# Supplemental Material

Chengyu Zheng[1,2*] Jin Huang[1*] Honghua Chen[3†] Mingqiang Wei[1,2†]
[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics
[2]Shenzhen Research Institute, Nanjing University of Aeronautics and Astronautics
[3]School of Data Science, Lingnan University, Hong Kong SAR, China
zhengcy@nuaa.edu.cn; jinhuang.nuaa@gmail.com; chenhonghuacn@gmail.com; mqwei@nuaa.edu.cn

## 1. Ablation of Different Initial Transformations and Features

Our method takes both the initial transformation and 3D features as inputs, allowing us to experiment with different combinations of initial transformations and 3D features from various methods. The results are presented in Table 1. In the table, the method on the left side of the plus sign provides the initial transformation, while the method on the right side provides the 3D features. As shown in the table, the quality of the initial transformation has a significant impact on the registration results when using different initial transformations with the same features, which aligns with the limitations discussed previously. Furthermore, when the same initial transformation is used with different features, the choice of features also influences the registration outcomes. This is quite obvious. However, diffusion features, serving as supplementary features, can still further enhance the registration performance based on 3D features.

Table 1. Ablation of Different Initial Transformations And Features.

| | 3DMatch | | |
|---|---|---|---|
| | RR ↑ | RE ↓ | TE ↓ |
| FCGF [4] | 67.9 | 2.20 | 0.078 |
| GeoTransformer [8] | 92.0 | 1.72 | 0.062 |
| FCGF + GeoTransformer | 93.2 | 1.52 | 0.057 |
| GeoTransformer + GeoTransformer | 94.5 | 1.56 | 0.050 |
| FCGF + FCGF | 76.2 | 2.07 | 0.055 |
| GeoTransformer + FCGF | 90.9 | 2.08 | 0.069 |

## 2. The Impact of Initial Transformation on GeoTransformer

In addition, we have also verified the impact of initial transformations on point cloud registration methods, as shown in Table 2. From the results, it can be observed that initial transformations can indeed optimize point cloud registration methods, albeit with less significant improvement compared to the use of diffusion features. Although theoretically, geometry-based 3D feature descriptors should be insensitive to transformations, this insight inspires us to potentially leverage them in the future to further enhance optimization effects. Alternatively, we can aim to suppress this sensitivity and improve the generalization capability of 3D feature descriptors. The plus sign in the table has the same meaning as in Table 1.

Table 2. Ablation of Components.

| | 3DMatch | | |
|---|---|---|---|
| | RR ↑ | RE ↓ | TE ↓ |
| GeoTransformer [8] | 92.0 | 1.72 | 0.062 |
| GeoTransformer + GeoTransformer | 92.7 | 1.65 | 0.057 |
| our | **94.5** | **1.56** | **0.050** |

Table 3. Registration results based on FCGF [4].

| Methods | RR↑ | RE↓ | TE↓ |
|---|---|---|---|
| FCGF [4] | 67.9 | 2.20 | 0.078 |
| SC$^2$-PCR [3] | 93.1 | 2.09 | 0.065 |
| PointDSC$^\dagger$ [1] | 91.8 | 2.10 | 0.065 |
| VBReg$^\dagger$ [6] | 82.7 | 2.14 | 0.067 |
| MAC [15] | **93.7** | 2.02 | 0.060 |
| FastMAC [16] | 92.6 | **2.00** | 0.064 |
| our | 76.2 | 2.07 | **0.055** |

## 3. Limitations

While our method achieves certain improvements over certain low-accuracy approaches, its performance is inferior to other outlier removal methods, as indicated in Table 3. $^\dagger$ represents the learning-based method. Specifically, we use FCGF to generate the initial transformations and evaluate our method with other outlier removal approaches. Although the registration recall achieved by our method was

