# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

**Renlang Huang, Yufan Tang, Jiming Chen, and Liang Li***

The State Key Laboratory of Industrial Control Technology.
College of Control Science and Engineering, Zhejiang University, Hangzhou, China.

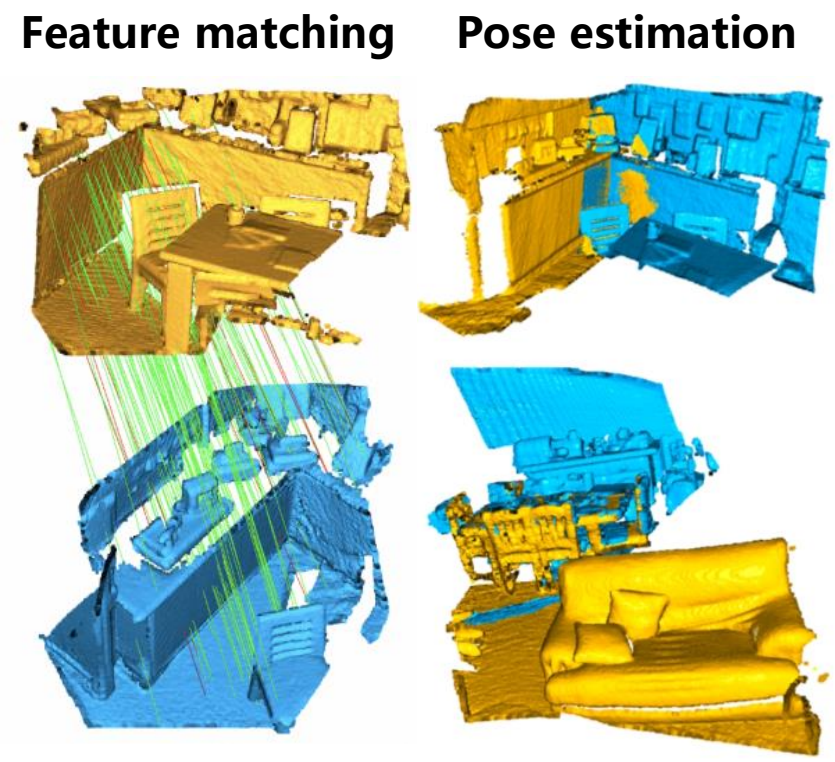
Code Release: http://github.com/RenlangHuang/CAST

# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

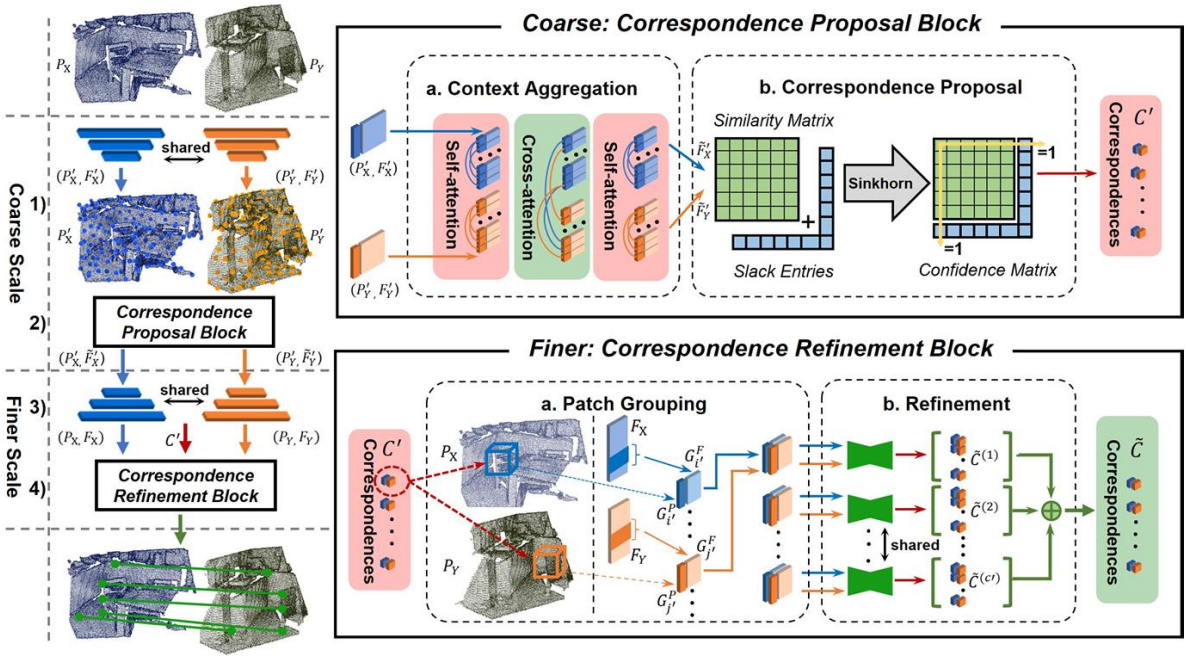
ZHEJIANG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

- **Point cloud registration aims at finding an optimal rigid transformation between two partially overlapped point clouds.**



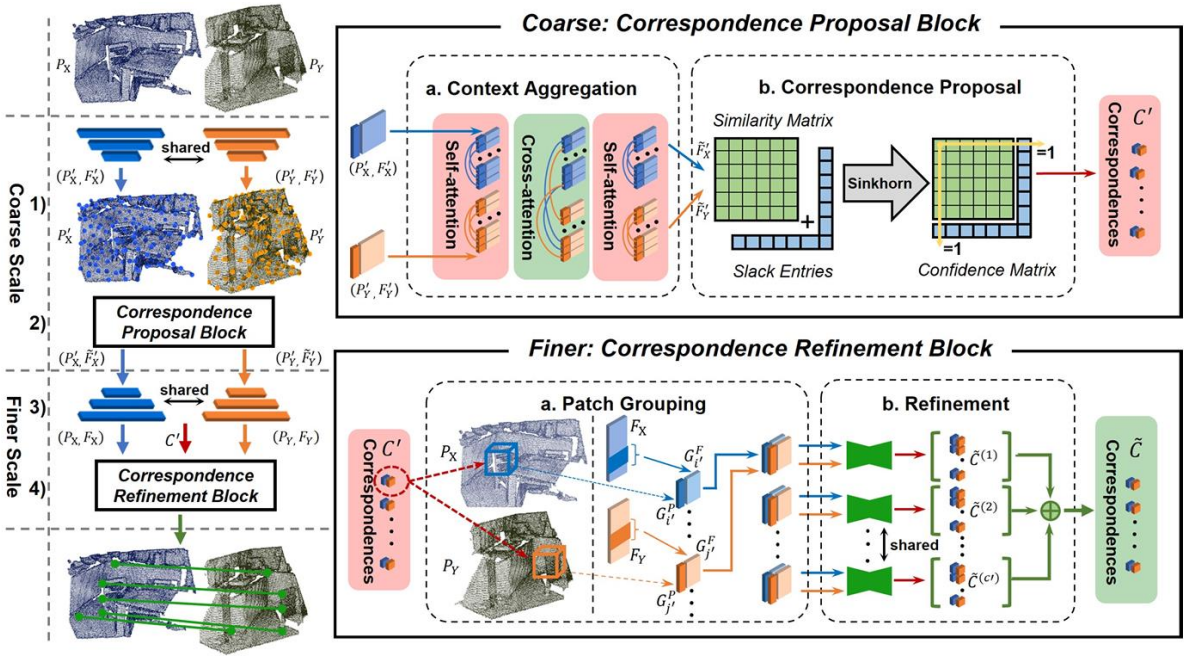

- **Coarse-to-fine matching has showcased great superiority in 2D-2D, 3D-3D, and even 2D-3D feature matching.**



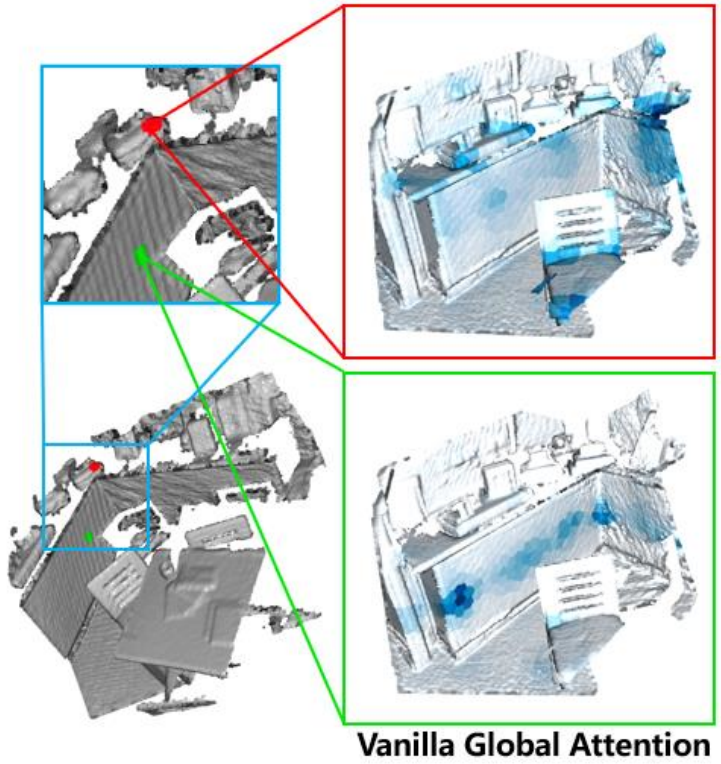

Hao Yu, *et al*. CoFiNet: Reliable Coarse-to-fine Correspondences for Robust Point Cloud Registration, NeurIPS 2021.

# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

- **Coarse-to-fine matching has showcased great superiority in 2D-2D, 3D-3D, and even 2D-3D feature matching.**




- **Coarse matching is sparse and loose without consideration of geometric consistency.**

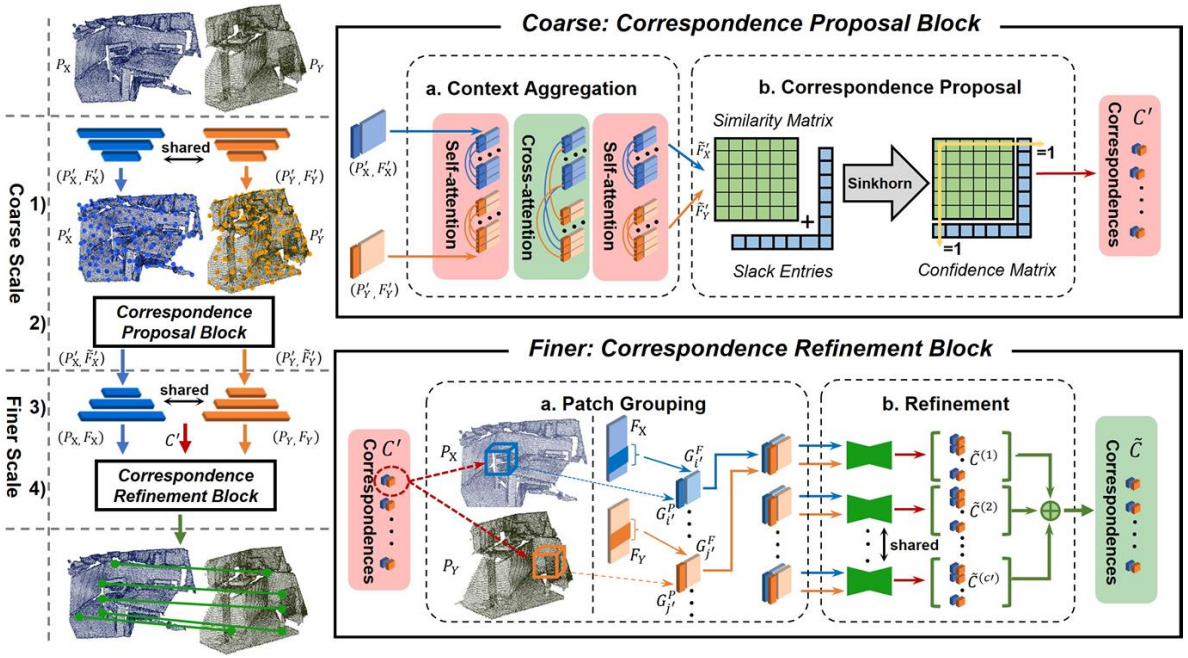


# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

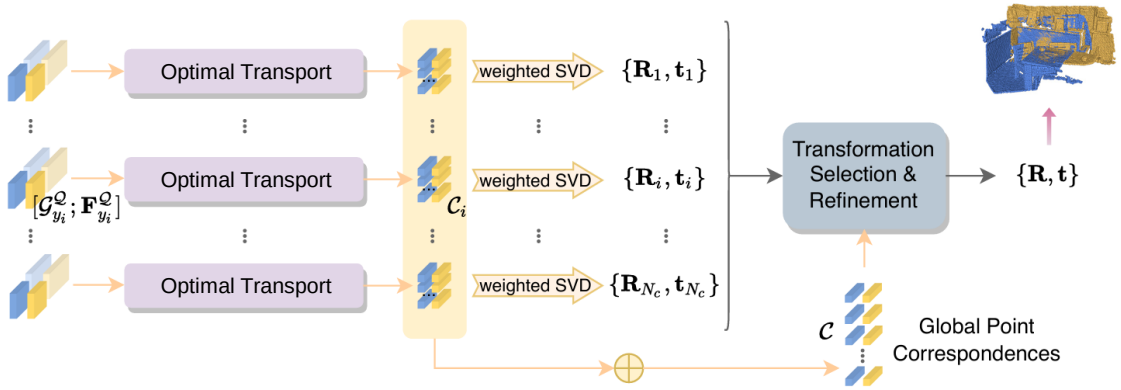
ZHEJIANG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

- **Coarse-to-fine matching has showcased great superiority in 2D-2D, 3D-3D, and even 2D-3D feature matching.**



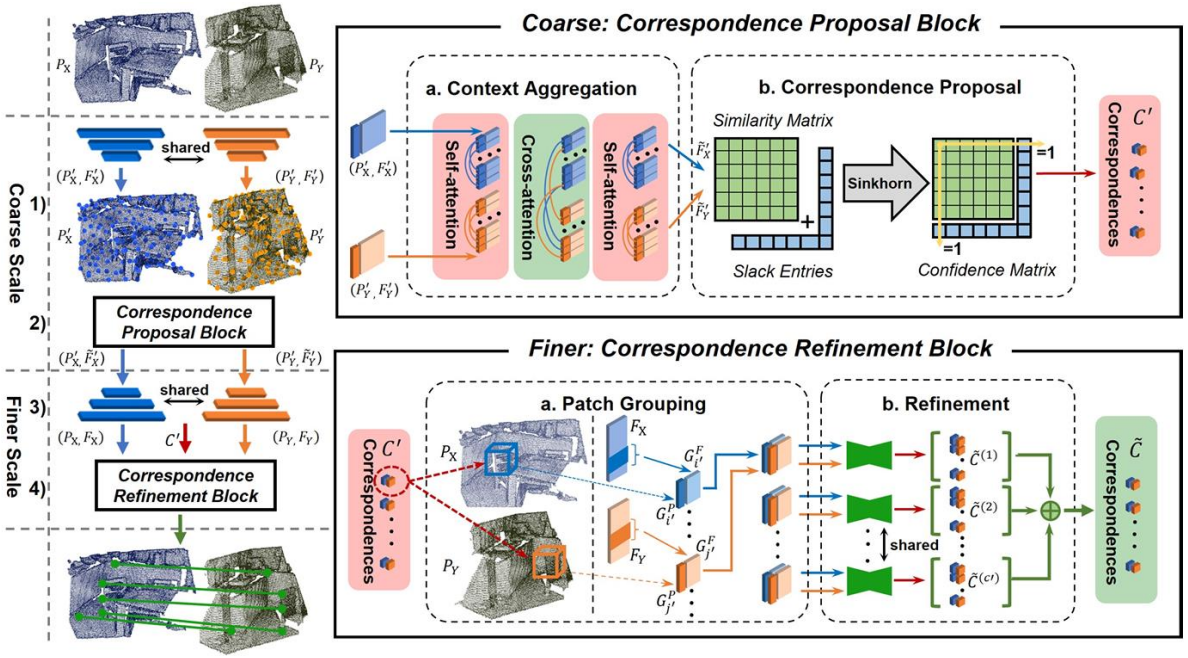

- **Coarse matching is sparse and loose without consideration of geometric consistency.**
- **Fine matching relies on ineffective optimal transport and hypothesis-and selection methods (RANSAC, etc) for consistent pose estimation.**



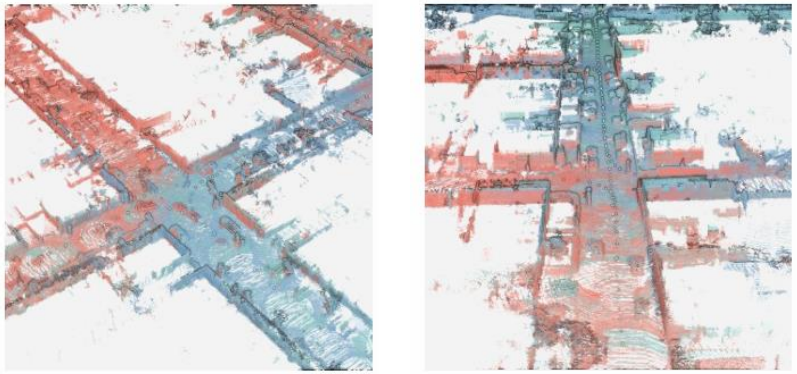

Optimal transport + hypothesis-and-selection pipeline

# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

- **Coarse-to-fine matching has showcased great superiority in 2D-2D, 3D-3D, and even 2D-3D feature matching.**



