Supplementary for "GM-R^2: Generative Matching Learning for Unsupervised Geometric Representation and Registration"

Haobo Jiang[1], Liang Yu[2], Jianmin Zheng[*1]

[1]Nanyang Technological University, Singapore, [2]Alibaba Group

{haobo.jiang, ASJMZheng}@ntu.edu.sg, liangyu.yl@alibaba-inc.com

A. Theoretical Analysis of Generative Matching Learning

As demonstrated in Sec. 3.2.1 in the main manuscript, the optimization objective of our *Generative Matching Learning* under GM-R^2 paradigm can be formulated as the problem of maximizing the log-likelihood over the dataset D with paired point clouds $(\mathbf{P}, \mathbf{Q})$ and corresponding RGB images $(\mathbf{I}^P, \mathbf{I}^Q)$:

$$\max_{\boldsymbol{\theta}} \mathbb{E}_{(\mathbf{I}^P, \mathbf{I}^Q, \mathbf{P}, \mathbf{Q}) \sim \mathcal{D}} \left[\log p_{\boldsymbol{\omega}}(\mathbf{I}^P, \mathbf{I}^Q \mid g_{\boldsymbol{\theta}}(\mathbf{P}), g_{\boldsymbol{\theta}}(\mathbf{Q})) \right]. \tag{1}$$

In practice, we map the paired point clouds into the range maps $\tilde{\mathbf{d}}^{PQ} = (\mathbf{d}^P, \mathbf{d}^Q)$ via our proposed AFoV-ERP, and the coupled latent image representation of the paired images can be represented as $\tilde{\mathbf{x}}^{PQ} = [\tilde{\mathbf{x}}^P; \tilde{\mathbf{x}}^Q]$. Moreover, we leverage the ControlNet encoder as the geometric feature encoder $g_{\boldsymbol{\theta}}(\cdot)$. Consequently, the log-likelihood optimization objective can be rewritten as: $\mathbb{E}_{p_{data}}[\log p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}^{PQ} \mid \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))]$.

Our diffusion process progressively introduces the Gaussian noise into this clean image pair, which gradually converts them into the noise distribution $\mathbf{x}_T^{PQ} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, forming a Markov chain $\mathbf{x}^{PQ} = \mathbf{x}_0^{PQ} \rightarrow \mathbf{x}_1^{PQ} \cdots \rightarrow \mathbf{x}_T^{PQ}$. As demonstrated in [2], the random variable $\mathbf{x}_t^{PQ} \sim q(\mathbf{x}_t^{PQ} | \mathbf{x}_{t-1}^{PQ}) := \mathcal{N}(\mathbf{x}_t^{PQ}; \sqrt{1-\beta_t}\mathbf{x}_{t-1}^{PQ}, \beta_t \mathbf{I})$ can also be expressed in a closed form $\mathbf{x}_t^{PQ} \sim q(\mathbf{x}_t^{PQ} | \mathbf{x}_0^{PQ})$, which can be formulated as follows:

$$\mathbf{x}_t^{PQ} \sim \mathcal{N}(\mathbf{x}_t^{PQ}; \sqrt{\bar{\alpha}_t}\mathbf{x}_0^{PQ}, (1-\bar{\alpha}_t)\mathbf{I}). \tag{2}$$

Here, the diffusion coefficients $\bar{\alpha}_t = \Pi_{s=0}^t \alpha_s = \Pi_{s=0}^t (1-\beta_s)$, and β_s indicates the noise coefficient determined by such as a linear schedule [2] or a cosine schedule [3]. Then, the variational lower bound of the log-likelihood optimization objective can be represented as:

$$\begin{aligned}
&\mathbb{E}_{p_{data}}[\log p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}^{PQ} | \, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))] \\
\geq &\mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ} | \tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{0:T}^{PQ} | \, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_{1:T}^{PQ} | \tilde{\mathbf{x}}_0^{PQ})} \right] \\
= &\mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ} | \tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ}) \prod_{t=1}^T p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ} | \tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{\prod_{t=1}^T q(\tilde{\mathbf{x}}_t^{PQ} | \tilde{\mathbf{x}}_{t-1}^{PQ})} \right] \\
= &\mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ} | \tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ}) p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ} | \tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta})) \prod_{t=2}^T p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ} | \tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_1^{PQ} | \tilde{\mathbf{x}}_0^{PQ}) \prod_{t=2}^T q(\tilde{\mathbf{x}}_t^{PQ} | \tilde{\mathbf{x}}_{t-1}^{PQ}, \tilde{\mathbf{x}}_0^{PQ})} \right] \\
= &\mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ} | \tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ}) p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ} | \tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_1^{PQ} | \tilde{\mathbf{x}}_0^{PQ})} + \log \prod_{t=2}^T \frac{p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ} | \tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_t^{PQ} | \tilde{\mathbf{x}}_{t-1}^{PQ}, \tilde{\mathbf{x}}_0^{PQ})} \right] \\
= &\mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ} | \tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ}) p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ} | \tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_1^{PQ} | \tilde{\mathbf{x}}_0^{PQ})} + \log \prod_{t=2}^T \frac{p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ} | \tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{\frac{q(\tilde{\mathbf{x}}_{t-1}^{PQ} | \tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ}) q(\tilde{\mathbf{x}}_t^{PQ} | \tilde{\mathbf{x}}_0^{PQ})}{q(\tilde{\mathbf{x}}_{t-1}^{PQ} | \tilde{\mathbf{x}}_0^{PQ})}} \right]
\end{aligned} \tag{3}$$

