HybridReg: Robust 3D Point Cloud Registration with Hybrid Motions

Keyu Du[1*], **Hao Xu**[2,3*], **Haipeng Li**[1,4], **Hong Qu**[1], **Chi-Wing Fu**[2,3], **Shuaicheng Liu**[1,4†]
[1]University of Electronic Science and Technology of China
[2]Department of Computer Science and Engineering, CUHK
[3]Institute of Medical Intelligence and XR, CUHK
[4]Megvii Technology
{dukeyu@std., lihaipeng@std., hongqu@, liushuaicheng@}uestc.edu.cn, {xuhao, cwfu}@cse.cuhk.edu.hk

Abstract

Scene-level point cloud registration is very challenging when considering dynamic foregrounds. Existing indoor datasets mostly assume rigid motions, so the trained models cannot robustly handle scenes with non-rigid motions. On the other hand, non-rigid datasets are mainly object-level, so the trained models cannot generalize well to complex scenes. This paper presents HybridReg, a new approach to 3D point cloud registration, learning uncertainty mask to account for hybrid motions: rigid for backgrounds and non-rigid/rigid for instance-level foregrounds. First, we build a scene-level 3D registration dataset, namely HybridMatch, designed specifically with strategies to arrange diverse deforming foregrounds in a controllable manner. Second, we account for different motion types and formulate a mask-learning module to alleviate the interference of deforming outliers. Third, we exploit a simple yet effective negative log-likelihood loss to adopt uncertainty to guide the feature extraction and correlation computation. To our best knowledge, HybridReg is the first work that exploits hybrid motions for robust point cloud registration. Extensive experiments show HybridReg's strengths, leading it to achieve state-of-the-art performance on both widely-used indoor and outdoor datasets.

Code — https://github.com/hxwork/HybridReg_PyTorch

Introduction

3D point cloud registration is vital for diverse applications, attracting incredible attention from industry and academia. Recent data-driven methods have shown remarkable success, particularly in indoor scenes (Bai et al. 2020; Huang et al. 2021; Yu et al. 2021), thanks to large-scale datasets like 3DMatch (Zeng et al. 2017), SIRA-PCR (Chen et al. 2023a), and PointRegGPT (Chen et al. 2024).

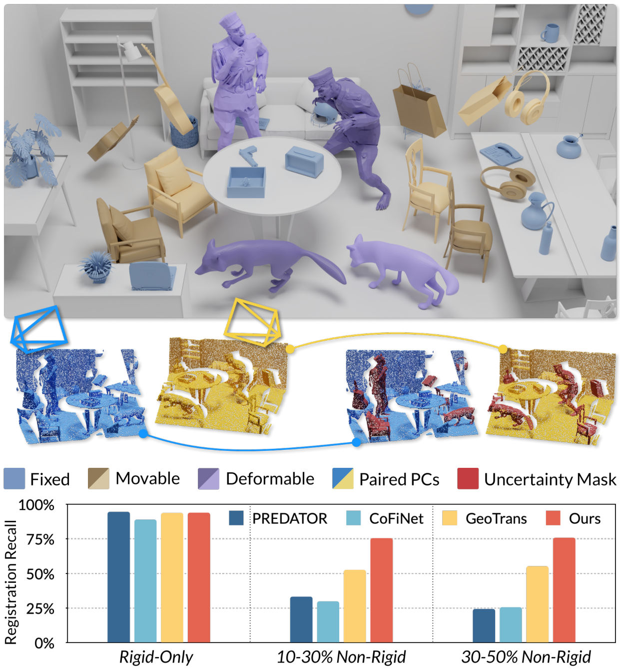

Figure 1: We build HybridMatch, the first large-scale indoor synthetic dataset with hybrid motions for 3D point cloud registration (top). To improve the robustness, we further design HybridReg, integrating probabilistic uncertainty masks into correspondence estimation (middle), achieving significant boosts (below) in challenging hybrid-motion scenes.

Typically, the input pair of point clouds are partially overlapping, with (i) rigid background motions due to the camera motion, and (ii) non-rigid foreground motion due to moving or deforming objects. To align the background, algorithms should eliminate the foreground interference and determine a 3D rigid transformation. However, most existing indoor datasets assume rigid motions in the whole scene, *neglecting* common non-rigid foreground motions. So, the first challenge is the lack of a comprehensive benchmark for model training and evaluation on scenes with hybrid motions.

Though directly training on real-captured datasets can alleviate this issue, data collection is highly time-consuming, and data annotation is expensive and error-prone. An appealing alternative to overcome the data scarcity is to leverage simulated data, *e.g.*, Lepard (Li and Harada 2022) introduces 4DMatch, an object-level synthetic dataset for non-rigid point cloud registration. Yet, models trained on object-level data have proven to struggle to generalize to complex indoor scenes (Qin et al. 2022; Huang et al. 2021), as ob-

---

[*] These authors contributed equally.

[†] Corresponding author.

jects are simpler and have strong shape priors. To sum up, the second challenge is how to design a scene-level benchmark for 3D point cloud registration with hybrid motions, to efficiently account for both data collection and annotations.

Further, even with such appropriate data, algorithms still need to handle outliers in non-overlapping regions, caused not only by different captured views but also by the dynamic foregrounds. While powerful feature descriptors help address the former, they lack specific designs to handle the latter, as foreground points with similar local geometric structures may be outliers due to inconsistent motions with the background. Directly applying prior methods to hybrid-motion scenes leads to heavy performance degradation. Hence, the third challenge is how to integrate feature extraction and correspondence matching with precise guidance to account for hybrid motions.

In this work, to address the first two challenges, we draw inspiration from FlowNet (Dosovitskiy et al. 2015) and SIRA-PCR (Chen et al. 2023a) to create the first large-scale indoor synthetic dataset with hybrid motions, named HybridMatch. Specifically, beyond the rigid backgrounds provided by 3D-FRONT (Fu et al. 2021a), we simulate non-rigid motions by applying instance-level rigid motions to objects from ShapeNet (Chang et al. 2015) on the surfaces of some furniture and in mid-air, combined with deforming objects from DeformingThings4D (Li et al. 2021).

For the third challenge, training existing models on simulated point cloud pairs proves insufficient due to the increased difficulty in correspondence matching caused by hybrid motions. As shown in Fig. 1, their performances degrade as the ratio of non-rigid motion increases. To address it, inspired by the 2D motion estimation method (Truong et al. 2023), we propose a probabilistic model named **HybridReg** to learn uncertainty masks that address **Hybrid** motions in robust 3D point cloud **Reg**istration. Rather than deterministically predicting correspondences, our reformulated probabilistic model computes the correspondence relationships while simultaneously learning a robust confidence map to model correspondence-wise uncertainty. This map indicates the reliability of the correspondence predictions, which is crucial for accurately solving the 3D transformation. To facilitate training, we assume the output follows a Laplace distribution and employ a simple yet effective negative log-likelihood (NLL) loss to incorporate probabilistic uncertainty into the loss function, enhancing robust correspondence matching. Our main contributions are three-fold:

- We construct HybridMatch, the first large-scale indoor synthetic dataset with hybrid and diverse motions for 3D point cloud registration.
- We design HybridReg, a probabilistic model to learn uncertainty masks, guided by an effective negative log-likelihood loss, to account for hybrid motions and enhance feature extraction and correspondence matching.
- Extensive qualitative and quantitative comparisons on both widely-used indoor and outdoor datasets demonstrate the state-of-the-art performance of our approach.