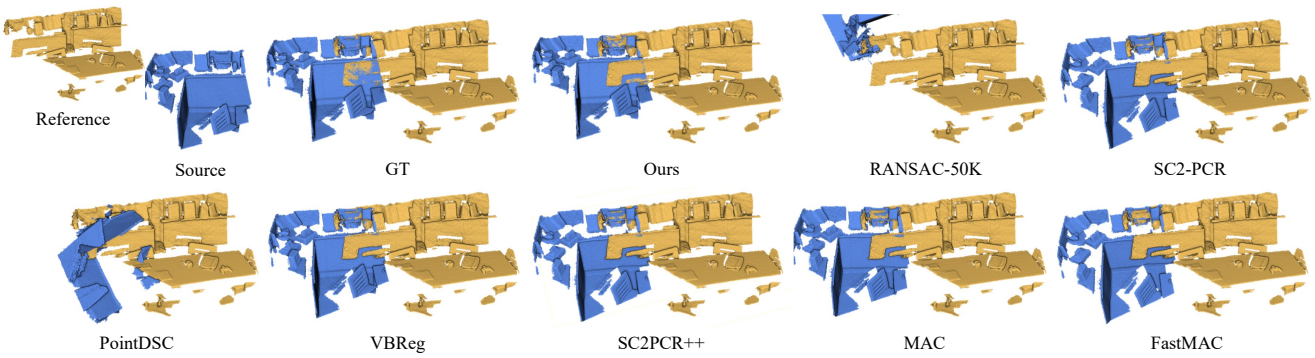


Figure 1. Registration visualization results.

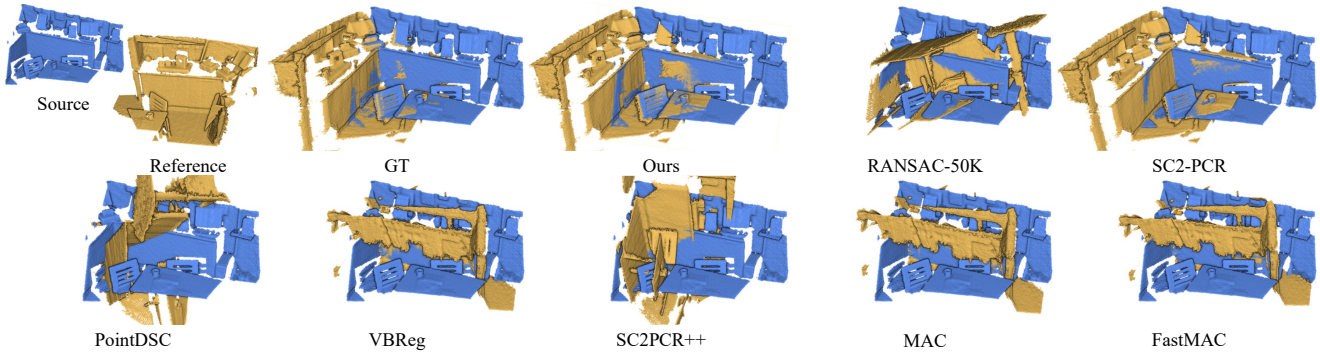


Figure 2. Registration visualization results.

lower than expected, the results remain competitive when evaluated on the RR and ER metrics. Furthermore, our approach still achieves an 8.3% improvement over the baseline FCGF approach. These results are consistent with what we mentioned in the introduction: incorrect initial transformations can lead to large parallax, which in turn affects the accuracy of correspondences obtained through diffusion features. This motivates us to explore multi-view approaches in the future to reduce parallax and mitigate the impact of erroneous initialization. Additionally, our method requires more time compared to other approaches due to the necessity of executing multiple reverse diffusion processes. However, recent research in accelerated sampling techniques and parallel computation architectures suggests that future implementations could achieve a balance between computational efficiency and generative performance.

## 4. Visualizations

We show more registration results in Figure 1, Figure 2, Figure 3 and Figure 4. These images demonstrate the registration results of our method on other paired point cloud examples. As shown, our approach achieves visually compelling results comparable to those of other refinement methods.

