Multi-view supervision for single-view reconstruction via differentiable ray consistency.

Summary

1. Reconstruction of 3D shape given only 2D observations.

2. The loss function is:

$$L(x;(O,C)) = \sum_{r \in R} L_r(x)$$

view consistency loss

per-ray consistency term.

3. $L_r(x)$ captures if the inferred 3D model x correctly explains the observation associated with the specific ray r.

Ray-tracing in a probabilistic occupancy grid

1. To define $L_r(x)$, they examine the ray r as it travels across the voxel grid with occupancy probabilities x.

2. A probabilistic occupancy grid induces a distribution of events that can occur to the ray r

3. The loss $L_r(x)$ characterizes the incompatibility of those events with the available observations O_r.

4. Assume ray r passes through N_r voxels.

5. The events associated with this ray correspond to it either terminating at one of these N_r voxels or passing through.

6. So $z_r = i$ indicates the ray terminates at voxel i.

7. $z_r = i$ iff the previous voxels in the path are all unoccupied and the i^{th} voxel is occupied.

8. Assuming an independent distribution of occupancies where the prediction x_i^r corresponds to the probability of the i^{th} voxel on the path of the ray r being empty, then:

$$p(z_r = i) = \begin{cases} (1 - x_i^r) \prod_{j=1}^{i-1} x_j^r, & \text{if } i \leq N_r \\ \prod_{j=1}^{N_r} x_j^r & \text{if } i = N_r + 1 \end{cases}$$

Event Cost Functions

1. Each events $z_r = i$ induces a prediction, namely the distance d_i^r the ray travels before terminating.

2. We can define a cost function between the induced prediction and the observation. $\psi_r(i)$.

3. $\psi_r(i)$ assigns cost to event $z_r = i$ based on whether it produces prediction inconsistent with o_r.

4. e.g. if depth observation is available, $\psi_r(i)$ is the distance b/t observed distance and event-induced distance

$$\psi_r^{depth}(i) = | d_i^r - d_r |$$

Ray-consistency loss

1. $$L_r(x) = \mathop{E}_{z_r}[\psi_r(z_r)] = \sum_{i=1}^{N_r+1} \psi_r(i)\, p(z_r = i)$$