# Supplementary Material for "Robust Outlier Rejection for 3D Registration with Variational Bayes"

Haobo Jiang[1], Zheng Dang[2], Zhen Wei[2], Jin Xie[*1], Jian Yang[*1], and Mathieu Salzmann[*2]

[1]PCA Lab, Nanjing University of Science and Technology, China
[2]CVLab, EPFL, Switzerland
{jiang.hao.bo, csjxie, csjyang}@njust.edu.cn
{zheng.dang, zhen.wei, mathieu.salzmann}@epfl.ch

## A. Proof of Variational Lower Bound

We propose two forms of variational lower bounds on our variational non-local network, including the point cloud-wise lower bound (§A.1) and the point-wise lower bound (§A.2). The former demonstrates the point-cloud distribution, while the latter shows a more detailed distribution for each point. For clarity, we apply the former in our paper.

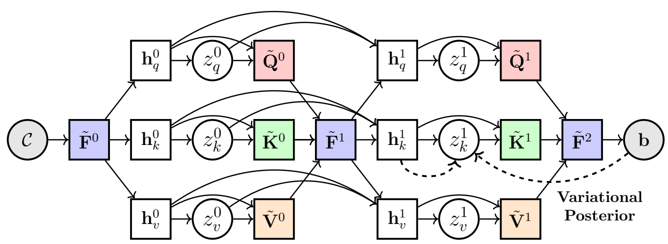


Figure 1. Probabilistic graphical model for our variational non-local network. For simplicity, we just demonstrate two iterations. The white circles indicate the random features and the white squares denote the deterministic hidden features. The solid line represents the inlier/outlier prediction process (generative process) and the dashed line denotes the label-dependent posterior encoder (inference model). We just show the inference model for $z_k^1$.

### A.1. Point Cloud-wise Variational Lower Bound

Given the putative correspondences $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_N\}$ and their inlier/outlier labels $\mathbf{b} = \{b_1, b_2, ..., b_N \mid b_i \in \{0, 1\}\}$, we first inject a set of random feature variables $z_{q,k,v}^{<L} = \{z_q^{l,i}, z_k^{l,i}, z_v^{l,i}\}_{0 \le l < L, 1 \le i \le N}$ into its log-likelihood correspondence labels $p_\theta(\mathbf{b} \mid \mathcal{C})$ and the initial variational lower bound can be derived using the Jensen's inequality as follows:

$$\begin{aligned}
&\ln p_\theta(\mathbf{b} \mid \mathcal{C}) = \ln \int_{z_{q,k,v}^{<L}} p_\theta(\mathbf{b}, z_{q,k,v}^{<L} \mid \mathcal{C}) \\
=&\ln \int_{z_{q,k,v}^{<L}} q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b}) \frac{p_\theta(\mathbf{b}, z_{q,k,v}^{<L} \mid \mathcal{C})}{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} = \ln \mathbb{E}_{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} \left[ \frac{p_\theta(\mathbf{b}, z_{q,k,v}^{<L} \mid \mathcal{C})}{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} \right] \\
\ge&\mathbb{E}_{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} \left[ \ln \frac{p_\theta(\mathbf{b}, z_{q,k,v}^{<L} \mid \mathcal{C})}{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} \right] \overset{(1)}{=} \mathbb{E}_{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} \left[ \ln \frac{p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) \cdot p_\theta(z_{q,k,v}^{<L} \mid \mathcal{C})}{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} \right],
\end{aligned} \tag{1}$$

where step (1) is based on the chain rule in probability theory. Next, based on the defined conditional dependencies of random variables in our probabilistic graphical model (Fig. 2), the chain rule is also used to factorize the posterior distribution