$$
\begin{aligned}
&= \mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ}) p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ}|\tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_1^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} + \log \prod_{t=2}^{T} \frac{p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{\frac{q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ}) q(\tilde{\mathbf{x}}_t^{PQ}|\tilde{\mathbf{x}}_0^{PQ})}{q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_0^{PQ})}} \right] \\
&= \mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ}) p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ}|\tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{\cancel{q(\tilde{\mathbf{x}}_1^{PQ}|\tilde{\mathbf{x}}_0^{PQ})}} + \log \frac{\cancel{q(\tilde{\mathbf{x}}_1^{PQ}|\tilde{\mathbf{x}}_0^{PQ})}}{q(\tilde{\mathbf{x}}_T^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} + \log \prod_{t=2}^{T} \frac{p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ})} \right] \\
&= \mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ}) p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ}|\tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_T^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} + \sum_{t=2}^{T} \log \frac{p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ})} \right] \\
&= \mathbb{E}_{p_{data}, q(\tilde{\mathbf{x}}_{1:T}^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ}) p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ}|\tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_T^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} + \sum_{t=2}^{T} \log \frac{p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ})} \right] \\
&= \mathbb{E} \left[\log p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ}|\tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta})) \right] + \mathbb{E}_{q(\tilde{\mathbf{x}}_T^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p(\tilde{\mathbf{x}}_T^{PQ})}{q(\tilde{\mathbf{x}}_T^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} \right] + \sum_{2}^{T} \mathbb{E}_{q(\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_0^{PQ})} \left[\log \frac{p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))}{q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ})} \right] \\
&= \mathbb{E} \left[\log p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_0^{PQ}|\tilde{\mathbf{x}}_1^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta})) - \mathrm{D}_{\mathrm{KL}}(q(\tilde{\mathbf{x}}_T^{PQ}|\tilde{\mathbf{x}}_0^{PQ}) || p(\tilde{\mathbf{x}}_T^{PQ})) - \sum_{t>1} \mathrm{D}_{\mathrm{KL}}(q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ}) || p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))) \right],
\end{aligned}
\tag{4}
$$

where the third term is the core loss (termed denoising matching loss) for denoising network training. As derived in [2], the posterior distribution $q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ})$ can be formulated as:

$$
q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ}) = \mathcal{N}\left(\tilde{\mathbf{x}}_{t-1}^{PQ}; \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} \tilde{\mathbf{x}}_0^{PQ} + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} \tilde{\mathbf{x}}_t^{PQ}, \tilde{\beta}\mathbf{I} \right). \tag{5}
$$

Furthermore, through the reparameterization trick, $\tilde{\mathbf{x}}_0^{PQ} = \frac{\tilde{\mathbf{x}}_t^{PQ} - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0}{\sqrt{\bar{\alpha}_t}}$ ($\boldsymbol{\epsilon}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$), Eq. 5 can be rewritten as:

$$
q(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \tilde{\mathbf{x}}_0^{PQ}) = \mathcal{N}\left(\tilde{\mathbf{x}}_{t-1}^{PQ}; \frac{1}{\sqrt{\alpha_t}} \tilde{\mathbf{x}}_t^{PQ} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}} \boldsymbol{\epsilon}_0, \tilde{\beta}\mathbf{I} \right). \tag{6}
$$

Also, the denoising network (i.e., prior distribution) can be modeled as:

$$
p_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_{t-1}^{PQ}|\tilde{\mathbf{x}}_t^{PQ}, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta})) = \mathcal{N}\left(\tilde{\mathbf{x}}_{t-1}^{PQ}; \frac{1}{\sqrt{\alpha_t}} \tilde{\mathbf{x}}_t^{PQ} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}} \boldsymbol{\epsilon}_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_t^{PQ}, t, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta})), \tilde{\beta}\mathbf{I} \right). \tag{7}
$$

