-augmented generation. *Automat. Constr.*, *179*, 106490, 2025.

[18] W. Liu, X. Tao, C. Mao, and W. He. Scheduling optimization for production of prefabricated components with parallel work of serial machines. *Automat. Constr.*, *148*, 104770, 2023.

[19] W. Liu, Z. Wu, X. Tao, H. H. Kwok, and J. C. Cheng. Automated simulation modeling and sustainable performance evaluation for flat layout design. *Automat. Constr.*, *180*, 106556, 2025.

[20] Yin, C., Yue, H., Han, Q., Hu, D., Liang, Z., Lin, F., ... & Cheng, J. C. (2026). Industrial3D: A Terrestrial LiDAR Point Cloud Dataset and CrossParadigm Benchmark for Industrial Infrastructure. *arXiv preprint arXiv:2603.28660*.