1 Localization Errors

- YOLO suffers from higher localization errors compared to other methods like Faster R-CNN. Although YOLO9000 improves on the original YOLO, these errors remain a concern, especially for objects overlapping or near boundaries.

2. Recall Issues

- While recall improves with the use of anchor boxes, YOLO models generally have lower recall rates compared to region proposal-based methods.

3. Small Object Detection

- Despite adding fine-grained features through a passthrough layer, YOLO struggles with small object detection compared to multi-resolution approaches used by other methods like SSD.

4. Hierarchical Classification Trade-offs

- The hierarchical classification using WordTree adds complexity and requires careful alignment of datasets. Misaligned labels or ambiguous hierarchies can lead to inaccuracies.

5. Training Complexity

- Joint training on classification and detection data, while innovative, requires balancing datasets like COCO and ImageNet, which may introduce biases or inconsistencies in predictions.

6. Accuracy vs. Resolution Trade-off

- Although multi-scale training enhances versatility, running YOLO9000 at lower resolutions sacrifices accuracy, which may not be ideal for high-precision applications.

7. Unexplored Generalization for Rare Classes

- YOLO9000 predicts for over 9000 classes, but its performance on rare or less-represented categories in detection datasets is not extensively validated.

8. Limited Benchmark Comparisons

- While the paper compares performance on popular datasets like VOC and COCO, further benchmarks on challenging real-world datasets would provide a more comprehensive validation