remote sensing

MDPI

Article

Facade-Aware LiDAR Depth Supervision for Gaussian Splatting from Handheld Mobile Mapping

Ahmed Elamin *, Moathe El-Rabbany and Ahmed El-Rabbany

Department of Civil Engineering, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada; me@torontomu.ca (M.E.-R.); rabbany@torontomu.ca (A.E.-R.)
* Correspondence: agelamin@torontomu.ca

Highlights

What are the main findings?

- In the handheld LiDAR scanning of building facades, most LiDAR depths fall on the ground, so conventional depth supervision spends many Gaussian surfels on the ground and gives no reliable gain in the facade completeness of the splat model.
- Facade-aware LiDAR depth supervision (FA-Depth) gave the most complete facade and the highest F-score of all 2DGS configurations on two platforms, as verified against an independent TLS survey.

What are the implications of the main findings?

- Facade-focused LiDAR supervision enables Gaussian splatting from handheld mobile mapping to produce more complete facade records for documentation and digital twins.
- The gains were largest where 2DGS covered the facade poorly, which suggests FA-Depth is most valuable for sparse or low-resolution handheld captures.

Abstract

Handheld mobile mapping platforms are increasingly used to document buildings, and Gaussian splatting can turn their imagery into photorealistic three-dimensional (3D) models. However, the geometric reliability of these models for facade assessment remains uncertain. As such platforms record pose, camera imagery, and LiDAR together, supervising Gaussian splatting using LiDAR depth measurements is the natural way to improve the model geometry. However, as handheld platforms are ground-based, most of the LiDAR points projected into each image fall on the ground near the operator. As such, supervision that weights all pixels equally pulls the model towards the ground rather than the building, and much of its benefit is lost for the facade, which is the surface of interest. This paper introduces a facade-aware LiDAR depth supervision (FA-Depth) model for 2D Gaussian Splatting (2DGS), which excludes ground returns and gives more weight to distant facade returns. The model is evaluated against a terrestrial laser scanning (TLS) survey of the facade of a commercial building, using data from two commercial handheld LiDAR-camera platforms and repeated training runs. On both platforms, FA-Depth moved the primitives of the model from the ground to the facade and gave the most complete facade of all the 2DGS configurations tested, improving mean completeness by 9% and 36% over 2DGS without depth supervision. Either depth term improved surfaces extracted from the rendered depth, but only FA-Depth improved the facade completeness of the splat model itself. Rendering quality was also improved by a margin that depended on the platform.

Check for updates

Academic Editors: Fabiana Di Ciaccio and Erica Isabella Parisi

Received: 15 October 2026
Revised: 24 November 2026
Accepted: 25 November 2026
Published: 28 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: Gaussian splatting; LiDAR depth supervision; handheld mobile mapping; terrestrial laser scanning; building facade; geometric accuracy

1. Introduction

Reality capture has become a routine part of civil infrastructure assessment, and building facade inspection depends on three-dimensional (3D) records that are both interpretable and measurable. Visual fidelity supports the identification of surface defects such as cracking, spalling and sealant failure. Metric fidelity supports the quantification of out-of-plane deformation and change over time. Terrestrial laser scanning (TLS) delivers the latter with millimetre-level precision but produces static products without view-dependent appearance. Mobile and handheld laser scanning systems are now widely used to survey built and urban environments and cultural heritage [1,2]. Handheld simultaneous localization and mapping (SLAM) platforms now record pose, camera imagery and LiDAR together during a short walk. For example, Matellon et al. [3] compared three handheld SLAM scanners with a TLS survey and reported a mean cloud-to-cloud distance of approximately 6 mm on a building facade for the most accurate device. Strecha et al. [4] reported maximum errors of approximately 10–15 cm for the SHARE S20 with real-time kinematic positioning against independent checkpoints. Whether the imagery recorded by such platforms can produce photorealistic models that also support measurement is therefore a practical question for the profession.

Neural radiance fields (NeRF), introduced by Mildenhall et al. [5], represent a scene as a continuous volumetric function of density and view-dependent colour. Nerfacto, distributed within the Nerfstudio framework of Tancik et al. [6], combines this formulation with the scene contraction of Mip-NeRF 360 [7] and the hash encoding of Instant-NGP [8] for unbounded outdoor scenes. Three-dimensional Gaussian Splatting (3DGS), proposed by Kerbl et al. [9], replaced ray marching with rasterized anisotropic Gaussian primitives and achieved real-time rendering at comparable visual quality. Several variants were then developed to recover surfaces. Huang et al. [10] introduced 2D Gaussian Splatting (2DGS), which collapses each primitive to an oriented disc and adds depth-distortion and normal-consistency regularization. Because each disc has a well-defined depth and orientation, 2DGS is a natural target for supervision with measured depth. Guédon and Lepetit [11] proposed SuGaR, which regularizes Gaussians towards a surface, extracts a triangle mesh and refines Gaussians bound to the mesh. Gaussian Opacity Fields [12] and planar Gaussian splatting [13] pursued the same goal through level-set extraction and planar primitives, respectively. Reviews of this rapidly growing field are available in Ref. [14].

Independent geometric assessment of these representations against surveying-grade references remains comparatively limited. Petrovska et al. [15] evaluated NeRF and TLS against a structured-light reference and reported a TLS standard deviation of 1.68 mm against an NeRF deviation of 18.61 mm. Petrovska and Jutzi [16] placed a window and vegetation in front of a statue and reported mean deviations of 2.27 mm for multi-view stereo (MVS) and 8.52 mm for Nerfacto behind the window, while the radiance fields reached a completeness of up to 93%. Remondino et al. [17] reported a cloud-to-cloud root mean square error (RMSE) of 2.50 mm for photogrammetry against 5.50 mm for NeRF on a laser-scanned statue, although NeRF performed better on transparent objects. Murtiyoso and Grussenmeyer [18] compared Nerfacto with MVS on heritage facades and reported standard deviations of ±3.5 and ±5.0 cm. For Gaussian splatting, Murtiyoso and Macher [19] compared a heritage facade with TLS and reported an average standard deviation of approximately 2 cm. Billi et al. [20] reported a standard deviation of 3.00 cm for

a mesh-based Gaussian method on a cathedral portal, identical to that of MVS. Suwardhi et al. [21] reported a mean deviation of 15.2 cm between 3DGS and TLS and treated 3DGS as a visualization layer. McNally et al. [22] evaluated 3DGS and 2DGS on two buildings against TLS and reported a mean cloud-to-cloud distance of 0.037 m for 2DGS trained from structure-from-motion (SfM) poses on one of the buildings, measured in one direction only. Tao et al. [23] benchmarked handheld LiDAR-camera data against TLS and reported F-scores at 10 cm of 0.31–0.68 for Nerfacto against 0.54–0.85 for LiDAR SLAM. These studies show that radiance-field geometry is less accurate than direct measurement. However, most of them are image-only, and the one that uses platform LiDAR [22] uses it for poses and initialization only, not as depth supervision.

Other studies use LiDAR in radiance-field and Gaussian pipelines, either as a starting point or during training itself. LIV-GaussMap [24] and Gaussian-LIC [25] use LiDAR-inertial SLAM poses and initialize Gaussians from the LiDAR map, with Gaussian-LIC adding triangulated visual points. DrivingGaussian [26] uses an LiDAR prior in driving scenes. LetsGo [27] initializes Gaussians from an LiDAR mesh captured by a handheld scanner and adds an LiDAR depth regularizer. Urban Radiance Fields [28] and Street Gaussians [29] supervise depth with projected LiDAR, and TCLC-GS [30] supervises it with dense depth rendered from an LiDAR-camera mesh. LI-GS [31] applied LiDAR depth and normal losses to 2D Gaussian surfels and reported a Chamfer-L1 distance of 4.51 cm against 16.28 cm for 2DGS. However, its reference was a denoised point cloud from the same platform. ARSGaussian [32] constrained aerial 3DGS with LiDAR and reported an RMSE of 0.327 m, computed against the LiDAR points themselves. Such methods are therefore usually tested with rendering metrics, or against references built from the training sensor, rather than against an independent survey. One exception is SiLVR [33], which supervised an NeRF with LiDAR depth and normals from a handheld unit. It reported an accuracy of 8–11 cm against TLS, compared with 6–8 cm for the LiDAR map and 14–76 cm for Nerfacto.

Depth supervision is not uniformly beneficial. Depth-supervised NeRF [34] and dense depth priors [35] improved reconstruction from few views. Turkulainen et al. [36] added sensor depth to 2DGS on indoor scenes and reduced the Chamfer-L1 distance from 0.0687 to 0.0275. In the same study, monocular depth increased it to 0.0795, and depth supervision reduced the mesh F-score of Nerfacto from 0.721 to 0.697. Xiao et al. [37] reported a loss of 0.52 dB in the peak signal-to-noise ratio (PSNR), a pixel-level measure of rendering quality in decibels (dB), when monocular depth supervision was applied naively to Gaussian splatting. Zhang et al. [38] weighted depth supervision by a confidence map and reported a gain of up to 2.31 dB. How depth residuals are weighted therefore matters. For a ground-based platform, this raises a specific question, because much of each LiDAR depth map lies on the pavement near the operator rather than on the building.

Glass adds a further complication. LiDAR returns at glazing include mirrored points [39], and transparent surfaces may return no measurement at all [40], and radiance fields tend to reconstruct reflections as geometry behind the glass [41,42]. These offsets matter for monitoring. In brick walls, crack widths below 1 mm are classed as very slight damage and wider cracks as slight or moderate damage [43], and TLS deformation monitoring of a bridge load test required sub-millimetre accuracy for deformations of a few millimetres [44]. A systematic offset of several centimetres could therefore be misread as structural movement, or could mask it.

Research is still lacking on three points. First, mobile mapping platforms record LiDAR alongside the imagery, but it is not known whether supervising Gaussian splatting with this LiDAR improves facade geometry. From a ground-based platform, most of the valid pixels of every LiDAR depth map, which is the LiDAR cloud projected into a training image, lie on the pavement close to the operator, because the platform is carried a short height above

the ground and the density of LiDAR returns decreases with range. In addition, handheld surveys are usually carried out at some distance from the building, for example, to limit the obstruction and multipath of global navigation satellite system (GNSS) signals near the facade, so the facade lies farther from the platform than the pavement. The effect of this imbalance on the reconstruction has not been examined. Second, LiDAR-integrated Gaussian splatting has rarely been evaluated against an independent survey, so its accuracy and completeness on building facades remain largely unknown. Third, no evaluation has separated glazed from opaque surfaces, although glass is where both LiDAR and image-based reconstruction are least reliable and where a systematic offset could be mistaken for deformation.

This paper introduces a facade-aware LiDAR depth supervision (FA-Depth) model for 2DGS and evaluates it against TLS. Its main contributions are as follows:

- FA-Depth is developed to exclude ground returns using a locally fitted ground surface and to assign a range-based weight to each remaining depth residual.
- The effect of ground returns, which dominate handheld LiDAR depth maps, on the uniformly weighted LiDAR depth supervision of Gaussian splatting is analyzed.
- LiDAR-supervised Gaussian splatting is evaluated against an independent TLS survey on two handheld platforms. The parameters of FA-Depth are selected using the data of one platform and then applied without any change to the data of the other platform.
- The splat model and the surfaces extracted from it are evaluated as two different products. In addition, the glazed and opaque parts of the facade are evaluated individually, with the glass delineated independently of the LiDAR.

These contributions differ from existing work as follows. Existing LiDAR-supervised Gaussian and radiance-field methods give the same importance to all LiDAR depth pixels [28–31]. Alternatively, the importance given to each LiDAR depth pixel is adjusted according to the depth residual or a confidence measure [29,38]. In contrast, FA-Depth is designed for the ground-dominated depth maps of handheld capture, and requires no additional network, sensor or manual labelling. Existing LiDAR-supervised Gaussian methods were evaluated against references that were built from the training sensor [31,32]. Similarly, earlier comparisons with TLS-assessed representations trained without LiDAR depth supervision [15–23] or an LiDAR-supervised NeRF [33]. In this study, LiDAR-supervised Gaussian splatting is evaluated against an independent TLS survey on two platforms, with the parameters selected on one platform and then applied without change to the other. Earlier studies of glass placed a window in front of an object [16] or examined transparent objects [17]. In this study, the glazing of a full-scale facade is evaluated separately from the opaque surfaces.

This paper is organized as follows: Section 2 describes the study site and the data; Section 3 presents the proposed method, FA-Depth; Section 4 describes the experimental setup and the evaluation protocol; Section 5 reports the results; Section 6 discusses them, including the limitations of the study; and Section 7 concludes.

2. Study Site and Data

The complete workflow is shown in Figure 1. Data from both platforms were processed by the same procedures. In addition, every rule that depends on the data, such as the split of the images into training and held-out test views (Section 2.3) or the reference depth used for range weighting (Section 3.4), was applied identically to each platform.

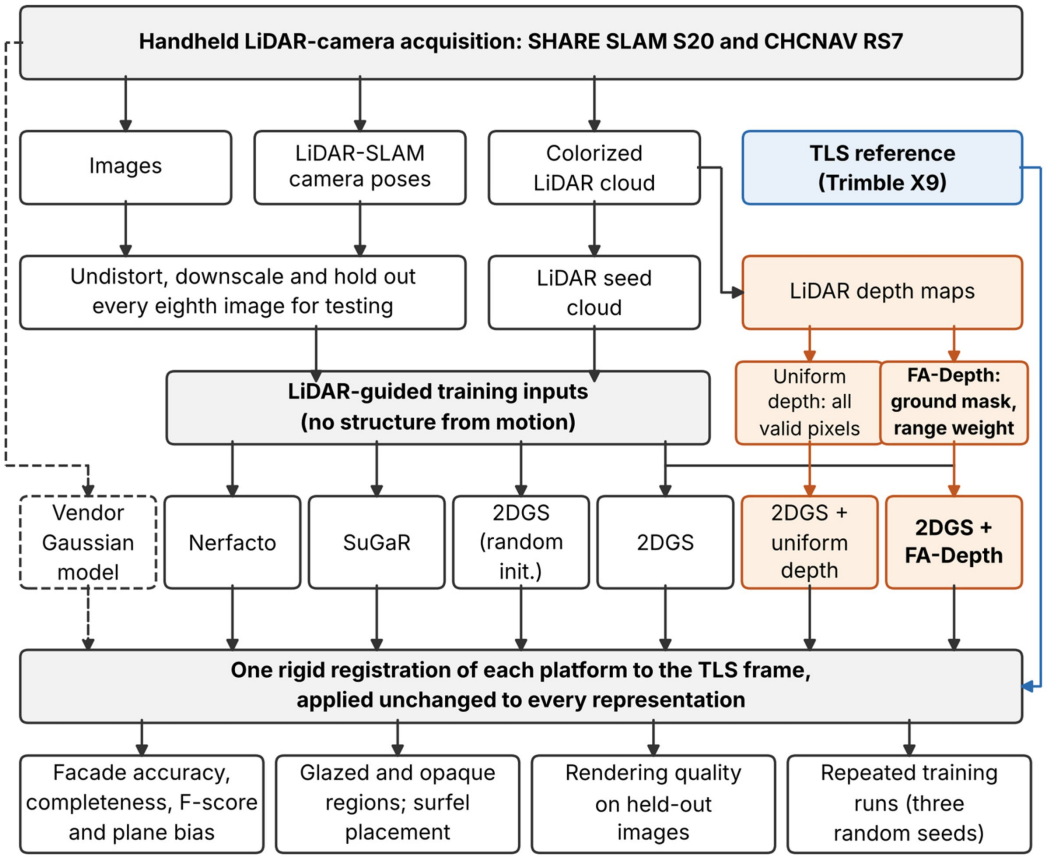

Figure 1. Workflow of the LiDAR-guided reconstruction, the facade-aware LiDAR depth supervision (FA-Depth) and the evaluation. The vendor Gaussian models are shown dashed because they were produced by the proprietary platform software rather than from the prepared training inputs. Grey boxes are data and steps shared by all methods, white boxes are individual processing steps and reconstruction methods, orange boxes show the LiDAR depth supervision examined in this study, with the proposed FA-Depth in bold, and the blue box is the independent TLS reference.

