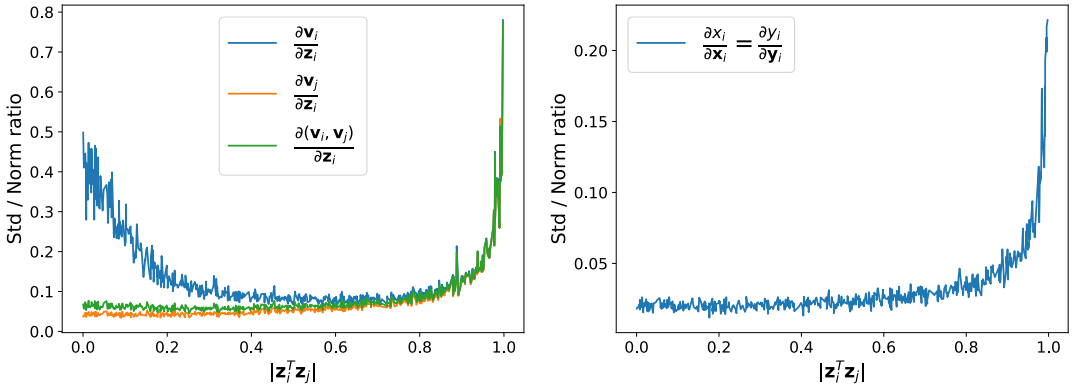


(a) Gradient with respect to $\boldsymbol{z}_i$.

(b) Gradient with respect to $\boldsymbol{x}_i$ (or $\boldsymbol{y}_i$).

Figure 4: Ratio between the variance of the gradient and its norm as a function of the similarity between the viewing directions $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$.

## A. Gradient of Common-Line loss at co-planar poses

In this section, we show how the presence of images with co-planar poses affect our common-lines based loss and, more specifically, its gradient.

Consider two images $o_i$ and $o_j$ and their predicted poses $R_i = (\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i) \in \mathbb{R}^{3\times 3}$ and $R_j = (\boldsymbol{x}_j, \boldsymbol{y}_j, \boldsymbol{z}_j) \in \mathbb{R}^{3\times 3}$. Note we assume $R_* \in \mathbb{R}^{3\times 3}$ rather than $R_* \in \mathrm{SO}(3)$ since our loss depends on $R_i, R_j$ parameterized as matrices in $\mathbb{R}^{3\times 3}$ and the following arguments are independent of how our neural networks predict elements of $\mathrm{SO}(3)$ (although the particular choice of parameterization can affect how this gradient is back-propagated through the model).

Note that the loss $\mathcal{L}(R_i, R_j)$ from Eq. 17 or Eq. 18 only depends on the coordinates in Eq. 16, i.e.:

$$x_i = \boldsymbol{l}_{ij}^T \boldsymbol{x}_i \quad y_i = \boldsymbol{l}_{ij}^T \boldsymbol{y}_i \quad x_j = \boldsymbol{l}_{ij}^T \boldsymbol{x}_j \quad y_j = \boldsymbol{l}_{ij}^T \boldsymbol{y}_j$$

For convenience, define $\boldsymbol{v}_i = (x_i, y_i)^T \in \mathbb{R}^2$ and $\boldsymbol{v}_j = (x_j, y_j)^T \in \mathbb{R}^2$ and write $\mathcal{L}_{ij} = \mathcal{L}(R_i, R_j)$. Then, using the chain rule:

$$\frac{\partial \mathcal{L}_{ij}}{\partial R_i} = \frac{\partial \boldsymbol{v}_i}{\partial R_i}\frac{\partial \mathcal{L}_{ij}}{\partial \boldsymbol{v}_i} + \frac{\partial \boldsymbol{v}_j}{\partial R_i}\frac{\partial \mathcal{L}_{ij}}{\partial \boldsymbol{v}_j} \tag{21}$$

Hence, in this section, we can focus only on the partial derivatives $\frac{\partial \boldsymbol{v}_i}{\partial R_i}$ and $\frac{\partial \boldsymbol{v}_j}{\partial R_i}$. We will study the partial derivatives with respect to $\boldsymbol{x}_i, \boldsymbol{y}_i$ and $\boldsymbol{z}_i$ independently.

