SAFER: SURFACE-ADAPTIVE FEATURE FIELDS FOR CROSS-SOURCE POINT CLOUD REGISTRATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Point clouds captured by heterogeneous LiDAR sensors exhibit divergent sampling patterns and densities, causing the same physical surfaces to be observed at shifted and non-uniform locations. Existing methods typically encode local geometry with superpoints and their sampled neighborhoods, which can entangle stable surface structure with sensor-dependent sampling variation. Consequently, representations may vary with the sensing mechanism, even when the underlying physical geometry is unchanged. To address this, we propose SAFER, a cross-source point cloud registration framework that lifts superpoint features into a continuous, surface-adaptive feature field by assigning each superpoint an anisotropic Gaussian basis in its local tangent-normal frame. The resulting normalized field affinity provides a geometry-aware prior that encourages compatible support interactions while down-weighting geometrically inconsistent neighborhoods. With this prior, we design Field-Aware Geometric Attention to guide contextual aggregation and introduce a Kernel-Guided Feature Adapter to perform kernel-conditioned descriptor refinement. An overlap-aware field objective further maximizes the consistency of ground-truth-aligned feature fields within their mutually observable regions. To alleviate the scarcity of large-scale cross-source data, we construct two benchmarks. *Forest Cross-Source* comprises ground-to-aerial LiDAR pairs in unstructured forest scenes, while *Indoor Multi-LiDAR* covers heterogeneous spinning and solid-state LiDARs in indoor environments. Extensive experiments demonstrate that SAFER improves registration performance across diverse sensor configurations, with notable benefits under severe sampling discrepancies.

1 INTRODUCTION

Point cloud registration is a fundamental task in 3D vision, aiming to estimate a rigid transformation that aligns two partially overlapping point clouds into a shared coordinate system (Ao et al., 2021; 2022). As 3D sensing modalities proliferate, point clouds are increasingly acquired from heterogeneous sources rather than a single homogeneous sensor (Wu et al., 2026; Zhao et al., 2025a). Cross-source point cloud registration extends this conventional task to data captured by different sensors, and is essential for multi-sensor fusion and large-scale 3D perception applications.

Cross-source registration remains challenging due to the divergent physical sampling processes of heterogeneous sensors. Unlike same-source scenarios, where scans share relatively consistent acquisition patterns, cross-source pairs can exhibit substantial differences in point density, distribution, and noise characteristics (Huang et al., 2021b; 2023). As illustrated in Fig. 1, even after ground-truth alignment, cross-source point clouds can sample the same physical structure at shifted and non-uniform locations, leading to larger nearest-neighbor distances than same-source pairs. For instance, an Ouster OS0 with 128 beams captures structured, dense rings, whereas Livox solid-state sensors emit non-repetitive, irregular patterns (Li et al., 2022). These discrepancies cause the same physical surface to be represented by different sampling patterns, such that representations may vary with the sensing mechanism even when the underlying geometry remains unchanged.

Recent studies have investigated cross-source registration from complementary perspectives. SPEAL (Xiong et al., 2024) incorporates skeletal geometric priors, Cross-PCR (Zhao et al., 2025b) develops density-robust feature learning and hierarchical matching, and Cross3DReg (Xu et al., 2025) exploits complementary multi-modal information. Like many same-source registration meth-

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.

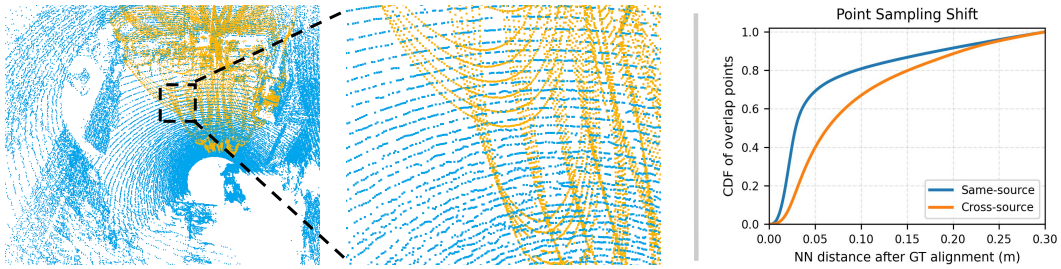


Figure 1: **Motivation of SAFER.** 1) *Left:* ground-truth-aligned point clouds captured by an Ouster OS0 and a Livox Avia, together with a zoomed region exhibiting shifted and non-uniform samples. 2) *Right:* The cumulative distribution function (CDF) shows that cross-source pairs retain larger nearest-neighbor distances than same-source pairs after alignment, indicating stronger sampling displacement across heterogeneous sensors. These observations motivate SAFER to associate each coarse superpoint with a continuous anisotropic spatial support that provides a surface-adaptive geometric prior for feature learning.

ods, these approaches generally follow a coarse-to-fine pipeline in which superpoints summarize local geometry before fine-level refinement (Qin et al., 2022). For coarse correspondence estimation, however, each superpoint is still commonly represented by a descriptor extracted from its sampled neighborhood. Because both the superpoint and its neighborhood depend on the acquisition process, the resulting representation can entangle stable surface structure with sensor-dependent sampling variation. Consequently, descriptor consistency may deteriorate when corresponding physical regions are observed under substantially different ray patterns, densities, and local sample locations.

This sensitivity becomes particularly severe at the superpoint level. Even when two superpoints represent the same physical surface, heterogeneous sensors may produce different local neighborhoods because of distinct ray configurations and sampling densities. The resulting descriptors therefore become less compatible despite the underlying geometric consistency. To address this, we retain superpoints as efficient coarse primitives but associate each of them with a continuous anisotropic spatial support. Rather than relying solely on features anchored to discrete sampled locations, this support extends the superpoint features over a local smooth surface-adaptive region. It is then used to regulate contextual feature aggregation and descriptor refinement, reducing sensitivity to sensor-induced variations in support centers and neighborhood composition and thereby improving over conventional coarse-to-fine correspondence pipeline.