2.1. Study Site and Acquisition Platforms

The study site is a two-storey commercial retail building situated within a large open asphalt parking lot. The elevation selected for evaluation is a long, articulated facade combining full-height glazed storefront bays, light-toned precast cladding panels, brick piers and a concrete base course, with projecting metal canopies over the entrances. The evaluated facade spans approximately 65 m horizontally and 9.8 m vertically. This elevation was selected because it combines a reflective, weakly textured glazed component with a diffuse, textured brick component within a single nominally planar surface. This is the configuration under which image-based reconstruction is expected to degrade non-uniformly.

Two handheld LiDAR-camera platforms were used on 5 July 2026, and their characteristics are summarized in Table 1. These two platforms were selected because they represent two different designs of low-cost handheld LiDAR-camera systems, with different LiDAR sensors and camera geometries, while both deliver posed images, a colourized LiDAR cloud and a vendor Gaussian model, which are the inputs and benchmarks required in this study. The SHARE SLAM S20 combines a Livox MID-360 LiDAR with a stereo camera pair and was processed in SHARE PointClouds Studio [45]. Its images were delivered undistorted and were modelled as pinhole images. The CHCNAV RS7 combines a Hesai LiDAR with two opposed fisheye cameras and was processed in CHC CoPre [46]. Its LiDAR was delivered after a vendor range filter of 0.8–20 m. Both routes comprised walked circuits

of the building. As can be seen in Figure 2, the SHARE route circled the building twice in opposite directions, and the RS7 route was walked out and back. Revisits were frequent on both routes. For 97% of the SHARE camera positions (1552 of 1599) and 95.5% of the RS7 camera positions (1321 of 1383), the route passed within 6 m of the same location again at least 40 m earlier or later along the path. These statistics show that revisit constraints were available to the SLAM solutions, although it was not clear whether either commercial software took advantage of loop closure.

Table 1. Acquisition and reference datasets. The first row shows the two handheld platforms.

Item	SHARE SLAM S20	CHCNAV RS7	TLS Reference
Platform			—
Acquisition date	5 July 2026	5 July 2026	13 November 2025
Scanner	Livox MID-360	Hesai, single echo [1]	Trimble X9
Cameras	Stereo pair, 6324 × 8432 px, undistorted	Two opposed fisheye, 3000 × 3000 px	—
Camera used	Left	Camera1	—
Posed images	1599	1383	—
Camera path length	459.6 m	459.7 m	—
Colourized points	49,922,468	42,880,261	246,588,089
Processing software	SHARE PointClouds Studio 2.5.0	CHC CoPre 2.11.2	Trimble Perspective
Vendor Gaussian model	1,280,132 Gaussians, SH degree 0	6,884,691 Gaussians, SH degree 3	—
Registration	—	—	9 stations, mean error 1.3 mm

[1] Vendor range filter of 0.8–20 m applied to the delivered cloud. SH, spherical harmonics.

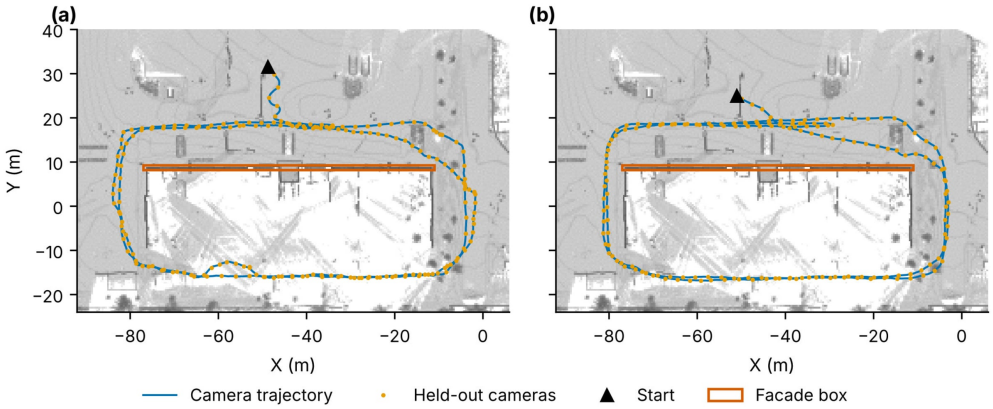

Figure 2. Camera trajectories in the TLS frame, drawn over a top view of the TLS reference: (**a**) SHARE SLAM S20, 1599 camera poses, of which 200 were held out for testing; (**b**) CHCNAV RS7, 1383 camera poses, of which 173 were held out. The facade box used for evaluation is outlined.

2.2. Terrestrial Laser Scanning Reference

The reference dataset is a TLS point cloud acquired with a Trimble X9 instrument [47] and registered in Trimble Perspective [48]. Registration comprised one registration set of nine stations, with an average registration error of 1.3 mm, an average overlap of 59% and an average consistency of 99%; per-station errors ranged from 0.7 to 1.6 mm. The registered cloud contains 246,588,089 points and spans approximately 358 × 310 × 31 m, making it considerably larger than either mobile mapping scene and expressed in a different coordinate system. Registration and an explicit definition of the overlapping region were therefore mandatory. This station registration error describes the internal consistency of the TLS survey only. It is approximately two orders of magnitude smaller than the facade distances reported here, whereas the alignment of each mobile platform to this reference is assessed separately in Sections 4.4 and 5.1.

The TLS reference was surveyed on 13 November 2025, approximately eight months before the mobile acquisitions. In principle, any scene change over this interval would be indistinguishable from reconstruction error. For the evaluated facade, the SHARE and RS7 LiDAR clouds agree with the TLS reference to Chamfer-L1 distances of 0.047 and 0.045 m, respectively, where the Chamfer-L1 distance is the mean nearest-neighbour distance defined in Section 4.5.1. Material change to the facade would appear directly in these comparisons between independent direct measurements. The observed agreement therefore bounds any such change well below the facade distances of the learned representations, which range from 0.107 to 0.306 m, and the bound is conservative because the same figures also include sensor noise and residual registration error. The agreement was also uniform along the facade. In six segments of 11 m, the signed offset of each LiDAR cloud from the TLS facade plane stayed between −0.027 and +0.036 m (Section 5.1), so no localized change larger than a few centimetres was detected in any part of the evaluated facade.

2.3. LiDAR-Guided Initialization

Gaussian and radiance-field pipelines are conventionally initialized from a sparse point cloud and camera poses recovered by structure from motion (SfM). In the architecture adopted here, both quantities are taken directly from the mobile platform, and no SfM reconstruction is performed at any stage. Both platforms report camera-to-world poses in the OpenCV convention, in which the camera x-axis points right, the y-axis points down and the z-axis points forward. Before conversion, each convention was confirmed by projecting the LiDAR into images across the route and checking that building edges, mullions and poles coincided with their image counterparts. Poses were converted to the conventions required by the reconstruction frameworks [49,50] according to Equation (1).

$$R_{\mathrm{GL}} = R_{\mathrm{CV}} \cdot \mathrm{diag}(1, -1, -1), R_{\mathrm{w2c}} = R_{\mathrm{c2w}}^{\mathrm{T}}, t_{\mathrm{w2c}} = -R_{\mathrm{w2c}} c \quad (1)$$

where R_{CV} is the camera-to-world rotation in the OpenCV convention, R_{GL} is the corresponding rotation in the OpenGL convention used by Nerfstudio, R_{c2w} is the camera-to-world rotation, R_{w2c} and t_{w2c} are the world-to-camera rotation and translation required by COLMAP, and c is the camera centre.

For the SHARE platform, left-camera images were downscaled by a factor of four to 1581 × 2108 pixels. For the RS7 platform, one fisheye camera (Camera1) was used. Its images were undistorted to a centred pinhole model with the Kannala–Brandt coefficients supplied by the vendor [51], giving a field of view of 122.9°, and were downscaled by a factor of two to 1500 × 1500 pixels. This factor gives the pixel count closest to that of the SHARE images. On both platforms, every eighth frame in file order was held out for testing, which yielded 1399 training and 200 test images for SHARE and 1210 training and 173 test images for RS7. Approximately 400,000 points were sampled at a uniform stride from each

colourized LiDAR cloud and written directly in COLMAP format as the initialization cloud, namely 402,601 points for SHARE and 400,751 points for RS7. All Gaussian representations therefore began from metrically scaled geometry in the platform's own frame.

3. Proposed Method: Facade-Aware LiDAR Depth Supervision (FA-Depth)

Depth supervision adds a loss term that penalizes the difference between the depth rendered by the model and an independently measured depth. In this study, the measured depth is the platform's own colourized LiDAR, projected into each training view. This section first describes the base representation, the LiDAR depth maps and the simplest form of supervision, then the imbalance that motivates FA-Depth, and finally FA-Depth itself. The parameters of FA-Depth are of three kinds: the reference depth m of the range weighting is computed from the data of each platform; the depth weight λ is selected on validation views of one platform only (Section 4.3); and all other parameters, namely the ground-surface parameters, the 0.25 m height tolerance and the clipping bounds of the range weight, were set once before the final training runs and applied unchanged to both platforms.

3.1. Base Representation: 2D Gaussian Splatting

2DGS [10] is the base of the proposed method. It replaces the volumetric Gaussian of 3DGS with an oriented two-dimensional disc, or surfel, so that flattening one dimension encourages primitives to lie on surfaces. It supplements photometric optimization with depth-distortion and normal-consistency terms, which were used at the default values of the official code: $\lambda_{\text{dist}} = 0$ and $\lambda_{\text{normal}} = 0.05$. The depth-distortion weight therefore remained zero throughout training in every configuration, so that the 2DGS baseline is the default configuration of the official code and the depth term is the only difference between the configurations; the larger depth-distortion weights used in the official mesh-benchmark scripts were not applied. 2DGS was chosen over 3DGS for two reasons. Each surfel has a well-defined depth and orientation, so the depth it renders can be compared directly with a measured depth, whereas the depth of a volumetric Gaussian is spread along the viewing ray. In addition, its primitives are designed to lie on surfaces, which suits a large, nominally planar facade. Each 2DGS model was trained for 30,000 iterations at spherical-harmonic degree three.

3.2. LiDAR Depth Maps

For each 2DGS training view, the colourized LiDAR cloud was transformed into camera coordinates and projected with the camera intrinsics. Points beyond 50 m were discarded. A single-pixel depth buffer is not sufficient for this task because the cloud is sparse relative to the image, so background points can appear through gaps between fore-ground samples. Each point was therefore drawn as a disc whose radius, clip $(\lceil 0.10\, f/z \rceil, 1, 10)$ pixels, shrinks with depth z for focal length f, and the nearest depth was kept in each pixel. Isolated pixels lying more than max (1.0 m, 0.10 z) behind the local minimum depth were discarded. The maps were stored losslessly in millimetres with an explicit validity mask. At load time, the stored values were converted to metres. All depths, including those used for the ground mask, the range weights and the reference depth, are camera-axis depths z in the camera coordinate system rather than Euclidean ranges, on both platforms. They were inspected on at least ten views spanning each route, including views of the glazed bays, and no background was found to bleed through nearer walls, trees, poles or vehicles.

3.3. Uniform Depth Supervision and the Near-Field Imbalance

In its simplest form, referred to below as uniform depth, every valid LiDAR pixel contributes equally to the loss, with a uniform weight, as given in Equations (2) and (3).

$$L_{\text{depth}} = \frac{1}{|V|} \sum_{p \in V} \left| \widehat{D}_p - z_p \right| \tag{2}$$

$$L = L_{\text{2DGS}} + \lambda L_{\text{depth}} \tag{3}$$

where V is the set of valid LiDAR pixels in the image, z_p is the LiDAR depth at pixel p, $\hat{D}_p$ is the expected camera coordinate depth rendered by 2DGS at pixel p, L_{2DGS} is the unchanged photometric, depth-distortion and normal-consistency objective of 2DGS, and λ is the weight of the depth term, which is applied after a warm-up of 3000 iterations. The expected depth was chosen before training, because 2DGS also exposes a median surface depth that behaves differently. The expected depth is the opacity-weighted mean of the surfel depths along the pixel ray, normalized by the accumulated opacity. Pixels without a valid LiDAR depth or without a rendered depth do not contribute to the loss.

This form treats all pixels alike, but they are not alike: a handheld platform is carried close to the ground, so the lower part of every image views the pavement within a few metres of the operator, where LiDAR returns are densest, because the density of returns from a scanner with a fixed angular resolution decreases with range. In addition, handheld surveys are usually carried out at some distance from the building, for example, to limit the obstruction and multipath of GNSS signals near the facade, so the facade lies farther from the platform than the pavement. The statistics of the depth maps are reported in Table 2, and representative views are shown in Figure 3. Without any modification, the median valid depth was 1.42 m on SHARE and 2.00 m on RS7. Roughly half of every map therefore lay on the pavement within about two metres of the operator, while the facade occupied a smaller region at a greater range. Under uniform depth, the ground therefore dominates the loss. Here, the ground data are the ground pixels of the LiDAR depth maps, namely the valid pixels whose projected LiDAR point lies on the ground surface defined in Section 3.4, rather than the ground points of the LiDAR cloud or the ground visible in the images, because only valid depth pixels enter Equation (2). A median of 71.1% (SHARE) and 69.6% (RS7) of the valid pixels of each map were ground pixels (Table 2), and their effect on the placement of the surfels and on the facade is reported in Sections 5.2 and 5.4. LiDAR also covers glass poorly, because most window panes appear as holes without returns, and some returns pass through the glass to surfaces inside the building.

Table 2. Statistics of the LiDAR depth maps used for supervision. Values are medians over the 2DGS training views.

Supervision	SHARE Valid Fraction	SHARE Median Depth (m)	SHARE Weighted Median Depth (m)	RS7 Valid Fraction	RS7 Median Depth (m)	RS7 Weighted Median Depth (m)
Uniform depth (all valid pixels)	0.633	1.42	1.42	0.579	2.00	2.00
Ground mask only	0.185	4.82	4.82	0.176	5.32	5.32
Range weighting only	0.633	1.42	4.21	0.579	2.00	4.90
FA-Depth (ground mask and range weighting)	0.185	4.82	5.67	0.176	5.32	6.44

The weighted median is the median of the supervised depths weighted by the range weight of Equation (4). Valid fraction is the fraction of image pixels with a supervising LiDAR depth.

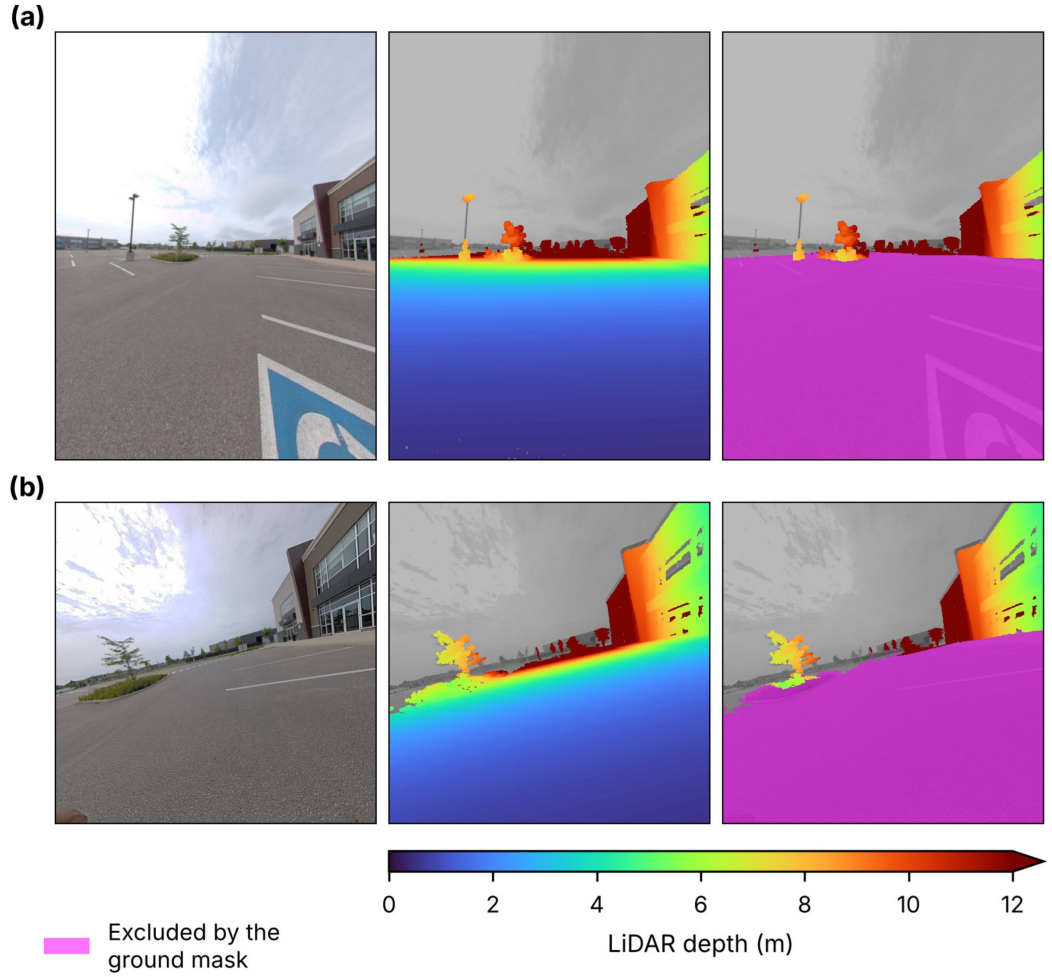

Figure 3. LiDAR depth maps for one storefront image from each platform: (**a**) SHARE; (**b**) RS7. From left to right, the image, the uniform depth map (all valid pixels) and the FA-Depth map, in which the pixels removed by the ground mask are shown in magenta. Depths beyond 12 m are shown in the darkest colour.

