Efficient differentiable implementation of Mean-Field inference for High-Order Potentials.

In this document, we concisely explain how Mean-Fields (MF) inference can be efficiently implemented with the High-Order potentials described in the main paper.

More precisely, for each pixel k and location i, we seek to compute efficiently the approximated natural gradient term

$$\widetilde{\nabla \eta_i}^k = -C^k \left(\mathbb{E}_Q \left[\Delta^k(Z) | Z_i = 1 \right] - \mathbb{E}_Q \left[\Delta^k(Z) | Z_i = 0 \right] \right) , \tag{1}$$

and then sum these terms for every pixel to obtain the approximated natural gradient $\widetilde{\nabla \eta_i}$.

Computing each natural gradient term Recall that $\Delta^k(Z)$ is a function of Z which takes value 0 if one of the "compatible explanations" for pixel k is present and 1 otherwise. Also, recall that we say that an explanation Z_i is compatible if a presence in Z_i gets projected on the camera plane in such a way that it matches the observation at pixel k produced by the discriminative model. Let $\mathcal{C}^k$ denote the list of indices $j \in \{1, \ldots, N\}$ such that a presence in Z_j is a compatible explanation for the observation at pixel k.

Let us consider pixel k and location i. If location i is not compatible with pixel k (i.e. $i \notin \mathcal{C}^k$), then the value taken by Z_i has no impact on $\Delta^k(Z)$ and therefore $\widetilde{\nabla \eta_i}^k = 0$.

Let us assume that $i \in \mathcal{C}^k$. Then,

$$\mathbb{E}_Q \left[\Delta^k(Z) | Z_i = 1 \right] = 0$$

and,

$$\begin{aligned} \mathbb{E}_Q \left[\Delta^k(Z) | Z_i = 0 \right] &= \prod_{j \in \mathcal{C}^k / i} (1 - Q(Z_j = 1)) \\ &= \frac{\prod_{j \in \mathcal{C}^k} (1 - Q(Z_j = 1))}{1 - Q(Z_i = 1)} , \end{aligned} \tag{2}$$

where the first equation the fact that the MF distribution Q is fully factorized.

Computing Updates for all variables in two steps Computing the gradient term of Eq. 2 directly would require a large multiplication for each pixel, which would be inefficient. However, we remark that the numerator of Eq. 2, doesn't depend on the chosen i, and its denominator doesn't depend on k. We therefore proceed using the two following steps

- For each pixel k, we compute

$$\delta_k = \prod_{j \in \mathcal{C}^k} (1 - Q(Z_j = 1)) . \tag{3}$$

- Then, for each variable index i, we compute the sum over all pixels

$$\widetilde{\nabla\eta_i} = \frac{1}{1 - Q(Z_i = 1)} \sum_{k|i \in \mathcal{C}^k} \delta_k \ . \tag{4}$$

Note that these operations are all differentiable with respect to the MF distribution Q and to the parameter C^k, which makes it possible to back-propagate the gradient through the MF iterations.

Furthermore, since the Gaussians were approximated for inference by constant terms, on a rectangular zones, the sum of Eq. 4, can be computed efficiently using integral images.