To this end, we propose SAFER, which lifts superpoint features into a continuous, surface-adaptive feature field by assigning each superpoint an anisotropic Gaussian support. The Gaussian covariance is first derived from the local PCA frame (Lim et al., 2021), and the tangential and normal scales are further refined via feature-conditioned residuals. Such spatial support is thus surface-aligned and adaptive. The normalized overlap between two Gaussian supports jointly captures their spatial displacement and support compatibility, remaining smooth under sampling shifts while suppressing geometrically incompatible interactions. Specifically, we propose the ***A**nisotropic **C**ontinuous **F**eature **F**ield (ACFF)* to model each superpoint with a surface-adaptive Gaussian support and measure field compatibility through normalized Gaussian overlap. We further design ***F**ield-**A**ware **G**eometric **A**ttention (FAGA)* and a ***K**ernel-**G**uided **F**eature **A**dapter (KGFA)* to use this affinity for contextual aggregation and descriptor refinement. Finally, we introduce an overlap-aware field objective to enforce consistency between ground-truth-aligned feature fields within their mutually observable regions. Experimental results on three challenging benchmarks. demonstrate that the proposed method yields superior performance across heterogeneous sensor configurations.

Our main contributions are summarized as follows:

- We propose SAFER, which introduces continuous anisotropic superpoint supports into a learning-based coarse-to-fine registration framework to reduce representation sensitivity to sensor-dependent sampling variation.

- We introduce ACFF to construct surface-adaptive Gaussian supports and normalized field affinity, FAGA and KGFA to guide contextual aggregation and descriptor refinement, and an overlap-aware field objective to enforce consistency between aligned feature fields.
- We construct two challenging cross-source registration benchmarks and conduct extensive experiments to validate the effectiveness of SAFER under heterogeneous sensor configurations.

2 Related Work

Learning-based Registration Methods. Learning-based registration methods can be broadly divided into direct and correspondence-based approaches. Direct methods (Aoki et al., 2019; Wang & Solomon, 2019; Yew & Lee, 2020) estimate rigid transformations through global feature regression or iterative soft matching, but often struggle with large-scale scenes and limited overlap. Correspondence-based methods first establish putative matches using learned descriptors and then recover transformations through robust estimation. DGR (Choy et al., 2020), GeoTransformer (Qin et al., 2022), Predator (Huang et al., 2021a), and PARE-Net (Yao et al., 2024) improve correspondence generation through outlier filtering, geometric attention, overlap awareness, or rotation-equivariant features. For cross-source registration, SPEAL (Xiong et al., 2024) incorporates skeletal geometric priors, while Cross-PCR (Zhao et al., 2025b) introduces density-robust feature extraction and a loose-to-strict matching strategy. BUFFER-X (Seo et al., 2025) extends BUFFER (Ao et al., 2023) with adaptive bootstrapping for zero-shot registration, and UGP (Zeng et al., 2025) employs BEV representations to improve generalization across LiDAR configurations. Nevertheless, many correspondence-based methods represent coarse structures using discrete point- or patch-level descriptors, whose consistency may be compromised by heterogeneous sampling patterns. SAFER introduces continuous anisotropic Gaussian supports into coarse feature learning while retaining the correspondence-based registration pipeline.

Probabilistic and Continuous Registration Methods. Probabilistic methods model point clouds as continuous distributions to reduce noise and correspondence ambiguity. CPD (Myronenko & Song, 2010) formulates registration as Gaussian mixture fitting, while DeepGMR (Yuan et al., 2020) and OGMM (Mei et al., 2023) improve distribution-level alignment through learned representations and overlap guidance. Other continuous methods embed point clouds in an RKHS (Clark et al., 2021), and Generalized-CVO (Zhang et al., 2026) further employs anisotropic kernels for correspondence-free alignment. Unlike these distribution- or field-alignment approaches, SAFER uses continuous anisotropic supports to enhance superpoint descriptors within a coarse-to-fine matching framework.

Cross-Source Point Cloud Registration Benchmarks. While same-source registration benefits from benchmarks such as 3DMatch (Zeng et al., 2017) and KITTI (Geiger et al., 2012), large-scale cross-source benchmarks remain limited. 3DCSR (Huang et al., 2021b) contains cross-source pairs but is relatively small for training deep networks. GermanyForest3D (Xiong et al., 2024) focuses on forest scenes, while Cross3DReg (Xu et al., 2025) covers several real-world cross-source settings. CS-Wild-Places (Griffiths et al., 2025) and TIERS (Sier et al., 2023) provide multi-platform or multi-sensor LiDAR sequences, yet were designed for place recognition or SLAM rather than pairwise registration. Based on CS-Wild-Places and TIERS, we construct two challenging cross-source registration benchmarks with large transformations, repetitive structures, and diverse sensor configurations.

3 Method

Problem Statement. Given two point clouds $\mathbf{P} = \{\mathbf{p}_i\}_{i=1}^{N}$ and $\mathbf{Q} = \{\mathbf{q}_j\}_{j=1}^{M}$ acquired by different sensors, cross-source registration aims to estimate a rigid transformation $\mathbf{T} = \{\mathbf{R}, \mathbf{t}\}$ with $\mathbf{R} \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$ that aligns $\mathbf{P}$ to $\mathbf{Q}$:

$$\min_{\mathbf{R},\mathbf{t}} \sum_{(\mathbf{p}_i,\mathbf{q}_j)\in\mathcal{C}} \rho\left(\|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_j\|_2\right), \tag{1}$$

where $\mathcal{C}$ denotes the inlier correspondence set and $\rho(\cdot)$ is a robust cost function. Due to divergent sampling patterns, densities, and noise characteristics across sources, $\mathcal{C}$ is unknown a priori and must

