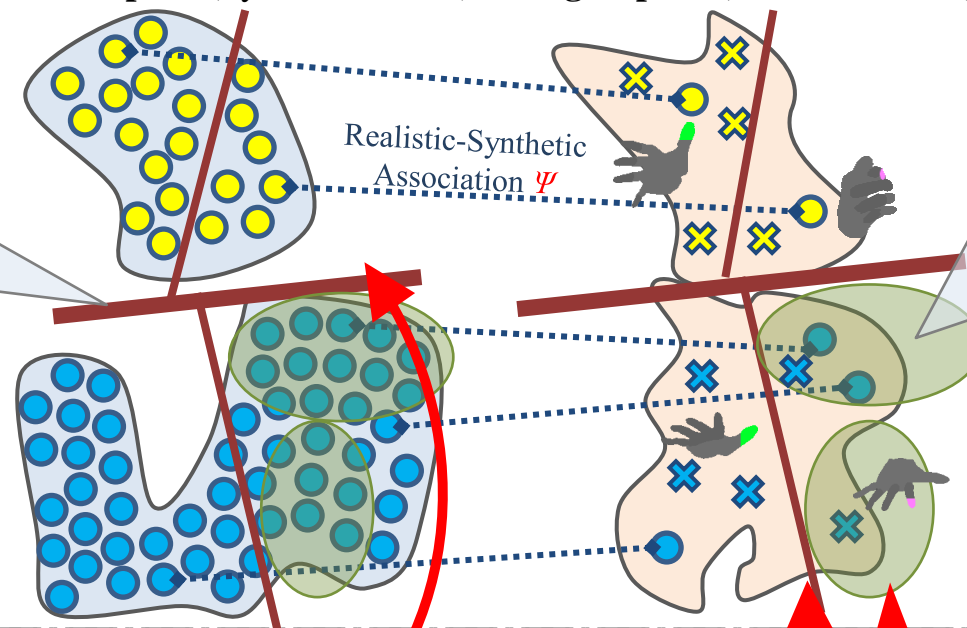


## Training Dataset $D$

**Transductive learning:** The realistic-synthetic fusion are learned by the transductive term  $Q_t$  throughout the whole forest.

Source space (Synthetic data  $S$ )    Target space (Realistic data  $R$ )



- Labelled datapoints
- ⊗ Unlabelled datapoints
- Tree split

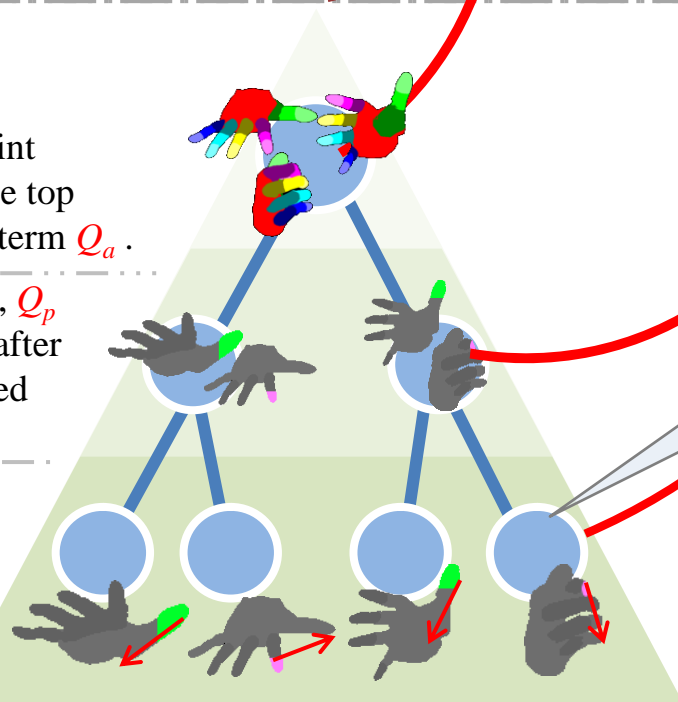
**Semi-supervised learning:** Labelled and unlabelled data are clustered via  $Q_u$ , by comparing appearances of patches.

## STR Forest

**Viewpoint Classification:** Viewpoint classification is first performed at the top levels, controlled by the viewpoint term  $Q_a$ .

**Joint Classification:** At mid levels,  $Q_p$  determines classification of joints, after most viewpoints have been classified successfully.

**Regression:** To describe the distribution of realistic data, nodes are optimised for data compactness via  $Q_v$  and  $Q_u$  towards the bottom levels.



**Voting scheme:** Each terminal node contains 3D vectors that vote for the 16 joint locations, each joint is modelled as an independent Gaussian distribution.



**Joint refinement:** Occluded joints are recovered from comparing with a dataset of hand poses.