Related Work

Rigid 3D point cloud registration. Most approaches are correspondence-based, evolved from traditional handcrafted features (Besl and McKay 1992; Rusinkiewicz and Levoy 2001; Yang, Li, and Jia 2013; Rusinkiewicz 2019; Rusu, Blodow, and Beetz 2009) and RANSAC (Fischler and Bolles 1981) to deep-learning-based feature descriptors (Choy, Park, and Koltun 2019; Gojcic et al. 2019; Bai et al. 2020; Ao et al. 2021; Poiesi and Boscaini 2022; Wang et al. 2022; Yu et al. 2021; Yew and Lee 2022; Wang et al. 2023; Yang et al. 2022; Li and Harada 2022; Ao et al. 2023; Yu et al. 2023; Mei et al. 2023; Liu et al. 2023b; Chen et al. 2023a; Liu et al. 2023a), including weighted SVD for efficient transformation derivation (Wang and Solomon 2019a,b; Li et al. 2020a; Yew and Lee 2020; Fu et al. 2021b) and deep robust estimators for transformation accuracy (Bai et al. 2021; Choy, Dong, and Koltun 2020; Pais et al. 2020; Jiang et al. 2023; Zhang et al. 2023; Hatem, Qian, and Wang 2023). Others are correspondence-free, directly regressing transformations, *e.g.*, (Aoki et al. 2019; Xu et al. 2021, 2022), yet they struggle with scalability and generalization when handling complex scenes (Qin et al. 2022). Existing methods are mainly designed for (i) indoor scenarios, such as 3DMatch (Zeng et al. 2017), 7Scenes (Shotton et al. 2013), SIRA-PCR (Chen et al. 2023b), and PointRegGPR (Chen et al. 2024), or (ii) object-level scenarios, *e.g.*, ModelNet40 (Chang et al. 2015), Stanford 3D Scan (Curless and Levoy 1996), and AutoSynth (Dang and Salzmann 2023). All of them, however, consider only rigid motions, ignoring the commonly existing non-rigid motions in practice, *e.g.*, moving humans/objects. In this work, we create the first large-scale indoor synthetic dataset with hybrid motions, enabling training and testing to effectively account for deformable foregrounds.

Non-rigid correspondence estimation is a major topic in dynamic reconstruction (Bozic et al. 2020b; Gao and Tedrake 2018; Newcombe, Fox, and Seitz 2015) and geometry processing (Huang et al. 2008; Groueix et al. 2018; Ovsjanikov et al. 2012), focusing on estimating non-rigid correspondences from deformable point clouds or manifold surfaces. DynamicFusion (Newcombe, Fox, and Seitz 2015) employs projective correspondence for real-time efficiency and is further utilized for dense correspondence learning in (Schmidt, Newcombe, and Fox 2016). VolumeDeform (Innmann et al. 2016) adopts SIFT (Lowe 2004) for robust non-rigid tracking. DeepDeform (Bozic et al. 2020b) learns sparse global correspondence for patches in deforming RGB-D sequences. (Li et al. 2020b) learn non-rigid features via a differentiable non-rigid alignment optimization. NNRT (Bozic et al. 2020a) estimates correspondence end-to-end with an outlier rejection network. Besides, scene flow estimation, *e.g.*, (Li, Kaesemodel Pontes, and Lucey 2021; Liu, Qi, and Guibas 2019; Wu et al. 2020), is a closely-related technique that delivers point-wise correspondence. The key difference between our task and these works is that non-rigid correspondences on moving or deforming foreground objects are outliers and should be filtered out for accurate background alignment in our scene-level task.

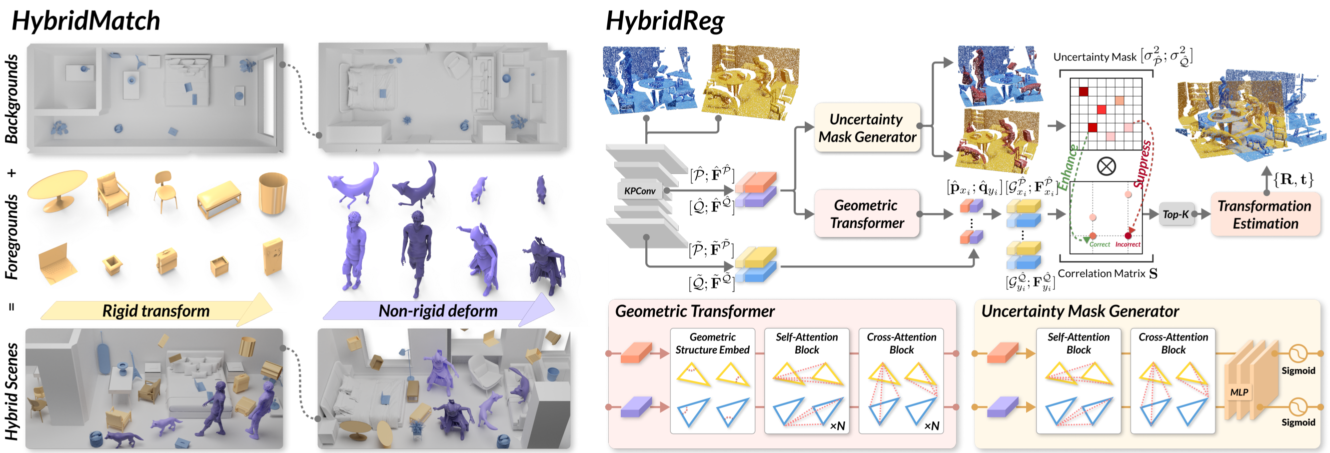

Figure 2: Our framework: (i) HybridMatch is a large-scale indoor synthetic dataset with hybrid motions for point cloud registration, containing diverse movable and deformable foreground objects; and (ii) HybridReg is designed to estimate uncertainty mask for enhancing discriminative feature extraction and accurate correspondence matching against hybrid motions.

Method

Overview

Fig. 2 gives an overview of our approach, which includes (i) HybridMatch, a large-scale indoor synthetic dataset with hybrid motions for 3D point cloud registration (Sec.); (ii) HybridReg, the method we designed to estimate probabilistic uncertainty masks and evaluate reliability and accuracy of the correspondence prediction against hybrid motions (Sec.); and (iii) our loss function design (Sec.).

HybridMatch Dataset

Existing indoor datasets typically contain only static scenes, yet dynamic scenes are common in the real world. To better simulate real-world environments, our HybridMatch includes not only rigid background motions but also a diverse range of rigid and non-rigid foreground motions for objects within the background scene, such as humans moving around, animals wandering, and objects or furniture shifting. This makes HybridMatch more representative of the complexities found in everyday indoor environments. To achieve this level of realism, we carefully design different strategies for various types of motion.

Rigid scene-level background motions. We construct the indoor scene background using 3D-FRONT, as it offers professionally-designed room layouts and a rich variety of furniture. To enhance the scene complexity, following (Chen et al. 2023b), models from the ShapeNet dataset are randomly scaled, rotated, and placed either mid-air in the rooms or on existing furniture items such as tables and beds. The rigid background motion is naturally introduced when adjusting the camera view in the point cloud rendering process.

Non-rigid/rigid object-level foreground motions. Given a room layout, we consider two distinct cases of injecting foreground object motions based on the rigidity of the object: (i) Deformable non-rigid objects, such as humans and animals. Inspired by (Li and Harada 2022), we employ models and offsets from the DeformingThings4D dataset, a synthetic collection, comprising thousands of animation sequences across various categories of humanoids and animals. For each scene, animation sequences are randomly selected; half feature a single human and the other half feature multiple humans and animals. To enhance the diversity in movement magnitude, we employ random offsets to the human and animal models in every frame. (ii) Movable rigid objects, such as chairs and cups. First, additional rigid models from ShapeNet are positioned around each deformable non-rigid object in (i), with random XY-plane offsets applied to each frame to simulate movement and introduce noise. To augment realism, we introduce subtle movements to movable furniture items, thereby enriching the scene's plausibility with hybrid motions. Care is taken to ensure that each foreground object insertion is collision-free, maintaining the physical plausibility of the scene.