3.4. Ground Masking and Range Weighting

FA-Depth modifies Equation (2) in two ways to redirect the supervision towards the facade. First, ground pixels are excluded. A local ground surface was fitted to the LiDAR cloud by dividing it into 1.0 m cells, taking the fifth-percentile height of each cell with at least 100 points as a ground candidate, and fitting a local plane within a 5 m radius by three iterations of Huber-weighted least squares [52] with a threshold of 0.15 m. Depth pixels whose LiDAR point lay no more than 0.25 m above the fitted surface were excluded. Second, each remaining residual is weighted by its range, according to Equations (4) and (5).

$$w_p = \mathrm{clip}\left(\frac{z_p}{m}, 0.25, 4.0\right) \tag{4}$$

$$\widetilde{w}_p = \frac{w_p}{\underset{q \in V'}{\mathrm{mean}} w_q} \tag{5}$$

where m is a reference depth, V′ is the set of valid, non-ground pixels in the image, and $\tilde{w}_p$ is the normalized weight. The clipping bounds of 0.25 and 4.0 limit each weight to between one quarter and four times that of a pixel at the reference depth, so that no pixel receives more than 16 times the weight of another; they were fixed before training and were not tuned. On SHARE, for example, every pixel farther than about 5.7 m therefore receives the same weight. The normalization keeps the mean weight in each image equal to one, so the

magnitude of the loss remains comparable with that of uniform depth. The reference depth m was defined as the median, over all training views, of the per-view median valid depth of the uniform depth maps. This is a property of the data and not a tuned value; it equals 1.4245 m for SHARE and 2.004 m for RS7. The FA-Depth loss is given by Equations (6) and (7). Every training image on both platforms contained valid non-ground pixels, at least 2.45% of the image on SHARE and 0.76% on RS7.

$$L_{\mathrm{FA}} = \frac{1}{|V'|} \sum_{p \in V'} \widetilde{w}_p \left| \widehat{D}_p - z_p \right| \tag{6}$$

$$L = L_{\mathrm{2DGS}} + \lambda L_{\mathrm{FA}} \tag{7}$$

As can be seen in Table 2, ground masking removed a median of 71.1% of the valid pixels on SHARE and 69.6% on RS7, and moved the median supervised depth to 4.82 and 5.32 m, respectively. Range weighting moved the weighted median further, to 5.67 and 6.44 m.

4. Experimental Setup and Evaluation

4.1. Reconstruction Methods

In addition to 2DGS (Section 3.1), three further representations were included as references, selected to span the design space rather than to sample it conveniently. SuGaR [11] represents Gaussian methods that extract an explicit mesh of the kind required by inspection and building information modelling workflows. It begins from a conventional Gaussian cloud, regularizes it towards a surface, extracts a triangle mesh, binds Gaussians to the mesh faces and refines them photometrically. The configuration used a 7000-iteration vanilla prerequisite, 15,000 coarse iterations, a surface level of 0.3, a decimation target of 200,000 vertices, six Gaussians per triangle and 2000 refinement iterations. Nerfacto [6] represents the radiance-field family, which keeps the results comparable with earlier geomatics evaluations [15,16]. It is a continuous neural radiance field and contains no explicit points or triangles, so geometry must be sampled from the field and cleaned. It was trained for 30,000 iterations, and 5,000,000 points with outlier removal were exported from the same checkpoint used for rendering. Finally, the Gaussian model produced by each platform's software represents what a practitioner obtains without additional processing. The vendor models were treated as commercial products rather than controlled configurations. The SHARE model contains 1,280,132 Gaussians with degree-zero spherical harmonics only. The RS7 model was produced by the vendor's Gaussian engine and contains 6,884,691 Gaussians with degree-three spherical harmonics. Its training images were crops taken at 170 of the 173 RS7 test timestamps, so its rendering scores are not held out and are excluded from the rendering comparison. The vendor documentation does not state whether LiDAR was used during its training.

Training settings are summarized in Table 3. The number of training views is not uniform across methods. 2DGS and SuGaR were trained on every second image of the non-test pool, namely 700 views for SHARE and 605 views for RS7, because these implementations cache complete images and the full set exceeded practical host memory. Nerfacto used all non-test images. This is a constraint imposed by implementation memory behaviour rather than an experimental design choice. All training ran on a workstation with an Intel Core i9-10900X processor, 125 GiB of memory and two NVIDIA RTX A6000 graphics cards with 48 GiB of memory each. The comparison that tests FA-Depth between the four 2DGS configurations used identical views for every configuration. In the comparison with the other representations, the larger view set of Nerfacto, if anything, favours Nerfacto, so the differences reported in Section 5.7 are conservative with respect to the Gaussian methods.

Table 3. Training settings and output primitives of the self-trained representations.

Setting	2DGS Configurations	Nerfacto	SuGaR
Training views (SHARE/RS7)	700/605	1399/1210	700/605
Resolution (SHARE/RS7)	1581 × 2108/1500 × 1500 px	Same	Same
Initialization	LiDAR seed, 402,601/400,751 points [1]	LiDAR-SLAM poses	LiDAR seed
Iterations	30,000	30,000	7000 + 15,000 + 2000
Main settings	SH degree 3; depth weight $\lambda = 0.50$ after 3000 iterations (uniform depth and FA-Depth)	4096 rays per batch	Surface level 0.3; 200,000 vertices; 6 Gaussians per triangle
Output, SHARE	272,611 ± 15,100 surfels [2]	4,998,194 exported points	166,280 vertices; 302,250 triangles; 1,813,500 Gaussians
Output, RS7	83,139 ± 4066 surfels [2]	4,996,231 exported points	150,673 vertices; 275,180 triangles; 1,651,080 Gaussians

[1] A random cloud of the same size was used for 2DGS-random. [2] 2DGS without a depth term, mean ± SD over three runs. SH, spherical harmonics.

4.2. Experimental Configurations

The four 2DGS configurations compared in this study are defined in Table 4. They differ only in the initialization and the depth term, so any difference between them can be attributed to those two factors. 2DGS, without a depth term, is the reference against which uniform depth and FA-Depth are compared. 2DGS-random isolates the contribution of the LiDAR seed cloud. Its random cloud contained the same number of points as the LiDAR seed, drawn uniformly within the box bounded by the 0.1 and 99.9 percentiles of the seed cloud coordinates, with random colours and a fixed random seed, while the LiDAR-SLAM poses were retained.

Table 4. The four 2DGS configurations compared in this study.

Configuration	Initialization	Depth Term	Ground Mask	Range Weighting	Runs (SHARE/RS7)
2DGS	LiDAR seed cloud	None	—	—	3/3
2DGS + uniform depth	LiDAR seed cloud	Equation (2)	No	No	3/2 [1]
2DGS + FA-Depth	LiDAR seed cloud	Equation (6)	Yes	Yes	3/2 [2]
2DGS-random	Random cloud	None	—	—	1/1

All configurations use the LiDAR-SLAM poses and identical training settings (Table 3). [1] The RS7 run with seed 20260817 failed due to memory exhaustion. [2] The RS7 run with seed 20260816 failed due to memory exhaustion.

Training 2DGS is not fully repeatable, because its rasterizer, which runs on the graphics card, has no deterministic mode, so two runs with identical settings can give slightly different models. 2DGS, uniform depth and FA-Depth were therefore each trained three times on each platform, with the random seeds 20260816, 20260817 and 20260818, and results are reported as the mean and standard deviation (SD) of the runs. With so few runs, statistical tests are not meaningful. Instead, a difference between two configurations is described as separated and treated as reliable only when every run of one configuration is better than every run of the other. Otherwise, the ranges are described as overlapping. On RS7, one FA-Depth run (seed 20260816) and one uniform depth run (seed 20260817) failed due to

memory exhaustion and were not repeated with other settings, so these two configurations have two runs each on RS7. Nerfacto, SuGaR and 2DGS-random were trained once each.

4.3. Parameter Selection on One Platform and Transfer to the Other

All settings were selected on the SHARE platform only, and then applied unchanged to RS7. This design uses the second platform as an independent test of whether FA-Depth generalizes, rather than as a second tuning set. Selection used a validation split of 70 of the 700 SHARE training views, namely every tenth view, which was excluded from training; the test views and the TLS reference were not used for any selection. For uniform depth, four weights, $\lambda \in \{0.05, 0.10, 0.20, 0.50\}$, were tested for both direct and inverse depth, and direct depth at $\lambda = 0.50$ was selected. For FA-Depth, ground masking alone, range weighting alone, and their combination were first trained at $\lambda = 0.50$, and the combination gave the highest validation PSNR (21.54 dB, against 21.39 and 21.24 dB). The combination was then trained at $\lambda = 1.0$ and 2.0. $\lambda = 2.0$ gave a validation PSNR only 0.05 dB higher than $\lambda = 0.50$ but used 3.4 times as many surfels, and its full training run failed due to memory exhaustion, as did the validation run at $\lambda = 1.0$. The selection rule was therefore amended, before any test view or TLS comparison had been computed, to select the smallest λ within 0.10 dB of the best validation PSNR among configurations whose full training completes, which selected $\lambda = 0.50$. For RS7, only the reference depth m was recomputed from the RS7 maps by the same definition. The facade geometry of the validation models, evaluated against the TLS reference only after all selections had been made, is reported in Section 5.4.

4.4. Registration to the TLS Frame

The colourized LiDAR of each platform was selected as its registration anchor because it is a direct geometric measurement expressed in the same platform-local frame as every image-based representation. Both clouds were prepared at 0.50 m. Fast Point Feature Histogram descriptors were computed [53], a coarse alignment was obtained by random sample consensus [54] at 1.5 m, and a robust point-to-plane iterative closest point (ICP) refinement [55,56] was applied at successively tighter thresholds of 3.00, 1.50, 0.75 and 0.40 m. All geometric processing used Open3D [57]. Convergence is reported in Table 5. The resulting rigid transform of each platform was applied unchanged to every representation from that platform. This matters for the evaluation, because if each representation were aligned to the TLS separately, the alignment would partly absorb that representation's own distortion and drift, and its geometry would appear more accurate than it is. With one common transform, the errors of each representation remain visible in the reported distances.

Table 5. Convergence of the robust point-to-plane ICP refinement for each platform.

Threshold (m)	SHARE Fitness	SHARE Inlier RMSE (m)	RS7 Fitness	RS7 Inlier RMSE (m)
3.00	0.940	0.741	0.898	0.810
1.50	0.867	0.437	0.806	0.409
0.75	0.783	0.294	0.734	0.259
0.40	0.673	0.223	0.672	0.212

4.5. Evaluation

4.5.1. Facade Geometry

Evaluation was confined to a box enclosing the main building facade, defined in the TLS frame and spanning −77.0 to −11.0 m in x, 8.1 to 9.2 m in y and −1.0 to 8.8 m in z. Because the box is defined in the TLS frame, the facade reference is identical for both

platforms. All representations were clipped to this box and voxelized at 0.10 m to limit source-density bias, after which the TLS reference contributed 77,325 points. Gaussian representations were evaluated through their primitive centres, because the splat model itself, rather than a mesh extracted from it, is the product used for visualization and digital twins, and its primitives are what depth supervision redistributes (Section 4.5.3). The SuGaR mesh was converted to a comparable point set by drawing 1,000,000 seeded uniform surface samples before voxelization. Only the facade was evaluated; because FA-Depth removes the ground supervision, it is not intended for whole-scene reconstruction.

Because each surfel has a spatial extent and an opacity, the 2DGS configurations were also evaluated on surfaces extracted from their rendered depth. For each run, the expected depth was rendered for every training view and fused by truncated signed distance function (TSDF) fusion with the official 2DGS implementation [10], with a voxel size of 0.02 m, a truncation distance of 0.08 m and a maximum depth of 30 m; only depth within 1 m of the evaluation box was integrated. The resulting mesh was post-processed with the cluster filter of the official implementation, transformed with the same rigid transform as the other representations, clipped to the evaluation box and sampled with 1,000,000 seeded uniform surface points before voxelization, as for SuGaR. The other representations were evaluated through their native outputs, namely exported points for Nerfacto and mesh samples for SuGaR, so the comparison between methods in Section 5.7 reflects the product that each method delivers. Finally, the sensitivity of the centre-based results to the evaluation settings was assessed with voxel sizes of 0.05 and 0.20 m and with evaluation boxes 0.6 and 1.6 m deep.

Two complementary errors were measured: accuracy indicates how close the reconstructed points lie to the reference, and completeness indicates how much of the reference is covered by the reconstruction. Both are needed, because a sparse reconstruction can be accurate where it exists while leaving much of the facade uncovered. For a candidate point set C and the TLS reference set R, accuracy is the candidate-to-reference nearest-neighbour distance (d_{acc}) and completeness is the reference-to-candidate nearest-neighbour distance (d_{comp}), following the multi-view stereo evaluation convention of Seitz et al. [58], as given in Equation (8).

$$d_{\mathrm{acc}}(c) = \min_{r \in R} \|c - r\|, \; d_{\mathrm{comp}}(r) = \min_{c \in C} \|r - c\| \tag{8}$$

where c and r are individual points, and the norm is the Euclidean distance in metres. The Chamfer-L1 distance (CD_{L1}), which summarizes accuracy and completeness in a single value, is given by Equation (9).

$$\mathrm{CD}_{\mathrm{L1}} = \frac{1}{2}\left[\operatorname*{mean}_{c \in C} d_{\mathrm{acc}}(c) + \operatorname*{mean}_{r \in R} d_{\mathrm{comp}}(r)\right] \tag{9}$$

The F-score at a distance threshold τ is the harmonic mean of precision (P) and recall (Rc) [59], as given in Equation (10).

$$F(\tau) = \frac{2\, P(\tau)\, \mathrm{Rc}(\tau)}{P(\tau) + \mathrm{Rc}(\tau)} \tag{10}$$

where precision $P(\tau)$ is the fraction of candidate points within τ of the reference, and recall $Rc(\tau)$ is the fraction of reference points within τ of the candidate. Thresholds of 0.05, 0.10, 0.20 and 0.50 m were evaluated. A robust plane was additionally fitted to the TLS facade points, and the mean signed distance of each representation along its normal was expressed relative to that of the TLS reference. This plane bias exposes systematic offsets that unsigned distances cannot reveal [60]. A negative bias indicates a representation lying behind the TLS plane, that is, inside the building.

4.5.2. Delineation of Glazed and Opaque Regions

The facade was partitioned into glazed and opaque regions by delineating the glass manually, independently of the LiDAR. An orthomosaic of the facade was produced by projecting the SHARE camera images onto the TLS facade plane at a ground sampling distance of 0.02 m and aligning it to an orthoimage of the TLS reference, and 233 glass panes, excluding frames and mullions, were outlined on it as rectangles. TLS points inside a pane were labelled glazed, which gave 13,133 glazed points (17.0%) and 64,192 opaque points. The partition was applied unchanged to both platforms, so that the same reference points are compared, and candidate points take the class of their nearest TLS point. A partition derived from the density of LiDAR returns with Otsu's method [61] was also tested, but it recovered only 13% of the annotated glass points, because low return density also occurs at occlusions and oblique surfaces; the manual partition was therefore used.

4.5.3. Capacity Diagnostic

A 2DGS model has a finite number of surfels, and depth supervision can change where they are placed. To measure this, the surfel centres of each 2DGS model were described by three quantities, namely the distance to the nearest training camera, the number of surfels inside the facade box, and the fraction lying within the 0.25 m ground band of Section 3.4.

