3DR-REG: RETHINKING COLOR POINT CLOUD REGISTRATION AS 3D RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Color point cloud registration still largely follows the feature-matching paradigm, in which point cloud features are matched before estimating a rigid transformation. Recent 3D reconstruction foundation models offer a different way to organize cross-view information by directly inferring structured scene geometry. We introduce 3DR-Reg, which reformulates color point cloud registration as 3D reconstruction. Rather than matching independently learned descriptors, 3DR-Reg reconstructs a shared 3D scene from the color information of two input point clouds using a pretrained 3D reconstruction foundation model. These predictions enable registration through camera poses, cross-view correspondences, or a shared reference frame. To improve correspondence-based registration, we further introduce a geometry-guided correspondence refinement module that resolves ambiguous correspondences through geometric consistency and global context. Experiments on 3DMatch and 3DLoMatch demonstrate state-of-the-art registration performance, with mean rotation/translation errors of $0.69°$/2.14 cm and $3.77°$/13.15 cm, respectively. Across multiple frozen 3D reconstruction models, reconstruction predictions support accurate registration without learning registration-specific representations. For the first time, we demonstrate that a 3D reconstruction foundation model can directly serve as the basis for color point cloud registration.

1 INTRODUCTION

Color point cloud registration aims to estimate the relative pose between two point clouds, which is a fundamental component of 3D computer vision (Zeng et al., 2017; Yuan et al., 2023; Mu et al., 2024). Compared with geometry-only registration, RGB-D data provides additional color information for establishing cross-view correspondences and recovering the rigid transformation.

Most learning-based registration methods follow a common paradigm: extract point features, establish correspondences by feature matching, and estimate the rigid transformation with a robust geometric solver (Choy et al., 2019; 2020; Yuan et al., 2023; Mu et al., 2024). Recent advances in convolutional networks and attention mechanisms have substantially improved the discriminability of these representations (Choy et al., 2019; Yew & Lee, 2022). For color point clouds, PointMBF performs multi-scale visual-geometric fusion, while ColorPCR incorporates color into hierarchical feature extraction and superpoint matching (Yuan et al., 2023; Mu et al., 2024). Despite their effectiveness, these methods still rely on learning representations that make corresponding regions distinguishable enough to be matched.

This feature-matching paradigm becomes fragile when overlap is limited or different regions exhibit similar geometry and color. Even visual foundation models have so far been used mainly as stronger feature extractors for correspondence estimation (Chen & Qu, 2025). Their transferable representations improve matching, but do not explicitly recover the shared 3D structure that relates the two views. As a result, ambiguous correspondences remain a major source of registration failure. This raises a natural question: *Can registration directly exploit cross-view 3D structure, rather than infer it indirectly through feature matching?*

Recent 3D reconstruction foundation models make this possibility increasingly practical. Given multiple color images, these models can directly predict geometrically structured outputs such as point maps and camera poses (Wang et al., 2024; 2025a; Leroy et al., 2024). Such predictions

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

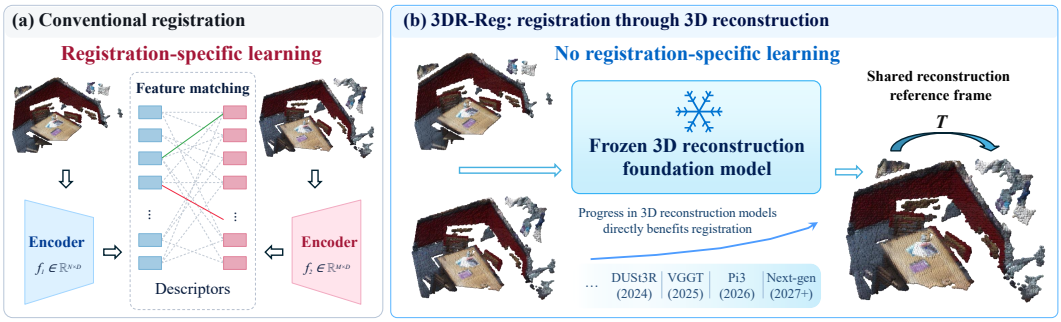


Figure 1: Comparison of color point cloud registration paradigms. (a) Conventional learning-based methods learn registration-specific descriptors and infer cross-view structure indirectly through feature matching. (b) 3DR-Reg uses a frozen 3D reconstruction foundation model to directly predict cross-view 3D structure in a shared reference frame. This formulation enables registration without learning registration-specific representations and allows registration to benefit from progress in 3D reconstruction.

already encode how color point clouds are related in 3D, and therefore provide a natural source of registration cues before explicit point-cloud matching. However, reconstruction predictions are not metrically perfect: errors in predicted geometry can make spatially close reconstructed points inconsistent with the input RGB-D geometry. The key challenge is thus not only to obtain cross-view geometry, but to reliably transfer it to accurate point cloud registration.

As illustrated in Figure 1, we address this challenge with **3DR-Reg**, a framework that formulates color point cloud registration through the predictions of a pretrained 3D reconstruction model. Given the color information of two input point clouds, a frozen reconstruction model predicts their shared 3D structure, including point maps and camera poses. These predictions support camera-pose registration, correspondence-based registration, and alignment in a shared reference frame. To handle reconstruction errors and ambiguous candidates, we introduce a compact geometry-guided correspondence refinement module that scores correspondences using rigid geometric consistency, camera-pose cues, and global context. In this way, reconstruction provides the cross-view structure, while registration-specific learning is confined to selecting geometrically reliable correspondences.

We evaluate 3DR-Reg on 3DMatch and 3DLoMatch (Zeng et al., 2017; Mu et al., 2024). The proposed framework achieves state-of-the-art registration performance among the compared methods, while direct camera-pose and point-map predictions already provide strong registration results without task-specific training. Experiments across multiple 3D reconstruction models further show that the formulation is not tied to a particular reconstruction model, and that improvements in reconstruction quality can translate into better registration. Ablation studies confirm that geometric consistency substantially improves correspondence quality, supporting our central premise that pretrained reconstruction models already contain useful cross-view structure for point cloud registration.

The main contributions of this paper are three-fold:

- We introduce a new formulation of color point cloud registration that directly exploits the cross-view 3D structure predicted by pretrained reconstruction models, instead of using foundation models only as feature extractors.
- We show that frozen reconstruction predictions, including camera poses and point maps, can be directly converted into registration estimates, and propose a lightweight geometry-guided refinement module to select rigidly consistent correspondences.
- Experiments on 3DMatch and 3DLoMatch demonstrate competitive state-of-the-art registration performance, applicability across multiple reconstruction foundation models, and the effectiveness of geometry-guided correspondence refinement.