$q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})$ and the prior distribution $p_\theta(z_{q,k,v}^{<L} \mid \mathcal{C})$ in Eq. 1:

$$
\begin{aligned}
q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b}) =& q_\phi(z_{q,k,v}^{L-1} \mid z_{q,k,v}^{<L-1}, \mathcal{C}, \mathbf{b}) \cdot q_\phi(z_{q,k,v}^{<L-1} \mid \mathcal{C}, \mathbf{b}) \\
=& q_\phi(z_{q,k,v}^{L-1} \mid z_{q,k,v}^{<L-1}, \mathcal{C}, \mathbf{b}) \cdot q_\phi(z_{q,k,v}^{L-2} \mid z_{q,k,v}^{<L-2}, \mathcal{C}, \mathbf{b}) \cdot q_\phi(z_{q,k,v}^{<L-2} \mid \mathcal{C}, \mathbf{b}) \\
& \ldots\ldots \\
=& \Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
p_\theta(z_{q,k,v}^{<L} \mid \mathcal{C}) =& p_\theta(z_{q,k,v}^{L-1} \mid z_{q,k,v}^{<L-1}, \mathcal{C}) \cdot p_\theta(z_{q,k,v}^{<L-1} \mid \mathcal{C}) \\
=& p_\theta(z_{q,k,v}^{L-1} \mid z_{q,k,v}^{<L-1}, \mathcal{C}) \cdot p_\theta(z_{q,k,v}^{L-2} \mid z_{q,k,v}^{<L-2}, \mathcal{C}) \cdot p_\theta(z_{q,k,v}^{<L-2} \mid \mathcal{C}) \\
& \ldots\ldots \\
=& \Pi_{l=0}^{L-1} p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}).
\end{aligned}
\tag{3}
$$

By inserting the factorized posterior (Eq. 2) and prior (Eq. 3) distributions into Eq. 1, we can drive the detailed variational lower bound as below:

$$
\begin{aligned}
&\ln p_\theta(\mathbf{b} \mid \mathcal{C}) = \ln \int_{z_{q,k,v}^{<L}} p_\theta(\mathbf{b}, z_{q,k,v}^{<L} \mid \mathcal{C}) \\
\geq& \mathbb{E}_{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} \left[ \ln \frac{p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) \cdot p_\theta(z_{q,k,v}^{<L} \mid \mathcal{C})}{q_\phi(z_{q,k,v}^{<L} \mid \mathcal{C}, \mathbf{b})} \right] \\
=& \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \ln \frac{p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) \cdot \Pi_{l=0}^{L-1} p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C})}{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \right] \\
=& \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \ln p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) - \ln \frac{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})}{\Pi_{l=0}^{L-1} p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C})} \right] \\
=& \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \ln p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) - \sum_{l=0}^{L-1} \ln \frac{q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})}{p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C})} \right] \\
=& \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \ln p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) \right] - \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \sum_{l=0}^{L-1} \ln \frac{q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})}{p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C})} \right] \\
=& \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \ln p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) \right] - \sum_{l=0}^{L-1} \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \ln \frac{q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})}{p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C})} \right] \\
=& \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \ln p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) \right] - \sum_{l=0}^{L-1} \mathbb{E}_{\Pi_{\tau=0}^{l} q_\phi(z_{q,k,v}^{\tau} \mid z_{q,k,v}^{<\tau}, \mathcal{C}, \mathbf{b})} \left[ \ln \frac{q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})}{p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C})} \right] \\
=& \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b})} \left[ \ln p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L}, \mathcal{C}) \right] - \sum_{l=0}^{L-1} \mathbb{E}_{\Pi_{\tau=0}^{l-1} q_\phi(z_{q,k,v}^{\tau} \mid z_{q,k,v}^{<\tau}, \mathcal{C}, \mathbf{b})} \left[ \mathrm{D}_{\mathrm{KL}} \left( q_\phi(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}, \mathbf{b}) || p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C}) \right) \right] \\
=& \,\mathrm{ELBO}(\theta, \phi)
\end{aligned}
\tag{4}
$$