4.5.4. Rendering Quality

Rendering quality was evaluated on the held-out views, namely 200 SHARE views at 1581 × 2108 pixels and 173 RS7 views at 1500 × 1500 pixels, using three standard image measures. PSNR is obtained from the mean squared error (MSE) between the rendered and reference images, as given in Equation (11).

$$\mathrm{PSNR} = 10\ \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \tag{11}$$

where MAX is the maximum pixel value, and MSE is the mean squared error over all pixels and colour channels, so higher values indicate smaller pixel differences. Because PSNR is logarithmic, an increase of 1 dB corresponds to a reduction of about 21% in the MSE, and an increase of 0.3 dB corresponds to a reduction of about 7%. The structural similarity index (SSIM) compares the local means, variances and covariance between the two images, rewarding agreement in local structure rather than in absolute pixel values [62]; it ranges from zero to one, where one denotes identical images. The learned perceptual image patch similarity (LPIPS) is the distance between deep features extracted from both images by a pretrained network, which agrees more closely with human judgement [63]; lower values indicate closer perceptual agreement. The Gaussian methods were evaluated with one common 3DGS-style implementation. Nerfacto was evaluated with the native Nerfstudio evaluator on both platforms. Its RS7 renders were also saved and evaluated with the common implementation. The RS7 operator's hand and body appear in many frames, covering a median of 2.3% of the image. The RS7 metrics were therefore also computed with the operator's mask excluded. Because every eighth image was held out along routes that revisit the same places, every held-out view had a training camera within 0.5 m, with a median distance of 0.28 m on SHARE and 0.32 m on RS7. The rendering results therefore measure interpolation between nearby training views rather than performance on spatially novel viewpoints.

5. Results

5.1. Registration Check on the Facade

Because the convergence statistics in Table 5 are computed over the whole scene, including the parking lot, vegetation and vehicles, they do not describe the alignment of the facade itself. A facade-specific check was therefore made with the colourized LiDAR of each platform inside the evaluation box (Section 4.5.1), using the same transform. The inlier RMSE at the 0.40 m threshold fell from 0.223 m for the whole scene to 0.076 m on the facade for SHARE, and from 0.212 to 0.059 m for RS7, and the median point-to-plane residual on the facade was 0.019 and 0.021 m, respectively. To assess residual registration error and SLAM drift along the facade, the facade was divided into six segments of 11 m. The signed offset of the LiDAR from the TLS facade plane ranged from −0.027 to +0.013 m for SHARE and from −0.017 to +0.036 m for RS7, and the two independent platforms differed by at most 0.042 m in any segment. Residual registration error and drift therefore contribute a few centimetres at most, which is small compared with the facade distances of the learned representations reported in Sections 5.2–5.7.

5.2. Effect of Uniform Depth Supervision

The effect of the depth term on facade geometry and rendering is summarized in Table 6, and its effect on where the surfels are placed is summarized in Table 7 and Figure 4. On both platforms, uniform depth drew the model towards the operator. On SHARE, the share of surfels within 5 m of the trajectory rose from 23.1% to 48.4%, the share in the ground band rose from 26.5% to 33.8%, and the number of surfels inside the facade box fell from 87,891 to 74,613, although the total number of surfels rose by 68%. On RS7, the same two shares rose from 31.5% to 62.0% and from 19.1% to 41.6%, and the median distance of the surfels to the trajectory halved, from 7.20 to 3.57 m.

Table 6. Effect of the depth term on facade geometry and rendering (mean ± SD of the runs). Distances are in millimetres unless stated. Facade geometry is given for the surfel centres of the splat model and for surfaces extracted from the rendered depth by TSDF fusion (Section 4.5.1).

Metric	SHARE 2DGS	SHARE Uniform Depth	SHARE FA-Depth	RS7 2DGS	RS7 Uniform Depth [1]	RS7 FA-Depth [1]
Facade geometry, surfel centres of the splat model (mm)						
Facade Chamfer-L1	110.6 ± 1.3	110.7 ± 6.0	106.8 ± 3.2	175.3 ± 2.1	159.2 ± 18.5 [a]	130.1 ± 11.0 [a]
Accuracy, mean	117.9 ± 2.2	106.3 ± 1.8 [a]	119.6 ± 5.4	133.6 ± 0.8	122.1 ± 0.7 [a]	121.9 ± 1.1 [a]
Completeness, mean	103.4 ± 1.6	115.1 ± 12.8 [b]	94.0 ± 4.9 [a]	217.0 ± 3.4	196.2 ± 37.7	138.3 ± 20.9 [a]
Completeness, 95th percentile	216 ± 7	255 ± 30 [b]	195 ± 13 [a]	528 ± 4	740 ± 277 [b]	428 ± 120 [a]
F-score at 0.10 m	0.575 ± 0.002	0.572 ± 0.041	0.620 ± 0.020 [a]	0.312 ± 0.007	0.434 ± 0.024 [a]	0.545 ± 0.012 [a]
F-score at 0.20 m	0.887 ± 0.004	0.883 ± 0.022	0.899 ± 0.007 [a]	0.682 ± 0.007	0.831 ± 0.012 [a]	0.863 ± 0.013 [a]
Plane bias, facade-wide	+11.3 ± 2.1	+36.6 ± 4.0 [b]	+12.4 ± 6.0	+30.9 ± 4.5	+47.1 ± 3.9 [b]	+33.0 ± 6.4
Facade points	49,784 ± 1926	41,360 ± 6109	63,430 ± 6495	17,024 ± 727	38,554 ± 5313	74,719 ± 4663
Facade geometry, surface extracted from the rendered depth (mm)						
Completeness, mean	251.2 ± 121.0	81.6 ± 6.3 [a]	82.7 ± 2.4 [a]	263.2 ± 29.7	156.5 ± 16.6 [a]	140.8 ± 8.9 [a]
Accuracy, mean	229.2 ± 35.8	124.0 ± 3.3 [a]	121.9 ± 4.3 [a]	185.8 ± 18.1	140.8 ± 4.0 [a]	144.5 ± 5.8 [a]
Chamfer-L1	240.2 ± 78.1	102.8 ± 3.0 [a]	102.3 ± 1.5 [a]	224.5 ± 23.9	148.7 ± 10.3 [a]	142.6 ± 7.3 [a]

Table 6. *Cont.*

Metric	SHARE 2DGS	SHARE Uniform Depth	SHARE FA-Depth	RS7 2DGS	RS7 Uniform Depth [1]	RS7 FA-Depth [1]
F-score at 0.10 m	0.401 ± 0.076	0.692 ± 0.009 [a]	0.678 ± 0.004 [a]	0.413 ± 0.031	0.349 ± 0.028 [b]	0.374 ± 0.027 [b]
Plane bias	−110.2 ± 16.5	+36.4 ± 9.3 [a]	+31.0 ± 11.1 [a]	−34.6 ± 28.1	+68.3 ± 15.2 [b]	+56.7 ± 3.7
Rendering, held-out views						
Held-out PSNR (dB)	19.88 ± 0.55	21.10 ± 0.19 [a]	22.12 ± 0.13 [a]	19.87 ± 0.25	20.18 ± <0.01 [a]	20.19 ± 0.01 [a]
Held-out SSIM	0.623 ± 0.004	0.622 ± 0.002	0.625 ± <0.001	0.563 ± 0.001	0.564 ± <0.001	0.564 ± <0.001
Held-out LPIPS	0.445 ± 0.006	0.449 ± 0.003	0.447 ± 0.001	0.512 ± 0.001	0.510 ± 0.001 [a]	0.508 ± <0.001 [a]

[a] Every run better than every 2DGS run (separated). [b] Every run worse than every 2DGS run. Values without a mark overlap with 2DGS; facade points are not ranked, and plane bias is ranked by its magnitude. [1] Two runs; the third failed through memory exhaustion.

Table 7. Where the 2DGS surfels are placed (means of the runs).

Metric	SHARE 2DGS	SHARE Uniform Depth	SHARE FA-Depth	RS7 2DGS	RS7 Uniform Depth	RS7 FA-Depth
Total surfels	272,611	457,967	908,251	83,139	472,300	1,025,662
Surfels inside the facade box	87,891	74,613	148,639	27,314	77,622	274,568
Surfels within 5 m of the trajectory (%)	23.1	48.4	39.6	31.5	62.0	46.1
Surfels in the ground band (%)	26.5	33.8	7.2	19.1	41.6	5.1
Median distance to the trajectory (m)	8.25	5.25	6.30	7.20	3.57	5.42

RS7 uniform depth and FA-Depth: means of two runs.

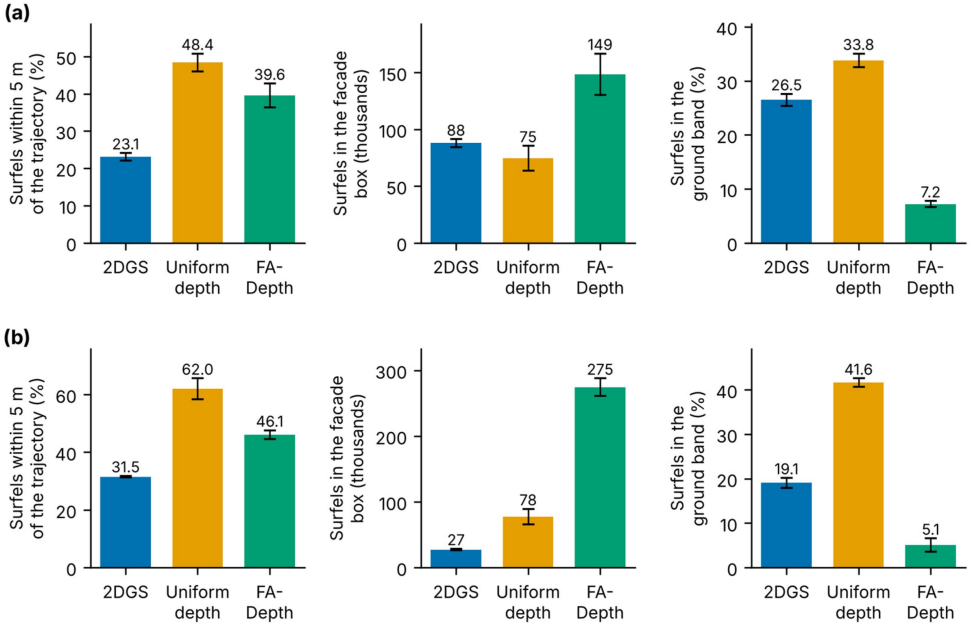

Figure 4. Where the 2DGS surfels are placed: (**a**) SHARE; (**b**) RS7. From left to right, the share of surfels within 5 m of the camera trajectory, the number of surfels inside the facade box and the share of surfels within the 0.25 m ground band. Bars show the means of the runs and whiskers one SD; on RS7, uniform depth and FA-Depth have two runs each.

The consequences for the facade depended on the platform. On SHARE, uniform depth improved the mean facade accuracy from 117.9 to 106.3 mm, but worsened the mean completeness from 103.4 to 115.1 mm and its 95th percentile from 216 to 255 mm, all with separated ranges. The gain in accuracy and the loss in completeness offset each other, so the facade Chamfer-L1 distance and the F-scores did not change beyond the spread of the runs. The model therefore placed surfels accurately where LiDAR returns were dense, namely on the pavement near the trajectory, and withdrew them from the building. On RS7, 2DGS contained only about 83,000 surfels and left much of the facade uncovered, and uniform depth raised this number to about 472,000. Some of the added surfels reached the facade, so the facade Chamfer-L1 distance improved from 175.3 to 159.2 mm, and the F-score at 0.10 m rose from 0.312 to 0.434, with separated ranges. However, the mean completeness did not improve beyond the spread of the runs, and its 95th percentile worsened from 528 to 740 mm. On both platforms, uniform depth also improved the held-out PSNR by 1.22 and 0.31 dB, and moved the facade-wide plane outwards, away from the building, to a plane bias of +36.6 and +47.1 mm against +11.3 and +30.9 mm for 2DGS. The RS7 uniform depth values are the means of two runs, because the third failed due to memory exhaustion.

5.3. FA-Depth on Both Platforms

With FA-Depth, the ground band fraction fell from 26.5% to 7.2% on SHARE and from 19.1% to 5.1% on RS7, and the number of surfels inside the facade box rose by factors of 1.7 and 10.1, respectively (Table 7). The total number of surfels grew faster than the number of facade surfels, so the fraction of surfels on the facade fell on both platforms. FA-Depth therefore redirected the model away from the ground, but it did not restore the distribution of 2DGS completely, since the median distance to the trajectory fell from 8.25 to 6.30 m on SHARE and from 7.20 to 5.42 m on RS7.

The gain in coverage was reliable on both platforms (Table 6 and Figure 5). On SHARE, FA-Depth improved mean facade completeness by 9.4 mm, from 103.4 to 94.0 mm, and raised the F-scores at 0.10 and 0.20 m by 0.045 and 0.012, respectively, with every FA-Depth run better than every 2DGS run. However, the mean facade accuracy was 1.7 mm worse, with overlapping ranges. The facade Chamfer-L1 distance improved by 3.8 mm on average, but the ranges overlapped: in one of the three pairs of runs with the same seed, 2DGS gave the lower value. The SHARE results therefore show a reliable gain in completeness, but only a marginal gain in the Chamfer-L1 distance. FA-Depth improved the held-out PSNR by 2.24 dB, from 19.88 to 22.12 dB, with separated ranges, while SSIM and LPIPS did not change beyond the spread of the runs. FA-Depth was also less stable than 2DGS. Its SD of the facade Chamfer-L1 distance was 2.4 times that of 2DGS, and its SD of the number of surfels was 3.4 times larger.

On RS7, 2DGS was weaker than on SHARE, with a mean facade Chamfer-L1 distance of 175.3 mm and only about 17,000 facade points. FA-Depth improved mean completeness by 36%, from 217.0 to 138.3 mm, and raised the F-scores at 0.10 and 0.20 m by 0.233 and 0.182, respectively, again with separated ranges. Unlike SHARE, the mean facade accuracy on RS7 also improved, from 133.6 to 121.9 mm, with separated ranges. This gain came from the least accurate points, since the 95th percentile of the accuracy distance fell from 0.377 to 0.317 m while the median rose slightly from 0.092 to 0.096 m. The facade Chamfer-L1 distance improved by 26%, from 175.3 to 130.1 mm. The held-out PSNR improved by 0.32 dB, from 19.87 to 20.19 dB, with separated ranges, although the gain was only 1.3 times the SD of 2DGS. With the operator mask excluded, the gain was 0.37 dB. On both platforms, the facade-wide plane bias did not improve, and the ranges overlapped.

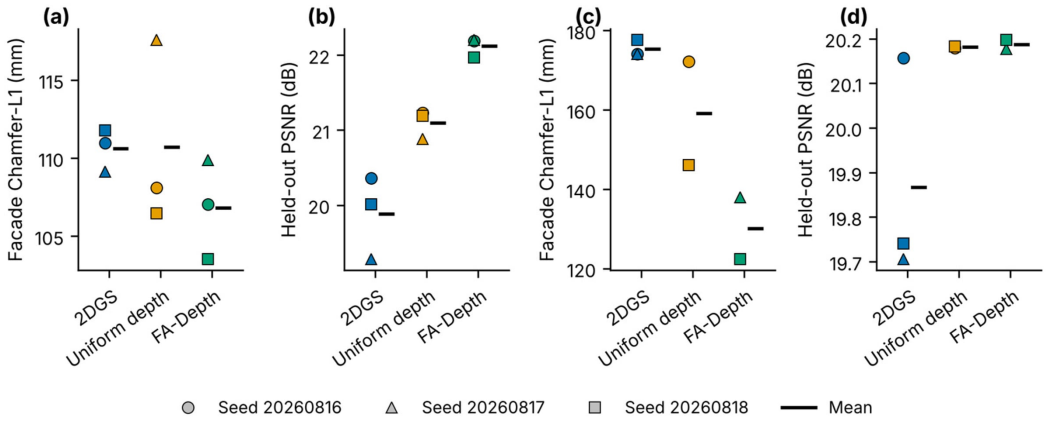

Figure 5. Facade Chamfer-L1 distance and held-out PSNR of each training run: (**a**,**b**) SHARE; (**c**,**d**) RS7. Marker shapes identify the random seed, and horizontal bars show the means. On RS7, uniform depth and FA-Depth have two runs each, because one run of each failed through memory exhaustion.