2 Related Work

Point Cloud Registration. Learning-based point cloud registration typically follows a correspondence-based pipeline: learning point representations, establishing cross-view correspondences, and robustly estimating a rigid transformation. Early methods focused on local descriptors and repeatable keypoints (Zeng et al., 2017; Choy et al., 2019; Bai et al., 2020; Gojcic et al., 2020), while later methods learned dense or hierarchical correspondences and incorporated broader geometric context (Choy et al., 2020; Yu et al., 2021; Yew & Lee, 2022; Qin et al., 2022). In parallel, robust estimators improve registration under low overlap and correspondence outliers through spatial compatibility, consensus sampling, and graph-based optimization (Huang et al., 2021; Bai et al., 2021; Chen et al., 2022; Yu et al., 2023; Zhang et al., 2024; Yan et al., 2025b;a). Despite substantial progress, these methods generally rely on representations and correspondence mechanisms developed specifically for registration.

Color Point Cloud Registration. Color point cloud registration augments geometric matching with color information from RGB-D data. Existing methods model photometric-geometric consistency or fuse visual and geometric features to improve correspondence estimation (El Banani et al., 2021; El Banani & Johnson, 2021; Wang et al., 2022; El Banani et al., 2023; Yuan et al., 2023; Mu et al., 2024; Xu et al., 2024), with recent work further addressing domain shift, low overlap, and photometric inconsistency (Shou et al., 2025). While these methods demonstrate the value of color information, they retain the same underlying paradigm: learn discriminative representations, infer correspondences, and estimate the rigid transformation from these correspondences.

Registration with Foundation Models. Recent work has introduced pretrained foundation models into point cloud registration. Visual foundation models provide transferable features that are projected or transferred to point descriptors and subsequently matched with geometric solvers (Chen & Qu, 2025; Jiang et al., 2025a; Vödisch et al., 2025). Generative models instead assist registration by producing synthetic training pairs, matching matrices, or cross-view images (Chen et al., 2024; Wu et al., 2024; Jiang et al., 2025b; Haitman et al., 2026). These methods reduce the dependence on registration-specific representation learning, but still use foundation models primarily to improve correspondence estimation. In contrast, our formulation directly uses the structured geometric outputs of a 3D reconstruction model, whether camera poses, point maps, or 3D point tracks, as registration cues.

3D Reconstruction Models. Classical 3D reconstruction systems recover camera geometry and scene structure through explicit feature matching and geometric optimization (Schonberger & Frahm, 2016; Huang et al., 2018; Yao et al., 2018; Sun et al., 2021), while neural scene representations provide learned alternatives for multi-view scene modeling (Mildenhall et al., 2021; Kerbl et al., 2023). More recently, feed-forward 3D reconstruction models directly predict structured cross-view geometry such as point maps and camera poses, with subsequent work extending these representations to image matching, scene priors, multi-view reconstruction, and dynamic scenes (Wang et al., 2024; Leroy et al., 2024; Wang et al., 2025a; Jang et al., 2025; Cabon et al., 2025; Sucar et al., 2025). These predictions explicitly encode geometric relations across views and therefore provide a natural interface to point cloud registration. Our work studies how such pretrained reconstruction models can be converted into accurate point cloud registration, while keeping the reconstruction model frozen.

3 Method

We denote two color point clouds as $\mathcal{P} = \{\mathbf{p}_i \in \mathbb{R}^3 \mid i = 1, \ldots, N\}$ with $\mathcal{P}_c = \{\mathbf{p}_{ci} \mid i = 1, \ldots, N\}$ and $\mathcal{Q} = \{\mathbf{q}_j \in \mathbb{R}^3 \mid j = 1, \ldots, M\}$ with $\mathcal{Q}_c = \{\mathbf{q}_{cj} \mid j = 1, \ldots, M\}$, in which $\mathcal{P}_c$ and $\mathcal{Q}_c$ denote the color of $\mathcal{P}$ and $\mathcal{Q}$, respectively. Our goal is to estimate the rigid transformation $\mathbf{T} = (\mathbf{R}, \mathbf{t}) \in \mathrm{SE}(3)$, where $\mathbf{R} \in \mathrm{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$, that aligns $\mathcal{P}$ to $\mathcal{Q}$:

$$\mathbf{q} = \mathbf{R}\mathbf{p} + \mathbf{t}. \tag{1}$$

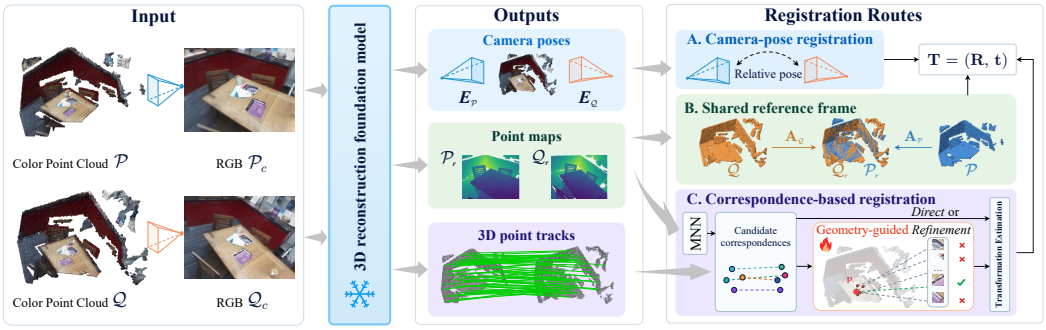


Figure 2: Overview of 3DR-Reg. We leverage a frozen 3D reconstruction foundation model for robust color point cloud registration through three registration routes. The color information $\mathcal{P}_c$ and $\mathcal{Q}_c$ is arranged in the pixel layout of the respective camera views and fed into the model. (A) Camera poses provide the relative transformation between the input point clouds. (B) Point maps enable alignment through a shared reference frame. (C) Point maps and 3D point tracks provide cross-view correspondences, which can be used directly for rigid transformation estimation or refined by our geometry-guided correspondence refinement module before estimation.

3.1 Overview

Our pipeline is illustrated in Fig 2. We employ a pretrained 3D reconstruction model to predict the cross-view geometry

$$Z = F_\theta(\mathcal{P}_c, \mathcal{Q}_c), \tag{2}$$

where F_θ represents a frozen 3D reconstruction model, and Z may contain camera poses, point maps, or 3D point tracks. These predictions describe how the color information of point clouds under specific views is related in 3D space. Then, we utilize these reconstruction predictions through three registration routes: camera poses directly provide a relative transformation; point maps and tracks provide cross-view correspondences between the two point clouds; and point maps also enable alignment through a shared reference frame.