First, note that:

$$\frac{\partial x_i}{\partial \boldsymbol{x}_i} = \frac{\partial y_i}{\partial \boldsymbol{y}_i} = \boldsymbol{l}_{ij} \tag{22}$$

$$\frac{\partial y_i}{\partial \boldsymbol{x}_i} = \frac{\partial x_i}{\partial \boldsymbol{y}_i} = 0 \tag{23}$$

and

$$\frac{\partial x_j}{\partial \boldsymbol{x}_i} = \frac{\partial y_j}{\partial \boldsymbol{x}_i} = \frac{\partial x_j}{\partial \boldsymbol{y}_i} = \frac{\partial y_j}{\partial \boldsymbol{y}_i} = 0 \tag{24}$$

Hence, we only need to consider the following quantities: $\frac{\partial x_i}{\partial \boldsymbol{x}_i}$, $\frac{\partial \boldsymbol{v}_i}{\partial \boldsymbol{z}_i} = (\frac{\partial x_i}{\partial \boldsymbol{z}_i}, \frac{\partial y_i}{\partial \boldsymbol{z}_i})$ and $\frac{\partial \boldsymbol{v}_j}{\partial \boldsymbol{z}_i} = (\frac{\partial x_j}{\partial \boldsymbol{z}_i}, \frac{\partial y_j}{\partial \boldsymbol{z}_i})$.

To understand how the similarity of $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$, we study the variance of the gradients when $R_j$ is perturbed by a small amount of noise, as a function of the similarity $|\boldsymbol{z}_i^T \boldsymbol{z}_j|$.

To do so, we sample 300 random pairs of rotations $R_i, R_j \in \mathrm{SO}(3)$ and compute their similarity $|\boldsymbol{z}_i^T \boldsymbol{z}_j|$. Then, for each pair, we generate 50 variations of $R_j$ by perturbing it with a small Gaussian noise with standard deviation $\sigma = 0.04$ in the quaternion space, and compute each gradient $\frac{\partial x_i}{\partial \boldsymbol{x}_i}$, $\frac{\partial \boldsymbol{v}_i}{\partial \boldsymbol{z}_i}$ and $\frac{\partial \boldsymbol{v}_j}{\partial \boldsymbol{z}_i}$.

For each pair, we compute the average norm of the gradients (Frobenious norm for the Jacobians) and the standard deviation (over the 50 samples) of each partial derivative, which we average to obtain a single number. In Fig. 4, we plot the ratio between the standard deviation and the average norm for each pair, as a function of the similarity $|\boldsymbol{z}_i^T \boldsymbol{z}_j|$.

Whenever $\boldsymbol{z}_i$ is close to $\pm \boldsymbol{z}_j$, the variance of the gradient is very close to its average norm (the ratio approaches 1); this is particularly true for the gradients of $\boldsymbol{z}_i$, see Fig. 4a but less severe for $\boldsymbol{z}_i$ and $\boldsymbol{y}_i$. That result suggests that the training process can be particularly unstable in this setting, especially since the gradient on $\boldsymbol{z}_i$ is necessary to leave this situation but it is also the most affected by that.

## B. Common-Line Loss Landscape

In this section, we provide a simple study of the common-line loss landscape. To do so, we compute the common line loss between two random images $o_i$ and $o_j$, respectively at the poses $R_i R_{\theta_1}$ and $R_j R_{\theta_2}$.

$R_i = (\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i)$ and $R_j = (\boldsymbol{x}_j, \boldsymbol{y}_j, \boldsymbol{z}_j)$ are the ground-truth poses of $o_i$ and $o_j$.

$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is a 2D rotation by $\theta \in [0, 2\pi)$.

Hence, $R_i R_\theta$ is the pose of $o_i$ rotated by $\theta$ around its projection axis $\boldsymbol{z}_i$ and simply corresponds to rotating the common line $\boldsymbol{v}_i = (\boldsymbol{x}_i^T \boldsymbol{l}_{ij}, \boldsymbol{y}_i^T \boldsymbol{l}_{ij})^T \in \mathbb{R}^2$ in the image $o_i$ by $\theta$. Note that if $\theta_1 = \theta_2 = \pi$, the loss is unchanged since the common line predicted is the same, only reflected.

In Fig. 5, we plot the common-line loss in Eq. 18 as a function of $\theta_1$ and $\theta_2$, for different random pairs $(i, j)$. Note that any possible predicted pair of common lines corresponds to a point in the figure (multiple choices of $R_i, R_j \in \mathrm{SO}(3)$ lead to the same common lines). This enables us to study the complete loss landscape for the simple case of $N = 2$ images. The right column of Fig. 5 highlights the global minima of the loss.

First, we note the expected periodicity of the loss by $\theta_1 = \theta_2 = \pi$ in all images.

In the first pair (first row), we also observe spurious global minima at $\theta_1 = 0$ and $\theta_2 = \pi$ (and the opposite), which corresponds to a reflection of the correct common line in only one of the two images. This is likely related to the spurious planar symmetry described in Levy et al. (2022a) and in Sec. 4, which motivated the use of a "symmetrized loss" in Levy et al. (2022a).

