Towards Self-Supervised Semantic Representation with a Viewpoint-Dependent Observation Model
Supplementary Material

Yuri Feldman and Vadim Indelman

This document provides supplementary material to [1]. Therefore, it should not be considered a self-contained document, but instead regarded as an appendix of [1]. Throughout this report, all notations and definitions are with compliance to the ones presented in [1].

A Derivation of the ELBO for the Viewpoint-Dependent Model

In the following we detail the derivation of Eq. (12) from Feldman and Indelman [1].

$$
\begin{aligned}
&KL(q_\phi(e \mid \mathcal{Z}_k, \mathcal{X}_k^{(rel)}) \parallel \mathbb{P}(e \mid \mathcal{Z}_k, \beta_k \mathcal{X}_k^{(rel)})) = \qquad (1)\\
&\mathop{\mathbb{E}}_{e\sim q_\phi}\{\log q_\phi(e \mid \mathcal{Z}_k, \mathcal{X}_k^{(rel)}) - \log \mathbb{P}(e, \mathcal{Z}_k, \beta_k \mid \mathcal{X}_k^{(rel)})\} + \log \mathbb{P}(\mathcal{Z}_k, \beta_k \mid \mathcal{X}_k^{(rel)})\\
&= KL(q_\phi(e \mid \mathcal{Z}_k, \mathcal{X}_k^{(rel)}) \parallel \mathbb{P}(e, \mathcal{Z}_k, \beta_k \mid \mathcal{X}_k^{(rel)})) + \log \mathbb{P}(\mathcal{Z}_k, \beta_k \mid \mathcal{X}_k^{(rel)}).
\end{aligned}
$$

Hence, we can write the evidence lower bound as

$$
\begin{aligned}
\log \mathbb{P}(\mathcal{Z}_k, \beta_k \mid \mathcal{X}_k^{(rel)}) &\geq -KL(q_\phi(e \mid \mathcal{Z}_k, \mathcal{X}_k^{(rel)}) \parallel \mathbb{P}(e, \mathcal{Z}_k, \beta_k \mid \mathcal{X}_k^{(rel)})) \qquad (2)\\
&= -KL(q_\phi(e \mid \mathcal{Z}_k, \mathcal{X}_k^{(rel)}) \parallel \mathbb{P}(e \mid \mathcal{X}_k^{(rel)}) \cdot \mathbb{P}(\mathcal{Z}_k, \beta_k \mid e, \mathcal{X}_k^{(rel)}))\\
&= -KL(q_\phi(e \mid \mathcal{Z}_k, \mathcal{X}_k^{(rel)}) \parallel \mathbb{P}(e)) + \mathop{\mathbb{E}}_{e\sim q_\phi}\{\log \mathbb{P}(\mathcal{Z}_k, \beta_k \mid e, \mathcal{X}_k^{(rel)})\},
\end{aligned}
$$

where $\mathbb{P}(e \mid \mathcal{X}_k^{(rel)}) = \mathbb{P}(e)$ since the semantic variable e and robot movement $\mathcal{X}_k^{(rel)}$ are independent in absence of measurements.

Y. Feldman is with the Department of Computer Science, Technion - Israel Institute of Technology, Haifa 32000, Israel, yurif@cs.technion.ac.il. V. Indelman is with the Department of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel, vadim.indelman@technion.ac.il.

B Derivation of the ELBO for the Viewpoint-Predictive Model

In this section we detail the derivation of Eq. (20) from Feldman and Indelman [1].

$$
\begin{aligned}
&KL(q_\phi(e \mid \mathcal{Z}_k) \parallel \mathbb{P}(e \mid \mathcal{Z}_{k+1}, \mathcal{Z}_k, \Delta\mathcal{X}_k)) = \\
&\underset{e \sim q_\phi}{\mathbb{E}} \{\log q_\phi(e \mid \mathcal{Z}_k) - \log \mathbb{P}(e, \mathcal{Z}_{k+1} \mid \mathcal{Z}_k, \Delta\mathcal{X}_k) + \log \mathbb{P}(\mathcal{Z}_{k+1} \mid \mathcal{Z}_k, \Delta\mathcal{X}_k)\} \\
&= KL(q_\phi(e \mid \mathcal{Z}_k) \parallel \mathbb{P}(e, \mathcal{Z}_{k+1} \mid \mathcal{Z}_k, \Delta\mathcal{X}_k)) + \log \mathbb{P}(\mathcal{Z}_{k+1} \mid \mathcal{Z}_k, \Delta\mathcal{X}_k).
\end{aligned}
\tag{3}
$$

Hence, we can write the evidence lower bound as

$$
\begin{aligned}
&\log \mathbb{P}(\mathcal{Z}_{k+1} \mid \mathcal{Z}_k, \Delta\mathcal{X}_k) \geq -KL(q_\phi(e \mid \mathcal{Z}_k) \parallel \mathbb{P}(e, \mathcal{Z}_{k+1} \mid \mathcal{Z}_k, \Delta\mathcal{X}_k)) \\
&= - \underset{e \sim q_\phi}{\mathbb{E}} \{\log q_\phi(e \mid \mathcal{Z}_k) - \log \frac{\mathbb{P}(\mathcal{Z}_{k+1}, \mathcal{Z}_k \mid e, \Delta\mathcal{X}_k) \cdot \mathbb{P}(e \mid \Delta\mathcal{X}_k)}{\mathbb{P}(\mathcal{Z}_k \mid \Delta\mathcal{X}_k)}\} \\
&= - \underset{e \sim q_\phi}{\mathbb{E}} \{\log q_\phi(e \mid \mathcal{Z}_k) - \log \mathbb{P}(e) - \log \mathbb{P}(\mathcal{Z}_{k+1}, \mathcal{Z}_k \mid e, \Delta\mathcal{X}_k)\} - \log \mathbb{P}(\mathcal{Z}_k \mid \Delta\mathcal{X}_k),
\end{aligned}
\tag{4}
$$

the last equation true since as before $\mathbb{P}(e \mid \Delta\mathcal{X}_k) = \mathbb{P}(e)$, as e and $\Delta\mathcal{X}_k$ are independent in absence of measurements, and since the last term is constant w.r.t. the expectation. We can further develop the expression to reach the final evidence lower bound:

$$
\begin{aligned}
&= -KL(q_\phi(e \mid \mathcal{Z}_k) \parallel \mathbb{P}(e)) + \\
&\underset{e \sim q_\phi}{\mathbb{E}} \{\log \mathbb{P}(\mathcal{Z}_{k+1} \mid \mathcal{Z}_k, e, \Delta\mathcal{X}_k) + \mathbb{P}(\mathcal{Z}_k \mid e)\} - \log \mathbb{P}(\mathcal{Z}_k \mid \Delta\mathcal{X}_k),
\end{aligned}
\tag{5}
$$

where $\mathbb{P}(\mathcal{Z}_k \mid e, \Delta\mathcal{X}_k) = \mathbb{P}(\mathcal{Z}_k \mid e)$ i.e. we can drop the conditioning on $\Delta\mathcal{X}_k$ since it only contains relative pose information and thus is by definition

References

[1] Y. Feldman and V. Indelman. Towards self-supervised semantic representation with a viewpoint-dependent observation model. In *Workshop on Self-Supervised Robot Learning, in conjunction with Robotics: Science and Systems*, July 2020.