Although reconstruction predictions support direct registration, the resulting cross-view correspondences can remain ambiguous. To address this, we further introduce a lightweight geometry-guided refinement module that evaluates candidate correspondences using the geometry and color of the input point clouds together with reconstruction predictions. The refined correspondences enable more robust point cloud registration. Only the correspondence refinement module is optimized, while the reconstruction model remains frozen.

3.2 From Reconstruction to Registration

Pretrained 3D reconstruction models commonly provide camera poses, point maps in a shared reference frame, or 3D point tracks (Wang et al., 2024; 2025a). Our formulation enables registration directly from any of these output types while keeping the reconstruction model frozen.

Camera-Pose Registration. Camera pose provides the most direct route to estimate the rigid transformation. Using the predicted world-to-camera transformations $\mathbf{E}_\mathcal{P}$ and $\mathbf{E}_\mathcal{Q}$ associated with the two color point clouds, we obtain the relative transformation as

$$\widetilde{\mathbf{T}} = \mathbf{E}_\mathcal{Q}\mathbf{E}_\mathcal{P}^{-1}. \tag{3}$$

The predicted rotation can be used directly, whereas the translation inherits the scale ambiguity of the reconstructed scene. Because the reconstruction is defined up to an unknown scale, we align the reconstructed geometry to the metric scale of the input point clouds and rescale the relative translation accordingly. This produces a complete registration estimate directly from the frozen reconstruction model.

Registration via Point-Map and Track Correspondences. Although camera poses provide a direct registration estimate, they represent cross-view geometry only through a single global transformation. Point maps instead retain dense per-point 3D structure and therefore support correspondence-based registration. The predicted point maps $\mathcal{P}_r$ and $\mathcal{Q}_r$ are expressed in a shared reconstruction reference frame. Points with small 3D distances in this frame are treated as candidate cross-view correspondences. We then apply mutual nearest-neighbor matching:

$$\mathcal{C}_{\text{map}} = \text{MNN}(\mathcal{P}_r, \mathcal{Q}_r). \tag{4}$$

Each point map has a one-to-one mapping to its input point cloud. Thus, each match $(u, v) \in \mathcal{C}_{\text{map}}$ gives an input point correspondence $(\mathbf{p}(u), \mathbf{q}(v))$. Predicted 3D tracks provide an alternative source of cross-view correspondences. In both cases, we estimate the rigid transformation from the resulting correspondence set $\mathcal{C}$ by minimizing the alignment error:

$$(\hat{\mathbf{R}}, \hat{\mathbf{t}}) = \underset{\mathbf{R}\in\text{SO}(3),\, \mathbf{t}\in\mathbb{R}^3}{\arg\min} \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{C}} \|\mathbf{R}\mathbf{p} + \mathbf{t} - \mathbf{q}\|_2^2 . \tag{5}$$

Registration in a Shared Reference Frame. Since both predicted point maps share a reference frame, they also enable registration without cross-view matching. We use the one-to-one mapping between each input point cloud and its point map to estimate the transformation into this frame.

Specifically, for input points $\mathbf{p}(u)$ and $\mathbf{q}(v)$ and their corresponding reconstructed points $\mathcal{P}_r(u)$ and $\mathcal{Q}_r(v)$, we estimate

$$\mathcal{P}_r(u) \approx a\,\mathbf{R}_{\mathcal{P}}\mathbf{p}(u) + \mathbf{t}_{\mathcal{P}}, \qquad \mathcal{Q}_r(v) \approx a\,\mathbf{R}_{\mathcal{Q}}\mathbf{q}(v) + \mathbf{t}_{\mathcal{Q}}, \qquad a > 0, \tag{6}$$

where a is the shared reconstruction scale, and $(\mathbf{R}_{\mathcal{P}}, \mathbf{t}_{\mathcal{P}})$ and $(\mathbf{R}_{\mathcal{Q}}, \mathbf{t}_{\mathcal{Q}})$ map the two input point clouds into the shared reconstruction reference frame. We estimate each rotation by applying SVD to the cross-covariance matrix between the centered input points and their reconstructed points. The shared scale is then estimated jointly from both point clouds, and the translations are recovered from their centroids. We represent the resulting transformations into the shared reference frame as

$$\mathbf{A}_{\mathcal{P}} = \begin{bmatrix} a\mathbf{R}_{\mathcal{P}} & \mathbf{t}_{\mathcal{P}} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \qquad \mathbf{A}_{\mathcal{Q}} = \begin{bmatrix} a\mathbf{R}_{\mathcal{Q}} & \mathbf{t}_{\mathcal{Q}} \\ \mathbf{0}^\top & 1 \end{bmatrix}. \tag{7}$$

The relative rigid transformation between the two input point clouds is then obtained by composing these transformations:

$$\widetilde{\mathbf{T}} = \mathbf{A}_{\mathcal{Q}}^{-1}\mathbf{A}_{\mathcal{P}}. \tag{8}$$

Sparse or inaccurate cross-view correspondences can reduce registration robustness. In contrast, we use the known one-to-one mappings to directly align the point-cloud coordinate systems through the shared reference frame, which avoids cross-view matching and its associated error propagation.

3.3 Geometry-guided Correspondence Refinement

Among the above registration routes, correspondence-based registration offers greater potential for precise alignment through explicit cross-view constraints. However, correspondences obtained from 3D reconstruction can be ambiguous, which limits registration accuracy. To address this problem, we propose a learnable geometry-guided correspondence refinement module that selects reliable correspondences according to the geometric consistency of the input point clouds.

Candidate Correspondence Construction. To account for reconstruction errors, we retain multiple candidate correspondences for each input point. For each $\mathbf{p}_i \in \mathcal{P}$, we find the nearest neighbors of its reconstructed point in $\mathcal{Q}_r$ and use the one-to-one mapping to obtain candidates $\mathbf{q}_j \in \mathcal{Q}$, with $j \in \mathcal{N}(i)$. A lightweight network refines distance-based initial scores using color differences, reconstruction confidence, and geometric information, including alignment residuals under the camera-pose transformation. We apply softmax to obtain correspondence probabilities p_{ij}.

Geometric Consistency Refinement. To resolve candidate ambiguities, we exploit the distance-preserving property of rigid transformations. For candidate correspondences (i, j) and (a, b), geometric consistency is quantified by the discrepancy between their within-cloud distances:

$$\delta_{ij,ab} = \big|\|\mathbf{p}_i - \mathbf{p}_a\|_2 - \|\mathbf{q}_j - \mathbf{q}_b\|_2\big|. \tag{9}$$

Each candidate is compared with reference correspondences sampled across $\mathcal{P}$, and their consistency is aggregated according to the current correspondence probabilities. The resulting scores guide candidate selection toward geometrically consistent correspondences. Finally, the highest-probability candidates are selected to recover the rigid transformation through Eq. 5.

