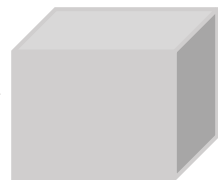


Input Image



FCNs
»»»



Shared

FCNs
»»»



Input Video



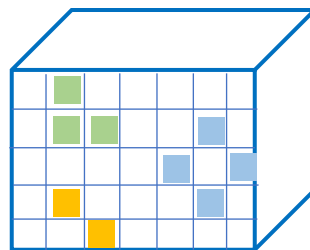
Instance
Segmentation

Classification
Mask Prediction

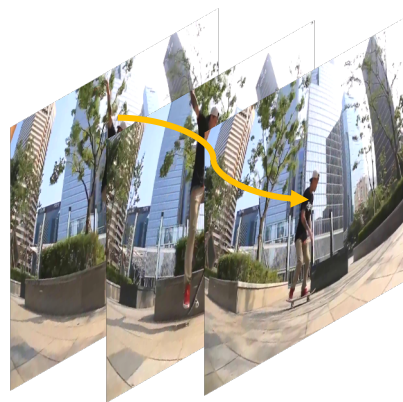
Image
Annotations



Embedding
branch

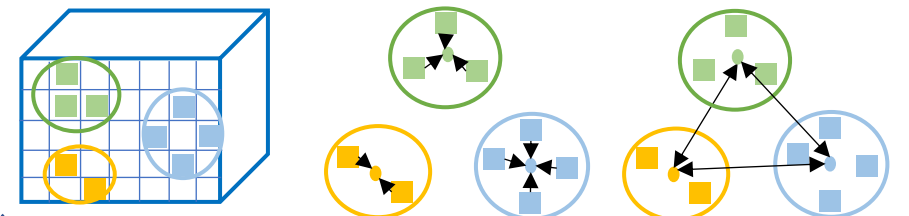


Instance Embedding

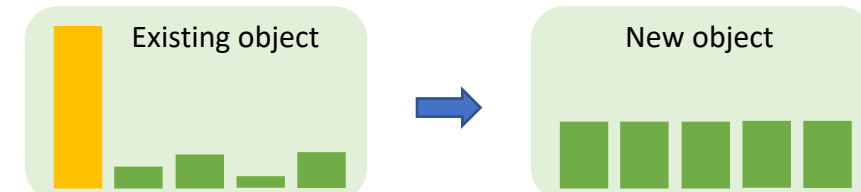


Video Instance Correspondence

(a) Instance Contrastive Loss



(b) Maximum Entropy Regularization



(c) Video Instance Correspondence