In our implementation, we use the deterministic hidden features $\{\mathbf{h}_{q,k,v}^{l}\}_{l=0}^{L-1}$ to summarize the historical information in previous iterations (*i.e.*, the condition parts of prior and posterior distributions) so that the lower bound can be rewritten as:

$$
\begin{aligned}
&\mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid \underbrace{z_{q,k,v}^{<l}, \mathcal{C}}_{\mathbf{h}_{q,k,v}^{l}}, \mathbf{b})} \left[ \ln p_\theta(\mathbf{b} \mid \underbrace{z_{q,k,v}^{<L}, \mathcal{C}}_{\tilde{\mathbf{F}}^{L}}) \right] - \sum_{l=0}^{L-1} \mathbb{E}_{\Pi_{\tau=0}^{l-1} q_\phi(z_{q,k,v}^{\tau} \mid z_{q,k,v}^{<\tau}, \mathcal{C}, \mathbf{b})} \left[ \mathrm{D}_{\mathrm{KL}} \left( q_\phi(z_{q,k,v}^{l} \mid \underbrace{z_{q,k,v}^{<l}, \mathcal{C}}_{\mathbf{h}_{q,k,v}^{l}}, \mathbf{b}) || p_\theta(z_{q,k,v}^{l} \mid \underbrace{z_{q,k,v}^{<l}, \mathcal{C}}_{\mathbf{h}_{q,k,v}^{l}}) \right) \right] \\
=& \mathbb{E}_{\Pi_{l=0}^{L-1} q_\phi(z_{q,k,v}^{l} \mid \mathbf{h}_{q,k,v}^{l}, \mathbf{b})} \left[ \ln y_\theta(\mathbf{b} \mid \tilde{\mathbf{F}}^{L}) \right] - \sum_{l=0}^{L-1} \mathbb{E}_{\Pi_{\tau=0}^{l-1} q_\phi(z_{q,k,v}^{\tau} \mid z_{q,k,v}^{<\tau}, \mathcal{C}, \mathbf{b})} \left[ \mathrm{D}_{\mathrm{KL}} \left( q_\phi(z_{q,k,v}^{l} \mid \mathbf{h}_{q,k,v}^{l}, \mathbf{b}) || p_\theta(z_{q,k,v}^{l} \mid \mathbf{h}_{q,k,v}^{l}) \right) \right]
\end{aligned}
\tag{5}
$$

where we use the correspondence features $\tilde{\mathbf{F}}^{L}$ of the last non-local iteration to summarize the condition parts of $p_\theta(z_{q,k,v}^{l} \mid z_{q,k,v}^{<l}, \mathcal{C})$. Also, to avoid ambiguity, we denote the label prediction model $p_\theta(\mathbf{b} \mid \tilde{\mathbf{F}})$ as $y_\theta(\mathbf{b} \mid \tilde{\mathbf{F}})$.

### A.2. Point-wise Variational Lower Bound

Furthermore, we extend Eq. 4 to a point-wise version. To this end, we rewrite the injected random variables $z_{q,k,v}^{<L}$ as $z_{q,k,v}^{<L,1:N} = \{z_q^{l,i}, z_k^{l,i}, z_v^{l,i}\}_{0 \le l < L, 1 \le i \le N}$ and we assume the points are independent. Thus, the prior and posterior distributions in Eq. 4 can be further divided as:

$$
\begin{aligned}
p_\theta(z_{q,k,v}^{l,1:N} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C}) &= \Pi_{i=1}^{N} p_\theta(z_{q,k,v}^{l,i} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C}) \\
q_\phi(z_{q,k,v}^{l,1:N} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C}, \mathbf{b}) &= \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{l,i} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C}, b_i)
\end{aligned}
\tag{6}
$$