Table 4. Results on different depth map generation strategies.

| Method | RR↑ | RE↓ | TE↓ |
|---|---|---|---|
| w/o depth | 92.0 | 1.72 | 0.06 |
| Generated from RGB images | 93.4 | 1.62 | 0.06 |
| Generated from point cloud (ours) | 94.5 | 1.56 | 0.05 |

## 5. Ablation of different depth map generation strategies

The depth maps used in our method are not acquired from dedicated depth sensors, but are projected from point cloud data already available during training and testing. This avoids any additional hardware cost and does not limit deployment feasibility.

To evaluate the effectiveness of the depth obtained by different depth map generation methods, we tested our method with depth maps estimated from single-view RGB images (see Table 4). The results show that while both depth variants outperform the depth-free baseline, depth maps generated from RGB images yield slightly lower performance compared to those projected from point clouds. This is expected, as our point cloud-derived depth maps are reconstructed from multi-view geometry, providing more accurate and detailed spatial cues than single-view RGB-

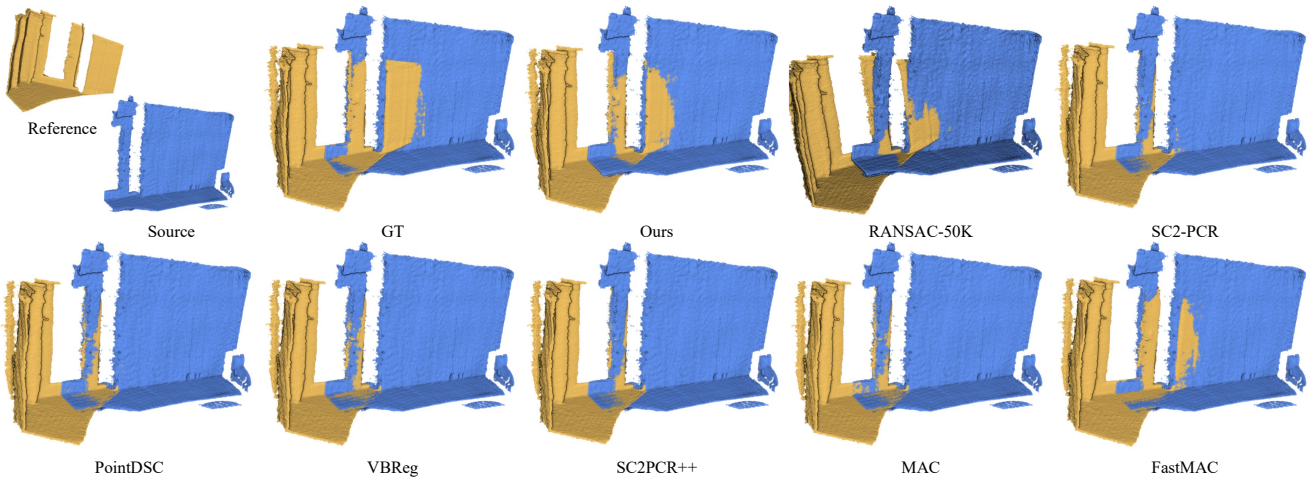


Figure 3. Registration visualization results.

