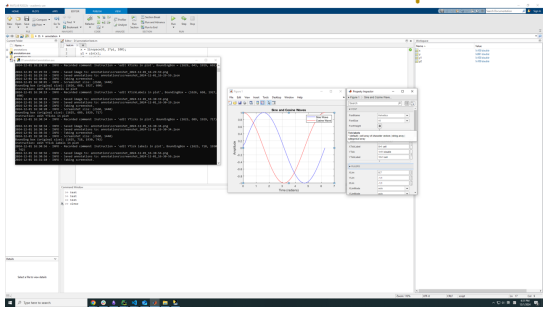
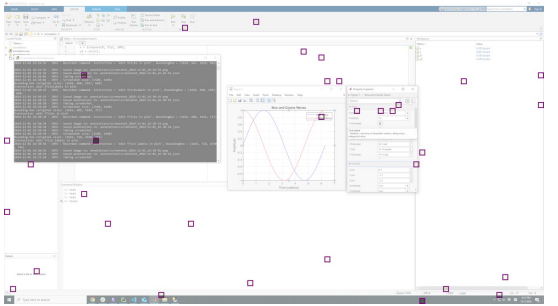
Instruction: edit font in plot

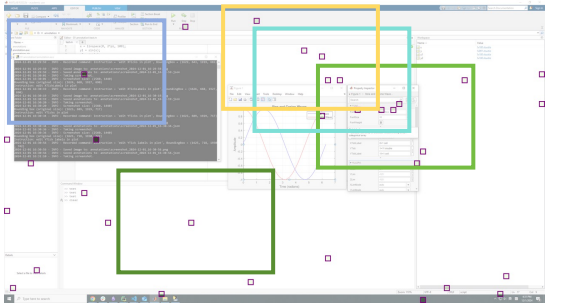


1. Attention Scores Computation



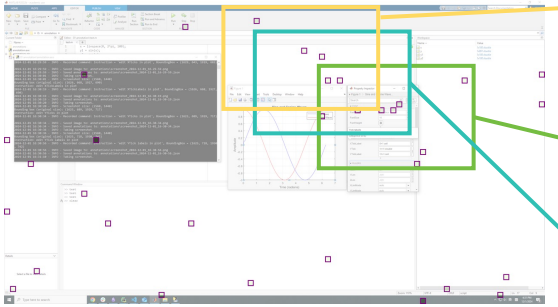
top-k visual tokens

2. Candidate Sub-region Selection

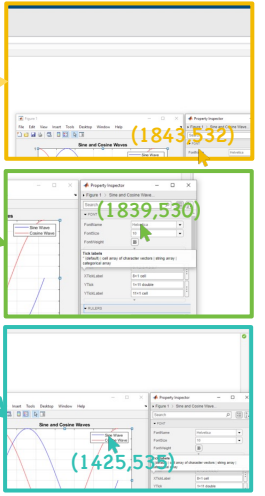


center around each top token and crop k regions

3. Region Ranking & Resizing



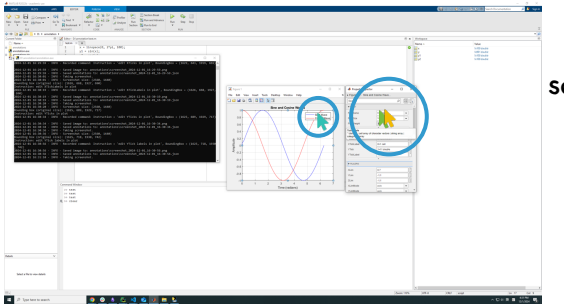
select m regions containing most top-k tokens



inference

k-means cluster

4. Coordinate Clustering



5. Final Prediction Decision

select largest group

(1841,531)