# A. Appendix.

## A.1. Optimal Transport and Sinkhorn Iteration

To ensure the integrity of this paper, we briefly introduce the derivation of the Sinkhorn Iteration algorithm which we emphasize not our contributions and belongs to textbook knowledge.

The mathematical formula of the Optimal Transport problem is defined in Eq. 1. This is a linear program which can be solved in polynomial time. For dense detectors, however, the resulting linear program is large, involving the square of feature dimensions with anchors in all scales. This issue can be addressed by a fast iterative solution, which converts the optimization target in Eq. 1 into a non-linear but convex form with an entropic regularization term $E$ added:

$$\min_{\pi} \quad \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\pi_{ij} + \gamma E(\pi_{ij}),$$

where $E(\pi_{ij}) = \pi_{ij}(\log \pi_{ij} - 1)$. $\gamma$ is a constant hyper-parameter controlling the intensity of regularization term. According to Lagrange Multiplier Method, the constraint optimization target in the above equation can be convert to a non-constraint target:

$$\begin{aligned} \min_{\pi} \quad & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\pi_{ij} + \gamma E(\pi_{ij}) + \\ & \alpha_j(\sum_{i=1}^{m}\pi_{ij} - d_j) + \beta_i(\sum_{j=1}^{n}\pi_{ij} - s_i), \end{aligned}$$

where $\alpha_j (j = 1, 2, ...n)$ and $\beta_i (i = 1, 2, ..., m)$ are Lagrange multipliers. By letting the derivatives of the optimization target equal to 0, the optimal plan $\pi^*$ is resolved as:

$$\pi^*_{ij} = \exp(-\frac{\alpha_j}{\gamma})\exp(-\frac{c_{ij}}{\gamma})\exp(-\frac{\beta_i}{\gamma}).$$

Letting $u_j = \exp(-\frac{\alpha_j}{\gamma}), v_i = \exp(-\frac{\beta_i}{\gamma}), M_{ij} = \exp(-\frac{c_{ij}}{\gamma})$, the following constraints can be enforced:

$$\begin{aligned} \sum_{i}\pi_{ij} &= u_j(\sum_{i} M_{ij}v_i) = d_j, \\ \sum_{j}\pi_{ij} &= (u_j\sum_{i} M_{ij})v_i = s_i. \end{aligned}$$

These two equations have to be satisfied simultaneously. One possible solution is to calculate $v_i$ and $u_j$ by repeating the following updating formulas sufficient steps:

$$u_j^{t+1} = \frac{d_j}{\sum_i M_{ij}v_i^t}, \quad v_i^{t+1} = \frac{s_i}{\sum_j M_{ij}u_j^{t+1}}.$$

The updating rule in the above equation is also known as the Sinkhorn-Knopp Iteration. After repeating this iteration $T$ times, the approximate optimal plan $\pi^*$ can be obtained:

$$\pi^* = diag(v)Mdiag(u).$$

$\gamma$ and $T$ are empirically set to 0.1 and 50. Please refer to our code for more details.