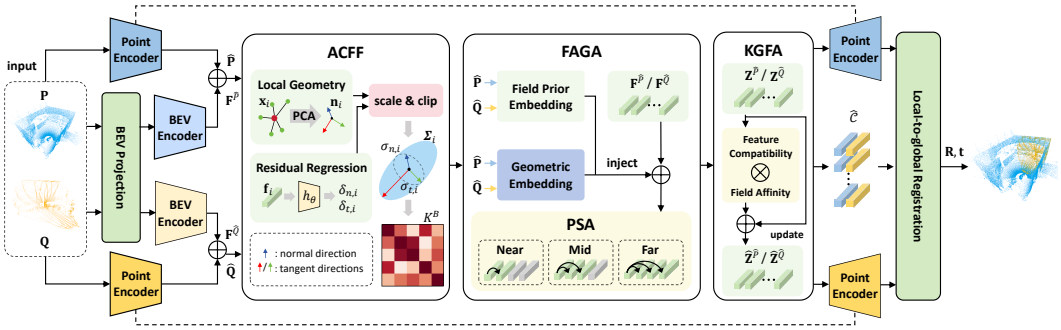


Figure 2: **Overall Pipeline of SAFER.** The backbone extracts superpoints and multi-level features from the input point clouds $\mathbf{P}$ and $\mathbf{Q}$. Building on this, the proposed ACFF represents each superpoint using a surface-adaptive anisotropic Gaussian support with feature-conditioned scale residuals. The resulting field affinity is then integrated into FAGA and KGFA for descriptor refinement. Finally, coarse correspondences are established from the refined descriptors, and the rigid transformation is estimated via local-to-global registration (Qin et al., 2022; Zeng et al., 2025).

be established through discriminative feature learning and robust matching prior to transformation estimation.

Overview and Notations. SAFER follows the coarse-to-fine registration pipeline of UGP (Zeng et al., 2025), as shown in Fig. 2. Given two point clouds $\mathbf{P}$ and $\mathbf{Q}$, we use a sparse 3D backbone to extract multi-level features. The points at the coarsest level are denoted as superpoints $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$, with associated features $\mathbf{F}^{\hat{P}} \in \mathbb{R}^{|\hat{\mathbf{P}}|\times d_t}$ and $\mathbf{F}^{\hat{Q}} \in \mathbb{R}^{|\hat{\mathbf{Q}}|\times d_t}$. Following UGP, BEV cues and the Progressive Self-Attention (PSA) mechanism are employed for coarse correspondence extraction. Building on this, SAFER introduces an anisotropic continuous feature field to refine coarse descriptors and enforce field-level alignment during training. Finally, coarse correspondences are established from the refined descriptors and passed to the point matching module and local-to-global registration for rigid transformation estimation.

3.1 Anisotropic Continuous Feature Field

At the coarse level, conventional methods treat each superpoint as a discrete token. Let $\mathbf{x}_i \in \mathbb{R}^3$ denote the coordinate of the i-th coarse superpoint and $\mathbf{f}_i \in \mathbb{R}^{d_t}$ its descriptor. In cross-source registration, corresponding local structures may be sampled at different positions due to sensor-dependent ray casting; two superpoints representing the same physical structure may therefore not be exactly co-located even after geometric alignment. To reduce sensitivity to such discrete sampling shifts, SAFER augments each discrete superpoint with a continuous field element, formally represented as a Gaussian $g_i(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{x}_i, \mathbf{\Sigma}_i)$, whose covariance $\mathbf{\Sigma}_i \in \mathbb{R}^{3\times3}$ encodes the spatial support of the superpoint. This continuous support reduces dependence on exact superpoint coincidence by allowing spatially shifted samples on the same surface to retain overlapping support.

Directly regressing an unrestricted $\mathbf{\Sigma}_i$ introduces coupled degrees of freedom and requires additional constraints for positive definiteness. We instead use local PCA to initialize a surface-aligned support. For each superpoint, the eigendecomposition of its k-nearest-neighbor covariance matrix yields eigenvalues $\lambda_{i,1} \le \lambda_{i,2} \le \lambda_{i,3}$ and eigenvectors $\mathbf{U}_i$. The eigenvector corresponding to $\lambda_{i,1}$ defines the local normal $\mathbf{n}_i$, while the eigenvalues initialize the tangential and normal scales: $\tilde{\sigma}_{t,i} = \sqrt{(\lambda_{i,2}+\lambda_{i,3})/2}$ and $\tilde{\sigma}_{n,i} = \sqrt{\lambda_{i,1}}$. Rather than using these raw PCA scales directly, a lightweight head h_θ predicts residuals $(\delta_{t,i}, \delta_{n,i})$, which modulate the initial scales as $\sigma_{*,i} = \mathrm{clip}(\tilde{\sigma}_{*,i}\exp(\alpha\tanh(\delta_{*,i})), \sigma_{\min}, \sigma_{\max})$, for $* \in \{t, n\}$, with α controlling the residual influence and clipping limiting excessive scale variation. The final covariance is assembled as

$$\mathbf{\Sigma}_i = \sigma_{t,i}^2(\mathbf{I} - \mathbf{n}_i\mathbf{n}_i^\top) + \sigma_{n,i}^2\mathbf{n}_i\mathbf{n}_i^\top + \epsilon\mathbf{I}. \tag{2}$$

This covariance is axisymmetric about $\mathbf{n}_i$: the PCA statistics initialize a bounded support aligned with the estimated tangent plane, rather than serving as descriptors themselves. The sign ambiguity of $\mathbf{n}_i$ is inconsequential, as only $\mathbf{n}_i\mathbf{n}_i^\top$ appears in Eq. equation 2. For non-degenerate neighbor-

hoods, the normal outer product transforms consistently under rotations. Feature-conditioned residuals and scale clipping further adapt the support while limiting excessive variation from irregular local sampling.