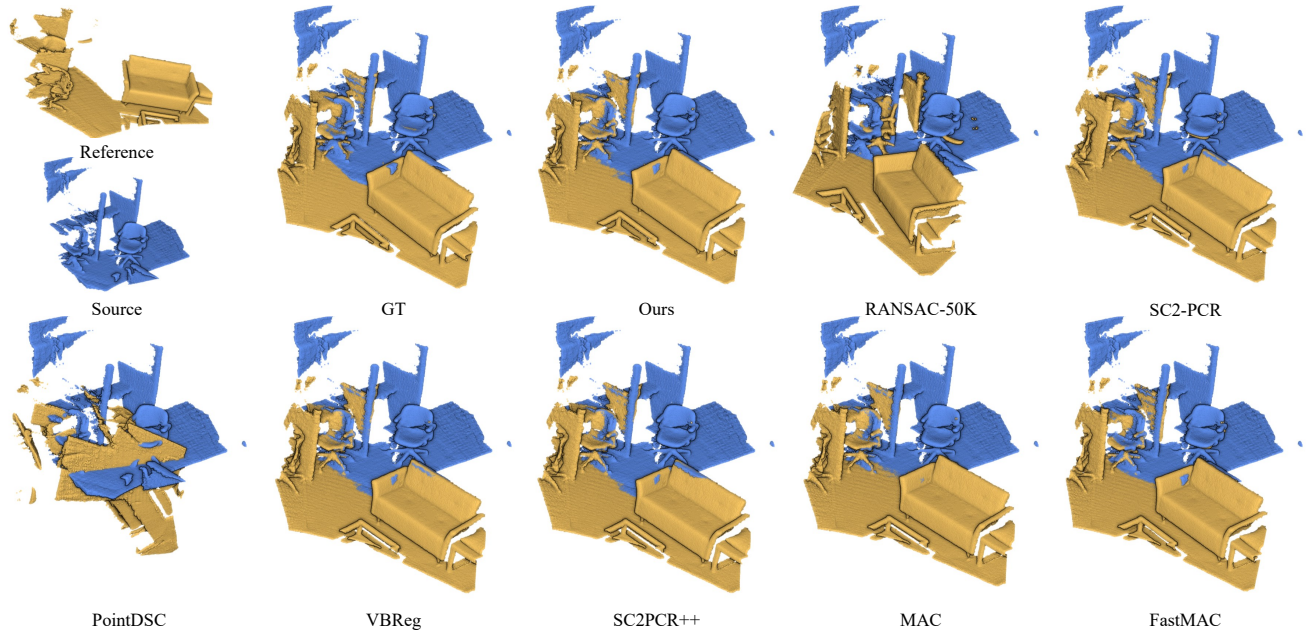


Figure 4. Registration visualization results.

based estimations. Nonetheless, this experiment demonstrates that our method remains effective with estimated depths, making it adaptable to real-world settings where RGB input is available.

## 6. Ablation of Running Time

Our runtime analysis on 3DMatch (Table 5) shows higher latency due to ControlNet's iterative denoising steps (2× slower than ZeroReg). However, this trade-off enables significant performance gains (Table 6). Notably, recent research has increasingly focused on optimizing diffusion processes, including fast samplers (e.g., Analytic-DPM [2]) and timestep distillation techniques (e.g., SDXL Turbo [9]), demonstrating the potential to reduce time.

## 7. Comparison with other non-refine-based point cloud registration methods

We first compare our method with the zero-shot registration SOTA method, ZeroReg [14]. While both ZeroReg and our method adopt zero-shot point cloud registration frameworks, our approach is distinguished by the integration of diffusion models as 2D image feature extractors. This strategy is validated in semantic correspondence estimation [12, 13]. Experiments on 3DMatch (Table 6) show that

our approach yields a 10.1% improvement over ZeroReg, demonstrating the effectiveness of diffusion features.

Furthermore, for the supervised SOTA method PSReg [5], we do not exceed its performance (Table 6) which is consistent with the limitation of training-free refinement frameworks. Specifically, PSReg employs end-to-end supervised training, whereas our method refines correspondences from the GeoTranformer baseline, whose performance is slightly inferior to PSReg. Critically, our method is model-agnostic: it can be seamlessly integrated with any SOTA registration backbone (e.g., replacing GeoTransformer with stronger baselines), offering flexibility to bridge this performance gap in future work.

Table 5. Running time on 3DMatch.

| | MAC | ZeroReg | Ours |
|---|---|---|---|
| Time(s) | 16.4 | 11.2 | 23.1 |

Table 6. Comparison with ZeroReg and PSReg on 3DMatch.

| Metric | ZeroReg | PSReg | MAC | Ours |
|---|---|---|---|---|
| RR↑ | 84.4 | 95.7 | 93.8 | 94.5 |

## 8. Comparison on other datasets

Following the Buffer-X framework [10], we conducted a cross-dataset evaluation on ETH [7] and WOD [11]. Despite the unavailability of the original test sets in Buffer-X, experimental results on subsets from publicly available ETH and WOD datasets demonstrate a +1.8% RR improvement on ETH and +1.5% RR on WOD (Table 7), which verifies the robustness of our method across sensors and scenes.