Registration data preparation. To produce paired data for point cloud registration, we follow (Chen et al. 2023b) to use a virtual perspective camera to render a depth map for each viewpoint, subsequently converting these maps into point clouds. We limit the depth range and randomly remove half of the large flat surfaces, such as walls and floors, to emphasize the foreground. To ensure scene complexity, high-quality camera views are selected. The viewing range is set to [0°,360°] horizontally and [0°,45°] vertically, with views uniformly sampled at intervals of 30°and 15°, respectively. Ground-truth correspondences of the overlapping point cloud pairs are easily established, given the accurate camera poses. Following 3DMatch, we categorize our data into HybridMatch (>30% overlap ratio) and HybridLoMatch (10-30% overlap ratio). Based on the proportion of non-rigid motions, each set is further divided into *10-30% non-rigid* and *30-50% non-rigid* splits. To validate our data, each pair is accompanied by a corresponding pair without any animations, forming a *rigid-only* split. For each split, we create a validation/test set, consisting of 100/1,000 pairs.

HybridReg

Problem formulation. Given the input point cloud pair $\mathcal{I} = (\mathcal{P}, \mathcal{Q})$, where $\mathcal{P} \in \mathbb{R}^{N \times 3}$ is the source point point with N points and $\mathcal{Q} \in \mathbb{R}^{M \times 3}$ is the target with M points. A network F with parameters θ is employed to estimate the correspondences $\mathcal{C}$, *i.e.*, $\mathcal{C} = F(\mathcal{I}; \theta)$. Then, the 3D transformation $\{\mathbf{R} \in SO(3), \mathbf{t} \in \mathbb{R}^3\}$ is derived from $\mathcal{C}$.

Feature extraction. We first downsample the source and target points to obtain superpoints $\hat{\mathcal{P}}$ and $\hat{\mathrm{Q}}$ as in (Qin et al. 2022). Utilizing KPConv-FPN (Thomas et al. 2019), we extract features of the point clouds, with the associated learned features denoted as $\mathbf{F}^{\hat{\mathcal{P}}} \in \mathbb{R}^{\hat{N} \times d}$ and $\mathbf{F}^{\hat{\mathcal{Q}}} \in \mathbb{R}^{\hat{M} \times d}$.

Geometric transformer. Following (Qin et al. 2022), we employ geometric self-attention mechanisms to learn global correlations among superpoints within each point cloud. Here, we describe the computation for $\hat{\mathcal{P}}$, and the same process applies to $\hat{\mathcal{Q}}$. Given the input feature matrix $\mathbf{X} \in \mathbb{R}^{\hat{N} \times d_t}$, the self-attention feature $\mathbf{Z} \in \mathbb{R}^{\hat{N} \times d_t}$ is computed as the cumulative weighted sum of all projected input features: $\mathbf{Z}_i = \sum_{j=1}^{\hat{N}} a_{i,j}(\mathbf{X}_j \mathbf{W}^V)$, where the weight coefficient $a_{i,j} = \mathrm{Softmax}(\frac{(\mathbf{X}_i \mathbf{W}^Q)(\mathbf{X}_j \mathbf{W}^K + \mathbf{R}_{i,j} \mathbf{W}^R)^T}{\sqrt{d_t}})$. $\mathbf{R}_{i,j} \in \mathbb{R}^{d_t}$ represents the geometric structure embedding, which includes pairwise distances and angular information among points. $\mathbf{W}$ denotes the projection matrix, whose superscripts Q, K, V, and R correspond to queries, keys, values, and geometric structure embeddings, respectively. Then, to encode inter-frame geometric consistency, cross-attention mechanisms are adopted. The cross-attention feature $\mathbf{Z}^{\hat{\mathcal{P}}} \in \mathbb{R}^{\hat{N} \times d_t}$ for $\hat{\mathcal{P}}$ is computed as $\mathbf{Z}_i^{\hat{\mathcal{P}}} = \sum_{j=1}^{\hat{M}} a_{i,j}(\mathbf{X}_j^{\hat{\mathcal{Q}}} \mathbf{W}^V)$, where the weight coefficient $a_{i,j} = \mathrm{Softmax}(\frac{(\mathbf{X}_i^{\hat{\mathcal{P}}} \mathbf{W}^Q)(\mathbf{X}_j^{\hat{\mathcal{Q}}} \mathbf{W}^K)^T}{\sqrt{d_t}})$.

Uncertainty mask generation. Previous methods typically predict correspondences $\mathcal{C}$ directly by $\mathcal{C} = F(\mathcal{I}; \theta)$, however, it does not provide confidence measures. Instead, our objective is to estimate the conditional probability density function $p(\mathcal{C}|\mathcal{I}; \theta)$, which allows us to capture the uncertainty associated with the predictions. To achieve this, we design a network to predict the parameters $\Phi(\mathcal{I}; \theta)$ of a family of distributions, *i.e.*,

$$p(\mathcal{C}|\mathcal{I}; \theta) = p(\mathcal{C}|\Phi(\mathcal{I}; \theta)) = \Pi_{i,j} p(c_{i,j}|\varphi_{i,j}(\mathcal{C}; \theta)). \tag{1}$$

For tractable estimation, we assume conditional independence between the predictions for the i-th source point and the j-th target point. Here, $c_{i,j} \in \{0, 1\}$ is the ground-truth correspondence relationship, and $\varphi_{i,j} \in \mathbb{R}^n$ denotes the predicted parameters, including both the predicted correspondence relationship and the confidence. For simplicity, we omit the subscript i, j in the subsequent discussion.

Compared to the direct prediction approach $\mathcal{C} = F(\mathcal{I}; \theta)$, the parameters $\Phi(\mathcal{I}; \theta)$ of the predictive distribution encode additional information about the correspondence, including its uncertainty. Following previous works (Ilg et al. 2018; Kendall and Gal 2017; Truong et al. 2023), this is done by predicting the variance of the correspondence. In these cases, the predictive density $p(c|\varphi)$ is modeled using Laplace distributions:

$$\mathcal{L}(c|\sigma^2) = \frac{1}{\sqrt{2\sigma^2}} e^{-\sqrt{\frac{2}{\sigma^2}}|c-\mu|}, \tag{2}$$

where the mean μ and variance σ^2 are predicted by the network as $(\mu, \sigma^2) = \Phi(\mathcal{I}; \theta)$ for each point pair. σ^2 serves as the uncertainty mask, indicating higher confidence for matching pairs likely to overlap and lower confidence for non-overlapping or foreground regions.

To predict the uncertainty mask, we first compute the cross-attention feature $\mathbf{G}^{\hat{\mathcal{P}}} \in \mathbb{R}^{\hat{N} \times d_t}$ for $\hat{\mathcal{P}}$ after obtaining the self-attention features $\mathbf{H}^{\hat{\mathcal{P}}}$ and $\mathbf{H}^{\hat{\mathcal{Q}}}$, *i.e.*,

$$\mathbf{G}_i^{\hat{\mathcal{P}}} = \sum_{j=1}^{\hat{M}} a_{i,j}(\mathbf{H}_j^{\hat{\mathcal{Q}}} \mathbf{W}^V), \tag{3}$$

where $a_{i,j} = \mathrm{Softmax}(\frac{(\mathbf{H}_i^{\hat{\mathcal{P}}} \mathbf{W}^Q)(\mathbf{H}_j^{\hat{\mathcal{Q}}} \mathbf{W}^K)^T}{\sqrt{d_t}})$ is the weight coefficient computed by applying a row-wise softmax on the cross-attention score. Similarly, we can derive the cross-attention feature $\mathbf{G}^{\hat{\mathcal{Q}}}$ for $\mathcal{Q}$. $\mathbf{G}^{\hat{\mathcal{P}}}$ and $\mathbf{G}^{\hat{\mathcal{Q}}}$ exhibit robust geometric characteristics and awareness of hybrid motions between the input point clouds. Finally, we use an MLP, Φ with parameters θ followed by a sigmoid layer, to regress the uncertainty masks, *i.e.*, $\sigma^2_{\hat{\mathcal{P}}}, \sigma^2_{\hat{\mathcal{Q}}} = \mathrm{Sigmoid}(\Phi(\mathbf{G}^{\hat{\mathcal{P}}}, \mathbf{G}^{\hat{\mathcal{Q}}}; \theta))$.