To measure the compatibility between two field elements, we adopt the normalized Gaussian overlap. Defining the square-root Gaussian basis $\psi_i(\mathbf{x}) = \sqrt{g_i(\mathbf{x})}$, the field affinity between superpoints i and j is their L_2 inner product:

$$K_{ij}^B = \langle \psi_i, \psi_j \rangle_{L_2} = \int \sqrt{g_i(\mathbf{x}) g_j(\mathbf{x})}\, d\mathbf{x}, \tag{3}$$

which is exactly the Bhattacharyya coefficient between two Gaussians and has the closed form:

$$K_{ij}^B = \exp\left(-\frac{1}{8} \delta_{ij}^\top \bar{\mathbf{\Sigma}}_{ij}^{-1} \delta_{ij} - \frac{1}{2} \log \frac{|\bar{\mathbf{\Sigma}}_{ij}|}{\sqrt{|\mathbf{\Sigma}_i||\mathbf{\Sigma}_j|}} \right), \tag{4}$$

where $\delta_{ij} = \mathbf{x}_i - \mathbf{x}_j$ and $\bar{\mathbf{\Sigma}}_{ij} = (\mathbf{\Sigma}_i + \mathbf{\Sigma}_j)/2$. The first term penalizes the center displacement under the average covariance, while the second penalizes scale and orientation mismatch between the two supports. Consequently, superpoints with similar centers but incompatible covariance structures can receive lower affinity, whereas spatially shifted samples with substantially overlapping supports can retain relatively high affinity. In practice, the fields are never evaluated densely; the closed-form affinity K_{ij}^B serves as a continuous geometric prior for field-aware attention, kernel-guided descriptor refinement, and field-level supervision.

3.2 Field-Aware Geometric Attention

To incorporate continuous spatial support into context aggregation, we introduce Field-Aware Geometric Attention (FAGA), which injects anisotropic field affinity directly into intra-cloud self-attention. The structure and computational graph of FAGA are detailed in Fig. 3. Given an input feature matrix $\mathbf{F} \in \mathbb{R}^{N \times d_t}$, where $N = |\hat{\mathbf{P}}|$ is the number of superpoints, let $\mathbf{f}_i \in \mathbb{R}^{d_t}$ denote the feature vector of the i-th superpoint. The output feature matrix $\mathbf{Z} \in \mathbb{R}^{N \times d_t}$ is computed as the weighted sum of projected input features:

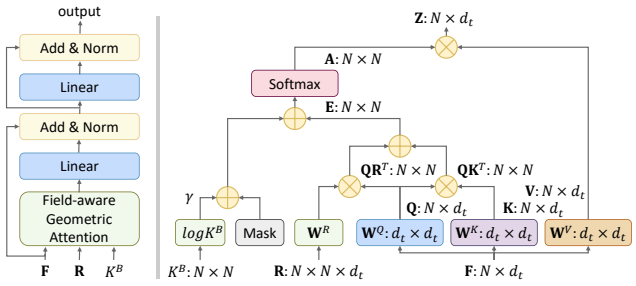


Figure 3: The structure (left) and computational graph (right) of Field-Aware Geometric Attention.

$$\mathbf{z}_i = \sum_{j=1}^{N} a_{ij} \left(\mathbf{f}_j \mathbf{W}^V \right), \tag{5}$$

where a_{ij} is the weight coefficient computed by a row-wise softmax on the attention score e_{ij}. Specifically, the field-aware attention logit e_{ij} is computed as:

$$e_{ij} = \frac{\left(\mathbf{f}_i \mathbf{W}^Q \right) \left(\mathbf{f}_j \mathbf{W}^K + \mathbf{r}_{ij} \mathbf{W}^R \right)^\top}{\sqrt{d_t}} + \gamma \log(K_{ij}^B + \epsilon) + m_{ij}. \tag{6}$$

Here, $\mathbf{r}_{ij} \in \mathbb{R}^{d_t}$ represents the relative geometric embedding between superpoints i and j (Qin et al., 2022). K_{ij}^B is the continuous field affinity defined in Eq. equation 4, γ is a scalar controlling the influence of continuous support priors, and ϵ ensures numerical stability. m_{ij} denotes the self-attention mask. For valid pairs, $m_{ij} = 0$; for masked pairs, $m_{ij} = -\infty$. The projection matrices $\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V, \mathbf{W}^R \in \mathbb{R}^{d_t \times d_t}$ project features and geometric embeddings into query, key, value, and position representations, respectively.

Through this formulation, FAGA combines feature compatibility, relative geometric information, and field-support compatibility within the attention score. When superpoints share compatible surface supports, their attention weight is boosted, whereas interactions between geometrically incompatible structures are effectively suppressed.

3.3 KERNEL-GUIDED FEATURE ADAPTER

Although FAGA introduces the field prior inside the transformer, the resulting coarse descriptors may still retain variations associated with local sampling patterns. We therefore introduce a lightweight Kernel-Guided Feature Adapter (KGFA) after the coarse transformer, which performs an additional residual descriptor refinement, aggregating neighboring superpoints weighted by both feature compatibility and field affinity.

Let $\mathbf{z}_i$ be the output descriptor of superpoint $\mathbf{x}_i$. KGFA computes adapter-specific query, key, and value projections from $\mathbf{z}_i$, using the same notation $\mathbf{q}_i$, $\mathbf{k}_j$, and $\mathbf{v}_j$ for simplicity. The aggregation weight from superpoint j to superpoint i is defined as

$$\beta_{ij} = \frac{K_{ij}^B \exp\left(\mathbf{q}_i^\top \mathbf{k}_j / \sqrt{d}\right)}{\sum_l K_{il}^B \exp\left(\mathbf{q}_i^\top \mathbf{k}_l / \sqrt{d}\right) + \epsilon}. \tag{7}$$

Here, the exponential term measures feature compatibility, while K_{ij}^B acts as a continuous geometric gate. Thus, even if two descriptors are similar, their aggregation is weakened when their anisotropic supports are incompatible.