Consequently, minimizing the KL divergence between the posterior distribution and the prior distribution in the denoising matching loss is equivalent to optimizing the parameters of noise network $\boldsymbol{\epsilon}_\theta(\tilde{\mathbf{x}}_t^{PQ}, t, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta}))$ to approach noise $\boldsymbol{\epsilon}_0$. The final loss function can be formulated as:

$$
\mathcal{L} = \mathbb{E}\left[\left\| \boldsymbol{\epsilon}_{\boldsymbol{\omega}}(\tilde{\mathbf{x}}_t^{PQ}, t, \mathrm{CN}_{\mathrm{enc}}(\tilde{\mathbf{d}}^{PQ}; \boldsymbol{\theta})) - \boldsymbol{\epsilon} \right\|_2^2 \right], \tag{8}
$$

which is equivalent to our practical denoising losses as in Eq. 12 in our manuscript.

B. More Visualization Results

In Fig. 1, we provide more qualitative comparison results on challenging low-overlap cases from ScanNet dataset [1].

References

[1] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *CVPR*, 2017. 2, 3

[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 2020. 1, 2

[3] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *ICML*, 2021. 1

[4] Runzhao Yao, Shaoyi Du, Wenting Cui, Canhui Tang, and Chengwu Yang. Pare-net: Position-aware rotation-equivariant networks for robust point cloud registration. In *ECCV*, 2024. 3

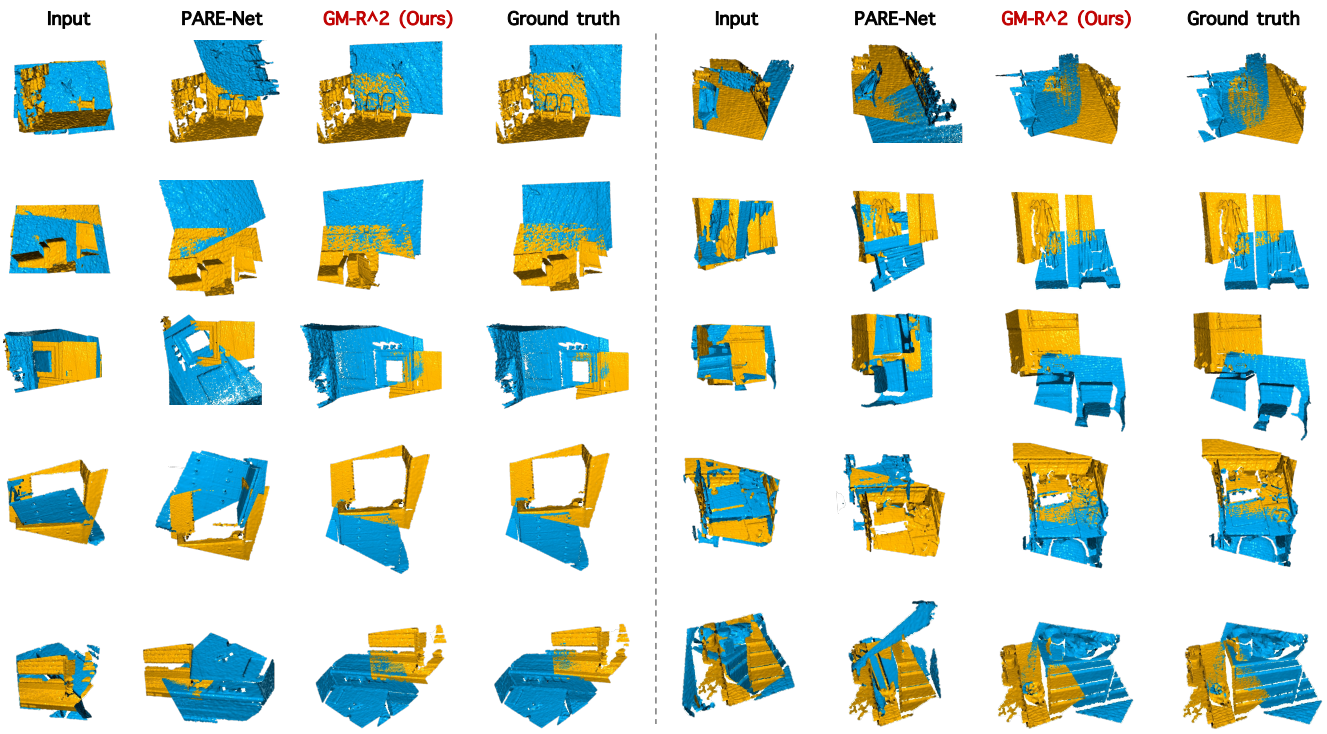

Figure 1. Qualitative comparison between the SOTA fully-supervised deep descriptor PARE-Net [4] and our unsupervised GM-R^2 descriptor in challenging low-overlap cases from **ScanNet** [1]. Our unsupervised method visually presents higher alignment precision.