Uncertainty-aware superpoint matching. To determine superpoint correspondences, we follow (Qin et al. 2022) to first normalize $\mathbf{H}^{\hat{\mathcal{P}}}$ and $\mathbf{H}^{\hat{\mathcal{Q}}}$ onto a unit hypersphere. Then, we compute the Gaussian correlation matrix $\mathbf{S} \in \mathbb{R}^{\hat{N} \times \hat{M}}$, where $s_{i,j} = \exp(-||\mathbf{h}_i^{\hat{\mathcal{P}}} - \mathbf{h}_i^{\hat{\mathcal{Q}}}||_2^2)$. Besides, a dual-normalization operation (Rocco et al. 2018; Sun et al. 2021) is performed on $\mathbf{S}$ to suppress ambiguous matches. To further enhance the robustness against incorrect matches caused by hybrid motions, we mask the correlation matrix $\mathbf{S}$ using our generated uncertainty masks. We then select the largest N_c entries as the superpoint correspondences, *i.e.*,

$$\hat{\mathcal{C}} = \{(\hat{\mathbf{p}}_{x_i}, \hat{\mathbf{q}}_{y_i})|(x_i, y_i) \in \mathrm{topk}_{x,y}(s_{x,y} \cdot (1 - \sigma^2_{x,y}))\}, \tag{4}$$

where $\hat{\mathbf{p}}_{x_i}$ and $\hat{\mathbf{q}}_{y_i}$ are superpoints in $\hat{\mathcal{P}}$ and $\hat{\mathcal{Q}}$, respectively.

Transformation estimation. To recover the 3D transformation between the inputs based on the estimated correspondences, we employ both the RANSAC algorithm and the deep robust estimator LGR proposed by (Qin et al. 2022).

Loss Functions

The loss function consists of an uncertainty mask loss $\mathcal{L}_{um}$ for uncertainty prediction, an overlap-aware circle loss $\mathcal{L}_{oc}$ for superpoint matching, and a point matching loss $\mathcal{L}_p$. Both $\mathcal{L}_{oc}$ and $\mathcal{L}_p$ are implemented following (Qin et al. 2022). The overall loss for training is $\mathcal{L} = \mathcal{L}_{um} + \mathcal{L}_{oc} + \mathcal{L}_p$.

Uncertainty mask loss $\mathcal{L}_{um}$. As customary in probabilistic regression (Ilg et al. 2018; Kendall and Gal 2017; Truong et al. 2023), we employ the negative log-likelihood loss to train the uncertainty mask generator:

$$-\log p(\mathcal{C}|\Phi(\mathcal{I}; \theta)) = -\sum_{i,j} \log p(c_{i,j}|\varphi_{i,j}(\mathcal{I}; \theta)), \tag{5}$$

Methods	HybridMatch									HybridLoMatch								
	rigid-only			*10-30% non-rigid*			*30-50% non-rigid*			*rigid-only*			*10-30% non-rigid*			*30-50% non-rigid*		
	RRE ↓	RTE ↓	RR ↑	RRE ↓	RTE ↓	RR ↑	RRE ↓	RTE ↓	RR ↑	RRE ↓	RTE ↓	RR ↑	RRE ↓	RTE ↓	RR ↑	RRE ↓	RTE ↓	RR ↑
PREDATOR†	1.7	6.7	98.7	4.6	19.2	31.6	5.6	24.5	26.6	2.0	7.6	89.9	9.7	34.9	6.2	8.2	30.4	5.0
CoFiNet†	1.3	5.2	99.2	3.8	12.9	33.6	4.0	17.8	28.2	1.8	7.2	93.0	7.5	25.0	5.0	6.2	29.4	5.8
GeoTrans†	0.8	3.2	98.9	3.7	13.6	27.6	4.0	17.6	22.6	1.1	4.2	91.3	5.7	18.3	5.2	5.7	25.5	1.4
PREDATOR‡	1.1	4.4	98.9	2.4	8.7	80.8	2.8	12.6	75.4	1.4	5.3	**94.6**	5.1	17.5	33.2	4.2	17.5	24.4
CoFiNet‡	1.4	5.8	98.2	2.8	10.6	82.4	3.3	14.6	78.0	1.8	7.3	88.9	5.1	18.0	29.8	4.7	23.1	25.6
GeoTrans‡	0.6	2.3	**99.3**	1.4	5.6	95.6	1.6	8.1	93.6	**0.8**	3.4	93.8	3.3	12.3	64.2	3.2	13.7	61.0
Our HybridReg‡	**0.5**	**2.1**	99.2	**1.2**	**5.4**	**97.4**	**1.5**	**7.1**	**96.6**	**0.8**	**3.3**	93.9	**3.0**	**12.2**	**75.6**	**2.8**	**12.3**	**73.6**

Table 1: Evaluation results on our HybridMatch and HybridLoMatch. †: the model is only trained on 3DMatch. ‡: the model is trained on our HybridMatch. The best and second-best results are marked in bold and underlined for better comparison.

# Samples	HybridMatch										HybridLoMatch									
	10-30% non-rigid					*30-50% non-rigid*					*10-30% non-rigid*					*30-50% non-rigid*				
	5000	2500	1000	500	250	5000	2500	1000	500	250	5000	2500	1000	500	250	5000	2500	1000	500	250
Feature Matching Recall (%) ↑																				
PREDATOR	95.2	96.0	96.2	96.2	95.4	94.0	94.2	93.6	93.2	94.0	66.8	68.4	68.4	70.2	67.4	62.8	64.0	64.8	63.8	61.8
CoFiNet	96.0	96.4	96.2	96.4	95.6	93.0	94.6	95.2	95.2	95.0	59.4	61.4	63.2	63.4	64.0	53.4	55.2	58.8	59.4	59.4
GeoTrans	**99.9**	**99.8**	**99.8**	**99.8**	**99.8**	99.4	99.4	**99.6**	99.4	99.4	**93.2**	**92.8**	**93.4**	92.2	92.0	91.4	91.4	90.4	91.6	90.8
Our HybridReg	99.8	**99.8**	**99.8**	**99.8**	**99.8**	**99.6**	**99.6**	**99.6**	**99.6**	**99.6**	92.8	**92.8**	92.8	**92.6**	**93.0**	**94.4**	**94.4**	**94.8**	**94.6**	**94.8**
Inlier Ratio (%) ↑																				
PREDATOR	32.4	34.8	36.5	36.7	35.1	29.4	31.6	33.1	32.9	31.8	11.4	12.2	13.0	13.2	12.6	9.8	10.5	11.1	11.0	10.9
CoFiNet	23.4	26.7	31.2	32.9	34.0	21.3	24.5	28.7	30.5	31.6	7.7	8.7	10.9	11.8	12.4	7.1	7.9	9.7	10.6	11.2
GeoTrans	36.5	36.5	36.4	36.7	36.4	33.6	33.6	33.7	33.6	33.6	18.2	18.3	18.3	18.3	18.4	33.6	17.3	17.4	17.3	17.2
Our HybridReg	**58.6**	**58.7**	**58.7**	**58.8**	**58.6**	**54.4**	**54.4**	**54.4**	**54.3**	**54.3**	**34.4**	**34.3**	**34.3**	**34.3**	**34.3**	**33.8**	**33.9**	**33.7**	**33.7**	**33.7**
Registration Recall (%) ↑																				
PREDATOR	65.2	73.4	79.2	80.8	80.6	57.2	69.8	74.6	75.4	71.4	19.6	26.8	33.2	32.4	30.6	15.6	21.2	24.2	22.8	24.4
CoFiNet	79.0	80.4	82.4	80.6	79.8	75.6	78.0	77.8	77.2	72.2	27.2	27.8	29.8	27.6	27.2	21.6	24.8	25.2	25.6	24.8
GeoTrans	92.8	91.4	89.0	88.6	90.0	91.0	86.8	88.6	86.8	86.4	52.6	51.4	54.4	49.2	51.8	52.4	55.2	43.8	50.0	50.8
Our HybridReg	**97.4**	**96.4**	**97.6**	**97.2**	**96.4**	**96.6**	**97.4**	**97.8**	**95.0**	**96.0**	**75.6**	**73.0**	**75.2**	**73.0**	**70.8**	**76.2**	**75.8**	**75.0**	**74.2**	**72.4**