The kernel-guided update is then added to the coarse descriptor through a scaled residual connection:

$$\hat{\mathbf{z}}_i = \mathbf{z}_i + \eta \mathbf{W}_o \sum_j \beta_{ij} \mathbf{v}_j, \tag{8}$$

where $\mathbf{W}_o$ is the output projection and η is a residual scale controlling the strength of KGFA. The refined descriptors are subsequently normalized and passed to the coarse-to-fine matching pipeline.

KGFA can be viewed as a field-consistent descriptor refinement step. It encourages information exchange among superpoints with compatible local supports while reducing propagation between geometrically inconsistent neighborhoods. The residual scale η keeps this refinement conservative, so that KGFA refines the coarse representation without overriding the transformer output.

3.4 LOSSES

The total training objective follows the baseline (Qin et al., 2022) with an additional overlap-aware field loss:

$$\mathcal{L} = \lambda_c \mathcal{L}_c + \lambda_f \mathcal{L}_f + \lambda_\phi \mathcal{L}_\phi. \tag{9}$$

To complement point-wise correspondence supervision, we introduce a soft field-level objective $\mathcal{L}_\phi$ that requires no additional one-to-one assignment between superpoints. Given the ground-truth transformation $\mathbf{T}^\star = \{\mathbf{R}^\star, \mathbf{t}^\star\}$, the source superpoints and their covariance matrices are first transformed into the reference frame:

$$\bar{\mathbf{x}}_i^P = \mathbf{R}^\star \mathbf{x}_i^P + \mathbf{t}^\star, \quad \bar{\mathbf{\Sigma}}_i^P = \mathbf{R}^\star \mathbf{\Sigma}_i^P (\mathbf{R}^\star)^\top. \tag{10}$$

We restrict field-level supervision to mutually observable regions and define the overlap set of $\mathcal{P}$ as

$$\mathcal{O}_P = \left\{ i \mid \min_j \left\| \bar{\mathbf{x}}_i^P - \mathbf{x}_j^Q \right\|_2 \le r_o \right\}, \tag{11}$$

where r_o is the overlap radius. $\mathcal{O}_Q$ is defined similarly. The overlap-aware cross-field inner product is then computed as

$$\langle \Phi_{\mathbf{P}}, \Phi_{\mathbf{Q}} \rangle_{\mathcal{O}} = \sum_{i \in \mathcal{O}_P} \sum_{j \in \mathcal{O}_Q} K^B \left(\bar{\mathbf{x}}_i^P, \mathbf{x}_j^Q; \bar{\mathbf{\Sigma}}_i^P, \mathbf{\Sigma}_j^Q \right) (\hat{\mathbf{z}}_i^P)^\top \hat{\mathbf{z}}_j^Q. \tag{12}$$

Based on this inner product, we jointly formulate the positive and negative components of the field loss. We define the square-root Gaussian basis as $\psi_i(\mathbf{x}) = \sqrt{\mathcal{N}(\mathbf{x}; \bar{\mathbf{x}}_i, \bar{\mathbf{\Sigma}}_i)}$, which defines the continuous feature fields $\Phi_{\mathbf{P}}$ and $\Phi_{\mathbf{Q}}$ over their respective overlap sets. The positive alignment loss maximizes the normalized correlation between the two aligned fields:

$$\mathcal{L}_{\text{pos}} = 1 - \frac{\langle \Phi_{\mathbf{P}}, \Phi_{\mathbf{Q}} \rangle_{\mathcal{O}}}{\sqrt{\langle \Phi_{\mathbf{P}}, \Phi_{\mathbf{P}} \rangle_{\mathcal{O}_P} \langle \Phi_{\mathbf{Q}}, \Phi_{\mathbf{Q}} \rangle_{\mathcal{O}_Q}} + \epsilon}. \tag{13}$$

To discourage trivially high field correlation, we construct a negative counterpart $\Phi_{\mathbf{P}}^{-}$ by cyclically permuting the source descriptors $\{\hat{\mathbf{z}}_i^P\}$ among the overlap superpoints while retaining their geometries $\{\bar{\mathbf{x}}_i^P, \bar{\mathbf{\Sigma}}_i^P\}$. Its normalized correlation with the reference field is denoted as S_ϕ^-, computed analogously to the fractional term in $\mathcal{L}_{\text{pos}}$. We penalize this negative correlation with a margin:

$$\mathcal{L}_{\text{neg}} = \max(0,\ S_\phi^- - \tau), \tag{14}$$

where τ is a preset margin. The overall field loss is compactly defined as the weighted sum of the positive and negative terms:

$$\mathcal{L}_\phi = \mathcal{L}_{\text{pos}} + \lambda_{\text{neg}}\mathcal{L}_{\text{neg}}. \tag{15}$$

By supervising the consistency of aligned continuous fields, $\mathcal{L}_\phi$ provides a soft many-to-many alignment signal that complements the discrete correspondence losses and is robust to cross-source ray-casting mismatches. The overlap restriction prevents non-overlapping regions from being erroneously forced to share similar descriptors, while the coarse and fine losses preserve the standard correspondence-based registration objective.

4 Experiments

4.1 Datasets and Experimental Setup

Cross-Source Datasets. Large-scale cross-source datasets for point cloud registration in outdoor and indoor scenes remain scarce. To bridge this gap, we construct two cross-source registration datasets: the *Forest Cross-Source* dataset with large translation, rotation, and repetitive structures, and the *Indoor Multi-LiDAR* dataset with diverse LiDAR sensors. Additionally, we evaluate on the *KITTI Cross-Source* dataset (Xiong et al., 2024).

The *Forest Cross-Source* dataset is derived from the CS-Wild-Places dataset (Griffiths et al., 2025), which features ground and aerial LiDAR submaps collected from forests in Brisbane, Australia. The ground data were captured by a handheld VLP-16 LiDAR, while the aerial data were acquired via UAVs equipped with VLP-32C or RIEGL VUX-120 LiDARs. We construct independent training and test splits for both QCAT and Samford and train a separate model for each scene.