Also, the point-wise label prediction model can be written as:

$$
\ln p_\theta(\mathbf{b} \mid z_{q,k,v}^{<L,1:N}, \mathcal{C}) = \ln \Pi_{i=1}^{N} p_\theta(b_i \mid z_{q,k,v}^{<L,1:N}, \mathcal{C}) = \sum_{i=1}^{N} \ln p_\theta(b_i \mid z_{q,k,v}^{<L,1:N}, \mathcal{C}) \tag{7}
$$

By inserting Eq. 6 and Eq. 7 into Eq. 4, we can achieve the following point-wise variational lower bound:

$$
\begin{aligned}
\text{ELBO}(\theta, \phi) = &\mathbb{E}_{\Pi_{l=0}^{L-1} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{l,i} | z_{q,k,v}^{<l,1:N}, \mathcal{C}, b_i)} \left[ \sum_{i=1}^{N} \ln p_\theta(b_i \mid z_{q,k,v}^{<L,1:N}, \mathcal{C}) \right] \\
&- \sum_{l=0}^{L-1} \mathbb{E}_{\Pi_{\tau=0}^{l} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{\tau,i} | z_{q,k,v}^{<\tau,1:N}, \mathcal{C}, b_i)} \left[ \ln \frac{\Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{l,i} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C}, b_i)}{\Pi_{i=1}^{N} p_\theta(z_{q,k,v}^{l,i} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C})} \right] \\
= &\mathbb{E}_{\Pi_{l=0}^{L-1} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{l,i} | z_{q,k,v}^{<l,1:N}, \mathcal{C}, b_i)} \left[ \sum_{i=1}^{N} \ln p_\theta(b_i \mid z_{q,k,v}^{<L,1:N}, \mathcal{C}) \right] \\
&- \sum_{l=0}^{L-1} \sum_{i=1}^{N} \mathbb{E}_{\Pi_{\tau=0}^{l} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{\tau,i} | z_{q,k,v}^{<\tau,1:N}, \mathcal{C}, b_i)} \left[ \ln \frac{q_\phi(z_{q,k,v}^{l,i} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C}, b_i)}{p_\theta(z_{q,k,v}^{l,i} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C})} \right] \\
= &\mathbb{E}_{\Pi_{l=0}^{L-1} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{l,i} | z_{q,k,v}^{<l,1:N}, \mathcal{C}, b_i)} \left[ \sum_{i=1}^{N} \ln p_\theta(b_i \mid z_{q,k,v}^{<L,1:N}, \mathcal{C}) \right] \\
&- \sum_{l=0}^{L-1} \sum_{i=1}^{N} \mathbb{E}_{\Pi_{\tau=0}^{l-1} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{\tau,i} | z_{q,k,v}^{<\tau,1:N}, \mathcal{C}, b_i)} \left[ \mathrm{D_{KL}} \left( q_\phi(z_{q,k,v}^{l,i} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C}, b_i) || p_\theta(z_{q,k,v}^{l,i} \mid z_{q,k,v}^{<l,1:N}, \mathcal{C}) \right) \right]
\end{aligned}
\tag{8}
$$

Similarity, we use the deterministic hidden feature $\mathbf{h}_{q,k,v}^{l,i}$ to summarize the historical information in previous iterations:

$$
\begin{aligned}
\text{ELBO}(\theta, \phi) = &\mathbb{E}_{\Pi_{l=0}^{L-1} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{l,i} | \underbrace{z_{q,k,v}^{<l,1:N}, \mathcal{C}}_{\mathbf{h}_{q,k,v}^{l,i}}, b_i)} \left[ \sum_{i=1}^{N} \ln p_\theta(b_i \mid \underbrace{z_{q,k,v}^{<L,1:N}, \mathcal{C}}_{\tilde{\mathbf{F}}^L}) \right] \\
&- \sum_{l=0}^{L-1} \sum_{i=1}^{N} \mathbb{E}_{\Pi_{\tau=0}^{l-1} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{\tau,i} | z_{q,k,v}^{<\tau,1:N}, \mathcal{C}, b_i)} \left[ \mathrm{D_{KL}} \left( q_\phi(z_{q,k,v}^{l,i} \mid \underbrace{z_{q,k,v}^{<l,1:N}, \mathcal{C}}_{\mathbf{h}_{q,k,v}^{l,i}}, b_i) || p_\theta(z_{q,k,v}^{l,i} \mid \underbrace{z_{q,k,v}^{<l,1:N}, \mathcal{C}}_{\mathbf{h}_{q,k,v}^{l,i}}) \right) \right] \\
= &\mathbb{E}_{\Pi_{l=0}^{L-1} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{l,i} | \mathbf{h}_{q,k,v}^{l,i}, b_i)} \left[ \sum_{i=1}^{N} \ln p_\theta(b_i \mid \tilde{\mathbf{F}}^L) \right] \\
&- \sum_{l=0}^{L-1} \sum_{i=1}^{N} \mathbb{E}_{\Pi_{\tau=0}^{l-1} \Pi_{i=1}^{N} q_\phi(z_{q,k,v}^{\tau,i} | z_{q,k,v}^{<\tau,1:N}, \mathcal{C}, b_i)} \left[ \mathrm{D_{KL}} \left( q_\phi(z_{q,k,v}^{l,i} \mid \mathbf{h}_{q,k,v}^{l,i}, b_i) || p_\theta(z_{q,k,v}^{l,i} \mid \mathbf{h}_{q,k,v}^{l,i}) \right) \right]
\end{aligned}
\tag{9}
$$

## B. Proof of Theorem 1

We let $\{^{(\kappa)}\mathcal{M}_i^{sac}\}_{i=1}^{J}$ be the randomly sampled *hypothetical inlier* subset in RANSAC, and let $\mathcal{C}_{in}$, $\mathcal{C}_{out}$ and $p_{in}$ be the inlier subset, outlier subset and the inlier ratio, respectively, $\mathcal{C} = \mathcal{C}_{in} \cup \mathcal{C}_{out}$, $p_{in} = |\mathcal{C}_{in}|/|\mathcal{C}|$. We also denote the inliers in seed subset $\mathcal{C}_{seed}$ as $\tilde{\mathcal{C}}_{in} = \mathcal{C}_{in} \cap \mathcal{C}_{seed}$. Then, we can derive the following theorem:

**Theorem 1.** *Assume the number of outliers in* ${}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_i}$ *(*$\mathbf{c}_i \in \tilde{\mathcal{C}}_{in}$*) follows a Poisson distribution* $Pois\,(\alpha \cdot \kappa)$*. Then, if* $\alpha < -\frac{1}{\kappa} \cdot \log\left[1-(1-p_{in}^{\kappa})^{J/|\tilde{\mathcal{C}}_{in}|}\right] \triangleq \mathcal{U}$*, the probability of our method achieving the inlier subset is greater than or equal to that of RANSAC.*

$$
\begin{aligned}
&P\Big(\max_{\mathbf{c}_i \in \mathcal{C}_{seed}} |{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_i} \cap \mathcal{C}_{in}| = \kappa\Big) \geq \\
&P\Big(\max_{1\leq i \leq J} |{}^{(\kappa)}\mathcal{M}_i^{sac} \cap \mathcal{C}_{in}| = \kappa\Big).
\end{aligned} \tag{10}
$$