Table 2: Evaluation results on HybridMatch and HybridLoMatch.

where $p(c|\varphi)$ can be substituted by the Laplace distribution as detailed in Eq. 2. Taking into account both source and target uncertainty masks, $\mathcal{L}_{um}$ is formulated as

$$\mathcal{L}_{um} = -\log(\sum\nolimits_{i,j} e^{-\log(2)-\alpha_{i,j}-\sqrt{2}e^{-\frac{1}{2}\alpha_{i,j}}\cdot|c_{i,j}-\mu_{i,j}|}), \quad (6)$$

where $\alpha_{i,j} = \log(\sigma^2_{\hat{\mathcal{P}}} \cdot \sigma^2_{\hat{\mathcal{Q}}})_{i,j}$. Naturally, we use $\mu_{i,j} = s_{i,j}$ in the Gaussian correlation matrix $\mathbf{S}$ to denote the correspondence prediction for the point pair i, j. To avoid division by zero, we use a numerically stable log-sum-exp function and introduce a regularization term $1 - \sigma^2_{\hat{\mathcal{P}}} \cdot \sigma^2_{\hat{\mathcal{Q}}}$.

Overlap-aware circle loss $\mathcal{L}_{oc}$. To prioritize high-overlap matches, we compute the overlap-aware circle loss on $\mathcal{P}$,

$$\mathcal{L}^{\mathcal{P}}_{oc} = \frac{1}{|\mathcal{A}|}\sum_{\mathcal{G}^{\mathcal{P}}_i \in \mathcal{A}} \log[1 + \sum_{\mathcal{G}^{\mathcal{Q}}_j \in \varepsilon^i_p} e^{\lambda^j_i \beta^{i,j}_p (d^j_i - \Delta_p)} \cdot \sum_{\mathcal{P}^{\mathcal{Q}}_k \in \varepsilon^i_n} e^{\beta^{i,k}_p (\Delta_n - d^k_i)}]. \quad (7)$$

Here, $\mathcal{A}$ is the set of anchor patches in $\mathcal{P}$ that have at least one positive patch in $\mathcal{Q}$. For each anchor patch $\mathcal{G}^{\mathcal{P}}_i \in \mathcal{A}$, ε^i_p and ε^i_n denote the sets of its positive and negative patches in $\mathcal{Q}$. d^j_i is the feature distance, and $\lambda^j_i = (o^j_i)^{\frac{1}{2}}$ and o^j_i represents the overlap ratio between $\mathcal{G}^{\mathcal{P}}_i$ and $\mathcal{G}^{\mathcal{Q}}_j$. $\beta^{i,j}_p = \gamma(d^j_i - \Delta_p)$ and $\beta^{i,k}_n = \gamma(\Delta_n - d^k_i)$ are the positive and negative weights, respectively. The margins Δ_p and Δ_n are set to 0.1 and 0.4, respectively. The same goes for the loss $\mathcal{L}^{\mathcal{Q}}_{oc}$ on $\mathcal{Q}$.

Point matching loss $\mathcal{L}_p$ is computed for the ith patch as

$$\mathcal{L}_{p,i} = -\sum_{(u,v)\in\mathcal{M}_i} \log \bar{z}^i_{u,v} - \sum_{u\in\mathcal{I}_i} \log \bar{z}^i_{u,m_i+1} - \sum_{v\in\mathcal{J}_i} \log \bar{z}^i_{n_i+1,v}, \quad (8)$$

where $\mathcal{M}_i$ denotes the set of ground-truth point correspondences extracted with a matching radius τ from each $\hat{\mathcal{C}}^*_i$ and $\{\hat{\mathcal{C}}^*_i\}$ denotes the randomly sampled N_g ground-truth point correspondences. $\bar{z}^i_{u,v}$ is the element in the u-th row and the v-th column of the soft assignment matrix $\bar{\mathbf{Z}}_i$. $\mathcal{I}_i$ and $\mathcal{J}_i$ denote the sets of unmatched points in the two patches.

Experiments

Datasets and Evaluation Metrics

Datasets. To evaluate effectiveness on hybrid motions, we use the proposed HybridMatch/HybridLoMatch, which is divided into three splits: *rigid-only* (11.3k/14k pairs), *10-30% non-rigid* (7.3k/8.5k pairs), and *30-50% non-rigid* (4k/5.5k pairs). To assess robustness in indoor scenes, we employ the 3DMatch dataset (Zeng et al. 2017), comprising 62 scenes, with 46/8/8 scenes used for training/validation/testing. 3DMatch and 3DLoMatch are categorized

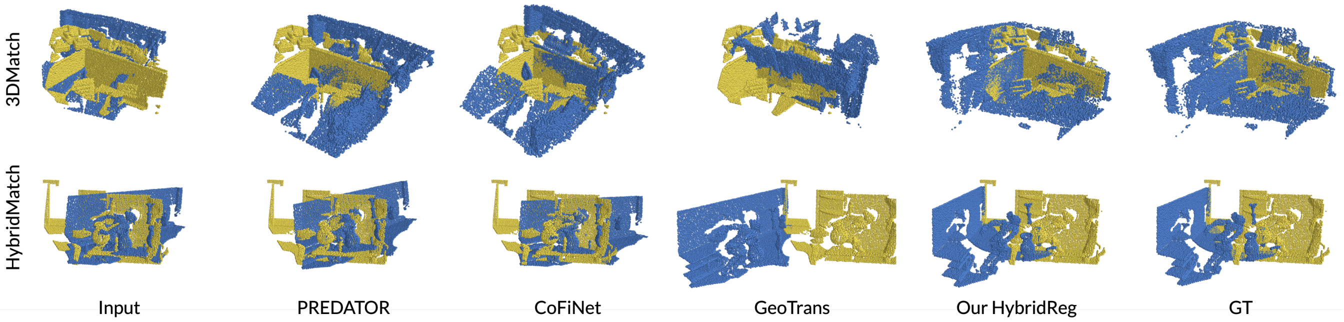

Figure 3: Qualitative comparison on HybridMatch and 3DMatch.

Methods	3DMatch			3DLoMatch		
	RRE ↓	RTE ↓	RR ↑	RRE ↓	RTE ↓	RR ↑
PerfectMatch	2.2	7.1	78.4	3.5	10.3	33.0
FCGF	<u>2.0</u>	6.6	85.1	3.2	10.0	40.1
D3Feat	2.2	6.7	81.6	3.4	10.3	37.3
PREDATOR	<u>2.0</u>	6.4	89.0	3.1	9.3	59.8
CoFiNet	<u>2.0</u>	6.2	89.3	3.3	9.4	67.5
GeoTrans	**1.6**	5.3	91.5	**2.5**	**7.4**	**74.0**
RegTR	**1.6**	**4.9**	<u>92.0</u>	<u>2.8</u>	7.7	64.8
Our HybridReg	**1.6**	<u>5.2</u>	**92.7**	**2.5**	<u>7.5</u>	<u>73.1</u>

Table 3: Registration results on 3DMatch and 3DLoMatch.