The *Indoor Multi-LiDAR* dataset is built upon the TIERS dataset (Sier et al., 2023), which provides point clouds from diverse LiDAR sensors including Ouster OS0 (128 channels), Ouster OS1 (64 channels), Velodyne VLP-16, Livox Horizon, and Livox Avia. This dataset comprises four cross-source configurations: OS0-to-OS1, OS0-to-Velodyne, OS0-to-Avia, and OS0-to-Horizon. We use indoor sequences 08, 10, and 11 for training, and sequence 06 for evaluation. Please refer to the supplementary material for details.

Evaluation Metrics. Following previous practices (Qin et al., 2022), we adopt Relative Rotation Error (RRE), Relative Translation Error (RTE), and Registration Recall (RR). RRE and RTE are averaged over successfully registered pairs. A successful registration requires RRE $< 5^\circ$, together with RTE $< 2\,\text{m}$ for *Forest Cross-Source* and *KITTI Cross-Source*, and RTE $< 0.5\,\text{m}$ for *Indoor Multi-LiDAR*.

Implementation Details. To train SAFER, we use the Adam optimizer (Kingma & Ba, 2014) with an initial learning rate of 10^{-4} and a weight decay of 10^{-6}. We train SAFER for 80 epochs with a batch size of 1 on a NVIDIA RTX 4090 GPU.

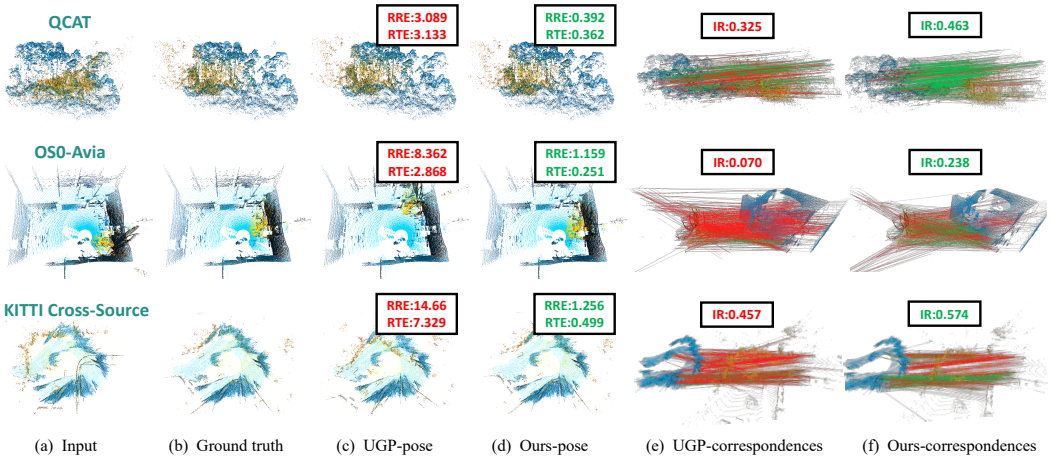
4.2 Comparison Experiments

Table 1 and Table 2 summarize the quantitative results across indoor and outdoor cross-source benchmarks. As visualized in Fig. 4, SAFER achieves higher Inlier Ratios (IRs) and lower pose errors by establishing more precise correspondences, mitigating spurious matches caused by sampling shifts.

On the *Forest Cross-Source* benchmark, SAFER achieves the highest RR of 97.4% on QCAT and 100.0% on Samford with reduced pose errors. This suggests that the proposed field prior can provide useful geometric regularization in cluttered environments with large pose shifts.

Table 1: Results on the *Forest Cross-Source* and *Indoor Multi-LiDAR* benchmarks. "–" denotes convergence failure.

Model	Forest Cross-Source						Indoor Multi-LiDAR											
	QCAT			Samford			OS0 to OS1			OS0 to Velo			OS0 to Avia			OS0 to Horizon		
	RRE	RTE	RR	RRE	RTE	RR	RRE	RTE	RR	RRE	RTE	RR	RRE	RTE	RR	RRE	RTE	RR
Predator		–			–		1.14	0.06	99.9	1.56	0.09	99.1		–			–	
GeoTransformer	0.37	0.29	95.9	0.16	0.31	**100.0**	0.92	0.06	**100.0**	1.29	**0.07**	**99.7**	**1.96**	0.16	88.4	2.74	0.21	67.4
BUFFER-X	1.93	0.51	69.8	1.10	0.28	96.4	1.38	0.08	99.8	2.29	0.15	86.5	2.88	0.20	54.4	2.90	0.22	39.7
PARE-Net	1.80	1.05	19.3	1.64	1.29	3.8	1.20	0.17	99.1	1.44	0.24	87.5	2.24	0.26	47.1	2.61	0.30	21.2
UGP	0.35	0.28	95.9	0.44	0.37	**100.0**	**0.89**	**0.05**	**100.0**	1.25	**0.07**	99.6	1.99	0.16	86.8	2.69	0.21	67.7
SAFER	**0.29**	**0.23**	**97.4**	**0.12**	**0.23**	**100.0**	**0.89**	**0.05**	**100.0**	**1.23**	**0.07**	**99.7**	**1.96**	**0.15**	**89.0**	**2.60**	**0.20**	**69.1**



(a) Input (b) Ground truth (c) UGP-pose (d) Ours-pose (e) UGP-correspondences (f) Ours-correspondences

Figure 4: Qualitative comparison of correspondences and registration results against UGP (Zeng et al., 2025).

On the challenging non-repetitive scanning pairs (OS0-to-Avia and OS0-to-Horizon), SAFER leads all baselines with RR of 89.0% and 69.1%, outperforming the strong baseline UGP by 2.2 and 1.4 percentage points (pp), respectively. These configurations couple mechanical spinning LiDARs with solid-state LiDARs, presenting drastic discrepancies in field of view, point density, and scanning patterns. Our gains align with the motivation that continuous anisotropic supports reduce sensitivity to heterogeneous sampling. For homogeneous spinning LiDAR pairs (OS0-to-OS1 and OS0-to-Velo), performance is nearly saturated; SAFER achieves 100.0% and 99.7% RR, while maintaining comparable pose accuracy, indicating that the additional field prior does not introduce an evident performance degradation on these relatively saturated settings.