*Proof.* The probability that RANSAC can achieve the inlier subset can be calculated via:

$$
\begin{aligned}
P\Big(\max_{1\leq i \leq J} |{}^{(\kappa)}\mathcal{M}_i^{sac} \cap \mathcal{C}_{in}| = \kappa\Big) &= 1 - P\Big(\max_{1\leq i \leq J} |{}^{(\kappa)}\mathcal{M}_i^{sac} \cap \mathcal{C}_{in}| < \kappa\Big) \\
&= 1 - P\Big(|{}^{(\kappa)}\mathcal{M}_1^{sac} \cap \mathcal{C}_{in}| < \kappa\Big) \cdots P\Big(|{}^{(\kappa)}\mathcal{M}_J^{sac} \cap \mathcal{C}_{in}| < \kappa\Big) \\
&\overset{(1)}{=} 1 - P\Big(|{}^{(\kappa)}\mathcal{M}_l^{sac} \cap \mathcal{C}_{in}| < \kappa\Big)^J \\
&= 1 - \Big(1 - P\Big(|{}^{(\kappa)}\mathcal{M}_l^{sac} \cap \mathcal{C}_{in}| = \kappa\Big)\Big)^J \\
&= 1 - \left(1 - \frac{C^{\kappa}_{|\mathcal{C}_{in}|}}{C^{\kappa}_{|\mathcal{C}|}}\right)^J \\
&= 1 - \left(1 - \frac{|\mathcal{C}_{in}| \cdots (|\mathcal{C}_{in}| - \kappa + 1)}{|\mathcal{C}| \cdots (|\mathcal{C}| - \kappa + 1)}\right)^J \\
&\leq 1 - \left(1 - \left(\frac{|\mathcal{C}_{in}|}{|\mathcal{C}|}\right)^{\kappa}\right)^J = 1 - (1 - p_{in}^{\kappa})^J,
\end{aligned} \tag{11}
$$

where step (1) is based on that random variables $\{|{}^{(\kappa)}\mathcal{M}_l^{sac} \cap \mathcal{C}_{in}| < \kappa\}_{1\leq l \leq J}$ are i.i.d. Analogously, the probability of our method achieving the inlier subset can be calculated via:

$$
\begin{aligned}
P\Big(\max_{\mathbf{c}_i \in \mathcal{C}_{seed}} |{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_i} \cap \mathcal{C}_{in}| = \kappa\Big) &\geq P\Big(\max_{\mathbf{c}_i \in \tilde{\mathcal{C}}_{in}} |{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_i} \cap \mathcal{C}_{in}| = \kappa\Big) \\
&= 1 - P\Big(\max_{\mathbf{c}_i \in \tilde{\mathcal{C}}_{in}} |{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_i} \cap \mathcal{C}_{in}| < \kappa\Big) \\
&= 1 - \Pi_{\mathbf{c}_i \in \tilde{\mathcal{C}}_{in}} P\Big(|{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_i} \cap \mathcal{C}_{in}| < \kappa\Big) \\
&\overset{(1)}{=} 1 - P\Big(|{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_l} \cap \mathcal{C}_{in}| < \kappa \mid \mathbf{c}_l \in \tilde{\mathcal{C}}_{in}\Big)^{|\tilde{\mathcal{C}}_{in}|} \\
&= 1 - \Big(1 - P\Big(|{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_l} \cap \mathcal{C}_{in}| = \kappa \mid \mathbf{c}_l \in \tilde{\mathcal{C}}_{in}\Big)\Big)^{|\tilde{\mathcal{C}}_{in}|} \\
&= 1 - \Big(1 - P\Big(|{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_l} \cap \mathcal{C}_{out}| = 0 \mid \mathbf{c}_l \in \tilde{\mathcal{C}}_{in}\Big)\Big)^{|\tilde{\mathcal{C}}_{in}|} \\
&\overset{(2)}{=} 1 - \big(1 - e^{-\alpha\cdot\kappa}\big)^{|\tilde{\mathcal{C}}_{in}|},
\end{aligned} \tag{12}
$$