Training Objective. We train the refinement module to select accurate and spatially distributed correspondences for rigid registration. Candidate selection is supervised by a cross-entropy loss. The target probability y_{ij} is determined by the alignment error under the ground-truth transformation, with smaller errors receiving higher probabilities:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j} y_{ij}\log p_{ij}. \tag{10}$$

To further improve alignment precision, the geometric loss directly penalizes the expected distance between corresponding points under the ground-truth transformation:

$$\mathcal{L}_{\text{geo}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\in\mathcal{N}(i)} \bar{p}_{ij}\left\|\mathbf{R}_{\text{gt}}\mathbf{p}_i + \mathbf{t}_{\text{gt}} - \mathbf{q}_j\right\|_2. \tag{11}$$

Here, $\bar{p}_{ij}$ denotes the normalized candidate probability. Correspondences concentrated in a small region provide limited spatial constraints for rigid registration. We therefore introduce a coverage loss to encourage correct correspondences across the point cloud. For each of the K spatial regions, c_k denotes the average probability of selecting a correct correspondence:

$$\mathcal{L}_{\text{cov}} = -\frac{1}{K}\sum_{k=1}^{K}\log c_k. \tag{12}$$

The three losses jointly supervise correspondence selection, geometric alignment, and spatial coverage:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_{\text{geo}}\mathcal{L}_{\text{geo}} + \lambda_{\text{cov}}\mathcal{L}_{\text{cov}}. \tag{13}$$

4 Experiments

4.1 Experimental Setup

Datasets. Following the evaluation protocol of ColorPCR (Mu et al., 2024), we evaluate on the 3DMatch (Zeng et al., 2017) and 3DLoMatch (Huang et al., 2021) benchmarks. The learning-based baselines and our geometry-guided correspondence refinement module are trained on 3DMatch only and evaluated on both benchmarks without retraining.

Implementation. We select the following 3D reconstruction foundation models: DUSt3R (Wang et al., 2024), MASt3R (Leroy et al., 2024), Fast3R (Yang et al., 2025), MUSt3R (Cabon et al., 2025), Pow3R (Jang et al., 2025), VGGT (Wang et al., 2025a), CUT3R (Wang et al., 2025b), MapAnything (Keetha et al., 2026), and Pi3 (Wang et al., 2026). Based on these reconstruction outputs, our registration routes include camera-pose registration (Cam.), registration via point-map or track correspondences (MNN or Trk.), and registration in a shared reference frame (SRF). Geometry-guided correspondence refinement (Ref.) further improves correspondence-based registration. Since VGGT provides the predictions required by all registration routes, we use it as the representative 3D reconstruction model unless otherwise specified. Specifically, we retain 16 candidate correspondences per point, perform three iterations of geometric consistency refinement, and set $\lambda_{\text{geo}} = \lambda_{\text{cov}} = 0.1$ in Eq. 13. Further implementation details are provided in Appendix A.

Evaluation metrics. To assess registration performance, we follow PointMBF (Yuan et al., 2023) and report Relative Rotation Error (RRE), Relative Translation Error (RTE), and rotation and translation accuracy. We evaluate rotation accuracy at $5°$, $10°$, and $45°$, and translation accuracy at 5, 10, and 25 cm. We define registration recall (RR) as the proportion of pairs with a mean point-cloud realignment error below 20 cm. Following ColorPCR (Mu et al., 2024), we further evaluate correspondence quality using inlier ratio (IR), the proportion of correspondences with an alignment error below 10 cm, and feature matching recall (FMR), the proportion of pairs with IR above 5%.

Table 1: Pairwise registration on 3DMatch. Gray shading indicates 3DR-Reg; bold denotes the best result in each column. ↑/↓ indicate that higher/lower values are better.

Method	Rotation (deg)					Translation (cm)				
	Accuracy↑			Error↓		Accuracy↑			Error↓	
	5	10	45	Mean	Med.	5	10	25	Mean	Med.
FPFH+RANSAC (Rusu et al., 2009)	93.4	99.0	99.8	2.68	1.94	42.5	83.5	97.6	7.57	5.63
FPFH+Colored ICP (Park et al., 2017)	66.1	67.7	71.3	6.00	0.77	56.7	63.8	67.5	11.44	2.11
GeoTransformer (Qin et al., 2022)	89.2	93.4	96.4	7.53	1.30	69.3	84.9	92.6	18.34	3.05
LLT (Wang et al., 2022)	95.0	98.0	99.5	1.97	1.42	77.2	91.5	97.9	5.21	3.86
PointMBF (Yuan et al., 2023)	98.2	99.4	99.8	1.20	0.70	86.1	95.8	99.3	3.38	2.09
ColorPCR (Mu et al., 2024)	96.7	98.5	99.0	3.33	1.39	66.8	91.5	98.2	8.89	3.46
DINOReg (Chen & Qu, 2025)	89.6	97.6	99.4	3.35	2.15	34.0	66.9	89.3	13.42	6.80
3DR-Reg (Ref.)	**99.8**	**99.9**	**100.0**	**0.69**	**0.53**	**93.6**	**99.2**	**100.0**	**2.14**	**1.64**

Table 2: Pairwise registration on 3DLoMatch.

Method	Rotation (deg)					Translation (cm)				
	Accuracy↑			Error↓		Accuracy↑			Error↓	
	5	10	45	Mean	Med.	5	10	25	Mean	Med.
FPFH+RANSAC (Rusu et al., 2009)	16.3	30.0	43.1	77.09	84.90	2.0	6.7	18.7	187.80	161.62
FPFH+Colored ICP (Park et al., 2017)	13.5	19.8	32.4	82.43	90.08	3.3	6.6	13.9	201.67	177.17
GeoTransformer (Qin et al., 2022)	8.5	13.8	23.7	104.00	108.42	2.9	6.3	12.4	242.50	232.77
LLT (Wang et al., 2022)	2.7	4.7	22.9	92.09	90.71	0.4	1.3	3.8	205.87	194.97
PointMBF (Yuan et al., 2023)	5.3	8.5	27.0	87.91	84.75	1.0	2.8	6.7	209.40	202.19
ColorPCR (Mu et al., 2024)	26.4	40.0	53.0	65.76	26.54	5.2	14.1	30.4	177.78	97.58
DINOReg (Chen & Qu, 2025)	60.4	81.0	**91.4**	14.04	3.97	14.2	38.0	67.5	45.23	14.02
3DR-Reg (Ref.)	**73.0**	**82.2**	86.0	**3.77**	**2.50**	**25.7**	**53.1**	**78.7**	**13.15**	**7.83**