As shown in Table 2, on the *KITTI Cross-Source* benchmark, SAFER yields a top RR of 97.6% and the lowest RTE of 0.48 m, outperforming UGP and SPEAL. Overall, these evaluations indicate that SAFER effectively adapts to various cross-source scenarios, with notable advantages under severe modality and sampling shifts.

Table 2: Results on the *KITTI Cross-Source benchmark.*

Model	RRE (°)	RTE (m)	RR (%)
Predator	5.06	2.59	34.2
GeoTransformer	1.87	0.63	96.8
BUFFER-X	**0.32**	1.26	18.0
PARE-Net	1.76	1.08	42.7
UGP	1.04	0.53	97.1
SPEAL	1.41	0.58	97.3
SAFER	0.94	**0.48**	**97.6**

4.3 Generalization Experiments

To evaluate model resilience against severe sensor domain shifts, we establish an unseen sensor-pair transfer protocol. Models are trained exclusively on cross-source pairs composed of mechanical spinning LiDARs (*e.g.*, OS0-to-OS1 or OS0-to-Velo) and directly evaluated on unseen configurations without fine-tuning. This setup presents a significant challenge: transferring representations

Table 3: **Cross-source generalization on unseen LiDAR-pair configurations.** Models are trained exclusively on mechanical spinning LiDAR pairs and directly evaluated on unseen sensor pairs without fine-tuning, *e.g.*, "OS1⇒Avia" denotes training on OS0-to-OS1 pairs and testing on OS0-to-Avia pairs.

Model	OS1⇒Velo			OS1⇒Avia			OS1⇒Horizon			Velo⇒OS1			Velo⇒Avia			Velo⇒Horizon		
	RRE	RTE	RR	RRE	RTE	RR	RRE	RTE	RR	RRE	RTE	RR	RRE	RTE	RR	RRE	RTE	RR
Predator	1.78	0.10	98.1	2.86	0.22	24.7	3.09	0.24	7.9	1.56	0.09	88.1	2.71	0.21	18.4	2.91	0.22	15.0
GeoTransformer	1.31	**0.07**	**99.7**	2.30	0.19	69.8	2.45	0.20	42.4	1.05	0.06	**100.0**	2.25	0.17	74.5	2.44	**0.19**	62.4
BUFFER-X	2.07	0.13	93.8	2.72	0.19	67.5	2.80	0.21	49.7	1.47	0.09	99.6	2.85	0.20	58.6	2.86	0.22	42.3
PARE-Net	1.74	0.13	94.8	2.90	0.20	17.8	3.51	0.26	2.2	1.48	0.11	99.9	3.24	0.21	6.3	3.23	0.23	3.9
UGP	1.30	**0.07**	**99.7**	**2.10**	**0.16**	70.5	**2.34**	**0.18**	47.1	1.10	0.06	**100.0**	2.28	**0.16**	72.8	2.41	**0.19**	63.3
SAFER	**1.26**	**0.07**	**99.7**	2.29	**0.16**	**73.9**	2.40	0.19	**56.3**	**0.96**	**0.05**	**100.0**	**2.14**	**0.16**	**80.2**	**2.39**	**0.19**	**64.6**

learned from regular, mechanical scan patterns to non-repetitive solid-state LiDARs (*e.g.*, Avia and Horizon), which exhibit drastic discrepancies in point density, field of view, and spatial sampling.

As summarized in Table 3, SAFER demonstrates competitive generalization, achieving the best RRs across all six unseen transfer scenarios. The improvement of SAFER is most visible in the most challenging transfers from spinning to solid-state LiDARs. For instance, on the OS1⇒Horizon task, SAFER boosts RR from 47.1% (UGP) to 56.3% (+9.2 pp). Similarly, on OS1⇒Avia and Velo⇒Avia, SAFER achieves clear RR gains, reaching 73.9% and 80.2%, respectively. Even on homogeneous transfers (OS1⇒Velo and Velo⇒OS1), SAFER maintains top-tier performance with near-saturated recall and superior pose accuracy.

Discussion. These empirical findings are consistent with our underlying motivation: conventional discrete representations tend to be sensitive to sensor-specific sampling patterns. By modeling surface-adaptive continuous spatial supports rather than rigid discrete neighborhoods, SAFER helps alleviate representation discrepancies induced by heterogeneous scanning mechanisms, thereby promoting feature compatibility across challenging domain shifts.

4.4 Ablation Studies

We conduct ablation studies on the OS0-to-Avia benchmark to evaluate the efficacy of the proposed designs under heterogeneous LiDAR sampling. We report registration errors and recall in Table 4. SAFER achieves the best overall performance, demonstrating the effectiveness of modeling surface-adaptive feature fields for cross-source registration. The improvement from the field loss indicates that continuous field-level supervision complements discrete correspondence learning. The full model further improves RR to 89.0%, which can be attributed to the joint effect of adaptive covariance prediction and field-aware feature refinement. Without residual covariance regression, the Gaussian support relies on fixed PCA-derived scales and yields weaker registration accuracy. Removing FAGA or KGFA also leads to lower recall, suggesting that field affinity is useful both for contextual aggregation and for post-transformer feature adaptation, where it helps preserve coherent structures while suppressing residual sampling noise.

Table 4: Ablation studies on the OS0-to-Avia benchmark.

Methods	RRE	RTE	RR
Baseline	1.99	0.16	86.8
+ Field Loss	**1.96**	**0.15**	87.4
W/o Residual Regression	1.98	**0.15**	87.9
W/o FAGA	1.97	**0.15**	88.8
W/o KGFA	1.98	**0.15**	88.0
Full	**1.96**	**0.15**	**89.0**

5 Conclusion

In this paper, we propose SAFER, a geometry-aware feature learning framework for cross-source point cloud registration. Our method associates each coarse superpoint with a continuous anisotropic Gaussian support constructed from local surface geometry. We further introduce FAGA and KGFA to integrate the resulting field affinity into contextual aggregation and descriptor refinement. Extensive experiments demonstrate that SAFER effectively improves registration robustness, particularly for heterogeneous LiDAR pairs with substantial sampling discrepancies.