where step (1) is based on that random variables $\{|{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_i} \cap \mathcal{C}_{in}| < \kappa\}_{\mathbf{c}_i \in \tilde{\mathcal{C}}_{in}}$ are i.i.d; Step (2) is based on our assumption of Poisson distribution: $P\Big(|{}^{(\kappa)}\tilde{\mathcal{M}}_{\mathbf{c}_i} \cap \mathcal{C}_{out}| = m \mid \mathbf{c}_i \in \tilde{\mathcal{C}}_{in}\Big) = \frac{(\alpha\cdot\kappa)^m e^{-\alpha\cdot\kappa}}{m!}$. Finally, we let $1 - (1 - e^{-\alpha\cdot\kappa})^{|\tilde{\mathcal{C}}_{in}|} \geq 1 - (1 - p_{in}^{\kappa})^J$ and can get that if $\alpha \leq -\frac{1}{\kappa} \cdot \log\left[1 - (1 - p_{in}^{\kappa})^{J/|\tilde{\mathcal{C}}_{in}|}\right]$, the inequality 10 holds. □

## C. Qualitative Evaluation

We first give some qualitative comparisons with PointDSC [1] (our baseline) on 3DLoMatch benchmark dataset in Fig. 2. As can be observed, in cases containing extremely low-overlapped regions (red box), our method can achieve more precise alignment. Those mainly benefit from our more discriminative correspondence embedding based on variational non-local network for more reliable inlier clustering. Also, we visualize the registration results on KITTI dataset in Fig. 3.

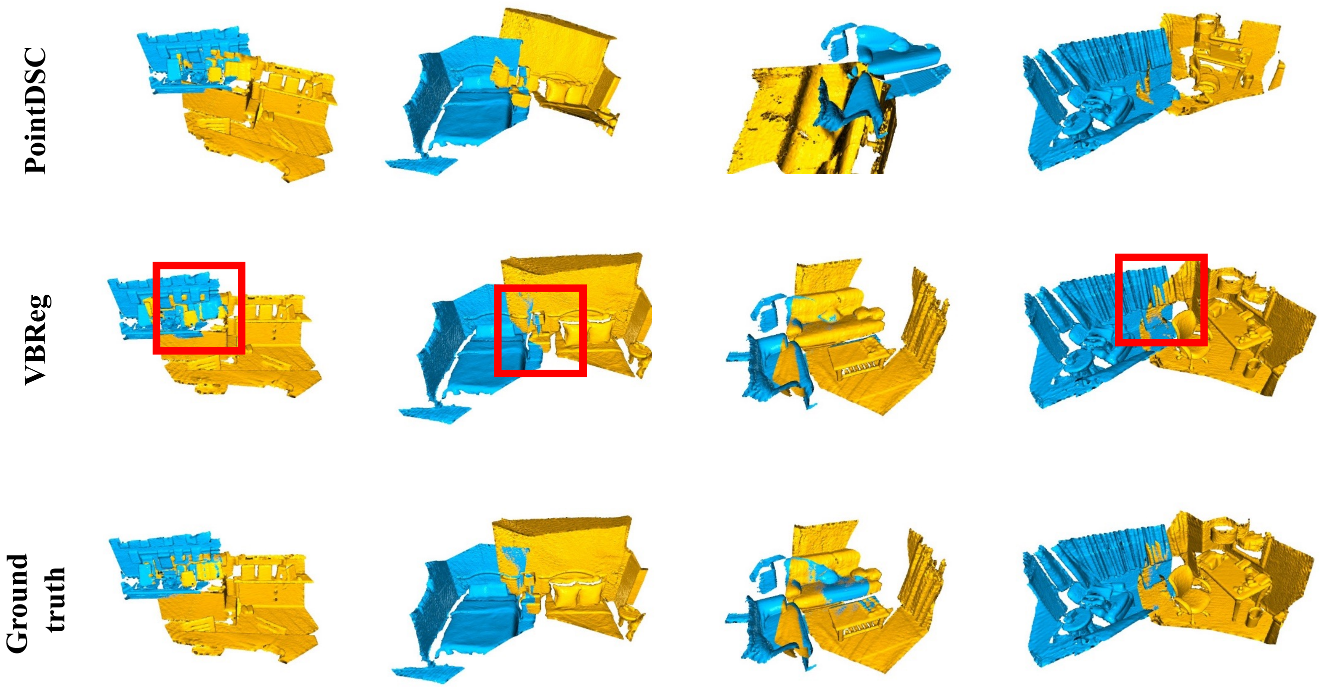


Figure 2. Qualitative comparison with PointDSC [1] (baseline) on 3DLoMatch benchmark [3].

