- Newly introduced **track queries** created by DETR keep track of the object by embedding the position (**autoregressive**). They are modified by decoder after each frame.
- Model performs attention on frame-level features and track queries for tracking implicitly in a unified way. For new objects, transformer spawns new track queries.
- Object queries interact in decoder self-attention layer where they can avoid looking for duplicates.
- At t = 0, N_{obj} boxes are produced, from these, boxes which are associated with an object are saved as **track queries**. Now N_{track} dynamic + N_{obj} static queries are passed from next timestamp.
- N_{track} queries follow existing objects while N_{obj} queries initialise **track queries** for new objects.
- While training, 2 adjacent frames are chosen. Frame t-1 gives N_{obj} preds which are matched to GT. The matched preds are passed on as track queries. Matching is done accordingly and set prediction loss.
- Track Augmentations :
    - Previous frame t-1 is sampled from frames around t to account for camera motion where there is substantial motion.
    - False negatives from frame t-1 are sampled with probability p_{FN} and track queries are removed before step ii to train **object queries** as it will trigger new detection. Also, ratio of False positives is kept high to train **track queries** with a balance.
    - FPs are added to track queries so that model is encouraged to stop track in case of occlusion. p_{FP}, the probability with which a track query gives rise to a false positive is chosen with a large likelihood to encourage occlusions.
- NMS is used as decoder self-attention cannot discriminate between highly overlapping tracks relating to the same target. So a high IoU NMS threshold of 0.9 is used.
- Single image track training : simulated track by spatial perturbations to get 2 adjacent frames.
- Track initialization filtering is used for public detection config by considering a box only if its predicted center falls within GT BB **or** IoU is above threshold.
- Public detections can be done using diff models. It was observed that CD filtering does not reflect on performance of public detection and hence was not right for TrackFormer. Hence IoU filter was applied.
-

Things to get back to :

- Is NMS applied only on track queries output