AI USE STATEMENT

We used AI-based tools to assist with early-stage literature retrieval and discovery, writing polishing, figure and table presentation refinement, and limited modifications of analysis scripts used for result organization and visualization. These tools were used only as auxiliary support. The research ideas, method design, experimental protocol, implementation decisions, empirical results, and conclusions were developed, verified, and approved by the authors.

REFERENCES

Sheng Ao, Qingyong Hu, Bo Yang, Andrew Markham, and Yulan Guo. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11753–11762, 2021.

Sheng Ao, Yulan Guo, Qingyong Hu, Bo Yang, Andrew Markham, and Zengping Chen. You only train once: Learning general and distinctive 3d local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3949–3967, 2022.

Sheng Ao, Qingyong Hu, Hanyun Wang, Kai Xu, and Yulan Guo. Buffer: Balancing accuracy, efficiency, and generalizability in point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1255–1264, 2023.

Yasuhiro Aoki, Hunter Goforth, Rangaprasad Arun Srivatsan, and Simon Lucey. Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7163–7172, 2019.

Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2514–2523, 2020.

William Clark, Maani Ghaffari, and Anthony Bloch. Nonparametric continuous sensor registration. *Journal of Machine Learning Research*, 22(271):1–50, 2021.

Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pp. 3354–3361. IEEE, 2012.

Ethan Griffiths, Maryam Haghighat, Simon Denman, Clinton Fookes, and Milad Ramezani. Hotformerloc: Hierarchical octree transformer for versatile lidar place recognition across ground and aerial views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6648–6658, 2025.

Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pp. 4267–4276, 2021a.

Xiaoshui Huang, Guofeng Mei, Jian Zhang, and Rana Abbas. A comprehensive survey on point cloud registration. *arXiv preprint arXiv:2103.02690*, 2021b.

Xiaoshui Huang, Guofeng Mei, and Jian Zhang. Cross-source point cloud registration: Challenges, progress and prospects. *Neurocomputing*, 548:126383, 2023.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Qingqing Li, Xianjia Yu, Jorge Pena Queralta, and Tomi Westerlund. Multi-modal lidar dataset for benchmarking general-purpose localization and mapping algorithms. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3837–3844. IEEE, 2022.

Hyungtae Lim, Minho Oh, and Hyun Myung. Patchwork: Concentric zone-based region-wise ground segmentation with ground likelihood estimation using a 3d lidar sensor. *IEEE Robotics and Automation Letters*, 6(4):6458–6465, 2021.

Guofeng Mei, Fabio Poiesi, Cristiano Saltori, Jian Zhang, Elisa Ricci, and Nicu Sebe. Overlap-guided gaussian mixture models for point cloud registration. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 4511–4520, 2023.

Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010.

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.

Minkyun Seo, Hyungtae Lim, Kanghee Lee, Luca Carlone, and Jaesik Park. Buffer-x: Towards zero-shot point cloud registration in diverse scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3851–3862, 2025.

Ha Sier, Qingqing Li, Xianjia Yu, Jorge Peña Queralta, Zhuo Zou, and Tomi Westerlund. A benchmark for multi-modal lidar slam with ground truth in gnss-denied environments. *Remote Sensing*, 15(13):3314, 2023.

Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3523–3532, 2019.

Jianshi Wu, Minghang Zhu, Dunqiang Liu, Wen Li, Sheng Ao, Siqi Shen, Chenglu Wen, and Cheng Wang. Leader: Learning reliable local-to-global correspondences for lidar relocalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9932–9942, June 2026.

Kezheng Xiong, Maoji Zheng, Qingshan Xu, Chenglu Wen, Siqi Shen, and Cheng Wang. Speal: Skeletal prior embedded attention learning for cross-source point cloud registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 6279–6287, 2024.

Zongyi Xu, Zhongpeng Lang, Yilong Chen, Shanshan Zhao, Xiaoshui Huang, Yifan Zuo, Yan Zhang, Qianni Zhang, and Xinbo Gao. Cross3dreg: Towards a large-scale real-world cross-source point cloud registration benchmark. *arXiv preprint arXiv:2509.06456*, 2025.

Runzhao Yao, Shaoyi Du, Wenting Cui, Canhui Tang, and Chengwu Yang. Pare-net: Position-aware rotation-equivariant networks for robust point cloud registration. In *European Conference on Computer Vision*, pp. 287–303. Springer, 2024.

Zi Jian Yew and Gim Hee Lee. Rpm-net: Robust point matching using learned features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11824–11833, 2020.

Wentao Yuan, Benjamin Eckart, Kihwan Kim, Varun Jampani, Dieter Fox, and Jan Kautz. Deepgmr: Learning latent gaussian mixture models for registration. In *European conference on computer vision*, pp. 733–750. Springer, 2020.

Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1802–1811, 2017.

Zhenxuan Zeng, Qiao Wu, Xiyu Zhang, Lin Yuanbo Wu, Pei An, Jiaqi Yang, Ji Wang, and Peng Wang. Unlocking generalization power in lidar point cloud registration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 22244–22253, 2025.

Ray Zhang, Marcus Greiff, Thomas Lew, and John Subosits. Generalized-cvo: Fast and correspondence-free local point cloud registration with second order riemannian optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2948–2958, 2026.

Guiyu Zhao, Sheng Ao, Ye Zhang, Kai Xu, and Yulan Guo. Progressive correspondence regenerator for robust 3d registration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 1210–1219, 2025a.

Guiyu Zhao, Zhentao Guo, Zewen Du, and Hongbin Ma. Cross-pcr: A robust cross-source point cloud registration framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 10403–10411, 2025b.