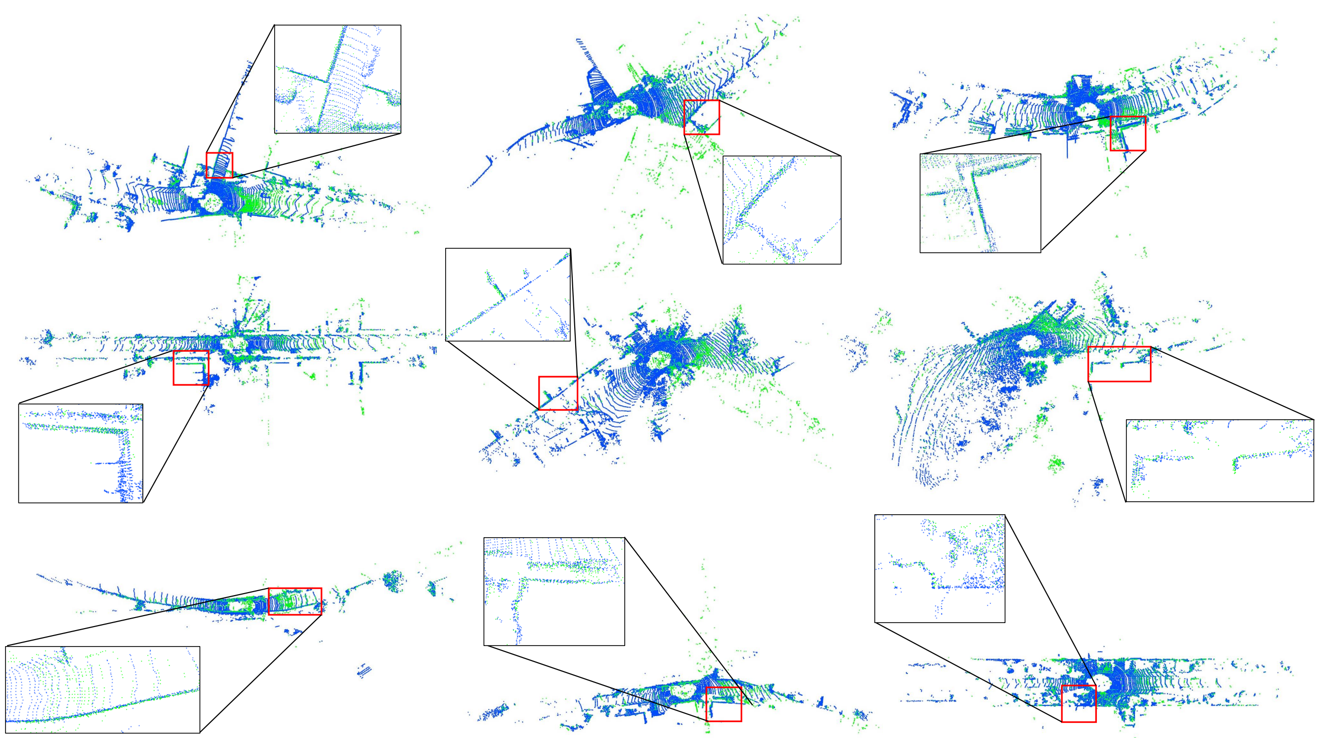

Figure 3. Registration visualization on KITTI benchmark [2].

## 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15859–15869, 2021. 4, 5

[2] 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*, pages 3354–3361. IEEE, 2012. 5

[3] 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*, pages 4267–4276, 2021. 5