Compared with uniform depth, FA-Depth gave better mean completeness on both platforms, by 21.1 mm on SHARE and 57.9 mm on RS7, and better F-scores on RS7, all with separated ranges. It also gave a higher PSNR on SHARE, by 1.02 dB, with separated ranges, whereas the two were indistinguishable on RS7. Uniform depth, in turn, gave the better mean accuracy on SHARE and the better 95th percentile of accuracy on both platforms. Redirecting the supervision from the ground to the facade therefore traded some accuracy for substantially more coverage.

The distance distributions in Figure 6 show the same pattern. On both platforms, the completeness curve of FA-Depth lies closer to that of the LiDAR, which is the ceiling of the supervision, whereas the accuracy curves of 2DGS and FA-Depth are close to each other on SHARE. The facade points in Figure 7 show the change in coverage directly. On RS7, the 2DGS points are sparse over most of the facade, and FA-Depth fills much of it, except at the western end. On SHARE, the difference is smaller, consistent with the smaller change in completeness. Under uniform depth, the facade points are concentrated near the ground. In the runs shown, 42% of the SHARE points and 32% of the RS7 points lay in the lowest two metres of the facade, against 25% and 27% for 2DGS and 26% of the TLS reference points. The SHARE uniform depth run shown is the one with the poorest completeness of its three runs. Held-out renders for both platforms are shown in Figure 8.

The same pattern held under different evaluation settings. On RS7, every FA-Depth run was better than every 2DGS run in completeness and in the F-score at 0.10 m for voxel sizes of 0.05, 0.10 and 0.20 m and for evaluation boxes 0.6, 1.1 and 1.6 m deep. On SHARE, this held at voxel sizes of 0.05 and 0.10 m and in the deeper box, whereas at 0.20 m the coarser voxels removed the advantage, and in the shallower box the completeness ranges overlapped by one run.

On surfaces extracted from the rendered depth (Table 6), both depth terms improved on 2DGS on both platforms. On SHARE, mean completeness improved from 251 mm for 2DGS to 82 and 83 mm for uniform depth and FA-Depth, and mean accuracy from 229 to 124 and 122 mm, respectively, with every run of either depth term better than every 2DGS run. On RS7, mean completeness improved from 263 to 157 and 141 mm, and mean accuracy from 186 to 141 and 145 mm, again with separated ranges, whereas the F-score at 0.10 m remained higher for 2DGS. On these extracted surfaces, the runs of uniform depth and FA-Depth overlapped on both platforms. The surfaces of 2DGS lay behind the TLS plane (−110 and −35 mm), and those of the depth-supervised models in front of it (+31 to +68 mm). With the median instead of the expected rendered depth, tested on one run

per configuration, the 2DGS surface was markedly more complete (74 mm on SHARE and 218 mm on RS7), while both depth terms remained more complete and more accurate on both platforms.

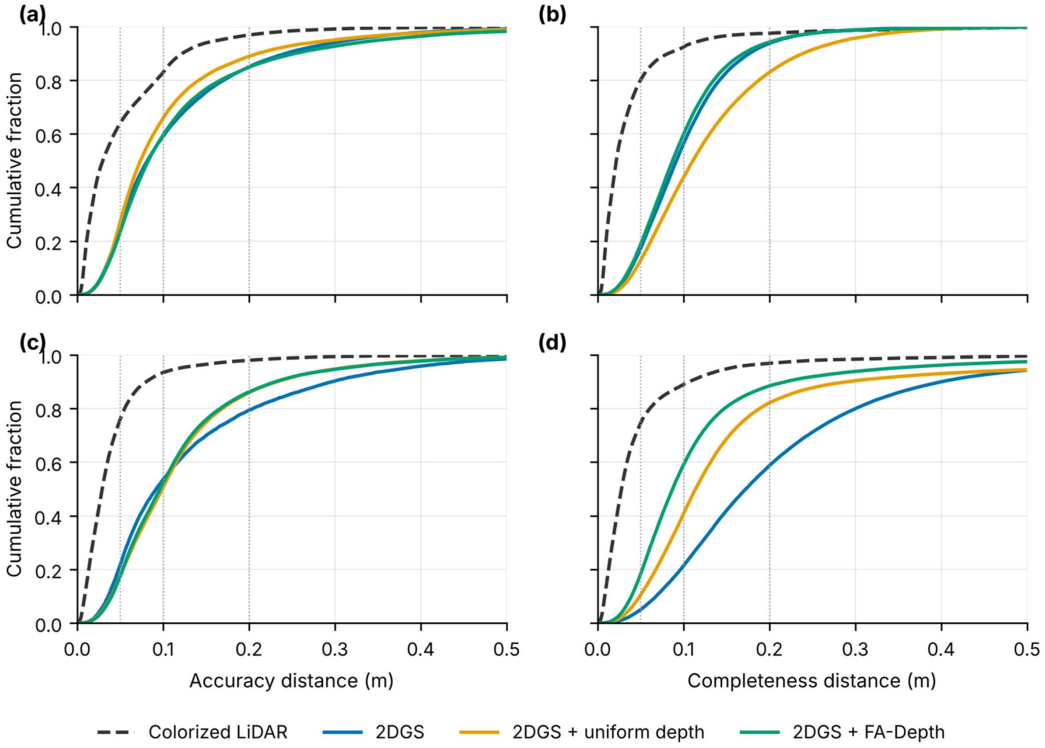

Figure 6. Cumulative distributions of the facade distances: (**a**) SHARE accuracy; (**b**) SHARE completeness; (**c**) RS7 accuracy; (**d**) RS7 completeness. Curves show one run of each configuration (seed 20260817 on SHARE and 20260818 on RS7, the only seed completed by all three RS7 configurations) and the colourized LiDAR, which is the ceiling of the depth supervision.

5.4. Contribution of Ground Masking and Range Weighting

To show the contribution of each component to the facade geometry, the three validation models of Section 4.3, each trained once without the validation views, were evaluated against the TLS reference after all selections had been made. Ground masking alone and the combination gave similar facade geometry, with mean completeness of 0.099 and 0.102 m and an F-score at 0.10 m of 0.603 for both, whereas range weighting alone gave the poorest completeness (0.110 m) and left 36.7% of the surfels in the ground band, against 8.4% and 6.5%, respectively. Ground masking is therefore the main contributor to the facade gain, and range weighting adds a gain in rendering and a further reduction in the ground share. This comparison also isolates the effect of the ground pixels of the depth maps. Range weighting alone keeps the ground pixels and only lowers their weight, whereas ground masking removes them, so the difference between the two, 0.110 against 0.099 m in mean completeness and 36.7% against 8.4% of the surfels in the ground band, is likely due mainly to the ground pixels. The ground pixels therefore limit the benefit of depth supervision for the facade rather than always degrading it, since uniform depth still raised the F-score at 0.10 m on RS7 (Section 5.2). The same evaluation showed that $\lambda = 2.0$ would have further improved facade completeness (0.078 m) and the F-score at 0.10 m (0.652), but with 3.0 million surfels, a model that could not be trained on the full view set. A geometric selection criterion would therefore not have changed the choice among the configurations that can be trained, and it would have required a geometric reference on the validation

views, which was deliberately excluded so that the TLS reference remained independent of all selections.

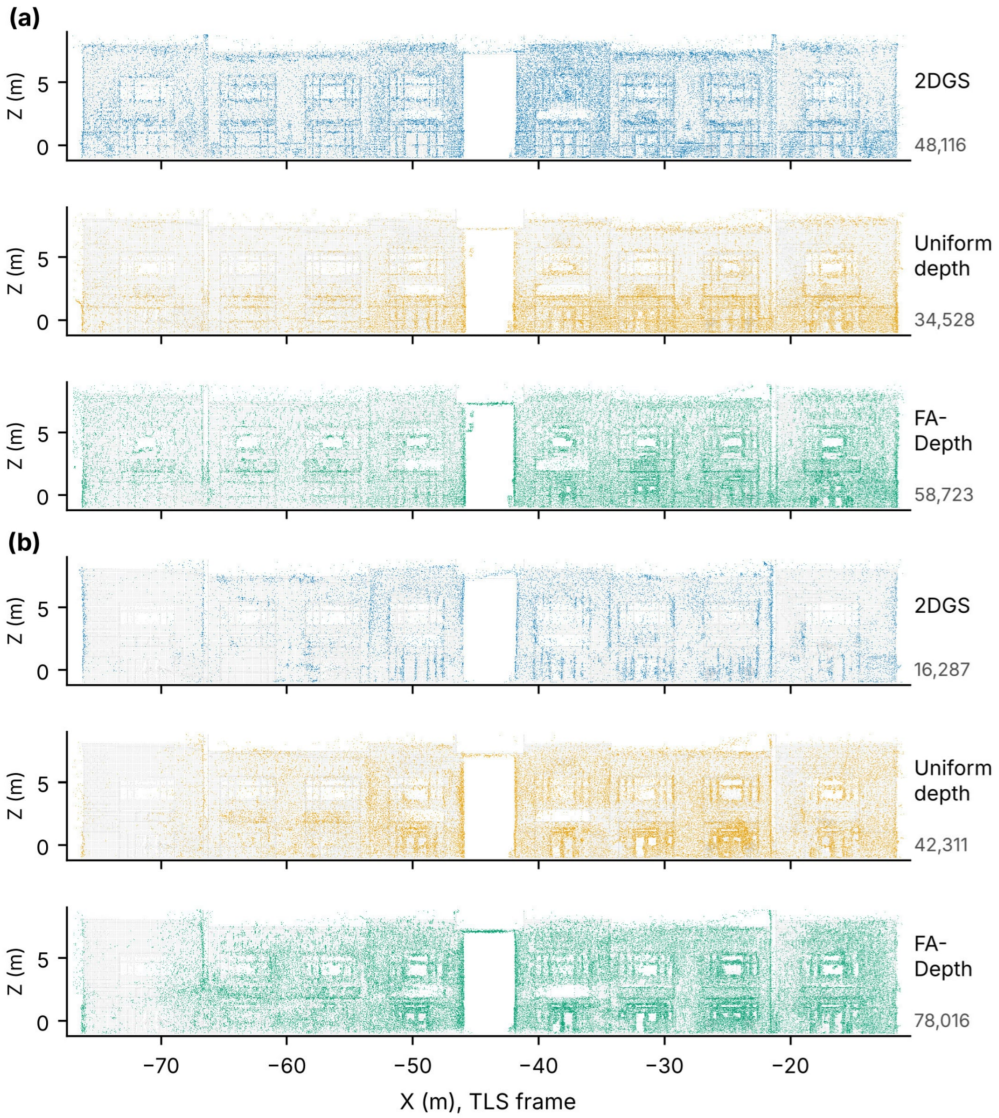

Figure 7. Facade points of the 2DGS configurations, viewed perpendicular to the facade: (**a**) SHARE; (**b**) RS7. In each row, the coloured points are the prepared facade points of one run, and the TLS reference is shown in grey behind them; the number of points is given on the right. Runs with seed 20260817 are shown on SHARE and with seed 20260818 on RS7. The differences are largest on RS7, where the 2DGS points are sparse over most of the facade, and FA-Depth fills it except at the western end and near the ground, where uniform depth concentrates its points.

5.5. Glazed and Opaque Regions

The regional results are reported in Table 8 and Figure 9. FA-Depth improved coverage in both regions on both platforms, with every FA-Depth run better than every 2DGS run in completeness and in the F-score at 0.10 m. On SHARE, the F-score at 0.10 m rose from 0.471 to 0.594 in the glazed region and from 0.585 to 0.618 in the opaque region. On RS7, it rose from 0.253 to 0.491 and from 0.314 to 0.552, respectively. The glazed Chamfer-L1 distance improved by 11% on SHARE and by 33% on RS7, with separated ranges. Because the LiDAR covers glass only partly, FA-Depth is more likely to have helped by freeing surfels from the ground and anchoring the surrounding frames, so that photometric optimization could fill the panes.

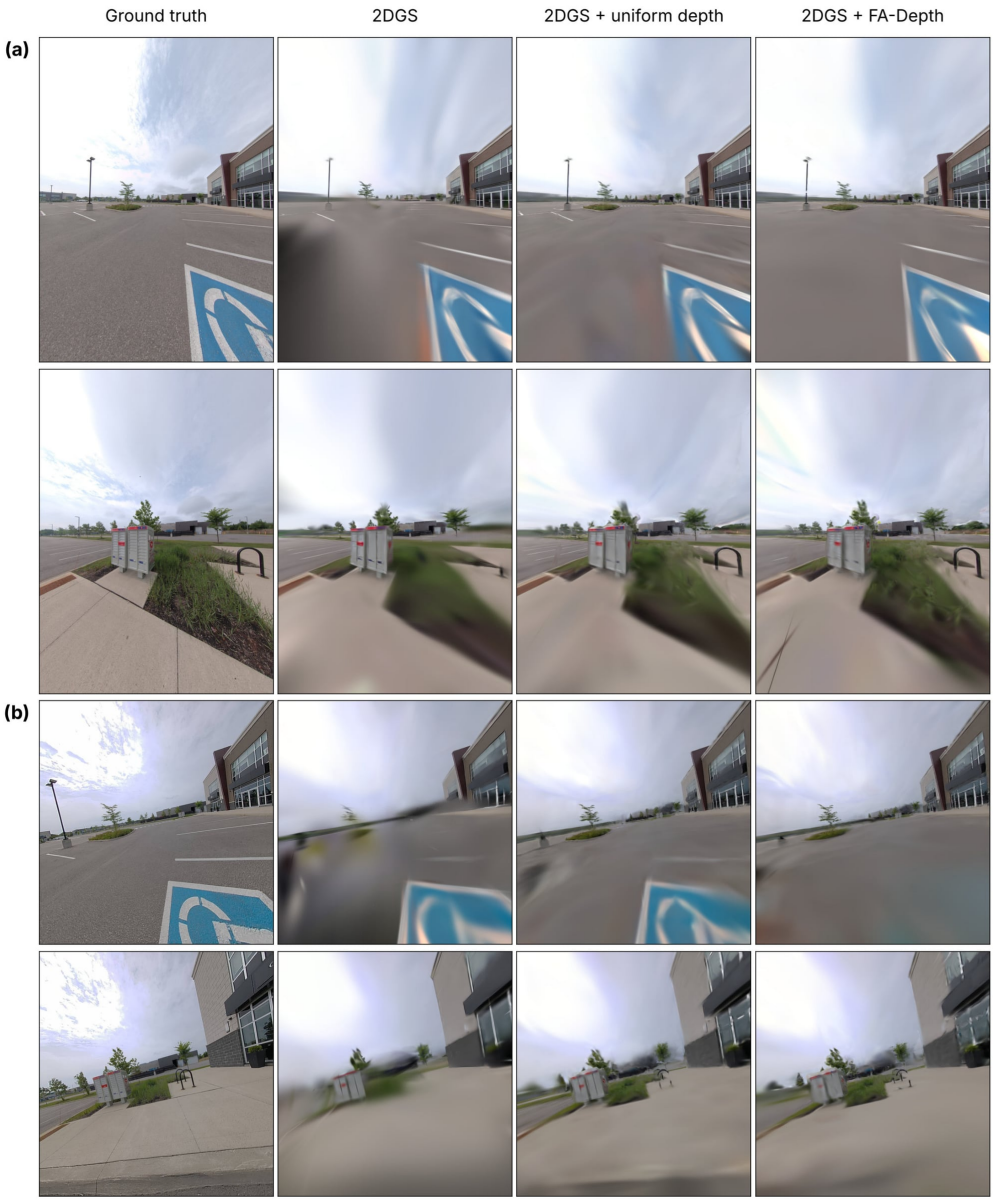

Figure 8. Held-out views rendered by the 2DGS configurations: (**a**) SHARE, runs with seed 20260817; (**b**) RS7, runs with seed 20260818, at the camera positions nearest to the SHARE views.

Table 8. Facade metrics by region, using the manually delineated glazed/opaque partition (means of the runs).