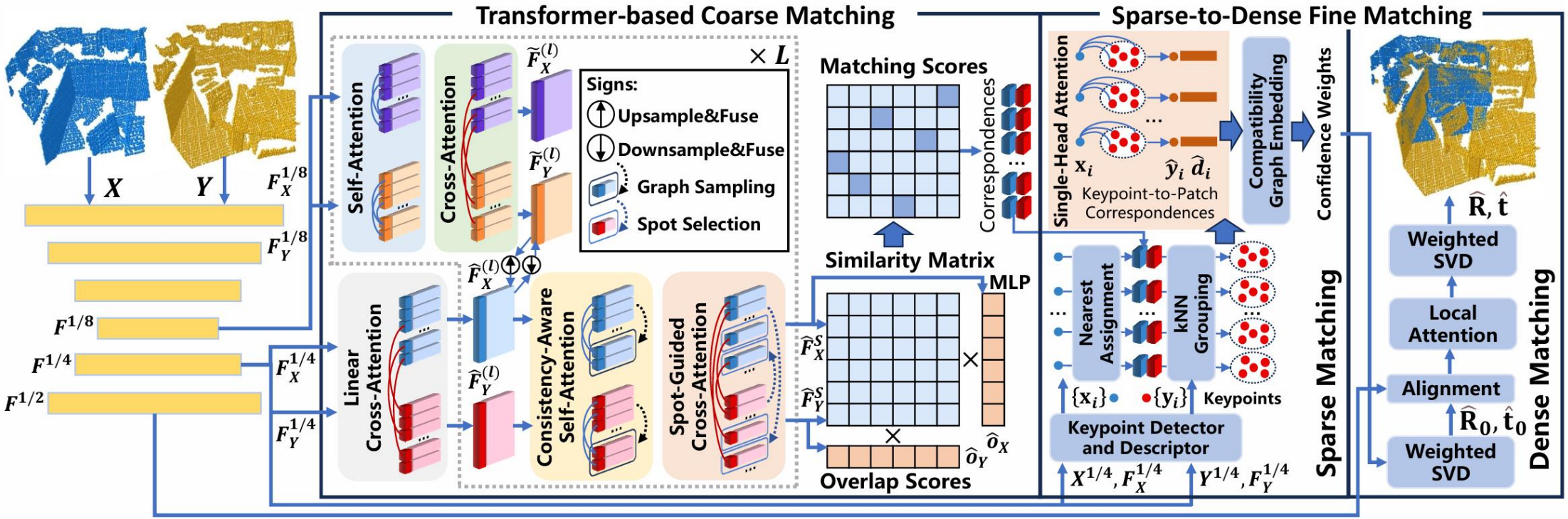

- **Coarse matching is sparse and loose without consideration of geometric consistency.**
- **Fine matching relies on ineffective optimal transport and hypothesis-and selection methods (RANSAC, etc) for consistent pose estimation.**
- **Therefore, existing methods are neither efficient nor scalable for real-time large-scale applications such as odometry in robotics.**



LiDAR odometry and mapping

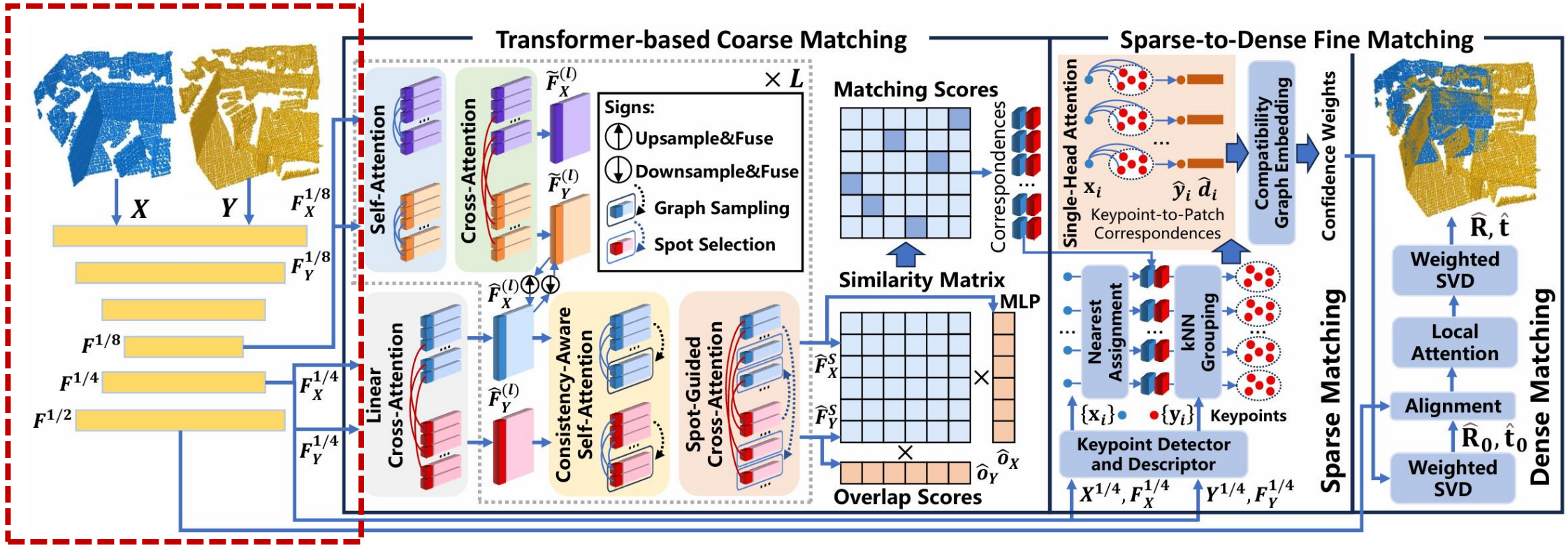
# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

- A novel **consistency-aware spot-guided Transformer (CAST)** with multi-scale feature fusion for much tighter coarse matching with a focus on geometric consistency.
- A lightweight and scalable **sparse-to-dense fine matching** module using both sparse keypoints and dense features for accurate registration without optimal transport and hypothesis-and-selection pipelines.



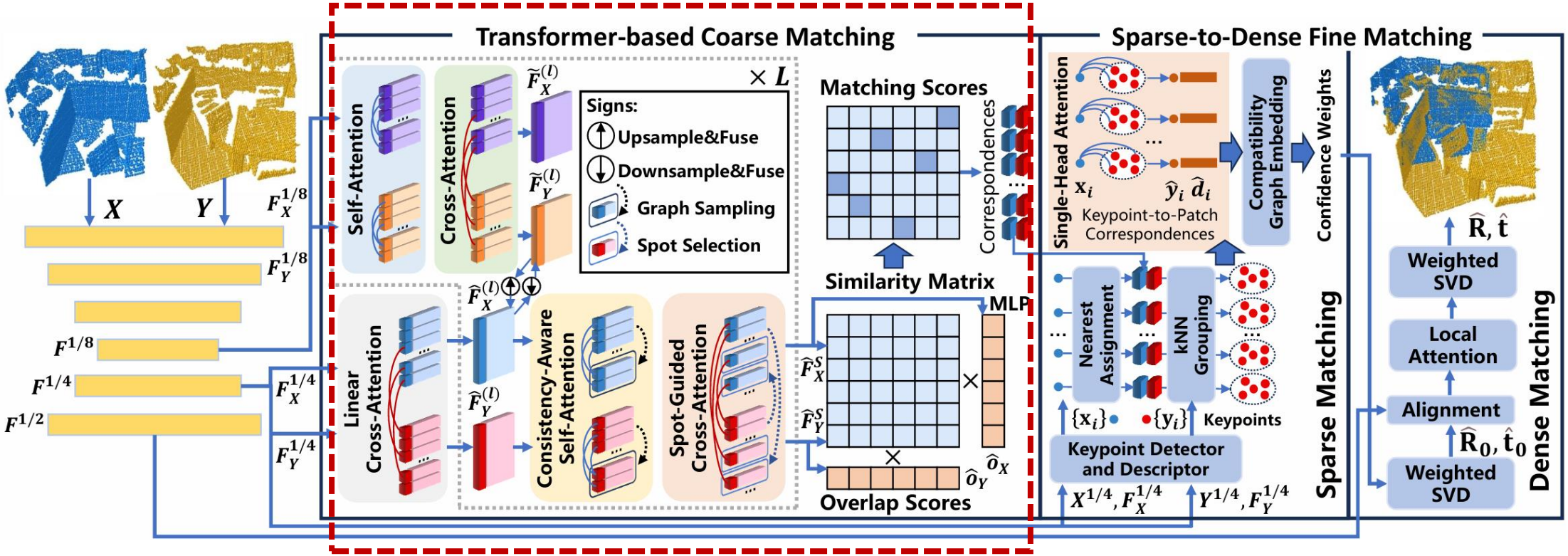
# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

- A novel **consistency-aware spot-guided Transformer (CAST)** with multi-scale feature fusion for much tighter coarse matching with a focus on geometric consistency.
- A lightweight and scalable **sparse-to-dense fine matching** module using both sparse keypoints and dense features for accurate registration without optimal transport and hypothesis-and-selection pipelines.



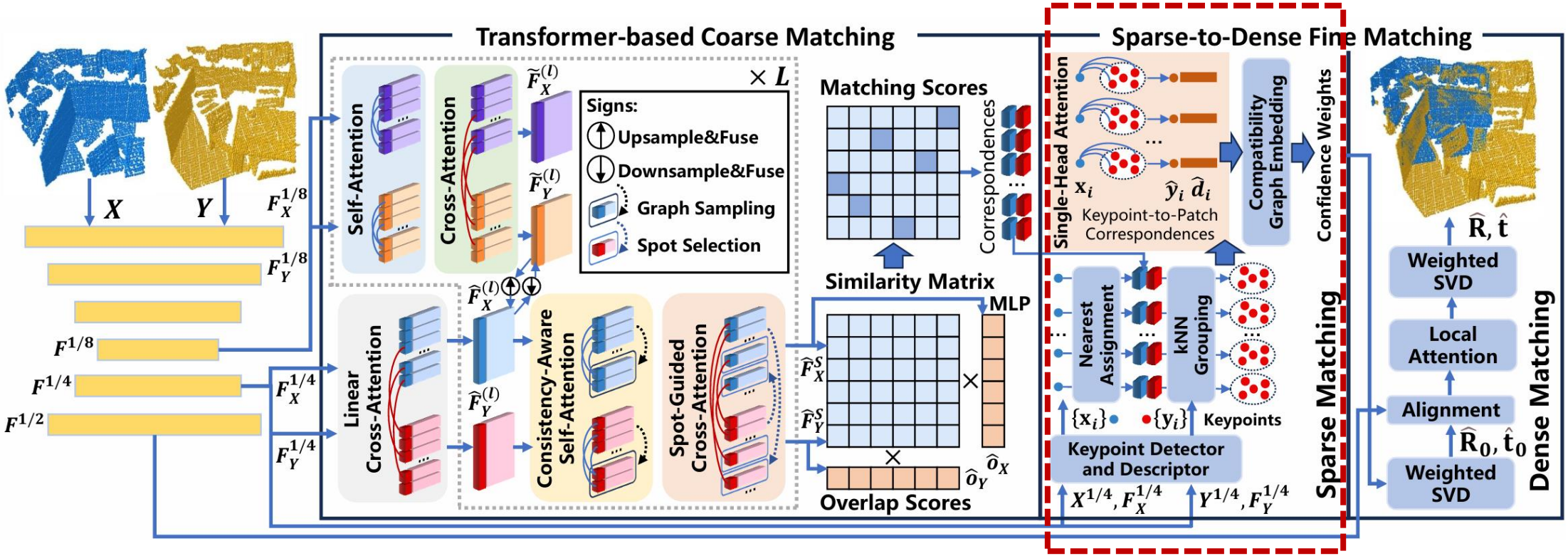
# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

- A novel **consistency-aware self-attention** module base on sparse sampling from the compatibility graph to enhance global consistency during feature aggregation.
- A novel **spot-guided cross-attention** module with a consistency-aware matching confidence criterion that can maintain local consistency without interfering with irrelevant areas.



