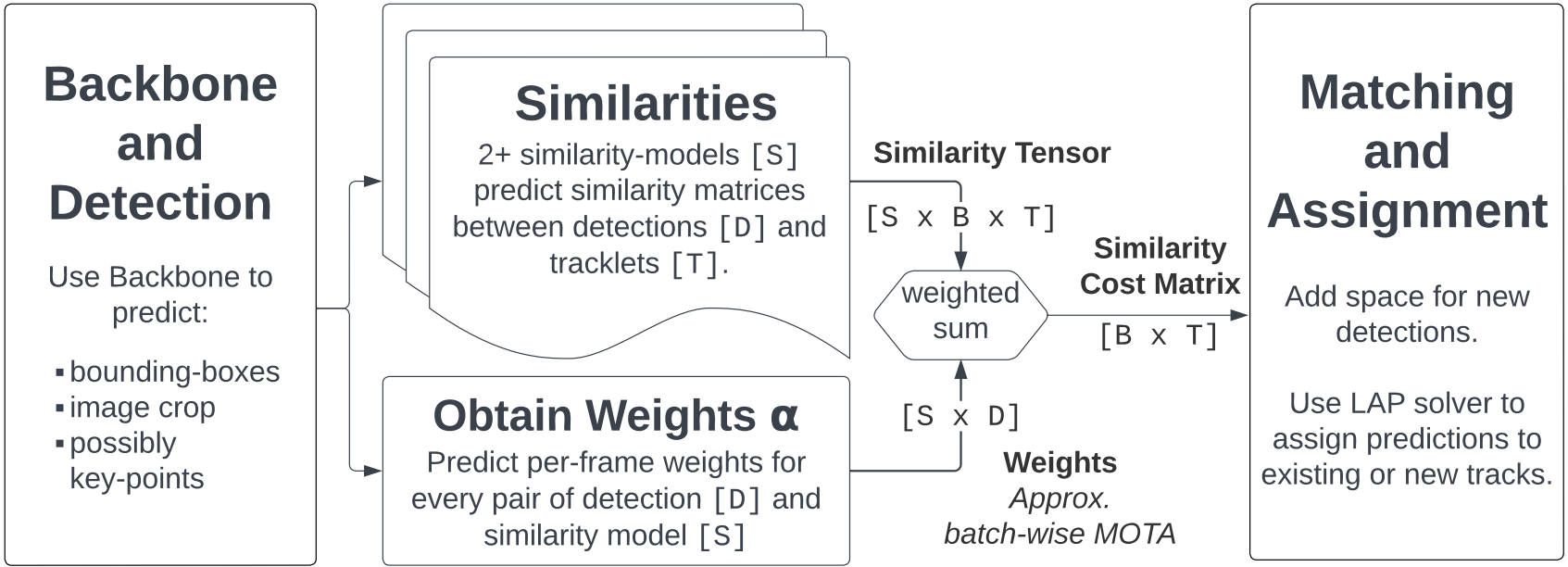
Backbone and Detection
Use Backbone to predict:
bounding-boxes
image crop
possibly key-points
Similarities
2+ similarity-models [S] predict similarity matrices between detections [D] and tracklets [T].
Obtain Weights α
Predict per-frame weights for every pair of detection [D] and similarity model [S]
Similarity Tensor
[S x B x T]
weighted sum
[S x D]
Weights
Approx. batch-wise MOTA
Similarity Cost Matrix
[B x T]
Matching and Assignment
Add space for new detections.
Use LAP solver to assign predictions to existing or new tracks.