Platform	Region	Configuration	Chamfer-L1 (m)	Completeness, Mean (m)	F-Score at 0.10 m	Plane Bias (m)
SHARE	Glazed	2DGS	0.128	0.128	0.471	−0.023
		FA-Depth	0.113	0.102	0.594	−0.027
	Opaque	2DGS	0.109	0.102	0.585	+0.018
		FA-Depth	0.107	0.095	0.618	+0.022
RS7	Glazed	2DGS	0.204	0.293	0.253	+0.021
		FA-Depth	0.137	0.138	0.491	−0.016
	Opaque	2DGS	0.178	0.220	0.314	+0.030
		FA-Depth	0.131	0.144	0.552	+0.047

The glazed region contains 13,133 TLS points and the opaque region 64,192. A negative plane bias indicates points behind the TLS plane. Over glazing, the supervising LiDAR itself lies 0.042 m (SHARE) and 0.013 m (RS7) behind the TLS plane.

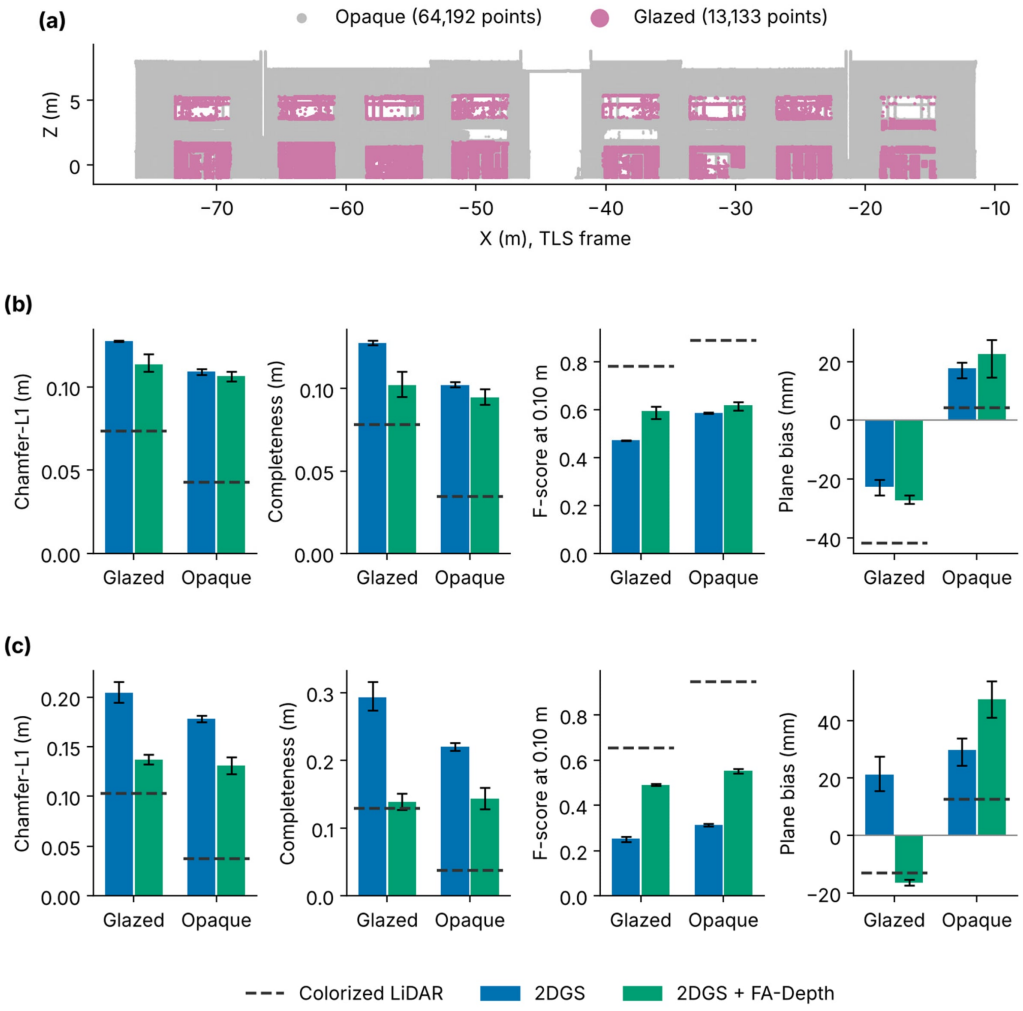

Figure 9. Glazed and opaque regions: (**a**) manual partition of the TLS facade reference into glazed (purple) and opaque (grey) points, delineated on a photo orthomosaic of the facade; (**b**) SHARE and (**c**) RS7 regional Chamfer-L1 distance, completeness, F-score at 0.10 m and plane bias for 2DGS and FA-Depth. Bars show the means of the runs and whiskers their range; dashed lines show the colourized LiDAR of each platform.

The plane bias changed little in either region. In the glazed region, it was −0.023 m for 2DGS and −0.027 m for FA-Depth on SHARE, and +0.021 and −0.016 m on RS7. In the opaque region, it was +0.018 and +0.022 m on SHARE and +0.030 and +0.047 m on RS7. The supervising LiDAR itself lay 0.042 m (SHARE) and 0.013 m (RS7) behind the TLS glazing, because some returns pass through the panes. FA-Depth therefore placed the glazing within 3 cm of the TLS plane on both platforms, at an offset comparable with that of its supervision. With the partition derived from LiDAR return density, the glazed offset of FA-Depth appeared larger (−0.111 and −0.064 m), because that partition selected mainly points where the LiDAR had few returns.

5.6. Role of the LiDAR Initialization

The initialization ablation is reported in Table 9. Without the LiDAR seed, the facade Chamfer-L1 distance increased by 62% on SHARE and by 88% on RS7. On both platforms, completeness degraded more than accuracy. On SHARE, mean accuracy worsened by 0.036 m and mean completeness by 0.101 m. On RS7, mean accuracy worsened by 0.034 m and mean completeness by 0.277 m. Without a geometric seed, densification follows the photometric gradient alone, so surfels remain locally accurate but fail to cover the weakly textured glazed bays. The seed therefore buys coverage more than precision, which mirrors the effect of FA-Depth in Section 5.3. On SHARE, the training-view PSNR rose from 19.64 dB for an earlier, unseeded 2DGS run to 20.06 dB for 2DGS-random, while the held-out PSNR of the same two runs fell from 19.55 to 18.14 dB, which indicates overfitting to the training views. The training-view figures come from the training log rather than from a controlled evaluation and are indicative only.

Table 9. Initialization ablation: 2DGS with the LiDAR seed cloud (mean of three runs) and 2DGS-random (single run).

Metric	SHARE 2DGS	SHARE 2DGS-Random	RS7 2DGS	RS7 2DGS-Random
Facade accuracy, mean (m)	0.118	0.154	0.134	0.167
Facade completeness, mean (m)	0.103	0.204	0.217	0.494
Facade Chamfer-L1 (m)	0.111	0.179	0.175	0.330
Facade F-score at 0.10 m	0.575	0.324	0.312	0.122
Facade F-score at 0.20 m	0.887	0.657	0.682	0.325
Facade plane bias (m)	+0.011	+0.040	+0.031	+0.016
Prepared facade points	49,784	22,748	17,024	5710
Total surfels	272,611	143,932	83,139	55,736
Held-out PSNR (dB)	19.88	18.14	19.87	19.24
Held-out SSIM	0.623	0.599	0.563	0.559
Held-out LPIPS	0.445	0.482	0.512	0.524

5.7. Comparison with Other Representations

Facade geometry for all representations is reported in Table 10, and the cloud-to-cloud (C2C) error maps are shown in Figure 10. The colourized LiDAR was the most accurate representation on both platforms, as expected of a direct measurement, with Chamfer-L1 distances of 0.047 m on SHARE and 0.045 m on RS7. Among the learned representations on SHARE, FA-Depth, 2DGS and uniform depth gave the lowest Chamfer-L1 distances, at 0.107, 0.111 and 0.111 m with overlapping ranges, followed by Nerfacto at 0.124 m, the SHARE vendor model at 0.143 m and the SuGaR mesh surface at 0.166 m. On RS7, FA-Depth gave 0.130 m, followed by uniform depth at 0.159 m, Nerfacto at 0.174 m, 2DGS at 0.175 m and the SuGaR mesh surface at 0.306 m. The difference between Nerfacto and 2DGS on RS7 lies within the spread of the 2DGS runs. Nerfacto covered the RS7 facade more completely, whereas 2DGS remained more accurate on both platforms. SuGaR remained last on both platforms. This can be attributed to its mesh extraction stage, which left holes in the mesh and smoothed the facade details, whereas the refined Gaussians used for rendering did not show these defects.

Table 10. Facade geometry against the TLS reference, voxelized at 0.10 m. Distances are in metres.

Platform	Representation	Points	Accuracy, Mean	Completeness, Mean	Chamfer-L1	F@0.10 m	F@0.20 m	Plane bias
SHARE	Colourized LiDAR [1]	91,862	0.054	0.040	0.047	0.873	0.972	−0.004
	2DGS [2]	49,784	0.118	0.103	0.111	0.575	0.887	+0.011
	2DGS + uniform depth [2]	41,360	0.106	0.115	0.111	0.572	0.883	+0.037
	2DGS + FA-Depth [2]	63,430	0.120	0.094	0.107	0.620	0.899	+0.012
	Nerfacto	85,430	0.142	0.105	0.124	0.482	0.844	+0.042
	Vendor model [3]	82,908	0.167	0.118	0.143	0.410	0.775	−0.042
	SuGaR mesh surface	88,411	0.174	0.157	0.166	0.369	0.693	+0.078
RS7	Colourized LiDAR [1]	82,791	0.042	0.048	0.045	0.911	0.974	+0.013
	2DGS [2]	17,024	0.134	0.217	0.175	0.312	0.682	+0.031
	2DGS + uniform depth [2]	38,554	0.122	0.196	0.159	0.434	0.831	+0.047
	2DGS + FA-Depth [2]	74,719	0.122	0.138	0.130	0.545	0.863	+0.033
	Nerfacto	91,540	0.177	0.172	0.174	0.404	0.711	+0.011
	Vendor model [3]	48,239	0.047	0.067	0.057	0.893	0.978	−0.008
	SuGaR mesh surface	54,711	0.229	0.384	0.306	0.195	0.424	+0.173

[1] Direct measurement and the ceiling of the depth supervision, not a competitor. [2] Means of the runs (SHARE: three runs each; RS7: three runs for 2DGS and two each for uniform depth and FA-Depth). [3] Commercial product with an undisclosed training procedure. The TLS facade reference contains 77,325 points for both platforms.

The RS7 vendor model reached a Chamfer-L1 distance of 0.057 m, within 12 mm of the RS7 LiDAR, which is far closer than any self-trained representation. Its training procedure is undisclosed, so it is reported as a commercial reference rather than as a competitor, and it is not known whether LiDAR constrained its geometry. The SHARE vendor model, by contrast, ranked below 2DGS and Nerfacto on the facade. The improvements reported for FA-Depth therefore refer to 2DGS trained from the same inputs and settings, which is the controlled comparison in this study, rather than to every available reconstruction product.

Signed plane bias varied considerably between representations, from −0.042 m for the SHARE vendor model to +0.173 m for the RS7 SuGaR mesh, against −0.004 and +0.013 m for the two LiDAR clouds. These offsets are systematic rather than random. They are not visible in the Chamfer-L1 distance, which is unsigned and therefore cannot reveal the direction of a consistent displacement of the whole representation along the surface normal.

The matched facade views in Figure 11 show the representations from an identical camera. On SHARE, the 2DGS panel reproduces the storefront mullions and canopy edges with the fewest structural artefacts, although elongated streaks are visible above and below the facade where surfels extend beyond the observed surface. The SHARE vendor and SuGaR panels show erosion of the glazed bays and the upper roof edge, and the SuGaR panel shows holes in the cladding where the mesh failed to close. On RS7, the ordering differs, and the vendor model shows the most complete facade. On RS7, the 2DGS splats reproduce the layout of the facade but blur the mullions and glazing, consistent with the lower completeness of 2DGS on this platform, and the SuGaR panel shows holes across most bays. The RS7 vendor model is shown as primitive centres.

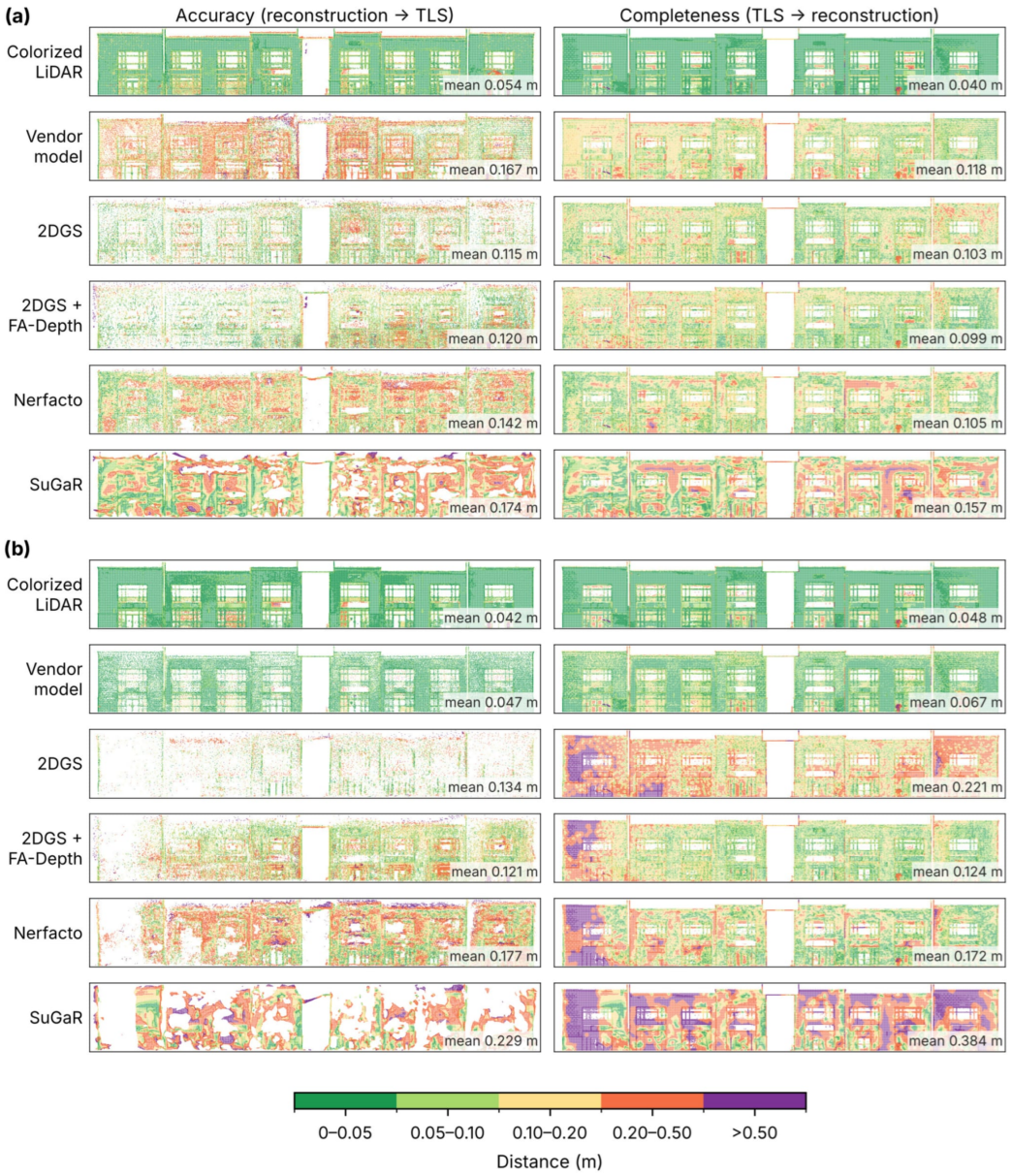

Figure 10. Cloud-to-cloud error maps within the facade box, viewed perpendicular to the facade and coloured in common distance classes: (**a**) SHARE; (**b**) RS7. For each representation, the left map shows accuracy (each reconstructed point coloured by its distance to the TLS reference) and the right map shows completeness (each TLS point coloured by its distance to the reconstruction); the mean distance is given in each map. 2DGS and FA-Depth are the runs with seed 20260817 on SHARE and 20260818 on RS7.