We also note that the landscape can vary a lot over different pairs. While the first pair has a smooth landscape with two clear global minima at expected locations $(0, 0)$ and $(\pi, \pi)$, other pairs show multiple global optima. In some cases, like the last row, the two locations $(0, 0)$ and $(\pi, \pi)$ are close but not exactly global optima.

We also compare the original formulation of the loss in Eq. 17 with the modified version in Eq. 18 which we use in our experiments. Fig. 6 shows similar plots obtained using Eq. 17 (pairs are randomly sampled and don't necessary match those in Fig. 5). When using the original loss in Eq. 17, the global optima often do not include the ground-truth $(0, 0)$ and $(\pi, \pi)$.

Finally, we emphasise that this study is limited to the case $N = 2$. However, during training, the loss is averaged over multiple pairs.

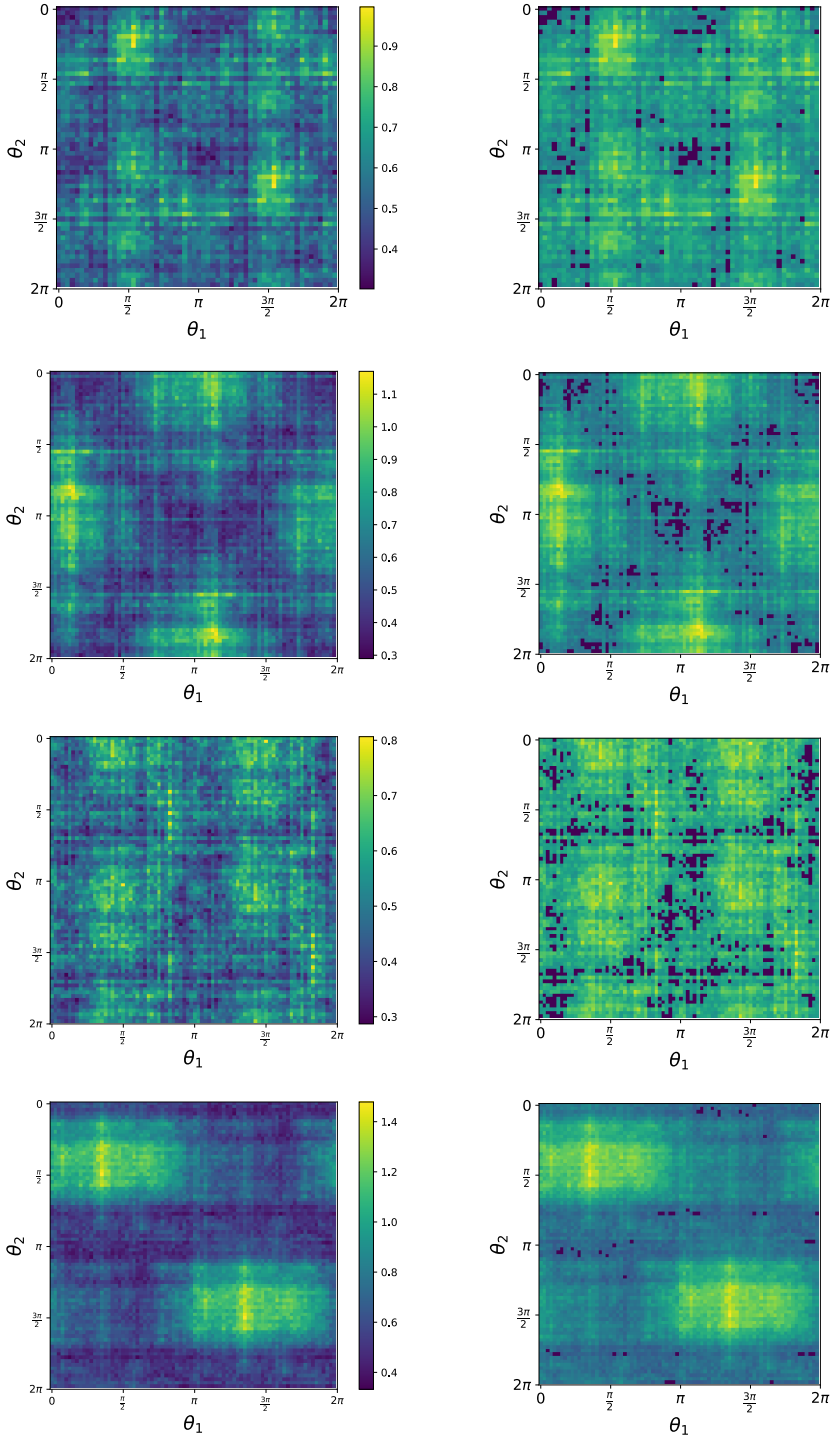


Figure 5: Common line loss using Eq. 18 between two random images $o_i$ and $o_j$ using the poses $R_i R_{\theta_1}$ and $R_j R_{\theta_2}$, with $\theta_1, \theta_2 \in [0, 2\pi)$. Each row is a different random pair $(i, j)$. In the right column, areas where $\mathcal{L} < 0.4$ are highlighted with a darker color.

