After doing transfer learning on the baseline model, we failed to find a way to improve the metric scores. The reason we know they can improve is because the paper’s from the competition are around 10% higher. Here are some areas that we want to investigate/share for this week’s check-in:

**Ensemble Method:**

We hope that our final model will come from the results of YOLO 11s, Faster R-CNN, and RT-DETR-L (transformer) combined. The plan would be to send the same image to each model and have the model predict the bounding box. Then, we could use non-max suppression to get a single predicted box for each polyp.

**Sequence Data:**

From EndoCV, we have single frame images coming from 6 hospitals along with 23 positive video sequences coming from some of the hospitals. EndoCV also provided negative sequence data coming from a different hospital. In our first attempts at transfer learning, we included the sequence data in our training, validation and test sets but used random sampling to take some frames from the sequences. Because we did not do the transfer learning on just the single frames, we are not sure how sequences are impacting performance. Thus, we want to take a step back and understand their impact.

So, we plan to look at the difference in the results when we train on a single frame versus containing sequences in our training.

**Static v Dynamic Augmentation:**

Somewhat related to the above topic, we want to check model performance when we use static or dynamic augmentation. Our thought right now is that in the dynamic augmentation, the model is not seeing enough images, especially different looking images. Thus, it makes sense to generate a larger dataset using static augmentation that has simple adjustments such as flipping, cropping, shearing, zoom-in and zoom-out.