

1. Provide your team background and organization description (if applicable).

I am Harshit Sheoran, I am 21 years old and I am pursuing my undergrad from Goldsmiths, University of London

I am super passionate about deep learning with computer vision so I do a lot of competitions online, from last 1 year I have been doing a lot of work outside kaggle.

2. Explain why you participated in the DigiLut challenge.

Playing around with a one of a kind dataset is always very very appealing to me plus a good prize is always a good motivator, large dataset size is a big pro for me as most of the time, larger datasets have less randomness involved

3. Describe how you built your winning model and elaborate on the technical and modeling choices you made.

My approach has two parts, the first part is ranking and second is detection

Each massive WSI is split into “structures” these are the lumps of tissue joined together in the image, on average an WSI has 50 of them

Ranking model takes a structure and predicts whether the structure has a bbox inside

Detection model is responsible to make the actual bbox prediction, for this, I take only the structures that have bbox inside them and train an object detection model on this data

For inference, I preprocess wsi into structures, each structure is given to Ranking and Detection models, for ranking I take the prediction and for Detection I take the confidence of the most confident bbox prediction, I average them and I select top-X structure for each WSI where X is the number of bbox sample_submission.csv file expects

Then, on the selected structures, I apply soft-NMS to the Detection model’s prediction for the final bbox

To use the extra unlabeled data that is provided, I make pseudo-labels by inferencing on it and selecting top X% ranked structures and bboxes (just like how I make test predictions), I use this data to train two more detection models and then re-predict the test



In the end, I have 4 folds from CoAT, 4 folds from MaxViT for ranking and 3 Object Detection Co-DINO models

4. What GPU/CPU/RAM resources you used to build your model

I have AMD TR 3960X CPU with 128GB ram (more is highly recommended for faster preprocessing) and 4 NVIDIA 3090 GPUs