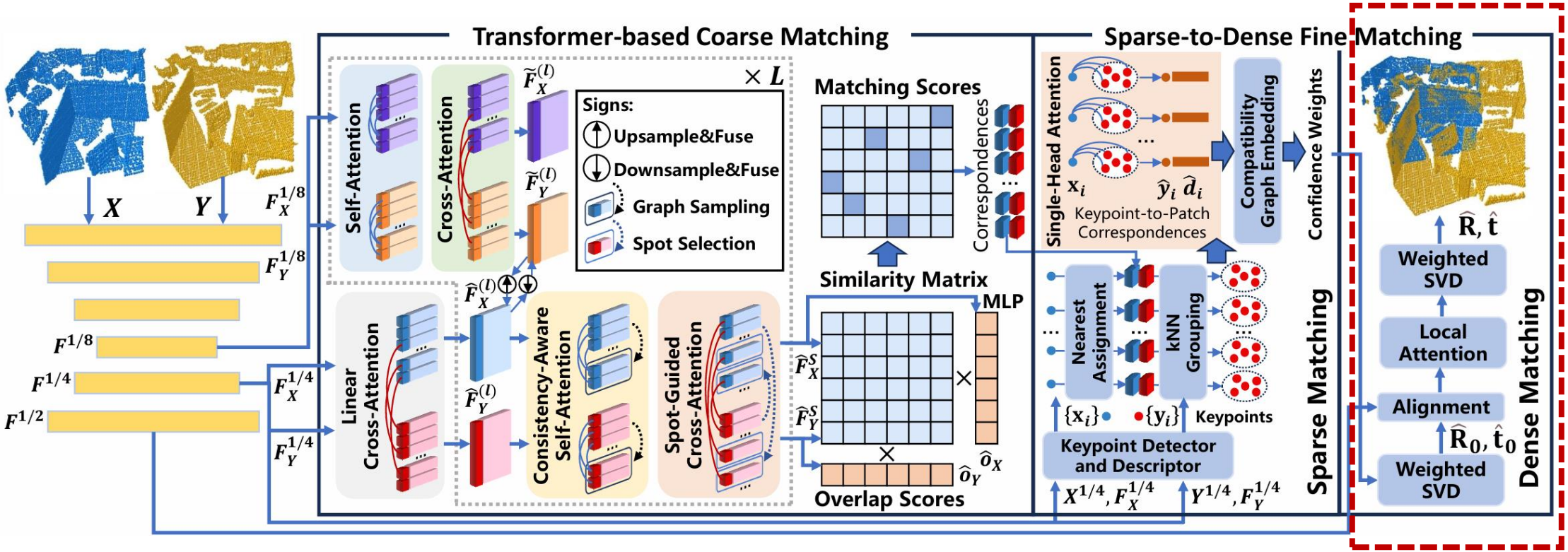
# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

- A novel **consistency-aware spot-guided Transformer (CAST)** with multi-scale feature fusion for much tighter coarse matching with a focus on geometric consistency.
- A lightweight and scalable **sparse-to-dense fine matching** module using both sparse keypoints and dense features for accurate registration without optimal transport and hypothesis-and-selection pipelines.



# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

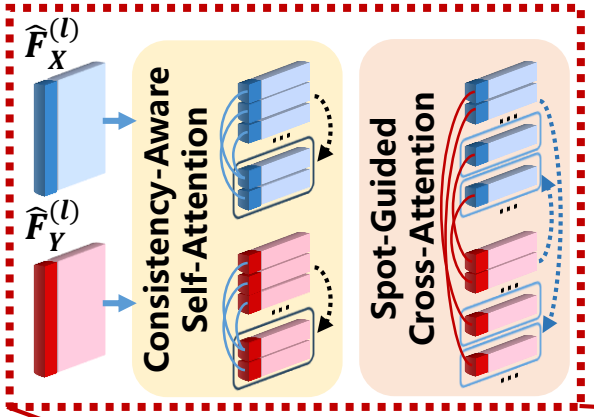
- A novel **consistency-aware spot-guided Transformer (CAST)** with multi-scale feature fusion for much tighter coarse matching with a focus on geometric consistency.
- A lightweight and scalable **sparse-to-dense fine matching** module using both sparse keypoints and dense features for accurate registration without optimal transport and hypothesis-and-selection pipelines.



# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

ZHEJIANG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

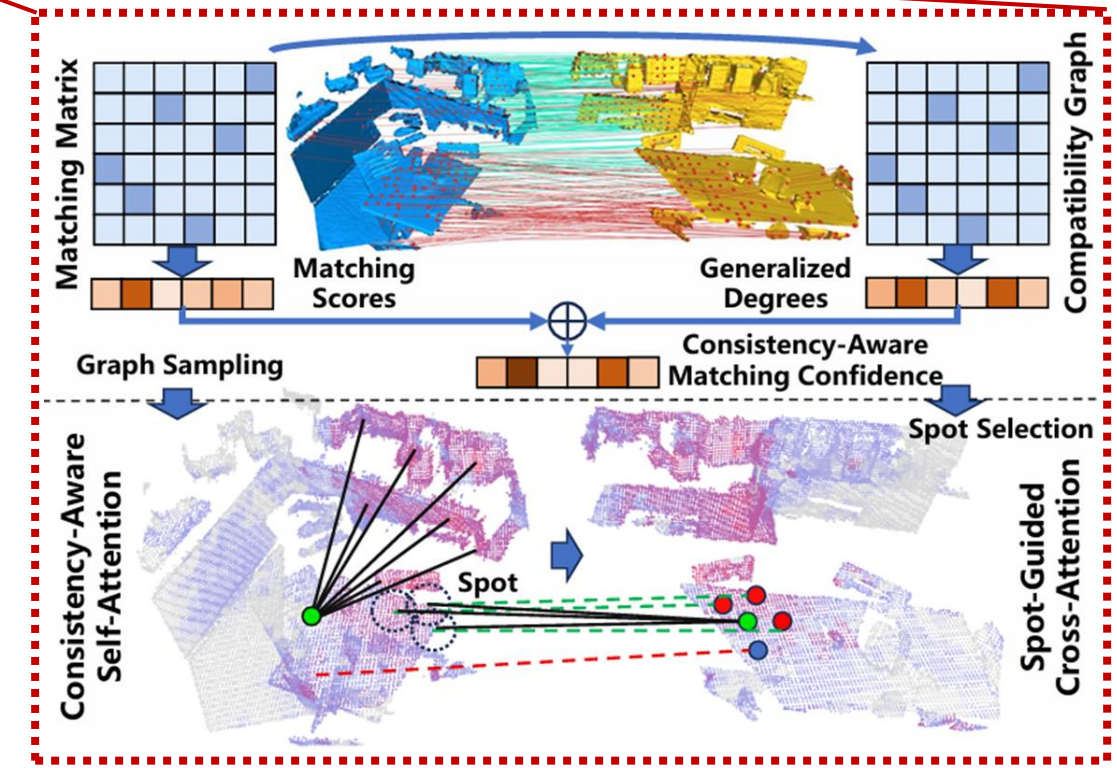


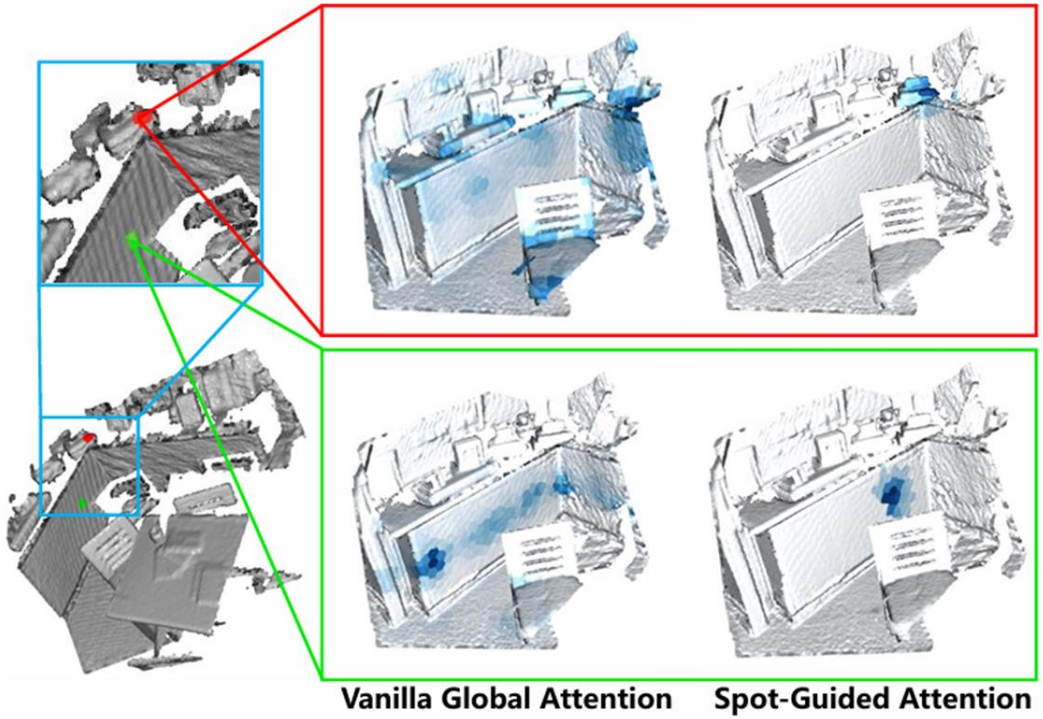

Dual softmax-based matching scores: $\mathbf{P}_{ij}^{(l)} = \underset{k\in\{1,\cdots,M'\}}{\text{softmax}}(\mathbf{S}_{kj}^{(l)})_i \underset{k\in\{1,\cdots,N'\}}{\text{softmax}}(\mathbf{S}_{ik}^{(l)})_j,\ \mathbf{S}^{(l)} = \hat{\mathbf{F}}_X^{(l)}(\hat{\mathbf{F}}_Y^{(l)})^\mathsf{T}.$

Estimated node correspondence set: $\mathcal{C}^{(l)} = \{(\mathbf{x}_i^S, \mathbf{y}_i^S) : \mathbf{x}_i^S \in \mathbf{X}^{1/4}, \mathbf{y}_i^S \in \mathbf{Y}^{1/4}\}.$

Geometric compatibility graph: $\beta_{ij} = \left[1 - d_{ij}^2/\sigma_c^2\right]^+, d_{ij} = \left|\|\mathbf{x}_i^S - \mathbf{x}_j^S\|_2 - \|\mathbf{y}_i^S - \mathbf{y}_j^S\|_2\right|.$

Spot selection for cross-attention: $\mathcal{S}(\mathbf{x}_i^S) = \bigcup_{\mathbf{x}_k^S \in \mathcal{N}_s(\mathbf{x}_i^S)} \mathcal{N}(\mathbf{y}_k^S)$

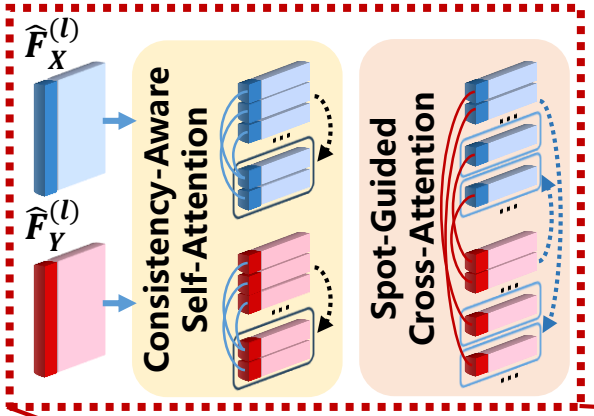


# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

ZHEJIANG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

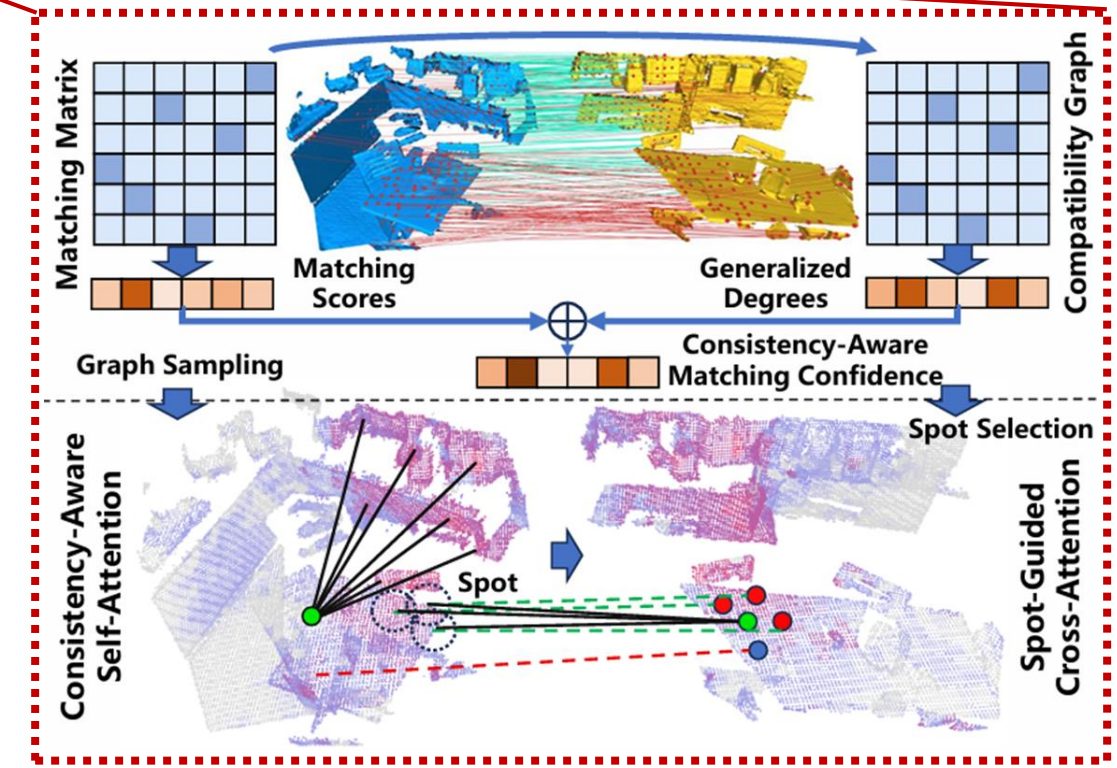


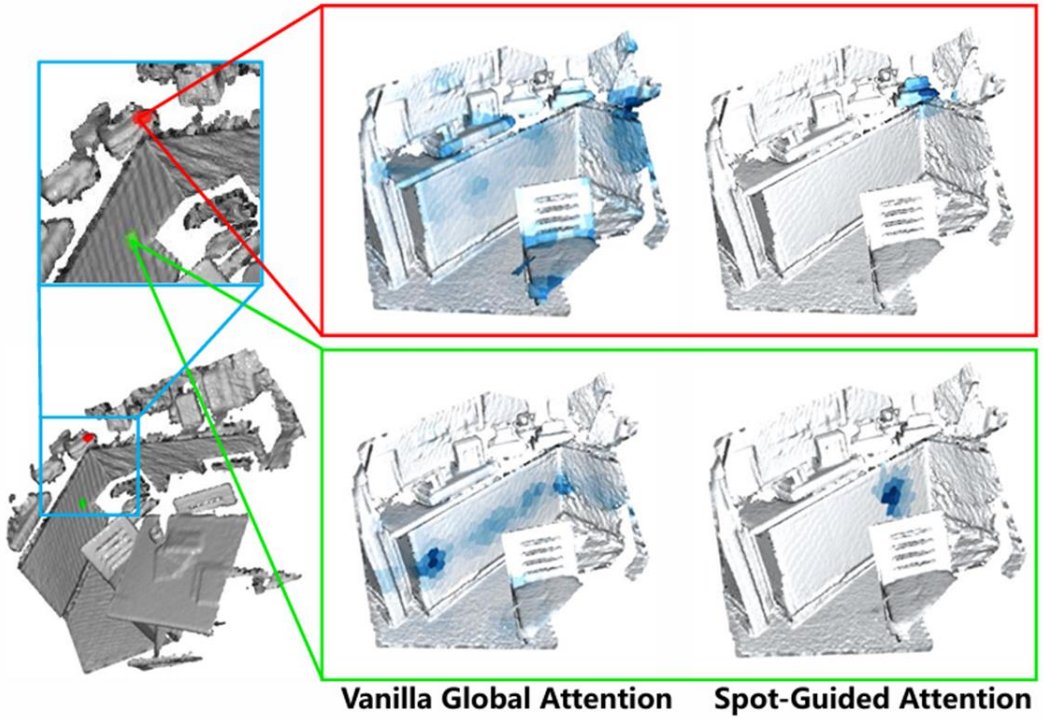

Dual softmax-based matching scores: $\mathbf{P}_{ij}^{(l)} = \underset{k\in\{1,\cdots,M'\}}{\text{softmax}}(\mathbf{S}_{kj}^{(l)})_i \underset{k\in\{1,\cdots,N'\}}{\text{softmax}}(\mathbf{S}_{ik}^{(l)})_j,\ \mathbf{S}^{(l)} = \hat{\mathbf{F}}_X^{(l)}(\hat{\mathbf{F}}_Y^{(l)})^\mathsf{T}.$

Estimated node correspondence set: $\mathcal{C}^{(l)} = \{(\mathbf{x}_i^S, \mathbf{y}_i^S) : \mathbf{x}_i^S \in \mathbf{X}^{1/4}, \mathbf{y}_i^S \in \mathbf{Y}^{1/4}\}.$

Geometric compatibility graph: $\beta_{ij} = \left[1 - d_{ij}^2/\sigma_c^2\right]^+, d_{ij} = \left|\|\mathbf{x}_i^S - \mathbf{x}_j^S\|_2 - \|\mathbf{y}_i^S - \mathbf{y}_j^S\|_2\right|.$