4.2 Comparison with State-of-the-Art Methods

Quantitative Comparison. Tables 1 and 2 present quantitative comparisons with representative geometric and color point cloud registration methods. On 3DMatch, our method achieves state-of-the-art performance across all evaluation metrics. We further evaluate registration performance on low-overlap point cloud pairs in 3DLoMatch. While most baselines suffer severe performance degradation, our method achieves the lowest rotation and translation errors and superior registration accuracy at strict thresholds. These results demonstrate high accuracy and robustness under challenging low-overlap conditions and highlight the effectiveness of using pretrained 3D reconstruction for accurate color point cloud registration. Additional comparisons of correspondence quality, registration recall, and joint pose accuracy are provided in Appendices B and C.

Parameter Efficiency. In Table 3, we compare total and trainable parameter counts across different methods. By reusing frozen reconstruction models, our direct registration routes require no additional training, while correspondence refinement introduces only 0.78 M trainable parameters, substantially fewer than the learning-based baselines. This allows 3DR-Reg to draw on the capacity of large pretrained models with only a lightweight module to learn for registration.

Table 3: Comparison of parameters.

Method	Total	Trainable
GeoTransformer (Qin et al., 2022)	9.83M	9.83M
LLT (Wang et al., 2022)	5.08M	5.08M
PointMBF (Yuan et al., 2023)	6.19M	6.19M
ColorPCR (Mu et al., 2024)	10.03M	10.03M
DINOReg (Chen & Qu, 2025)	33.82M	11.77M
3DR-Reg		
Cam./MNN/Trk./SRF	**0.4–1.3B**	**0M**
Ref.	**0.4–1.3B**	**0.78M**

4.3 Benefiting from Advances in 3D Reconstruction

To better understand the contribution of advances in 3D reconstruction, we evaluate our formulation across nine frozen reconstruction models. As shown in Table 4, our registration performance

Table 4: Registration across frozen reconstruction models. Within each registration group, bold and underlined values denote the best and second-best results, respectively.

Reconstruction model	3DMatch			3DLoMatch		
	RR↑ (%)	RRE↓ (°)	RTE↓ (cm)	RR↑ (%)	RRE↓ (°)	RTE↓ (cm)
Camera-Pose Registration						
VGGT (Wang et al., 2025a)	<u>99.71</u>	<u>0.81</u>	<u>2.99</u>	76.43	<u>3.90</u>	<u>19.73</u>
CUT3R (Wang et al., 2025b)	99.20	0.97	3.87	<u>76.96</u>	4.19	21.30
MapAnything (Keetha et al., 2026)	97.45	1.02	4.20	59.31	**3.69**	20.82
Pi3 (Wang et al., 2026)	**100.00**	**0.78**	**2.69**	**86.60**	4.33	**19.28**
Registration via Point-Map Correspondences						
DUSt3R (224 linear) (Wang et al., 2024)	99.92	0.87	2.64	78.88	7.05	25.20
DUSt3R (512 linear) (Wang et al., 2024)	**99.98**	0.81	2.54	84.29	4.54	18.15
DUSt3R (512 DPT) (Wang et al., 2024)	<u>99.96</u>	0.81	2.54	83.90	4.48	18.69
MASt3R (Leroy et al., 2024)	99.94	<u>0.80</u>	<u>2.51</u>	**85.56**	**3.93**	**15.73**
Fast3R (Yang et al., 2025)	99.80	0.84	2.60	70.99	14.38	43.87
MUSt3R (Cabon et al., 2025)	<u>99.96</u>	0.87	2.84	84.47	4.40	<u>16.83</u>
Pow3R (Jang et al., 2025)	99.92	<u>0.80</u>	**2.50**	83.62	4.46	18.54
VGGT (Wang et al., 2025a)	<u>99.96</u>	0.81	<u>2.51</u>	83.89	4.61	18.00
CUT3R (Wang et al., 2025b)	99.62	1.14	3.68	79.72	5.29	21.48
MapAnything (Keetha et al., 2026)	99.82	0.97	2.97	82.90	4.43	18.79
Pi3 (Wang et al., 2026)	<u>99.96</u>	**0.77**	**2.50**	<u>85.09</u>	<u>4.10</u>	17.11
Registration in a Shared Reference Frame						
DUSt3R (224 linear) (Wang et al., 2024)	99.86	1.21	4.08	80.45	7.45	29.62
DUSt3R (512 linear) (Wang et al., 2024)	99.92	<u>1.03</u>	<u>3.23</u>	88.43	5.11	20.71
DUSt3R (512 DPT) (Wang et al., 2024)	<u>99.96</u>	1.04	3.28	88.73	4.61	19.59
MASt3R (Leroy et al., 2024)	99.93	1.07	3.49	90.03	4.64	18.65
Fast3R (Yang et al., 2025)	99.63	1.38	4.26	70.75	15.83	50.71
MUSt3R (Cabon et al., 2025)	99.92	1.31	4.62	88.92	4.86	<u>18.07</u>
Pow3R (Jang et al., 2025)	99.83	1.14	3.64	89.13	5.08	22.15
VGGT (Wang et al., 2025a)	99.93	1.16	3.64	**91.28**	4.39	18.15
CUT3R (Wang et al., 2025b)	99.44	1.67	6.08	86.10	4.59	20.16
MapAnything (Keetha et al., 2026)	99.70	1.41	4.31	88.24	**4.18**	**16.96**
Pi3 (Wang et al., 2026)	**99.98**	**0.93**	**3.12**	<u>91.03</u>	<u>4.38</u>	19.78

generally improves with more expressive reconstruction architectures and stronger geometric representations. With the more recent reconstruction model Pi3, our formulation achieves the best or second-best performance on nearly all metrics across both benchmarks. These results show that our formulation can benefit from advances in 3D reconstruction without redesigning the registration procedure.

4.4 Qualitative Results

Figure 3 presents qualitative comparisons between our method and representative baselines. As overlap decreases, our method maintains consistently low registration errors, whereas PointMBF (Yuan et al., 2023) and ColorPCR (Mu et al., 2024) suffer substantial accuracy degradation, demonstrating high accuracy and robustness owing to the rich geometric priors of pretrained 3D reconstruction models.