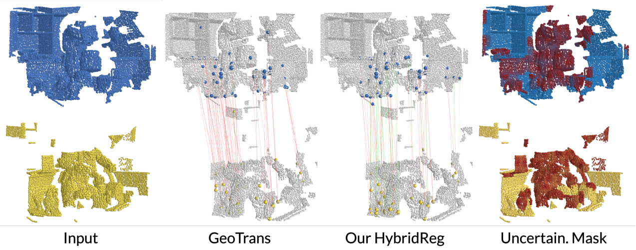

Figure 4: Visualization of correspondences and uncertainty masks, where unreliable non-rigid regions are marked in red.

based on >30% and 10%-30% overlap ratios. To evaluate the generalizability in outdoor scenes, we transfer models trained on 3DMatch to the ETH dataset (Pomerleau et al. 2012), including 4 scenes with 713 pairs of 132 point clouds.

Evaluation metrics. We adopt several metrics following (Qin et al. 2022; Chen et al. 2023b): (i) Registration Recall (RR), the fraction of successfully registered pairs whose transformation error RMSE<0.2m/0.2m/0.5m for HybridMatch/3DMatch/ETH; (ii) Inlier Ratio (IR), the fraction of inlier correspondences with residuals <0.1m/0.1m/0.2m for HybridMatch/3DMatch/ETH; (iii) Feature Matching Recall (FMR), the fraction of pairs with IR>5%; (iv) The median of the average Relative Rotation Error (RRE); and (v) The median of the average Relative Translation Error (RTE) for successfully registered pairs whose RMSE<0.2m/0.2m/0.5m for HybridMatch/3DMatch/ETH.

Implementation Details

To stabilize training, the uncertainty mask generator is first trained with L1 loss for 30 epochs, then fine-tuned with uncertainty mask loss for 20 epochs. Adam optimizer (Kingma and Ba 2015) is used with a batch size of 8 and a learning rate of 1e-4. All experiments run on 8 NVIDIA Tesla P40 GPUs. Please refer to our supp. material for more details.

Evaluation on HybridMatch

We compare the quantitative registration results of our method with recent state-of-the-art approaches in Tab. 1 and Tab. 2. As revealed in Tab. 1, either transferring pre-trained model weights from 3DMatch to our HybridMatch *rigid-only* split or direct training on our data yields high RR, validating our synthetic data. However, models trained on rigid-only data struggle with hybrid motions, particularly in scenarios with significant foreground motion and low overlap. Incorporating hybrid-motion data into training improves performance but remains inferior to rigid-only scenes, highlighting the need for algorithmic enhancements. Our HybridReg significantly outperforms other methods across almost all metrics, demonstrating robust registration capability for both rigid-only and challenging non-rigid scenarios. Specifically, our method achieves a 20.7% increase in RR compared to the second best on the most challenging test split (HybridLoMatch, *30-50% non-rigid*). Qualitative results in Fig. 3 also illustrate our method's superior performance. We visualize our uncertainty masks in Fig. 4, where the precise masks improve the correspondence accuracy.

Evaluation on Rigid Benchmark

Evaluation on indoor 3D(Lo)Match. Results on the rigid-only 3DMatch dataset are presented in Tab. 3. Our HybridReg achieves state-of-the-art performance on 3DMatch and competitive results on 3DLoMatch, highlighting its ability to preserve the strength on rigid-only scenes effectively. Fig. 3 also indicates the effectiveness of our method.

Evaluation on outdoor ETH. To assess generalizability, we transfer models trained on 3DMatch to the ETH dataset, following (Wang et al. 2023; Chen et al. 2023b). As shown in Tab. 5, our method outperforms the baseline (GeoTrans) and most other methods in RR, showcasing strong generalization

Models	HybridMatch						HybridLoMatch					
	10-30% non-rigid			*30-50% non-rigid*			*10-30% non-rigid*			*30-50% non-rigid*		
	RRE $\downarrow$	RTE $\downarrow$	RR $\uparrow$	RRE $\downarrow$	RTE $\downarrow$	RR $\uparrow$	RRE $\downarrow$	RTE $\downarrow$	RR $\uparrow$	RRE $\downarrow$	RTE $\downarrow$	RR $\uparrow$
(a) Baseline (3D-FRONT)	2.0	8.5	52.8	2.5	12.4	44.8	4.6	16.5	18.8	4.3	19.0	14.8
(b) + Non-rigid Foreground Motion	1.6	7.5	90.4	2.0	9.6	86.2	3.7	14.9	54.0	4.1	16.5	48.6
(c) + Rigid Foreground Motion	1.4	6.0	94.0	2.0	9.4	89.2	3.5	14.1	62.8	3.3	15.4	54.2
(d) + Delete Planes	**1.2**	**5.4**	**97.4**	**1.5**	**7.1**	**96.6**	**3.0**	**12.2**	**75.6**	**2.8**	**12.3**	**73.6**
(e) Baseline (GeoTrans)	1.4	5.6	95.6	1.6	8.1	93.6	3.3	12.3	64.2	3.2	13.7	61.0
(f) w/ Uncertainty Mask + L1 Loss	1.4	5.9	94.2	1.7	8.6	95.2	3.4	13.2	66.2	3.4	15.5	61.8
(g) w/ Uncertainty Mask + NLL Loss	**1.2**	**5.4**	97.2	**1.5**	7.2	**96.6**	**3.0**	**12.2**	75.5	**2.8**	12.4	73.2
(h) + Uncertainty-aware Correlation	**1.2**	**5.4**	**97.4**	**1.5**	**7.1**	**96.6**	**3.0**	**12.2**	**75.6**	**2.8**	**12.3**	**73.6**

Table 4: Ablation studies of our proposed HybridMatch (top) and HybridReg (bottom).

ETH					
# Samples	5000	2500	1000	500	250
PerfectMatch	81.4	73.5	59.3	46.5	35.0
D3Feat	59.1	50.4	49.7	44.6	29.1
FCGF	42.1	36.1	29.5	26.3	18.9
PREDATOR	74.7	72.9	67.7	60.3	51.7
CoFiNet	83.9	82.7	81.9	77.4	68.8
GeoTrans + RANSAC	80.9	76.8	73.3	72.8	70.8
GeoTrans + LGR	—		85.6	—	
Our HybridReg + LGR	—		**86.5**	—	

Table 5: Registration recall comparison on ETH. All models are trained on 3DMatch. —: results with different samples are not applicable for LGR since it uses all correspondences.

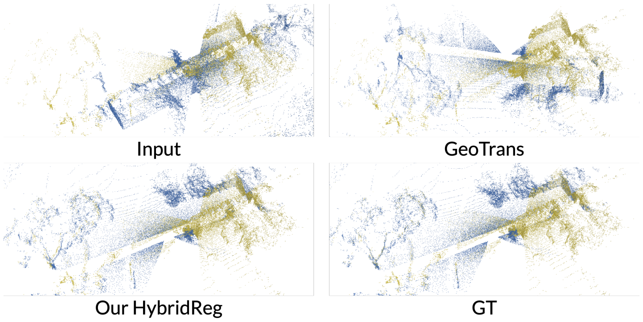

Figure 5: Qualitative comparison on ETH.

capability in outdoor scenes. The qualitative comparison in Fig. 5 shows our method's effectiveness in addressing challenging cases with complex, unseen structures.

Ablation Study

We conduct extensive experiments on our HybridMatch and HybridReg to evaluate the impact of each component. The hybrid-motion test set is fixed for consistent evaluation.

HybridMatch. We detail the construction of our HybridMatch in Tab. 4 (a-d). By incrementally incorporating components into our baseline, we assess their contributions to the model's performance on the test set. Comparing Rows (a) and (b), injecting non-rigid motions from DeformingThings4D significantly boosts registration performance. Rows (c-d) and (b) demonstrate that adding rigid motions to selected foreground objects (c) and removing planes to balance geometric structures (d) help the model focus on background motion, thereby improving performance.