Table 7. Comparison of additional datasets.

| Method | ETH | | | WOD | | |
|---|---|---|---|---|---|---|
| | RR↑ | RE↓ | TE↓ | RR↑ | RE↓ | TE↓ |
| GeoTR [8] | 93.9 | 8.41 | 0.613 | 90.1 | 8.16 | 4.661 |
| Ours | 95.7 | 6.70 | 0.544 | 91.6 | 8.37 | 3.908 |

## References


[1] Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. Pointdsc: Robust point cloud registration using deep spatial consistency. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 15859–15869, 2021. 1

[2] Fan Bao, Chongxuan Li, Jun Zhu, and Bo Zhang. Analytic-dpm: an analytic estimate of the optimal reverse variance in diffusion probabilistic models. In *International Conference on Learning Representations*, 2022. 3

[3] Zhi Chen, Kun Sun, Fan Yang, and Wenbing Tao. $Sc^2$-pcr: A second order spatial compatibility for efficient and robust point cloud registration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13211–13221, 2022. 1

[4] Christopher B. Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *IEEE/CVF International Conference on Computer Vision*, pages 8957–8965, 2019. 1

[5] Xiaoshui Huang, Zhou Huang, Yifan Zuo, Yongshun Gong, Chengdong Zhang, Deyang Liu, and Yuming Fang. Psreg: Prior-guided sparse mixture of experts for point cloud registration. In *AAAI*, pages 3788–3796, 2025. 4

[6] Haobo Jiang, Zheng Dang, Zhen Wei, Jin Xie, Jian Yang, and Mathieu Salzmann. Robust outlier rejection for 3d registration with variational bayes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1148–1157, 2023. 1

[7] François Pomerleau, Ming Liu, Francis Colas, and Roland Siegwart. Challenging data sets for point cloud registration algorithms. *The International Journal of Robotics Research*, 31(14):1705–1711, 2012. 4

[8] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11133–11142, 2022. 1, 4

[9] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, pages 87–103, 2024. 3

[10] Minkyun Seo, Hyungtae Lim, Kanghee Lee, Luca Carlone, and Jaesik Park. Buffer-x: Towards zero-shot point cloud registration in diverse scenes. *arXiv preprint arXiv:2503.07940*, 2025. 4

[11] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2443–2451, 2020. 4

[12] Luming Tang, Menglin Jia, Qianqian Wang, Cheng Perng Phoo, and Bharath Hariharan. Emergent correspondence from image diffusion. In *Advances in Neural Information Processing Systems*, 2023. 3

[13] Haiping Wang, Yuan Liu, Bing Wang, Yujing Sun, Zhen Dong, Wenping Wang, and Bisheng Yang. Freereg: Image-to-point cloud registration leveraging pretrained diffusion models and monocular depth estimators. In *International Conference on Learning Representations*, 2024. 3

[14] Weijie Wang, Guofeng Mei, Bin Ren, Xiaoshui Huang, Fabio Poiesi, Luc Van Gool, Nicu Sebe, and Bruno Lepri. Zero-shot point cloud registration. *arXiv preprint arXiv:2312.03032*, 2023. 3

[15] Jiaqi Yang, Xiyu Zhang, Peng Wang, Yulan Guo, Kun Sun, Qiao Wu, Shikun Zhang, and Yanning Zhang. MAC: max-

imal cliques for 3d registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10645–10662, 2024. 1

[16] Yifei Zhang, Hao Zhao, Hongyang Li, and Siheng Chen. Fastmac: Stochastic spectral sampling of correspondence graph. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17857–17867, 2024. 1