Rendering results on the held-out views are reported in Table 11, and qualitative examples are shown in Figure 12. On SHARE, FA-Depth gave the highest PSNR, at 22.12 dB, followed by uniform depth at 21.10 dB, SuGaR at 20.91 dB and Nerfacto at 20.59 dB, while the three LiDAR-seeded 2DGS configurations gave the best SSIM and LPIPS. On RS7, FA-Depth and uniform depth gave the highest PSNR, at 20.19 and 20.18 dB with overlapping ranges, and FA-Depth gave the best LPIPS, at 0.508. The margins were smaller on RS7.SuGaR, which gave the highest PSNR among the methods without a depth term on SHARE, gave the lowest PSNR on RS7. No method was best on all three measures on SHARE, so the three should be reported together rather than combined into a single ranking. Rendering quality did not follow geometry. For example, SuGaR rendered better than Nerfacto on

SHARE while its facade geometry was poorer, and 2DGS-random on RS7 rendered within 0.7 dB of 2DGS while its facade Chamfer-L1 distance nearly doubled.

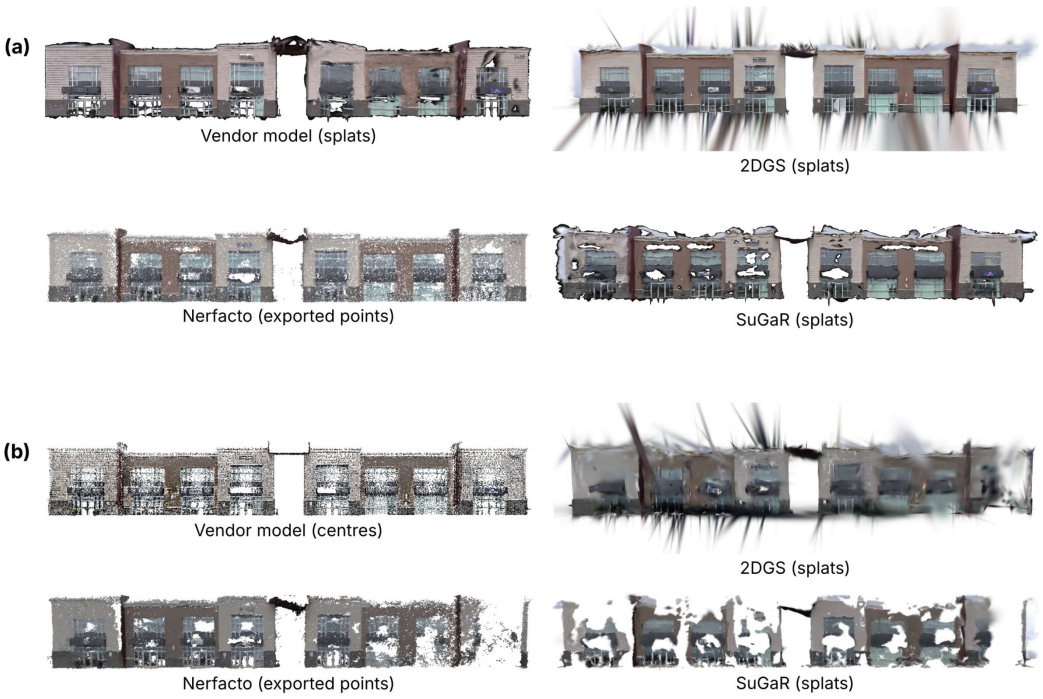

Figure 11. Matched facade comparison from an identical camera, rendered with the same Gaussian-splat viewer on both platforms: (**a**) SHARE vendor model, 2DGS and SuGaR as splats, with Nerfacto as exported points; (**b**) RS7 2DGS (run with seed 20260818) and SuGaR as splats, Nerfacto as exported points and the vendor model as primitive centres.

Table 11. Rendering quality on the held-out views: 200 SHARE views at 1581 × 2108 px and 173 RS7 views at 1500 × 1500 px.

Method	SHARE PSNR (dB)	SHARE SSIM	SHARE LPIPS	RS7 PSNR (dB)	RS7 SSIM	RS7 LPIPS	RS7 PSNR, Mask Excluded (dB) [1]
2DGS [2]	19.88	0.623	0.445	19.87	0.563	0.512	20.22
2DGS + uniform depth [2]	21.10	0.622	0.449	20.18	0.564	0.510	20.60
2DGS + FA-Depth [2]	22.12	0.625	0.447	20.19	0.564	0.508	20.58
2DGS-random	18.14	0.599	0.482	19.24	0.559	0.524	19.56
Nerfacto, native evaluator	20.59	0.595	0.570	19.43	0.544	0.653	—
Nerfacto, common evaluator	—	—	—	19.43	0.546	0.517	20.00
SuGaR	20.91	0.617	0.463	18.90	0.552	0.533	19.36
Vendor model	14.19	0.561	0.515	Not held-out [3]	—	—	—

[1] The vendor's operator mask is excluded from each test image. [2] Mean of the runs. [3] The RS7 vendor model was trained on crops at 170 of the 173 test timestamps. The SHARE vendor model stores degree-zero spherical harmonics only, which limits its view-dependent colour.



Figure 12. Held-out views from the storefront (top row of each part) and the mailbox (bottom row): (**a**) SHARE; (**b**) RS7, at the camera positions nearest to the SHARE views. 2DGS is the run with seed 20260817 on SHARE and 20260818 on RS7. The RS7 vendor model was trained on crops at these timestamps and was rendered without its separate sky model, so its views are not held out and its sky appears black.

6. Discussion

6.1. When LiDAR Depth Supervision Helps

Depth supervision is generally expected to improve Gaussian splatting, and it does when the sensor's returns lie on the surfaces being reconstructed. For example, Turkulainen et al. [36] improved the geometry of 2DGS on indoor scenes with sensor depth, where the sensor sees the surfaces of interest. From a handheld platform, this condition does not hold for the facade, because most LiDAR returns fall on the pavement near the operator. Uniform depth therefore spends much of the capacity of the model on the ground, and its benefit to the facade is diluted rather than lost. In the splat model, it gave no net facade gain on SHARE, where accuracy improved, but completeness worsened, so that the Chamfer-L1 distance and the F-scores did not change. On RS7, it improved the facade, but only because 2DGS placed very few surfels there, and the gain was well below that of FA-Depth. Where the supervision falls in the image therefore matters as much as whether supervision is used. Directing the same supervision to the facade recovered its benefit, and FA-Depth gave better facade completeness of the splat model than uniform depth on both platforms.

The capacity diagnostic shows the mechanism directly. Uniform depth moved the surfels of the model towards the trajectory and the ground on both platforms, raising the share in the ground band from 26.5% to 33.8% on SHARE and from 19.1% to 41.6% on RS7, whereas FA-Depth reduced it to 7.2% and 5.1%. The initialization ablation mirrors this result, because the LiDAR seed also buys coverage rather than precision: removing it degraded completeness about three times as much as accuracy on SHARE and eight times as much on RS7. The LiDAR seed and FA-Depth therefore both add coverage to the facade, whereas uniform depth spends it on the ground.

This finding is consistent with earlier work on weighting depth supervision. Street Gaussians [29] discarded the LiDAR pixels with the largest residuals, and Zhang et al. [38] weighted depth by confidence. The present results add a geometric reason for weighting, which applies to any ground-based platform. The same principle could be applied to other targets, for example, by masking vegetation or vehicles when the building is of interest, but this was not tested here.

FA-Depth also increased the total number of surfels by factors of 3.3 on SHARE and 12.3 on RS7 (Table 7). All configurations used identical densification settings and no cap, so this increase is a consequence of the depth term rather than a separate setting: 2DGS adds surfels where the view-space gradient of the loss remains high, and FA-Depth keeps that gradient on the facade. The number of surfels alone does not explain the facade gain. Uniform depth also increased the number of surfels by 68% on SHARE and 5.7 times on RS7, but placed the additional surfels near the trajectory, so that the number of surfels inside the facade box fell on SHARE and facade completeness did not improve on either platform. Under FA-Depth, the number of surfels inside the facade box rose by factors of 1.7 and 10.1. A fixed cap on the number of surfels was not imposed, because it would also restrict the densification through which both depth terms act, and it would change the 2DGS baseline.

When the facade was evaluated on surfaces extracted from the rendered depth by TSDF fusion (Table 6), uniform depth and FA-Depth gave similar completeness and accuracy on both platforms, and both improved on 2DGS. Both depth terms constrain the rendered depth on facade pixels, so the fused surfaces are similar. However, the rendered depth is an opacity-weighted mean along each ray, and the same depth can be produced by different arrangements of surfels. Under uniform depth, 33.8% and 41.6% of the surfels lay in the ground band, against 7.2% and 5.1% under FA-Depth, and the facade completeness of the splat model did not improve. The benefit of FA-Depth is therefore in the Gaussian splatting representation itself, which is the product used for visualization and digital twins, whereas for mesh extraction either depth term is sufficient.

6.2. Dependence on Facade Coverage

FA-Depth improved completeness and F-scores on both platforms with separated ranges. Its other effects depended on the platform. On SHARE, the rendering gain was large, at 2.24 dB, while the accuracy did not improve. On RS7, the rendering gain was small, at 0.32 dB, while the accuracy and the Chamfer-L1 distance improved substantially. These differences could be explained by the weaker 2DGS model on RS7. Its images were undistorted from a fisheye camera at a lower resolution, fewer views were available, the selected camera faced backwards for half of the route, and the operator appears in many frames. The RS7 2DGS model therefore contained only about 83,000 surfels, against about 273,000 on SHARE, and left more of the facade uncovered, which gave FA-Depth more room to improve the geometry. More generally, these results suggest that FA-Depth improves the geometry most where Gaussian splatting covers the facade poorly, for example, where views are few, distant or of low resolution, whereas where the facade is already well-covered, its gain is mainly in rendering. The accuracy gain on RS7 came from the tail of the distribution

rather than from the median, which suggests that the additional, well-placed surfels reduced the share of poorly placed ones rather than refining existing ones. The defensible general statement is therefore that FA-Depth reliably improves facade coverage, improves rendering by a margin that depends on how well the facade is already covered, and does not reliably improve the aggregate facade accuracy.

The configuration was selected on SHARE validation views and transferred to RS7 without further tuning. It is therefore possible that a different weight would suit RS7 better. However, retuning on RS7 would have removed the test of generalization that the transfer provides, and the clear replication of the coverage and capacity effects suggests that the configuration is not specific to one platform.

6.3. Accuracy Ceiling and Commercial Models

The direct sensor baseline remains substantially ahead. In terms of Chamfer-L1 distance, the colourized LiDAR was 2.3–2.4 times more accurate than 2DGS and FA-Depth on SHARE, and 2.9 times more accurate than FA-Depth on RS7. The supervising cloud also limits FA-Depth, which cannot make the model more accurate than the LiDAR where the LiDAR is dense and trusted. The RS7 vendor model came within 12 mm of its LiDAR, far closer than any self-trained representation, while the SHARE vendor model did not. Its training procedure is undisclosed, so the reason cannot be established here. It is plausible that LiDAR geometry constrains the vendor model, but this remains a hypothesis. The comparison does show that commercial handheld platforms are beginning to deliver Gaussian products whose facade geometry approaches that of the LiDAR, and that such products should be assessed against an independent reference before use.

The facade agreement measured here can be placed against earlier studies, with the caveat that the metrics differ. SiLVR [33] reported an accuracy of 8–11 cm against TLS for an LiDAR-supervised NeRF, which is of the same order as the mean facade accuracy of approximately 0.12 m reported here for FA-Depth on both platforms. Image-based Gaussian splatting of heritage facades has been reported with standard deviations of approximately 2–3 cm against TLS [19,20]. Those studies, however, report the spread of signed distances after alignment. The present study reports unsigned nearest-neighbour distances under one fixed registration per platform, which also includes residual drift. Direct numerical comparison is therefore not appropriate.

6.4. Implications for Facade Documentation and Monitoring

With the glass delineated independently of the LiDAR, the glazing of FA-Depth lay 0.027 m behind the TLS plane on SHARE and 0.016 m on RS7, close to the supervising LiDAR (0.042 and 0.013 m), and the opaque surfaces lay 0.022 and 0.047 m in front of it. A facade agreement of approximately 0.1 m in Chamfer-L1 distance with regional offsets of a few centimetres suits facade documentation, visual inspection and the detection of changes at the decimetre scale, which are the main uses of facade records in digital twins. For these uses, the completeness of the facade is the main requirement, because gaps in glazed bays and along upper edges are among the most visible defects of a splat model in a viewer or digital twin, and FA-Depth reduced them on both platforms. The results do not support the monitoring of structural deformation, which requires millimetre-level accuracy [43,64]. This limit applies to handheld mobile mapping as a whole rather than to FA-Depth, because the colourized LiDAR that supervises the model agreed with the TLS reference to 0.045–0.047 m in Chamfer-L1 distance (Table 10). For change detection, glazed and opaque regions should be assessed separately, and biases should be estimated against a stable reference and removed, before such representations are compared over time.

6.5. Practical Considerations

FA-Depth is specialized for the facade, and it is less stable than 2DGS. On SHARE, its SD of facade Chamfer-L1 distance was 2.4 times that of 2DGS. Its number of surfels grew by factors of 3.3 on SHARE and 12 on RS7, and it caused memory exhaustion at $\lambda = 1.0$ and 2.0 on SHARE and in one RS7 run at $\lambda = 0.50$. Uniform depth also exhausted the memory in one RS7 run. Training took approximately 8–17% longer on SHARE and 50–65% longer on RS7 than 2DGS. A cap on densification would likely improve stability, but it was not applied because it would have broken the one-variable comparison with 2DGS.

The computational record in Table 12 is also relevant to practice. The vendor pipelines required approximately six hours of processing on each platform. The self-trained 2DGS models required between 50 and 110 min on one graphics card, and SuGaR required approximately one hour. These figures are indicative, since they depend on the hardware and the number of training views.

Table 12. Processing time per run on the evaluation workstation, except where noted.

Process	SHARE	RS7
Vendor processing	22,069 s point cloud pipeline; Gaussian training failed three times, final duration not recorded	21,033 s in total, on the vendor's processing computer
2DGS (per run)	5621–5755 s	2999–3087 s
2DGS + uniform depth (per run)	5705–6300 s	4052–4059 s; one run failed after iteration 15,060
2DGS + FA-Depth (per run)	6227–6573 s	4636–4947 s; one run failed at iteration 14,320
2DGS-random	6267 s	3680 s
Nerfacto	2533 s	1696 s
SuGaR, including the 3DGS prerequisite	≈3948 s	3360 s

Workstation: Intel Core i9-10900X, 125 GiB memory, one NVIDIA RTX A6000 per run.

6.6. Limitations

The conclusions of this study are bounded by its design. The facade of one building was evaluated. However, it was captured by two platforms with different LiDAR sensors, camera geometries, routes and processing, which provide two independent datasets, and the parameters selected on one were transferred unchanged to the other. The facade is 65 m long and combines full-height glazed storefronts, which make up 17.0% of the reference points, with brick piers, precast cladding panels and projecting canopies. These surfaces differ in two properties that limit reconstruction, namely whether they return LiDAR measurements and whether their appearance stays the same from different viewpoints, as photometric matching between images assumes. Glazing is transparent and reflective, so LiDAR returns are missing or pass through it [39,40], and radiance fields tend to place reflections behind it [41,42], whereas brick and cladding are opaque and diffuse. This difference was measurable: the F-score at 0.10 m of 2DGS was 0.471 on the glazing against 0.585 on the opaque surfaces on SHARE, and 0.253 against 0.314 on RS7, and the supervising LiDAR lay 0.042 and 0.013 m behind the TLS glazing (Table 8). The evaluation therefore covers surfaces that are reconstructed well and surfaces that both LiDAR and images reconstruct poorly, and FA-Depth improved coverage on both (Section 5.5). On this basis, the findings are expected to apply to other facades that combine glazing with opaque cladding and are captured from the ground with a handheld platform, because the ground imbalance arises from the height at which such a platform is carried. Facades in narrow streets, behind dense vegetation, with strong relief, with a much larger share of glazing or on

taller buildings were not tested, and the size of the gain may differ there. Two RS7 runs failed due to memory exhaustion, so two configurations have two runs on RS7. The held-out views lie close to training views, so the rendering results do not measure performance on novel viewpoints. Finally, the vendor Gaussian models were produced by undisclosed procedures and were treated as commercial references rather than as controlled configurations.

7. Conclusions

This paper introduced a facade-aware LiDAR depth supervision (FA-Depth) approach for 2D Gaussian Splatting and evaluated it against a TLS survey of a two-storey commercial facade. Two handheld LiDAR-camera platforms, SHARE SLAM S20 and CHCNAV RS7, were used, and 2DGS with and without depth supervision was compared with Nerfacto, SuGaR, the vendor Gaussian models and the colourized LiDAR under one rigid registration per platform.