HybridReg. To understand the efficacy of components in HybridReg, we ablate each of them as shown in Tab. 4 (e-h). Initially, we integrate the uncertainty mask learning module into the baseline, employing two loss variants for supervision. Comparing Rows (f-g) and (e), the negative log-likelihood (NLL) loss outperforms the L1 loss in enhancing uncertainty modeling and learning capabilities, leading the model to better identify inlier correspondences, such as rigid background and immovable foreground. Further, comparing Rows (g) and (h), using the predicted uncertainty mask to refine the Gaussian correlation matrix yields an additional performance boost by suppressing incorrect matches.

Conclusion

We introduce HybridReg, a novel 3D point cloud registration framework to account for hybrid motions and to enhance robustness. To our best knowledge, we create the first large-scale indoor synthetic dataset, HybridMatch, tailored for point cloud registration with hybrid motions. In addition, we propose a probabilistic uncertainty mask learning module to mitigate the impact of outliers, guided by an effective negative log-likelihood loss, enhancing feature extraction and correspondence matching. Experimental quantitative and qualitative results manifest the state-of-the-art performance of HybridReg consistently on various indoor and outdoor benchmarks.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant Nos. 62372091, 62071097, 62236007, the Sichuan Science and Technology Program under Grant Nos. 2023NSFSC0462, 2023NSFSC0458, 2023NSFSC1972, and the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. T45-401/22-N).

References

Ao, S.; Hu, Q.; Wang, H.; Xu, K.; and Guo, Y. 2023. BUFFER: Balancing Accuracy, Efficiency, and Generalizability in Point Cloud Registration. In *CVPR*, 1255–1264.

Ao, S.; Hu, Q.; Yang, B.; Markham, A.; and Guo, Y. 2021. SpinNet: Learning a General Surface Descriptor for 3D Point Cloud Registration. In *CVPR*, 11753–11762.

Aoki, Y.; Goforth, H.; Srivatsan, R. A.; and Lucey, S. 2019. PointNetLK: Robust & Efficient Point Cloud Registration using PointNet. In *CVPR*, 7163–7172.

Bai, X.; Luo, Z.; Zhou, L.; Chen, H.; Li, L.; Hu, Z.; Fu, H.; and Tai, C.-L. 2021. PointDSC: Robust Point Cloud Registration Using Deep Spatial Consistency. In *CVPR*, 15859–15869.

Bai, X.; Luo, Z.; Zhou, L.; Fu, H.; Quan, L.; and Tai, C.-L. 2020. D3Feat: Joint Learning of Dense Detection and Description of 3D Local Features. In *CVPR*, 6359–6367.

Besl, P. J.; and McKay, N. D. 1992. A Method for Registration of 3D Shapes. *IEEE TPAMI*, 14(2): 239–256.

Bozic, A.; Palafox, P.; Zollhöfer, M.; Dai, A.; Thies, J.; and Nießner, M. 2020a. Neural non-rigid tracking. *NeurIPS*, 33: 18727–18737.

Bozic, A.; Zollhofer, M.; Theobalt, C.; and Nießner, M. 2020b. DeepDeform: Learning non-rigid RGB-D reconstruction with semi-supervised data. In *CVPR*, 7002–7012.

Chang, A. X.; Funkhouser, T.; Guibas, L. J.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. ShapeNet: An Information-Rich 3D Model Repository. *arXiv preprint arXiv:1512.03012*.

Chen, G.; Wang, M.; Yuan, L.; Yang, Y.; and Yue, Y. 2023a. Rethinking Point Cloud Registration as Masking and Reconstruction. In *ICCV*, 17717–17727.

Chen, S.; Xu, H.; Li, H.; Luo, K.; Liu, G.; Fu, C.-W.; Tan, P.; and Liu, S. 2024. PointRegGPT: Boosting 3D Point Cloud Registration using Generative Point-Cloud Pairs for Training. In *ECCV*.

Chen, S.; Xu, H.; Li, R.; Liu, G.; Fu, C.-W.; and Liu, S. 2023b. SIRA-PCR: Sim-to-Real Adaptation for 3D Point Cloud Registration. In *ICCV*, 14394–14405.

Choy, C.; Dong, W.; and Koltun, V. 2020. Deep Global Registration. In *CVPR*, 2514–2523.

Choy, C.; Park, J.; and Koltun, V. 2019. Fully Convolutional Geometric Features. In *ICCV*, 8958–8966.

Curless, B.; and Levoy, M. 1996. A Volumetric Method for Building Complex Models from Range Images. In *ACM SIGGRAPH*, 303–312.

Dang, Z.; and Salzmann, M. 2023. AutoSynth: Learning to Generate 3D Training Data for Object Point Cloud Registration. In *ICCV*, 9009–9019.

Dosovitskiy, A.; Fischer, P.; Ilg, E.; Hausser, P.; Hazirbas, C.; Golkov, V.; Van Der Smagt, P.; Cremers, D.; and Brox, T. 2015. FlowNet: Learning Optical Flow With Convolutional Networks. In *ICCV*, 2758–2766.

Fischler, M. A.; and Bolles, R. C. 1981. Random Sample Consensus: A Paradigm for Model Fitting With Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6): 381–395.

Fu, H.; Cai, B.; Gao, L.; Zhang, L.-X.; Wang, J.; Li, C.; Zeng, Q.; Sun, C.; Jia, R.; Zhao, B.; et al. 2021a. 3D-FRONT: 3D Furnished Rooms with Layouts and Semantics. In *ICCV*, 10933–10942.

Fu, K.; Liu, S.; Luo, X.; and Wang, M. 2021b. Robust Point Cloud Registration Framework Based on Deep Graph Matching. In *CVPR*, 8893–8902.

Gao, W.; and Tedrake, R. 2018. SurfelWarp: Efficient Non-Volumetric Single View Dynamic Reconstruction. In *RSS*.

Gojcic, Z.; Zhou, C.; Wegner, J. D.; and Wieser, A. 2019. The Perfect Match: 3D Point Cloud Matching with Smoothed Densities. In *CVPR*, 5545–5554.

Groueix, T.; Fisher, M.; Kim, V. G.; Russell, B. C.; and Aubry, M. 2018. 3D-CODED: 3D correspondences by deep deformation. In *ECCV*, 230–246.

Hatem, A.; Qian, Y.; and Wang, Y. 2023. Point-TTA: Test-Time Adaptation for Point Cloud Registration Using Multi-task Meta-Auxiliary Learning. In *ICCV*, 16494–16504.

Huang, Q.-X.; Adams, B.; Wicke, M.; and Guibas, L. J. 2008. Non-rigid registration under isometric deformations. In *Comput. Graph. Forum*, volume 27, 1449–1457.

Huang, S.; Gojcic, Z.; Usvyatsov, M.; Wieser, A.; and Schindler, K. 2021. PREDATOR: Registration of 3D Point Clouds with Low Overlap. In *CVPR*, 4267–4276.

Ilg, E.; Cicek, O.; Galesso, S.; Klein, A.; Makansi, O.; Hutter, F.; and Brox, T. 2018. Uncertainty estimates and multi-hypotheses networks for optical flow. In *ECCV*, 652–667.

Innmann, M.; Zollhöfer, M.; Nießner, M.; Theobalt, C.; and Stamminger, M. 2016. VolumeDeform: Real-time volumetric non-rigid reconstruction. In *ECCV*, 362–379.

Jiang, H.; Dang, Z.; Wei, Z.; Xie, J.; Yang, J.; and Salzmann, M. 2023. Robust Outlier Rejection for 3D Registration With Variational Bayes. In *CVPR*, 1148–1157.

Kendall, A.; and Gal, Y. 2017. What uncertainties do we need in bayesian deep learning for computer vision? *NeurIPS*, 30.

Kingma, P. D.; and Ba, L. J. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

Li, J.; Zhang, C.; Xu, Z.; Zhou, H.; and Zhang, C. 2020a. Iterative Distance-Aware Similarity Matrix Convolution with Mutual-Supervised Point Elimination for Efficient Point Cloud Registration. In *ECCV*, 378–394.