4.5 Ablation Studies

Correspondence Refinement. To assess the effectiveness of correspondence refinement, we compare Ref. with correspondences obtained directly from 3D point tracks (Trk.) or through mutual nearest-neighbor matching of predicted point maps (MNN). As shown in Table 5, both direct routes provide reliable correspondences on 3DMatch, but their correspondence quality deteriorates under low overlap. Our refinement module achieves higher inlier ratios on both benchmarks, with more

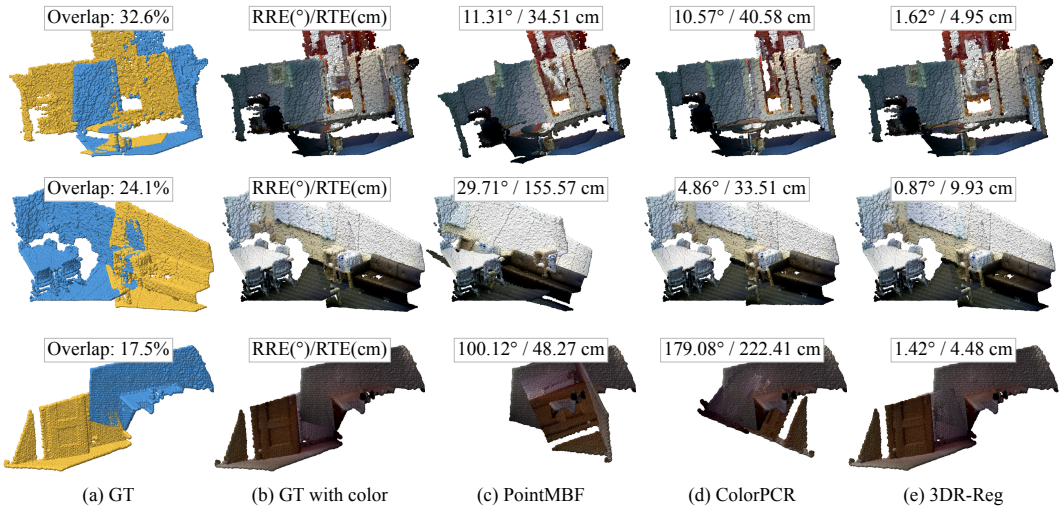


Figure 3: Qualitative comparison of color point cloud registration results.

Table 5: Evaluation of correspondence refinement against direct correspondences and component ablations on 3DMatch and 3DLoMatch.

Configuration	3DMatch			3DLoMatch		
	IR↑ (%)	FMR↑ (%)	RR↑ (%)	IR↑ (%)	FMR↑ (%)	RR↑ (%)
MNN	94.7	**99.8**	**100.0**	50.1	78.8	83.9
Trk.	88.7	99.7	99.9	30.6	72.4	81.8
w/o geometric consistency	99.3	**99.8**	**100.0**	64.7	78.6	90.7
w/o camera pose information	99.3	99.7	99.9	57.7	65.5	90.8
Ref.	**99.4**	**99.8**	**100.0**	**67.2**	**83.7**	**91.0**

pronounced improvements on 3DLoMatch. These improvements are accompanied by higher FMR and RR under low overlap, demonstrating the importance of refining reconstruction correspondences for reliable registration.

Refinement Components. We further investigate the contributions of geometric consistency and camera pose information by removing each from the refinement module. Removing either reduces correspondence quality, while omitting camera pose information leads to a larger decrease in IR and FMR on 3DLoMatch. Geometric consistency evaluates compatibility between candidate correspondences, whereas camera pose information constrains their alignment. Incorporating both produces the most reliable correspondences, supporting their complementary roles in correspondence refinement.

5 Conclusion

In this work, we propose 3DR-Reg, which uses pretrained 3D reconstruction models for color point cloud registration. Different from methods that learn registration-specific representations, our method uses predictions from frozen reconstruction models for camera-pose registration, correspondence-based registration, and alignment in a shared reference frame. Experiments on 3DMatch and 3DLoMatch demonstrate state-of-the-art registration accuracy and applicability across multiple 3D reconstruction models. Direct registration routes require no additional task-specific training, while lightweight geometry-guided refinement further improves correspondence precision. Furthermore, we believe future reconstruction advances can benefit registration through this formulation.

AI USE STATEMENT

Generative AI tools were used to correct grammatical errors, improve the wording and readability of the manuscript, and assist with manuscript formatting and compilation checks. All AI-assisted content was reviewed by the authors, who take full responsibility for the manuscript, including its claims and references.

ETHICS STATEMENT

This work studies color point cloud registration using existing public benchmarks and released pre-trained models. It does not involve new data collection or experiments with human participants. Applications involving captured indoor environments should respect privacy and the permitted uses of the underlying data.

REPRODUCIBILITY STATEMENT

To support reproducibility, Section 3 describes the proposed method and training objective, and Section 4.1 specifies the datasets, reconstruction models, and evaluation metrics. Appendix A provides additional details on preprocessing, registration, network architecture, and training. Anonymized source code is provided in the supplementary material.

REFERENCES

Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3d local features. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6358–6366. IEEE, 2020.

Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. Pointdsc: Robust point cloud registration using deep spatial consistency. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15854–15864. IEEE, 2021.

Yohann Cabon, Lucas Stoffl, Leonid Antsfeld, Gabriela Csurka, Boris Chidlovskii, Jerome Revaud, and Vincent Leroy. Must3r: Multi-view network for stereo 3d reconstruction. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1050–1060. IEEE, 2025.

Congjia Chen and Yufu Qu. Dinoreg: Strong point cloud registration with vision foundation model. *arXiv preprint arXiv:2509.24370*, 2025.

Suyi Chen, Hao Xu, Haipeng Li, Kunming Luo, Guanghui Liu, Chi-Wing Fu, Ping Tan, and Shuaicheng Liu. Pointreggpt: Boosting 3d point cloud registration using generative point-cloud pairs for training. In *European Conference on Computer Vision*, pp. 272–289. Springer, 2024.

Zhi Chen, Kun Sun, Fan Yang, and Wenbing Tao. Sc 2-pcr: A second order spatial compatibility for efficient and robust point cloud registration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13211–13221. IEEE, 2022.

Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8957–8965. IEEE, 2019.

Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2511–2520. IEEE, 2020.

Mohamed El Banani and Justin Johnson. Bootstrap your own correspondences. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6413–6422. IEEE, 2021.

Mohamed El Banani, Luya Gao, and Justin Johnson. Unsupervisedr&r: Unsupervised point cloud registration via differentiable rendering. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7125–7135. IEEE, 2021.