Uniform LiDAR depth supervision was dominated by the pavement near the operator, with median supervised depths of 1.42 and 2.00 m. It raised the share of surfels in the ground band from 26.5% to 33.8% and from 19.1% to 41.6% on the two platforms, respectively, and reduced facade completeness on SHARE. FA-Depth reduced the same share to 7.2% and 5.1%, and gave better facade completeness of the splat model than uniform depth on both platforms. It improved mean facade completeness from 0.103 to 0.094 m and from 0.217 to 0.138 m, and raised the F-score at 0.10 m from 0.575 to 0.620 and from 0.312 to 0.545, with separated ranges. Held-out PSNR was improved by 2.24 and 0.32 dB. Aggregate facade accuracy was improved on one platform only. The aggregate plane bias did not improve, and, with the glass delineated independently of the LiDAR, the glazing of FA-Depth lay within 3 cm of the TLS plane. On surfaces extracted from the rendered depth, both depth terms improved completeness and accuracy over 2DGS, whereas only FA-Depth improved the facade completeness of the splat model itself. These results indicate that LiDAR depth supervision improves the facade coverage of Gaussian splatting from handheld mobile mapping when it is directed at the facade rather than the ground.

Author Contributions: Conceptualization, A.E. and A.E.-R.; methodology, A.E.; software, A.E.; validation, A.E.; formal analysis, A.E.; investigation, A.E. and M.E.-R.; data curation, A.E. and M.E.-R.; writing—original draft preparation, A.E.; writing—review and editing, M.E.-R. and A.E.-R.; visualization, A.E.; supervision, A.E.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The analysis scripts, including the registration, geometry preparation, metric and figure code, are available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Di Stefano, F.; Chiappini, S.; Gorreja, A.; Balestra, M.; Pierdicca, R. Mobile 3D Scan LiDAR: A Literature Review. *Geomat. Nat. Hazards Risk* **2021**, *12*, 2387–2429. [CrossRef]
2. Tanduo, B.; Martino, A.; Balletti, C.; Guerra, F. New Tools for Urban Analysis: A SLAM-Based Research in Venice. *Remote Sens.* **2022**, *14*, 4325. [CrossRef]
3. Matellon, A.; Maset, E.; Beinat, A.; Visintini, D. Surface Reconstruction from SLAM-Based Point Clouds: Results from the Datasets of the 2023 SIFET Benchmark. *Remote Sens.* **2024**, *16*, 3439. [CrossRef]
4. Strecha, C.; Hughes, R.; Cucci, D.A.; Dotti, A.; Coelho Lopes, M.; Rothenbühler, P. Geometric and Visual SLAM: The Accuracy of Modern Handheld LiDAR Scanners. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2026**, *XLIX-B1-2026*, 193–200. [CrossRef]

5. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Computer Vision—ECCV 2020*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12346, pp. 405–421. [CrossRef]
6. Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Kerr, J.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; et al. Nerfstudio: A Modular Framework for Neural Radiance Field Development. In *Proceedings of the ACM SIGGRAPH 2023 Conference, Los Angeles, CA, USA*; ACM: New York, NY, USA, 2023; pp. 1–12. [CrossRef]
7. Barron, J.T.; Mildenhall, B.; Verbin, D.; Srinivasan, P.P.; Hedman, P. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA*; IEEE: Piscataway, NJ, USA, 2022; pp. 5460–5469. [CrossRef]
8. Müller, T.; Evans, A.; Schied, C.; Keller, A. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.* **2022**, *41*, 1–15. [CrossRef]
9. Kerbl, B.; Kopanas, G.; Leimkühler, T.; Drettakis, G. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* **2023**, *42*, 1–14. [CrossRef]
10. Huang, B.; Yu, Z.; Chen, A.; Geiger, A.; Gao, S. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. In *Proceedings of the ACM SIGGRAPH 2024 Conference Papers, Denver, CO, USA*; ACM: New York, NY, USA, 2024; pp. 1–11. [CrossRef]
11. Guédon, A.; Lepetit, V. SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA*; IEEE: Piscataway, NJ, USA, 2024; pp. 5354–5363. [CrossRef]
12. Yu, Z.; Sattler, T.; Geiger, A. Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes. *ACM Trans. Graph.* **2024**, *43*, 1–13. [CrossRef]
13. Chen, D.; Li, H.; Ye, W.; Wang, Y.; Xie, W.; Zhai, S.; Wang, N.; Liu, H.; Bao, H.; Zhang, G. PGSR: Planar-Based Gaussian Splatting for Efficient and High-Fidelity Surface Reconstruction. *IEEE Trans. Vis. Comput. Graph.* **2025**, *31*, 6100–6111. [CrossRef] [PubMed]
14. Wu, T.; Yuan, Y.-J.; Zhang, L.-X.; Yang, J.; Cao, Y.-P.; Yan, L.-Q.; Gao, L. Recent Advances in 3D Gaussian Splatting. *Comput. Vis. Media* **2024**, *10*, 613–642. [CrossRef]
15. Petrovska, I.; Jäger, M.; Haitz, D.; Jutzi, B. Geometric Accuracy Analysis between Neural Radiance Fields (NeRFs) and Terrestrial Laser Scanning (TLS). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-1/W3-2023*, 153–159. [CrossRef]
16. Petrovska, I.; Jutzi, B. Vision through Obstacles—3D Geometric Reconstruction and Evaluation of Neural Radiance Fields (NeRFs). *Remote Sens.* **2024**, *16*, 1188. [CrossRef]
17. Remondino, F.; Karami, A.; Yan, Z.; Mazzacca, G.; Rigon, S.; Qin, R. A Critical Analysis of NeRF-Based 3D Reconstruction. *Remote Sens.* **2023**, *15*, 3585. [CrossRef]
18. Murtiyoso, A.; Grussenmeyer, P. Initial Assessment on the Use of State-of-the-Art NeRF Neural Network 3D Reconstruction for Heritage Documentation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-M-2-2023*, 1113–1118. [CrossRef]
19. Murtiyoso, A.; Macher, H. Gaussian Splatting for Facade Orthophoto Generation—Comparison with MVS and TLS. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *XLVIII-M-9-2025*, 1059–1064. [CrossRef]
20. Billi, D.; Delasse, C.; Murtiyoso, A.; Macher, H.; Grussenmeyer, P.; Caroti, G.; Piemonte, A. 3D Meshing of Challenging Surfaces Using Gaussian Splatting. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2026**, *XLIX-B2-2026*, 801–808. [CrossRef]
21. Suwardhi, D.; Agustina, L.K.; Andika, W.; Widyastuti, R.; Fadilah, W.A.; Murtiyoso, A.; Remondino, F. Integrating Terrestrial Laser Scanning and 3D Gaussian Splatting for Heritage Building Documentation and Monitoring. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2026**, *XLVIII-2/W12-2026*, 463–470. [CrossRef]
22. McNally, S.; Jabari, S.; McGrath, H.; Masry, M. A Comprehensive Evaluation of the Spatial Accuracy of Building Gaussian Splatting. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2026**, *XI-M-1-2026*, 31–38. [CrossRef]
23. Tao, Y.; Muñoz-Bañón, M.Á.; Zhang, L.; Wang, J.; Fu, L.F.T.; Fallon, M. The Oxford Spires Dataset: Benchmarking Large-Scale LiDAR-Visual Localisation, Reconstruction and Radiance Field Methods. *arXiv* **2024**, arXiv:2411.10546.
24. Hong, S.; He, J.; Zheng, X.; Zheng, C.; Shen, S. LIV-GaussMap: LiDAR-Inertial-Visual Fusion for Real-Time 3D Radiance Field Map Rendering. *IEEE Robot. Autom. Lett.* **2024**, *9*, 9765–9772. [CrossRef]
25. Lang, X.; Li, L.; Wu, C.; Zhao, C.; Liu, L.; Liu, Y.; Lv, J.; Zuo, X. Gaussian-LIC: Real-Time Photo-Realistic SLAM with Gaussian Splatting and LiDAR-Inertial-Camera Fusion. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Atlanta, GA, USA*; IEEE: Piscataway, NJ, USA, 2025; pp. 8500–8507. [CrossRef]
26. Zhou, X.; Lin, Z.; Shan, X.; Wang, Y.; Sun, D.; Yang, M.-H. DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA*; IEEE: Piscataway, NJ, USA, 2024; pp. 21634–21643.
27. Cui, J.; Cao, J.; Zhao, F.; He, Z.; Chen, Y.; Zhong, Y.; Xu, L.; Shi, Y.; Zhang, Y.; Yu, J. LetsGo: Large-Scale Garage Modeling and Rendering via LiDAR-Assisted Gaussian Primitives. *ACM Trans. Graph.* **2024**, *43*, 1–18. [CrossRef]

28. Rematas, K.; Liu, A.; Srinivasan, P.P.; Barron, J.T.; Tagliasacchi, A.; Funkhouser, T.; Ferrari, V. Urban Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA*; IEEE: Piscataway, NJ, USA, 2022; pp. 12932–12942.
29. Yan, Y.; Lin, H.; Zhou, C.; Wang, W.; Sun, H.; Zhan, K.; Lang, X.; Zhou, X.; Peng, S. Street Gaussians: Modeling Dynamic Urban Scenes with Gaussian Splatting. In *Computer Vision—ECCV 2024*; Springer: Cham, Switzerland, 2024. [CrossRef]
30. Zhao, C.; Sun, S.; Wang, R.; Guo, Y.; Wan, J.-J.; Huang, Z.; Huang, X.; Chen, Y.V.; Ren, L. TCLC-GS: Tightly Coupled LiDAR-Camera Gaussian Splatting for Autonomous Driving. In *Computer Vision—ECCV 2024*; Springer: Cham, Switzerland, 2025; pp. 91–106. [CrossRef]
31. Jiang, C.; Gao, R.; Shao, K.; Wang, Y.; Xiong, R.; Zhang, Y. LI-GS: Gaussian Splatting with LiDAR Incorporated for Accurate Large-Scale Reconstruction. *arXiv* **2024**, arXiv:2409.12899.
32. Yao, Y.; Zhang, B.; Zhang, W.; Gao, L.; Peng, D.; Li, B.; Wang, Y.; Wang, B. ARSGaussian: 3D Gaussian Splatting with LiDAR for Aerial Remote Sensing Novel View Synthesis. *ISPRS J. Photogramm. Remote Sens.* **2026**, *231*, 288–306. [CrossRef]
33. Tao, Y.; Bhalgat, Y.; Fu, L.F.T.; Mattamala, M.; Chebrolu, N.; Fallon, M. SiLVR: Scalable Lidar-Visual Reconstruction with Neural Radiance Fields for Robotic Inspection. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan*; IEEE: Piscataway, NJ, USA, 2024. [CrossRef]
34. Deng, K.; Liu, A.; Zhu, J.-Y.; Ramanan, D. Depth-Supervised NeRF: Fewer Views and Faster Training for Free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA*; IEEE: Piscataway, NJ, USA, 2022; pp. 12882–12891.
35. Roessle, B.; Barron, J.T.; Mildenhall, B.; Srinivasan, P.P.; Nießner, M. Dense Depth Priors for Neural Radiance Fields from Sparse Input Views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA*; IEEE: Piscataway, NJ, USA, 2022; pp. 12892–12901.
36. Turkulainen, M.; Ren, X.; Melekhov, I.; Seiskari, O.; Rahtu, E.; Kannala, J. DN-Splatter: Depth and Normal Priors for Gaussian Splatting and Meshing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Tucson, AZ, USA*; IEEE: Piscataway, NJ, USA, 2025; pp. 2421–2431.
37. Xiao, W.; Goan, E.; Santa Cruz, R.; Ahmedt-Aristizabal, D.; Salvado, O.; Fookes, C.; Lebrat, L. In Depth We Trust: Reliable Monocular Depth Supervision for Gaussian Splatting. *arXiv* **2026**, arXiv:2604.05715.
38. Zhang, Q.; Wysocki, O.; Urban, S.; Jutzi, B. CDGS: Confidence-Aware Depth Regularization for 3D Gaussian Splatting. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *XLVIII-2/W7-2024*, 189–196. [CrossRef]
39. Yun, J.-S.; Sim, J.-Y. Reflection Removal for Large-Scale 3D Point Clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA*; IEEE: Piscataway, NJ, USA, 2018; pp. 4597–4605.
40. Shao, W.; Guo, Z.; Zhang, B.; Xue, Y.; Ji, T.; Lao, Y. GRAR: Glass-Induced Reflection Artifact Removal in LiDAR Point Clouds. *arXiv* **2026**, arXiv:2606.10541.
41. Guo, Y.-C.; Kang, D.; Bao, L.; He, Y.; Zhang, S.-H. NeRFReN: Neural Radiance Fields with Reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA*; IEEE: Piscataway, NJ, USA, 2022; pp. 18409–18418.
42. Verbin, D.; Hedman, P.; Mildenhall, B.; Zickler, T.; Barron, J.T.; Srinivasan, P.P. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA*; IEEE: Piscataway, NJ, USA, 2022; pp. 5491–5500.
43. Korswagen, P.A.; Longo, M.; Rots, J.G. Experimental and Computational Study into the Onset of Light Damage of Dutch Masonry Structures. *Heron* **2021**, *66*, 91–114.
44. Zhou, Y.; Zhu, J.; Zhao, L.; Hu, G.; Xin, J.; Zhang, H.; Yang, J. High-Precision Monitoring Method for Bridge Deformation Measurement and Error Analysis Based on Terrestrial Laser Scanning. *Remote Sens.* **2024**, *16*, 2263. [CrossRef]
45. SHAREUAV Ltd. SHARE PointClouds Studio, Version 2.5.0 [Computer Software]. 2025. Available online: https://www.share3dcam.com/en/app.html (accessed on 24 November 2026).
46. CHC Navigation. *CoPre*, Version 2.11.2 [Computer Software]; CHC Navigation: Shanghai, China, 2026.
47. Trimble Inc. Trimble X9 3D Laser Scanning System. Available online: https://geospatial.trimble.com/en/products/hardware/trimble-x9 (accessed on 24 November 2026).
48. Trimble Inc. *Trimble Perspective [Computer Software]*; Trimble Inc.: Westminster, CO, USA. Available online: https://help.fieldsystems.trimble.com/perspective/2024.10.htm (accessed on 24 November 2026).
49. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2004.
50. Schönberger, J.L.; Frahm, J.-M. Structure-from-Motion Revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*; IEEE: Piscataway, NJ, USA, 2016; pp. 4104–4113.
51. Kannala, J.; Brandt, S.S. A Generic Camera Model and Calibration Method for Conventional, Wide-Angle, and Fish-Eye Lenses. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1335–1340. [CrossRef] [PubMed]
52. Huber, P.J. Robust Estimation of a Location Parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [CrossRef]

53. Rusu, R.B.; Blodow, N.; Beetz, M. Fast Point Feature Histograms (FPFH) for 3D Registration. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan*; IEEE: Piscataway, NJ, USA, 2009; pp. 3212–3217.
54. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395.
55. Besl, P.J.; McKay, N.D. A Method for Registration of 3-D Shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256. [CrossRef]
56. Chen, Y.; Medioni, G. Object Modelling by Registration of Multiple Range Images. *Image Vis. Comput.* **1992**, *10*, 145–155. [CrossRef]
57. Zhou, Q.-Y.; Park, J.; Koltun, V. Open3D: A Modern Library for 3D Data Processing. *arXiv* **2018**, arXiv:1801.09847.
58. Seitz, S.M.; Curless, B.; Diebel, J.; Scharstein, D.; Szeliski, R. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA*; IEEE: Piscataway, NJ, USA, 2006; Volume 1, pp. 519–528. [CrossRef]
59. Knapitsch, A.; Park, J.; Zhou, Q.-Y.; Koltun, V. Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction. *ACM Trans. Graph.* **2017**, *36*, 1–13. [CrossRef]
60. Lindenbergh, R.; Pietrzyk, P. Change Detection and Deformation Analysis Using Static and Mobile Laser Scanning. *Appl. Geomat.* **2015**, *7*, 65–74. [CrossRef]
61. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **1979**, *9*, 62–66. [CrossRef]
62. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
63. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA*; IEEE: Piscataway, NJ, USA, 2018; pp. 586–595.
64. Mukupa, W.; Roberts, G.W.; Hancock, C.M.; Al-Manasir, K. A Review of the Use of Terrestrial Laser Scanning Application for Change Detection and Deformation Monitoring of Structures. *Surv. Rev.* **2016**, *49*, 99–116. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.