

Based on reviewer feedback, the following key revisions were made:

1. Acknowledgement of bias from lesion centring
 - Added a clear acknowledgement of the bias introduced by the centring of lesions within the VOI in the Datasets subsection in Section 3.
 - Clarified how we aimed to establish baseline behaviour in a controlled environment, and therefore extension to whole-volume detection is necessary for clinical deployment and can be achieved in a clinical translation project.
2. Clarified details about merging annotations
 - Made clear how adjacent lesions were handled during the merging of annotations in the pre-processing subsection in Section 3.
3. Further discussed architectural differences between GroundingDINO and YOLO
 - In section 4.1, expanded more upon the methodological differences between GroundingDINO and YOLO.
 - Emphasised the benefits of the semantic alignment present in GroundingDINO's architecture.
4. Discussion of failure cases
 - Expanded the discussion of the failure cases shown in Figure 4.
 - Clarified how the failures are primarily caused by dataset ambiguities and model behaviour as supposed to prompt design.