Mohamed El Banani, Ignacio Rocco, David Novotny, Andrea Vedaldi, Natalia Neverova, Justin Johnson, and Ben Graham. Self-supervised correspondence estimation via multiview registration. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1216–1225. IEEE, 2023.

Zan Gojcic, Caifa Zhou, Jan D Wegner, Leonidas J Guibas, and Tolga Birdal. Learning multiview 3d point cloud registration. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1756–1766. IEEE, 2020.

Yuval Haitman, Amit Efraim, and Joseph M. Francos. C-genreg: Training-free 3d point cloud registration by multi-view-consistent geometry-to-image generation with probabilistic modalities fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3004–3013, June 2026.

Po-Han Huang, Kevin Matzen, Johannes Kopf, Narendra Ahuja, and Jia-Bin Huang. Deepmvs: Learning multi-view stereopsis. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2821–2830. IEEE, 2018.

Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 4265–4274. IEEE, 2021.

Wonbong Jang, Philippe Weinzaepfel, Vincent Leroy, Lourdes Agapito, and Jerome Revaud. Pow3r: Empowering unconstrained 3d reconstruction with camera and scene priors. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1071–1081. IEEE, 2025.

Haobo Jiang, Jin Xie, Jian Yang, Liang Yu, and Jianmin Zheng. Zero-shot rgb-d point cloud registration with pre-trained large vision model. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16943–16952. IEEE, 2025a.

Haobo Jiang, Jin Xie, Jianmin Zheng, et al. Generative point cloud registration. In *Forty-second International Conference on Machine Learning*, 2025b.

Nikhil Keetha, Norman Müller, Johannes Schönberger, Lorenzo Porzi, Yuchen Zhang, Tobias Fischer, Arno Knapitsch, Duncan Zauss, Ethan Weber, Nelson Antunes, et al. Mapanything: Universal feed-forward metric 3d reconstruction. In *2026 International Conference on 3D Vision (3DV)*, pp. 499–509. IEEE, 2026.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, July 2023. ISSN 1557-7368. doi: 10.1145/3592433. URL `http://dx.doi.org/10.1145/3592433`.

Vincent Leroy, Yohann Cabon, and Jérôme Revaud. Grounding image matching in 3d with mast3r. In *European conference on computer vision*, pp. 71–91. Springer, 2024.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Juncheng Mu, Lin Bie, Shaoyi Du, and Yue Gao. Colorpcr: Color point cloud registration with multi-stage geometric-color fusion. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21061–21070. IEEE, 2024.

Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Colored point cloud registration revisited. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 143–152. IEEE, 2017.

Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11143–11152, 2022.

Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *2009 IEEE international conference on robotics and automation*, pp. 3212–3217. IEEE, 2009.

Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4104–4113, 2016.

Yejun Shou, Haocheng Wang, Lingfeng Shen, Qian Zheng, Gang Pan, and Yanlong Cao. Unsupervised rgb-d point cloud registration for scenes with low overlap and photometric inconsistency. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 24868–24877. IEEE, 2025.

Edgar Sucar, Zihang Lai, Eldar Insafutdinov, and Andrea Vedaldi. Dynamic point maps: A versatile representation for dynamic 3d reconstruction. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7295–7305. IEEE, 2025.

Jiaming Sun, Yiming Xie, Linghao Chen, Xiaowei Zhou, and Hujun Bao. Neuralrecon: Real-time coherent 3d reconstruction from monocular video. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15593–15602. IEEE, 2021.

Niclas Vödisch, Giovanni Cioffi, Marco Cannici, Wolfram Burgard, and Davide Scaramuzza. Lidar registration with visual foundation models. *arXiv preprint arXiv:2502.19374*, 2025.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5294–5306. IEEE, 2025a.

Qianqian Wang, Yifei Zhang, Aleksander Holynski, Alexei A Efros, and Angjoo Kanazawa. Continuous 3d perception model with persistent state. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10510–10522. IEEE, 2025b.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20697–20709. IEEE, 2024.

Yifan Wang, Jianjun Zhou, Haoyi Zhu, Wenzheng Chang, Yang Zhou, Zizun Li, Junyi Chen, Jiangmiao Pang, Chunhua Shen, and Tong He. π^3: Permutation-equivariant visual geometry learning. In *International Conference on Learning Representations*, volume 2026, pp. 10481–10497, 2026.

Ziming Wang, Xiaoliang Huo, Zhenghao Chen, Jing Zhang, Lu Sheng, and Dong Xu. Improving rgb-d point cloud registration by learning multi-scale local linear transformation. In *European Conference on Computer Vision*, pp. 175–191. Springer, 2022.

Qianliang Wu, Haobo Jiang, Lei Luo, Jun Li, Yaqing Ding, Jin Xie, and Jian Yang. Diff-reg: Diffusion model in doubly stochastic matrix space for registration problem. In *European Conference on Computer Vision*, pp. 160–178. Springer, 2024.

Zongyi Xu, Xinqi Jiang, Xinyu Gao, Rui Gao, Changjun Gu, Qianni Zhang, Weisheng Li, and Xinbo Gao. Igreg: Image-geometry-assisted point cloud registration via selective correlation fusion. *IEEE Transactions on Multimedia*, 26:7475–7489, 2024.

Shaocheng Yan, Pengcheng Shi, Zhenjun Zhao, Kaixin Wang, Kuang Cao, Ji Wu, and Jiayuan Li. Turboreg: Turboclique for robust and efficient point cloud registration. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26371–26381. IEEE, 2025a.

Shaocheng Yan, Yiming Wang, Kaiyan Zhao, Pengcheng Shi, Zhenjun Zhao, Yongjun Zhang, and Jiayuan Li. Hemora: Unsupervised heuristic consensus sampling for robust point cloud registration. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1363–1373. IEEE, 2025b.

Jianing Yang, Alexander Sax, Kevin J Liang, Mikael Henaff, Hao Tang, Ang Cao, Joyce Chai, Franziska Meier, and Matt Feiszli. Fast3r: Towards 3d reconstruction of 1000+ images in one forward pass. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21924–21935. IEEE, 2025.

Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. Mvsnet: Depth inference for unstructured multi-view stereo. In *European conference on computer vision*, pp. 785–801. Springer, 2018.

Zi Jian Yew and Gim Hee Lee. Regtr: End-to-end point cloud correspondences with transformers. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6667–6676. IEEE, 2022.

Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34:23872–23884, 2021.

Junle Yu, Luwei Ren, Yu Zhang, Wenhui Zhou, Lili Lin, and Guojun Dai. Peal: Prior-embedded explicit attention learning for low-overlap point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17702–17711, 2023.