Spot selection for cross-attention: $\mathcal{S}(\mathbf{x}_i^S) = \bigcup_{\mathbf{x}_k^S \in \mathcal{N}_s(\mathbf{x}_i^S)} \mathcal{N}(\mathbf{y}_k^S)$

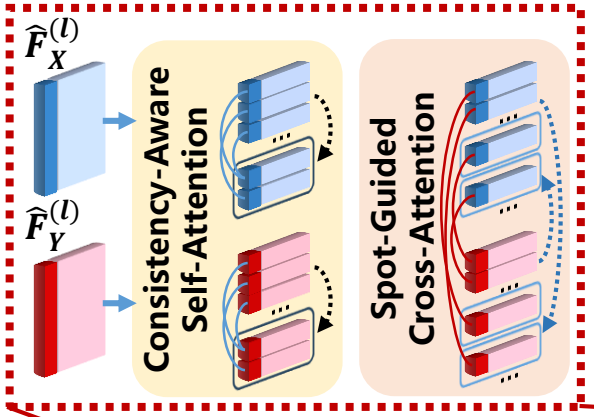


# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

ZHEJIANG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

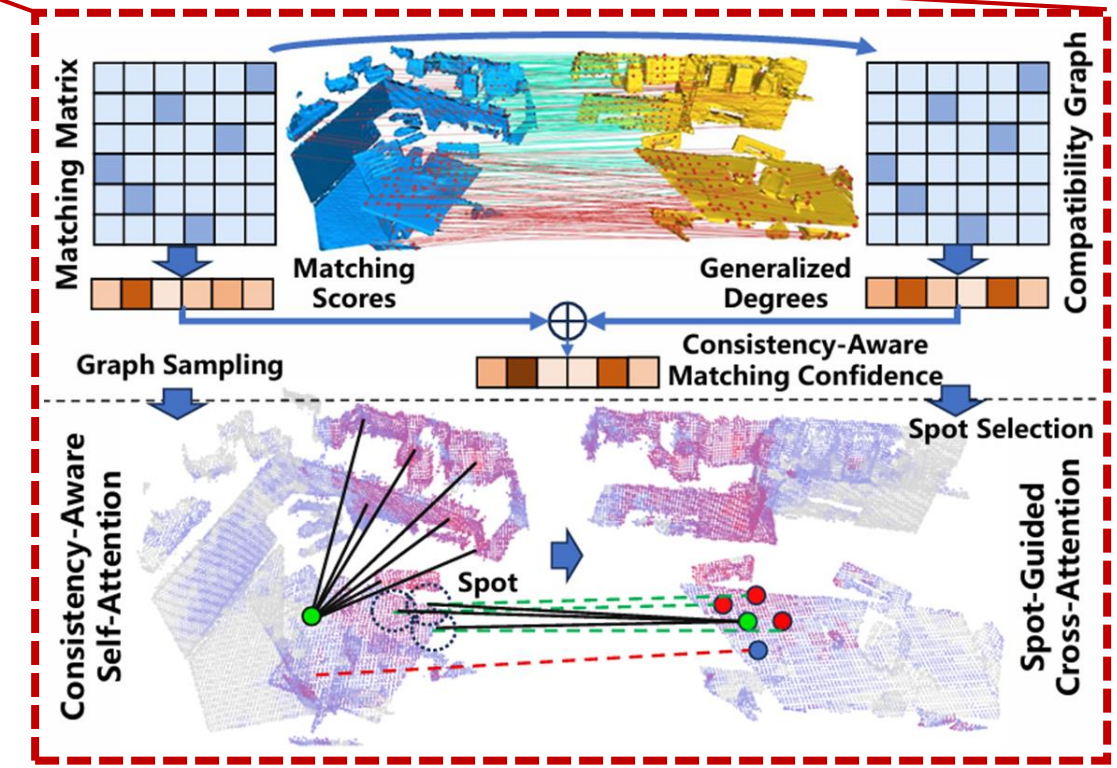


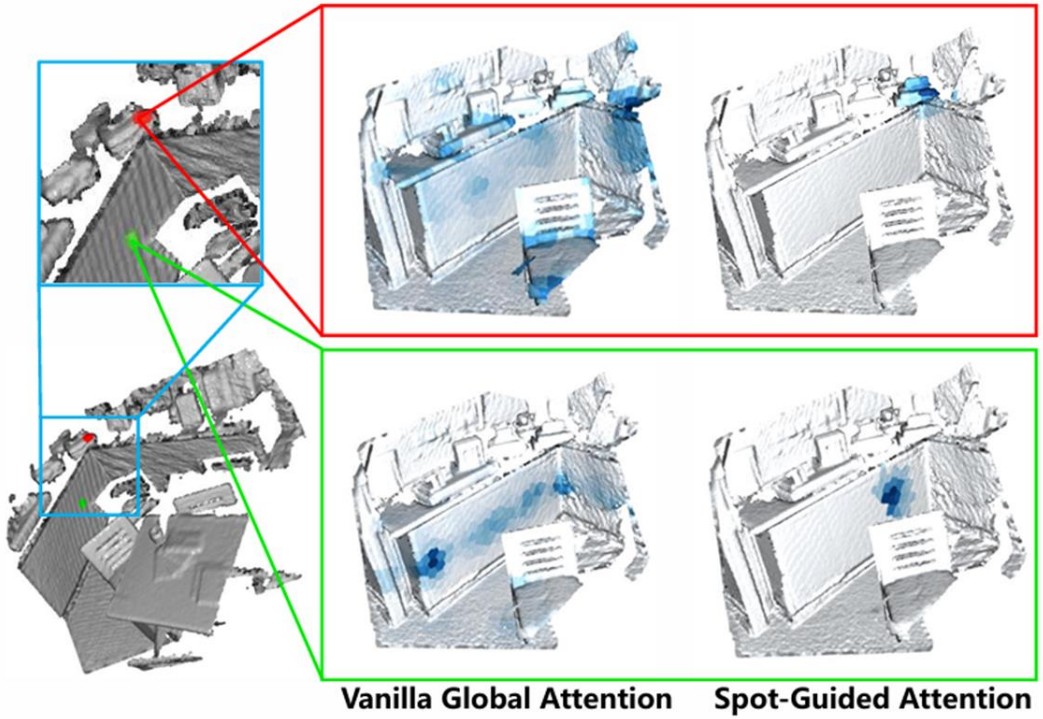

Dual softmax-based matching scores: $\mathbf{P}_{ij}^{(l)} = \underset{k\in\{1,\cdots,M'\}}{\text{softmax}}(\mathbf{S}_{kj}^{(l)})_i \underset{k\in\{1,\cdots,N'\}}{\text{softmax}}(\mathbf{S}_{ik}^{(l)})_j,\ \mathbf{S}^{(l)} = \hat{\mathbf{F}}_X^{(l)}(\hat{\mathbf{F}}_Y^{(l)})^\mathsf{T}.$

Estimated node correspondence set: $\mathcal{C}^{(l)} = \{(\mathbf{x}_i^S, \mathbf{y}_i^S) : \mathbf{x}_i^S \in \mathbf{X}^{1/4}, \mathbf{y}_i^S \in \mathbf{Y}^{1/4}\}.$

Geometric compatibility graph: $\beta_{ij} = \left[1 - d_{ij}^2/\sigma_c^2\right]^+, d_{ij} = \left|\|\mathbf{x}_i^S - \mathbf{x}_j^S\|_2 - \|\mathbf{y}_i^S - \mathbf{y}_j^S\|_2\right|.$

Spot selection for cross-attention: $\mathcal{S}(\mathbf{x}_i^S) = \bigcup_{\mathbf{x}_k^S \in \mathcal{N}_s(\mathbf{x}_i^S)} \mathcal{N}(\mathbf{y}_k^S)$