Li, X.; Kaesemodel Pontes, J.; and Lucey, S. 2021. Neural scene flow prior. *NeurIPS*, 34: 7838–7851.

Li, Y.; Bozic, A.; Zhang, T.; Ji, Y.; Harada, T.; and Nießner, M. 2020b. Learning to optimize non-rigid tracking. In *CVPR*, 4910–4918.

Li, Y.; and Harada, T. 2022. Lepard: Learning Partial Point Cloud Matching in Rigid and Deformable Scenes. *CVPR*.

Li, Y.; Takehara, H.; Taketomi, T.; Zheng, B.; and Nießner, M. 2021. 4Dcomplete: Non-rigid motion estimation beyond the observable surface. In *ICCV*, 12706–12716.

Liu, J.; Wang, G.; Liu, Z.; Jiang, C.; Pollefeys, M.; and Wang, H. 2023a. RegFormer: An efficient projection-aware transformer network for large-scale point cloud registration. In *ICCV*, 8451–8460.

Liu, Q.; Zhu, H.; Zhou, Y.; Li, H.; Chang, S.; and Guo, M. 2023b. Density-invariant Features for Distant Point Cloud Registration. In *ICCV*, 18215–18225.

Liu, X.; Qi, C. R.; and Guibas, L. J. 2019. FlowNet3D: Learning scene flow in 3D point clouds. In *CVPR*, 529–537.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *IJCV*, 60: 91–110.

Mei, G.; Tang, H.; Huang, X.; Wang, W.; Liu, J.; Zhang, J.; Van Gool, L.; and Wu, Q. 2023. Unsupervised Deep Probabilistic Approach for Partial Point Cloud Registration. In *CVPR*, 13611–13620.

Newcombe, R. A.; Fox, D.; and Seitz, S. M. 2015. DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time. In *CVPR*, 343–352.

Ovsjanikov, M.; Ben-Chen, M.; Solomon, J.; Butscher, A.; and Guibas, L. 2012. Functional maps: a flexible representation of maps between shapes. *ACM TOG*, 31(4): 1–11.

Pais, G. D.; Ramalingam, S.; Govindu, V. M.; Nascimento, J. C.; Chellappa, R.; and Miraldo, P. 2020. 3DRegNet: A Deep Neural Network for 3D Point Registration. In *CVPR*, 7193–7203.

Poiesi, F.; and Boscaini, D. 2022. Learning General and Distinctive 3D Local Deep Descriptors for Point Cloud Registration. *IEEE TPAMI*.

Pomerleau, F.; Liu, M.; Colas, F.; and Siegwart, R. 2012. Challenging Data Sets for Point Cloud Registration Algorithms. *IJRR*, 31(14): 1705–1711.

Qin, Z.; Yu, H.; Wang, C.; Guo, Y.; Peng, Y.; and Xu, K. 2022. Geometric Transformer for Fast and Robust Point Cloud Registration. In *CVPR*, 11143–11152.

Rocco, I.; Cimpoi, M.; Arandjelović, R.; Torii, A.; Pajdla, T.; and Sivic, J. 2018. Neighbourhood consensus networks. *NeurIPS*, 31.

Rusinkiewicz, S. 2019. A Symmetric Objective Function for LiDAR. *ACM TOG*, 38(4): 1–7.

Rusinkiewicz, S.; and Levoy, M. 2001. Efficient Variants of the ICP Algorithm. In *3DIM*, 145–152.

Rusu, R. B.; Blodow, N.; and Beetz, M. 2009. Fast Point Feature Histograms (FPFH) for 3D Registration. In *ICRA*, 3212–3217.

Schmidt, T.; Newcombe, R.; and Fox, D. 2016. Self-supervised visual descriptor learning for dense correspondence. *IEEE RAL*, 2(2): 420–427.

Shotton, J.; Glocker, B.; Zach, C.; Izadi, S.; Criminisi, A.; and Fitzgibbon, A. 2013. Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images. In *CVPR*, 2930–2937.

Sun, J.; Shen, Z.; Wang, Y.; Bao, H.; and Zhou, X. 2021. LoFTR: Detector-free local feature matching with transformers. In *CVPR*, 8922–8931.

Thomas, H.; Qi, C. R.; Deschaud, J.-E.; Marcotegui, B.; Goulette, F.; and Guibas, L. J. 2019. KPConv: Flexible and Deformable Convolution for Point Clouds. In *ICCV*, 6411–6420.

Truong, P.; Danelljan, M.; Timofte, R.; and Van Gool, L. 2023. PDC-Net+: Enhanced probabilistic dense correspondence network. *IEEE TPAMI*, 45(8): 10247–10266.

Wang, H.; Liu, Y.; Dong, Z.; and Wang, W. 2022. You Only Hypothesize Once: Point Cloud Registration With Rotation-Equivariant Descriptors. In *IEEE TMM*, 1630–1641.

Wang, H.; Liu, Y.; Hu, Q.; Wang, B.; Chen, J.; Dong, Z.; Guo, Y.; Wang, W.; and Yang, B. 2023. RoReg: Pairwise Point Cloud Registration with Oriented Descriptors and Local Rotations. *IEEE TPAMI*.

Wang, Y.; and Solomon, J. M. 2019a. Deep Closest Point: Learning Representations for Point Cloud Registration. In *ICCV*, 3523–3532.

Wang, Y.; and Solomon, J. M. 2019b. PRNet: Self-Supervised Learning for Partial-to-Partial Registration. In *NeurIPS*, 8814–8826.

Wu, W.; Wang, Z. Y.; Li, Z.; Liu, W.; and Fuxin, L. 2020. PointPWC-Net: Cost Volume on Point Clouds for (Self-) Supervised Scene Flow Estimation. In *ECCV*, 88–107.

Xu, H.; Liu, S.; Wang, G.; Liu, G.; and Zeng, B. 2021. OMNet: Learning Overlapping Mask for Partial-to-Partial Point Cloud Registration. In *ICCV*.

Xu, H.; Ye, N.; Liu, G.; Zeng, B.; and Liu, S. 2022. FINet: Dual Branches Feature Interaction for Partial-To-Partial Point Cloud Registration. In *AAAI*, volume 36, 2848–2856.

Yang, F.; Guo, L.; Chen, Z.; and Tao, W. 2022. One-inlier is first: Towards efficient position encoding for point cloud registration. *NeurIPS*, 35: 6982–6995.

Yang, J.; Li, H.; and Jia, Y. 2013. Go-ICP: Solving 3D Registration Efficiently and Globally Optimally. In *CVPR*, 1457–1464.

Yew, Z. J.; and Lee, G. H. 2020. RPM-Net: Robust Point Matching using Learned Features. In *CVPR*, 11824–11833.

Yew, Z. J.; and Lee, G. H. 2022. REGTR: End-to-end Point Cloud Correspondences with Transformers. In *CVPR*, 6677–6686.

Yu, H.; Li, F.; Saleh, M.; Busam, B.; and Ilic, S. 2021. CoFiNet: Reliable Coarse-to-Fine Correspondences for Robust Point Cloud Registration. *NeurIPS*, 34: 23872–23884.

Yu, J.; Ren, L.; Zhang, Y.; Zhou, W.; Lin, L.; and Dai, G. 2023. PEAL: Prior-Embedded Explicit Attention Learning for Low-Overlap Point Cloud Registration. In *CVPR*, 17702–17711.

Zeng, A.; Song, S.; Nießner, M.; Fisher, M.; Xiao, J.; and Funkhouser, T. 2017. 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions. In *CVPR*, 1802–1811.

Zhang, X.; Yang, J.; Zhang, S.; and Zhang, Y. 2023. 3D Registration With Maximal Cliques. In *CVPR*, 17745–17754.