Mingzhi Yuan, Kexue Fu, Zhihao Li, Yucong Meng, and Manning Wang. Pointmbf: A multi-scale bidirectional fusion network for unsupervised rgb-d point cloud registration. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17648–17659. IEEE, 2023.

Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 199–208. IEEE, 2017.

Yifei Zhang, Hao Zhao, Hongyang Li, and Siheng Chen. Fastmac: Stochastic spectral sampling of correspondence graph. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17857–17867. IEEE, 2024.

APPENDIX

A ADDITIONAL IMPLEMENTATION DETAILS

Reconstruction and Registration. We use the released reconstruction models with their original preprocessing and pretrained weights. Preprocessing resizes and crops color information in its acquisition pixel layout while preserving the one-to-one mapping between each input point cloud and its predicted point map. Point maps are expressed in a shared reference frame; predictions in separate camera frames are transformed using the corresponding camera poses. VGGT supplies all three output types used by our registration routes. We also use the native camera predictions of CUT3R, MapAnything, and Pi3 for Cam., and evaluate MNN and SRF across models that provide point maps. For Cam., we recover translation scale from the median ratio between input and predicted depths across both views. For SRF, we fit the input point clouds to their point maps with a shared scale using SVD and ten iterations of Huber reweighting with a 5 cm threshold. MNN retains mutual nearest neighbors within 10 cm after scale conversion. Trk. uses a 16×12 query grid and retains tracks with visibility above 0.5. Correspondence-based routes use RANSAC with a 5 cm inlier threshold and at most 4,096 hypotheses, followed by SVD on the inliers.

Correspondence Refinement. We sample up to 2,048 points from $\mathcal{P}$ and retain 16 nearest neighbors in the predicted point maps for each point. A point is included when its nearest candidate is within 15 cm after scale conversion. Candidate scores are initialized with a Gaussian distance prior of bandwidth 3 cm. Each candidate is represented by a 17-dimensional feature vector that describes reconstructed-point distances and offsets, color and pixel-coordinate differences, reconstruction confidence, geometric consistency, and alignment residuals under the camera-pose transformation. We encode these features into 192 dimensions and use four residual MLP blocks to combine candidate features with their averages over candidates and input points. Reference points are selected by farthest point sampling, with 256 points for initial geometric consistency features and 128 for the three geometric consistency refinement steps. During refinement, each reference point contributes its two highest-probability candidates to consistency evaluation. These comparisons use a 5 cm distance threshold and Gaussian bandwidth.

Training Details. We train the refinement module on 7,652 valid training pairs from 3DMatch and use 1,510 pairs for validation. The reconstruction model remains frozen throughout training. We use AdamW with an initial learning rate of 3×10^{-4}, weight decay of 10^{-4}, an effective batch size of eight, and cosine learning-rate decay over 80 epochs. All configurations are evaluated on both benchmarks without retraining. A candidate is considered correct when its ground-truth alignment error is below 5 cm. Classification includes a rejection target for points without correct candidates, while geometric and coverage supervision uses only points with correct candidates and their spatial regions. The geometric term normalizes probabilities over correspondence candidates alone. In the geometric loss, we cap alignment errors at 20 cm and normalize them by 5 cm. Target correspondence probabilities are obtained by normalizing Gaussian scores with a 1 cm bandwidth, based on each candidate's alignment error relative to the minimum candidate error. Spatial coverage uses eight regions defined by the coordinate-wise median of sampled points in each $\mathcal{P}$.

B CORRESPONDENCE QUALITY COMPARISON

To further assess the quality of reconstruction correspondences, we compare IR, FMR, and RR with existing methods. As shown in Table 6, MNN and Trk. provide reliable correspondences, while Ref. achieves the highest IR on both benchmarks and the highest RR on 3DLoMatch. Higher IR does not always yield higher FMR: correspondence precision within a pair and the proportion of pairs above the inlier threshold describe different aspects of matching quality. The improvements over direct correspondences support the effectiveness of refinement for reliable registration under low overlap.

We use the metric thresholds defined in Section 4.1. RRE and RTE are summarized over pairs with valid pose estimates, while rotation and translation accuracies use the full test set. IR is averaged over pairs with available correspondences, and FMR uses the full test set, with unavailable correspondences counted as failures. RR is computed over pairs with valid transformations and evaluable point-cloud alignment errors, then averaged equally across scenes.

Table 6: Correspondence quality and registration recall on 3DMatch and 3DLoMatch.

Method	3DMatch			3DLoMatch		
	IR↑ (%)	FMR↑ (%)	RR↑ (%)	IR↑ (%)	FMR↑ (%)	RR↑ (%)
FPFH+RANSAC (Rusu et al., 2009)	33.8	97.3	97.5	2.0	10.4	19.5
GeoTransformer (Qin et al., 2022)	48.8	96.6	82.2	3.8	20.7	10.4
LLT (Wang et al., 2022)	91.1	99.6	96.0	6.7	27.6	3.4
PointMBF (Yuan et al., 2023)	94.1	99.7	98.8	8.2	32.4	7.9
ColorPCR (Mu et al., 2024)	71.7	99.6	98.1	11.7	44.0	30.0
DINOReg (Chen & Qu, 2025)	53.6	99.5	96.2	29.8	**91.9**	72.9
3DR-Reg (Trk.)	88.7	99.7	99.9	30.6	72.4	81.8
3DR-Reg (MNN)	94.7	**99.8**	**100.0**	50.1	78.8	83.9
3DR-Reg (Ref.)	**99.4**	**99.8**	**100.0**	**67.2**	83.7	**91.0**

C JOINT POSE ACCURACY

To further evaluate registration accuracy, we measure the proportion of point cloud pairs with both $\mathrm{RRE} < 5^\circ$ and $\mathrm{RTE} < 10\,\mathrm{cm}$ over the full test set. As shown in Table 7, Ref. achieves the highest joint pose accuracy on both benchmarks, including the challenging low-overlap setting. These results complement the separate rotation and translation evaluations and demonstrate the effectiveness of pretrained 3D reconstruction for precise point cloud alignment.

Table 7: Joint pose accuracy on 3DMatch and 3DLoMatch, requiring $\mathrm{RRE} < 5^\circ$ and $\mathrm{RTE} < 10\,\mathrm{cm}$.

Method	3DMatch (%)↑	3DLoMatch (%)↑
ColorPCR (Mu et al., 2024)	91.07	11.74
DINOReg (Chen & Qu, 2025)	66.35	35.70
PointMBF (Yuan et al., 2023)	95.52	2.70
3DR-Reg (Cam.)	97.86	37.18
3DR-Reg (MNN)	98.37	45.43
3DR-Reg (Ref.)	**99.17**	**51.19**