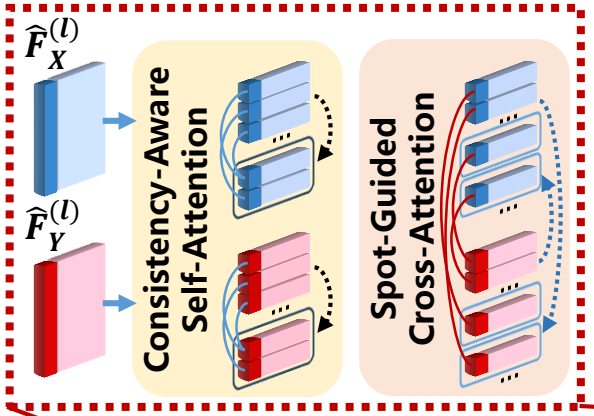


# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

ZHEJIANG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

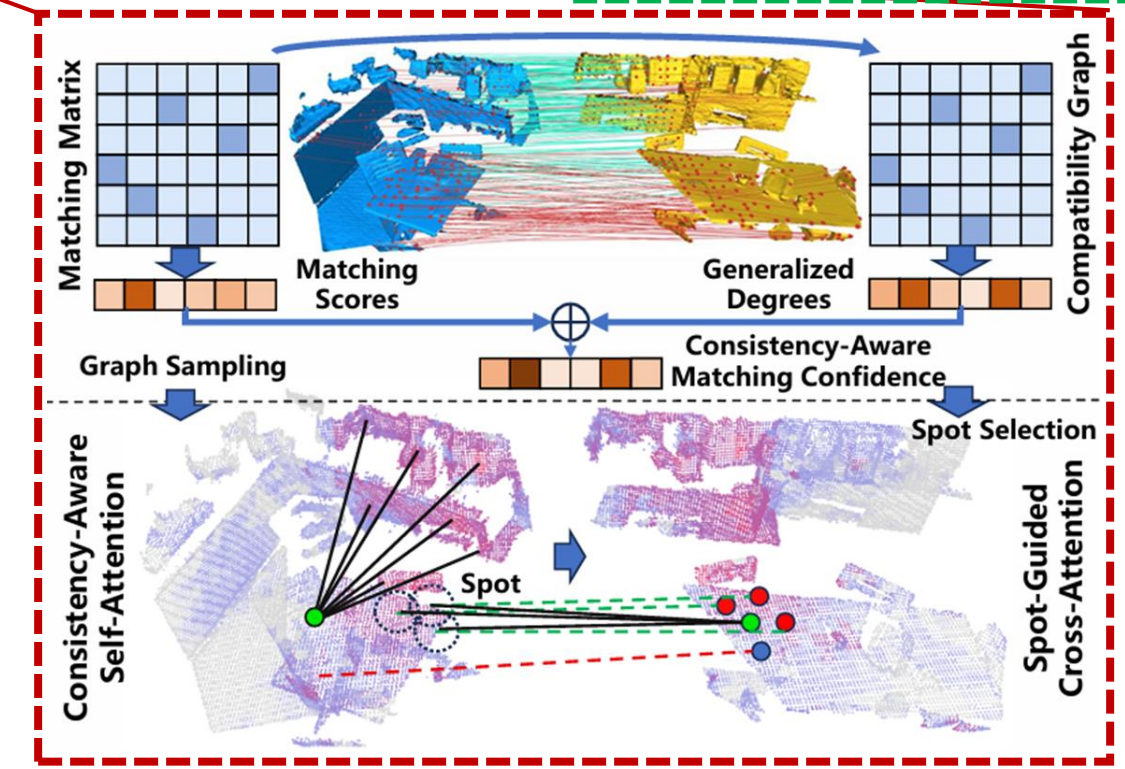


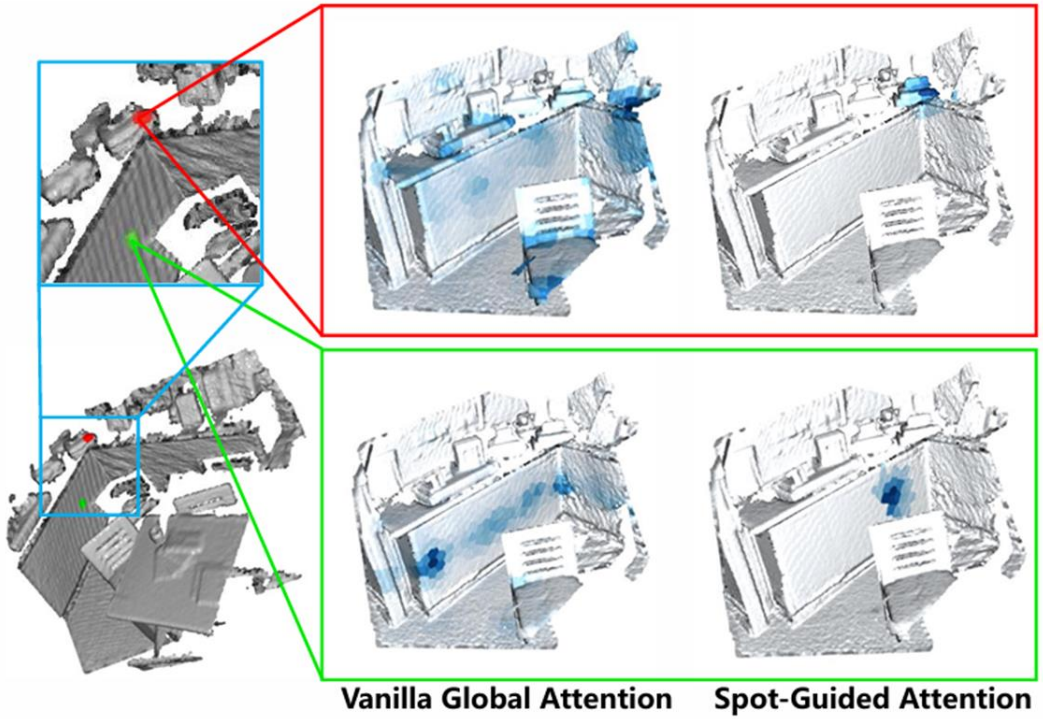

Dual softmax-based matching scores: $\mathbf{P}_{ij}^{(l)} = \underset{k\in\{1,\cdots,M'\}}{\text{softmax}}(\mathbf{S}_{kj}^{(l)})_i \underset{k\in\{1,\cdots,N'\}}{\text{softmax}}(\mathbf{S}_{ik}^{(l)})_j,\ \mathbf{S}^{(l)} = \hat{\mathbf{F}}_X^{(l)}(\hat{\mathbf{F}}_Y^{(l)})^\mathsf{T}.$

Estimated node correspondence set: $\mathcal{C}^{(l)} = \{(\mathbf{x}_i^S, \mathbf{y}_i^S) : \mathbf{x}_i^S \in \mathbf{X}^{1/4}, \mathbf{y}_i^S \in \mathbf{Y}^{1/4}\}.$

Geometric compatibility graph: $\beta_{ij} = \left[1 - d_{ij}^2/\sigma_c^2\right]^+, d_{ij} = \left|\|\mathbf{x}_i^S - \mathbf{x}_j^S\|_2 - \|\mathbf{y}_i^S - \mathbf{y}_j^S\|_2\right|.$

Spot selection for cross-attention: $\mathcal{S}(\mathbf{x}_i^S) = \bigcup_{\mathbf{x}_k^S \in \mathcal{N}_s(\mathbf{x}_i^S)} \mathcal{N}(\mathbf{y}_k^S)$



# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

**We evaluate our method on outdoor benchmarks KITTI and nuScenes using three metrics: Relative Translation Error (RTE), Relative Rotation Error (RRE), and Registration Recall (RR).**

**Performance highlights:**

- **Highest registration recalls (robustness)**
- **Significantly lower RTE than state-of-the-arts (accuracy)**

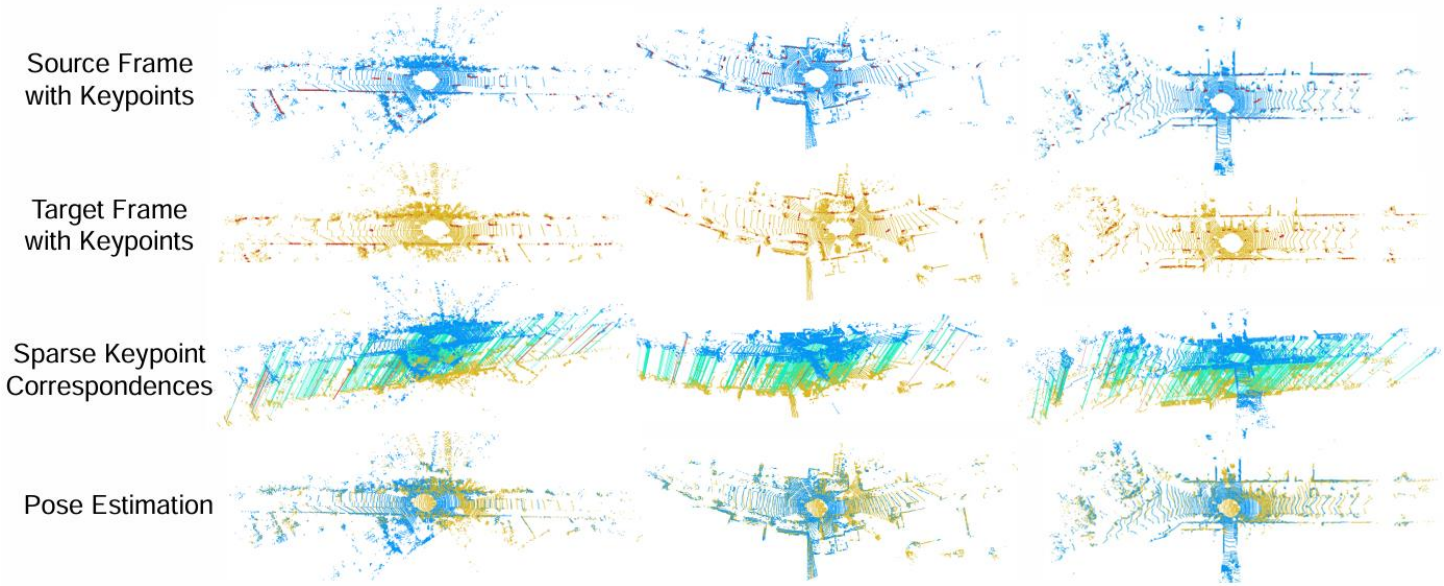



Table 1: Registration performance on KITTI odometry dataset.

| Model | Publication | RTE (cm) | RRE (°) | RR (%) |
|---|---|---|---|---|
| 3DFeat-Net | ECCV 2018 [25] | 25.9 | 0.25 | 96.0 |
| FCGF | ICCV 2019 [8] | 9.5 | 0.30 | 96.6 |
| D3Feat | CVPR 2020 [9] | 7.2 | 0.30 | 99.8 |
| SpinNet | CVPR 2021 [19] | 9.9 | 0.47 | 99.1 |
| Predator | CVPR 2021 [20] | 6.8 | 0.27 | 99.8 |
| CoFiNet | NeurIPS 2021 [11] | 8.2 | 0.41 | 99.8 |
| GeoTransformer | CVPR 2022 [12] | 6.8 | 0.24 | 99.8 |
| OIF-Net | NeurIPS 2022 [13] | 6.5 | **0.23** | 99.8 |
| PEAL | CVPR 2023 [29] | 6.8 | **0.23** | 99.8 |
| DiffusionPCR | CVPR 2024 [30] | 6.3 | **0.23** | 99.8 |
| MAC | CVPR 2023 [43] | 8.5 | 0.40 | 99.5 |
| RegFormer | ICCV 2023 [44] | 8.4 | 0.24 | 99.8 |
| CAST | | **2.5** | 0.27 | **100.0** |