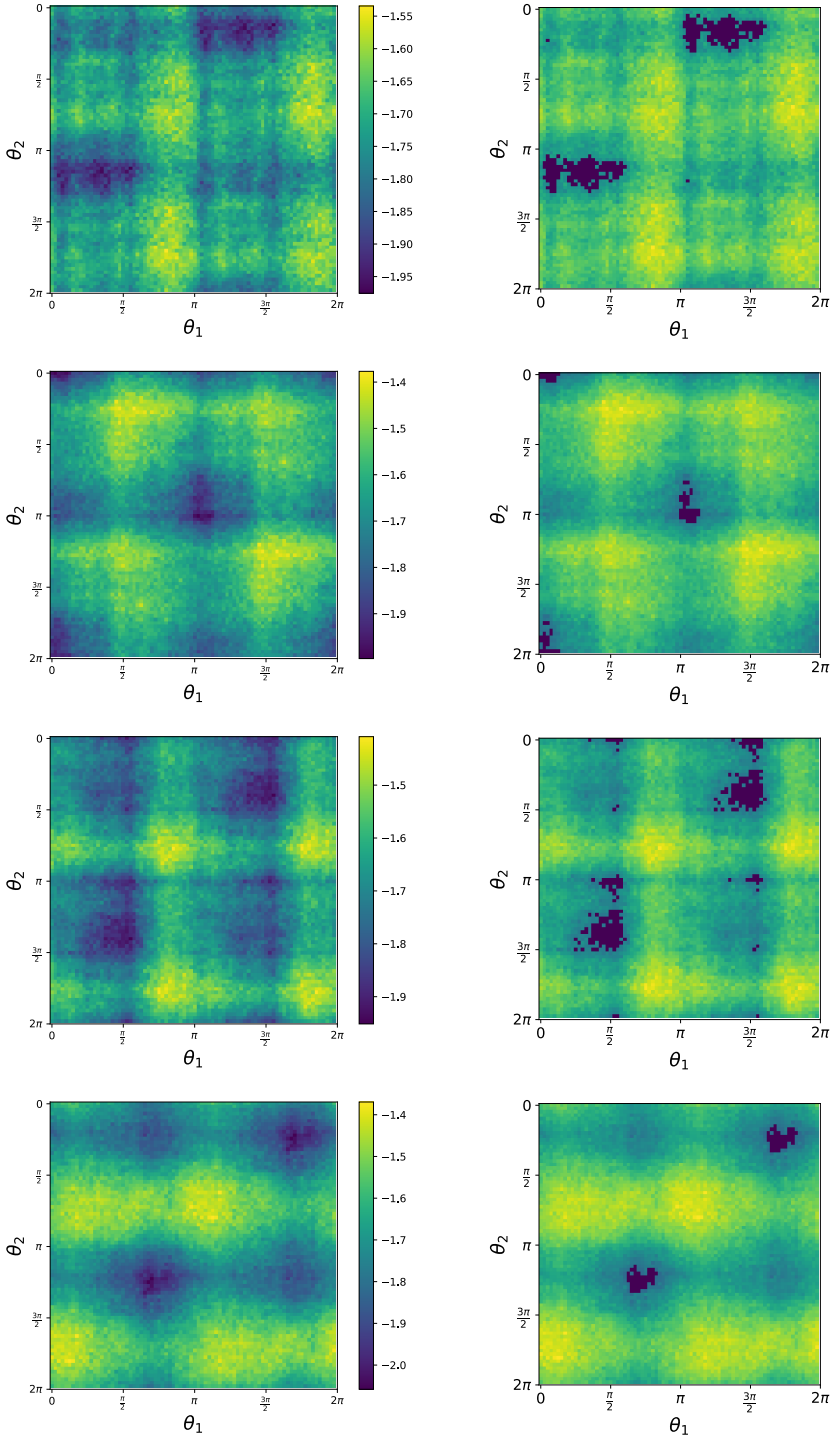


Figure 6: Common line loss using Eq. 17 between two random images $o_i$ and $o_j$ using the poses $R_i R_{\theta_1}$ and $R_j R_{\theta_2}$, with $\theta_1, \theta_2 \in [0, 2\pi)$. Each row is a different random pair $(i, j)$. In the right column, the points closer to global optima are highlighted with a darker color.