Table 2: Registration performance on nuScenes.

| Method | RTE (m) | RRE (°) | RR (%) |
|---|---|---|---|
| Point-to-Point ICP [42] | 0.25 | 0.25 | 18.8 |
| Point-to-Plane ICP [42] | 0.15 | 0.21 | 36.8 |
| FGR [45] | 0.71 | 1.01 | 32.2 |
| RANSAC [15] | 0.21 | 0.74 | 60.9 |
| DCP [26] | 1.09 | 2.07 | 56.8 |
| IDAM [27] | 0.47 | 0.79 | 88.0 |
| FMR [46] | 0.60 | 1.61 | 92.1 |
| DGR [31] | 0.21 | 0.48 | 98.4 |
| HRegNet [24] | 0.18 | 0.45 | **99.9** |
| CAST | **0.12** | **0.20** | **99.9** |

# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

ZHEJIANG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

## Performance highlights on indoor benchmarks:

- Superior robustness and efficiency
- Better accuracy than other registration baselines

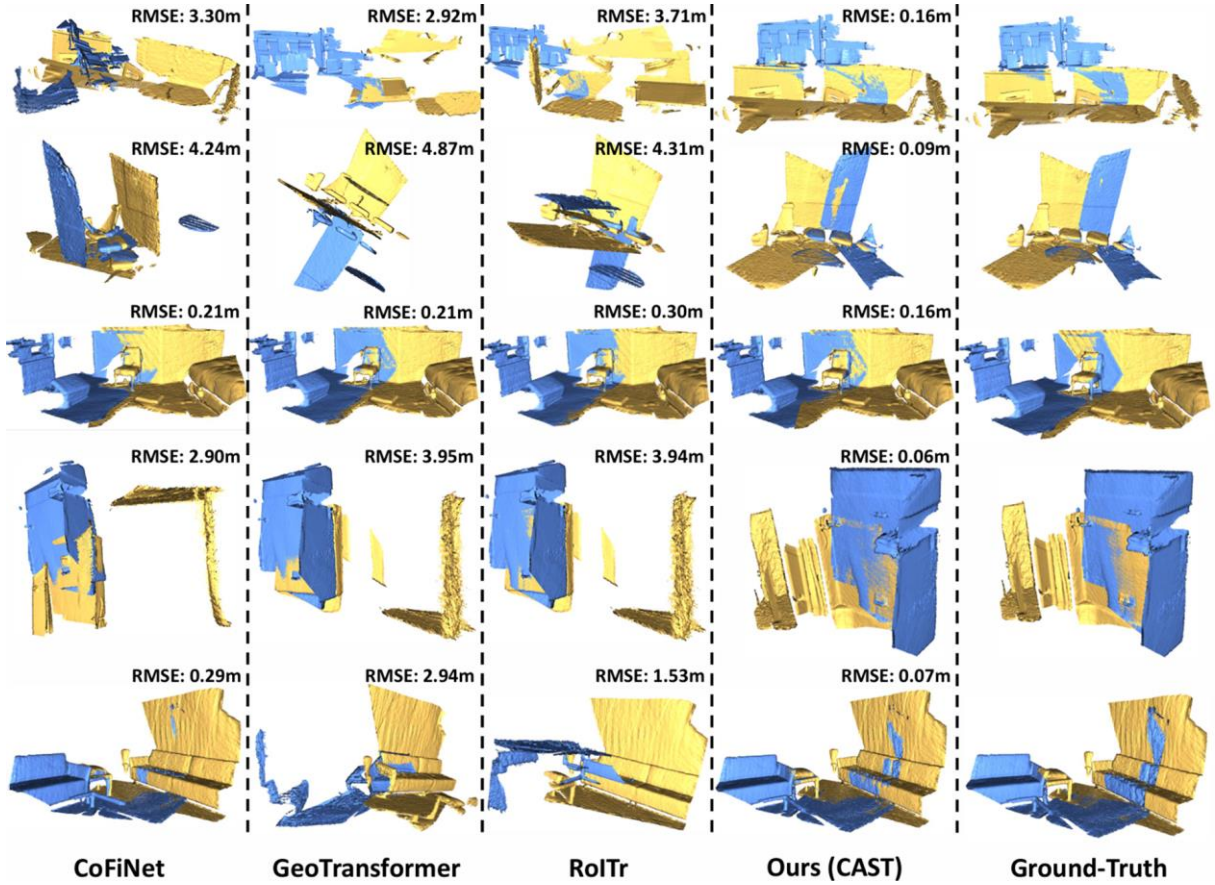



Table 3: Evaluation results on indoor RGBD point cloud datasets.

| | Dataset | 3DMatch | | | | | 3DLoMatch | | | | | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Registration Recall (%) | | | | | Registration Recall (%) | | | | | Time (s) |
| | Samples | 5000 | 2500 | 1000 | 500 | 250 | 5000 | 2500 | 1000 | 500 | 250 | All |
| descriptor-based | PerfectMatch [7] | 78.4 | 76.2 | 71.4 | 67.6 | 50.8 | 33.0 | 29.0 | 23.3 | 17.0 | 11.0 | - |
| | FCGF [8] | 85.1 | 84.7 | 83.3 | 81.6 | 71.4 | 40.1 | 41.7 | 38.2 | 35.4 | 26.8 | 0.271 |
| | D3Feat [9] | 81.6 | 84.5 | 83.4 | 82.4 | 77.9 | 37.2 | 42.7 | 46.9 | 43.8 | 39.1 | 0.289 |
| | SpinNet [19] | 88.6 | 86.6 | 85.5 | 83.5 | 70.2 | 59.8 | 54.9 | 48.3 | 39.8 | 26.8 | 90.804 |
| | YOHO [50] | 90.8 | 90.3 | 89.1 | 88.6 | 84.5 | 65.2 | 65.5 | 63.2 | 56.5 | 48.0 | 13.529 |
| | Predator [20] | 89.0 | 89.9 | 90.6 | 88.5 | 86.6 | 59.8 | 61.2 | 62.4 | 60.8 | 58.1 | 0.759 |
| correspondence-based | REGTR [47] | | | 92.0 | | | | | 64.8 | | | 0.382 |
| | CoFiNet [11] | 89.3 | 88.9 | 88.4 | 87.4 | 87.0 | 67.5 | 66.2 | 64.2 | 63.1 | 61.0 | 0.306 |
| | GeoTransformer [12] | 92.0 | 91.8 | 91.8 | 91.4 | 91.2 | 75.0 | 74.8 | 74.2 | 74.1 | 73.5 | 0.192 |
| | OIF-Net [13] | 92.4 | 91.9 | 91.8 | 92.1 | 91.2 | 76.1 | 75.4 | 75.1 | 74.4 | 73.6 | 0.555 |
| | RoITr [28] | 91.9 | 91.7 | 91.8 | 91.4 | 91.0 | 74.7 | 74.8 | 74.8 | 74.2 | 73.6 | 0.457 |
| | PEAL [29] | 94.4 | 94.1 | 94.1 | 93.9 | 93.4 | 79.2 | 79.0 | 78.8 | 78.5 | 77.9 | 2.074 |
| | BUFFER [48] | | | 92.9 | | | | | 71.8 | | | 0.290 |
| | SIRA-PCR [49] | 93.6 | 93.9 | 93.9 | 92.7 | 92.4 | 73.5 | 73.9 | 73.0 | 73.4 | 71.1 | 0.291 |
| | DiffusionPCR [30] | 94.4 | 94.3 | 94.5 | 94.0 | 93.9 | **80.0** | **80.4** | **79.2** | **78.8** | **78.8** | 1.964 |
| | CAST | | | **95.2** | | | | | 75.1 | | | **0.182** |

Table 9: Registration results on indoor RGBD point cloud datasets.

| Methods | 3DMatch | | | 3DLoMatch | | |
|---|---|---|---|---|---|---|
| | RR (%) | RTE (cm) | RRE (°) | RR (%) | RTE (cm) | RRE (°) |
| RANSAC-1M [15] | 88.42 | 9.42 | 3.05 | 9.77 | 14.87 | 7.01 |
| RANSAC-4M [15] | 91.44 | 8.38 | 2.69 | 10.44 | 15.14 | 6.91 |
| TEASER++ [58] | 85.77 | 8.66 | 2.73 | 46.76 | 12.89 | 4.12 |
| $SC^2$-PCR [59] | 93.16 | 6.51 | 2.09 | 58.73 | 10.44 | 3.80 |
| DGR [31] | 88.85 | 7.02 | 2.28 | 43.80 | 10.82 | 4.17 |
| PointDSC [32] | 91.87 | 6.54 | 2.10 | 56.20 | 10.48 | 3.87 |
| MAC [43] | 93.72 | 6.54 | 2.02 | 59.85 | 9.75 | 3.50 |
| FastMAC [60] | 92.67 | 6.47 | 2.00 | 58.23 | 10.81 | 3.80 |
| CAST | **96.48** | **5.64** | **1.71** | **76.13** | **8.47** | **2.75** |


浙江大學
ZHEJIANG UNIVERSITY


# A Consistency-Aware Spot-Guided Transformer for Versatile and Hierarchical Point Cloud Registration

**Renlang Huang, Yufan Tang, Jiming Chen, and Liang Li***

# Thanks for watching!

For more details, please refer to our paper.