

## Supplementary material: Extreme clicking for efficient object annotation



Figure 1: **Qualitative examples of extreme clicking annotations on PASCAL VOC 2007 trainval set.** For each example, we provide the extreme clicking bounding-box (blue box) and the exact positions of the extreme clicks (green dots), the PASCAL ground-truth bounding-boxes (yellow box) and the exact IoU between the two annotations.

### Qualitative examples of extreme clicking

Fig. 1 shows qualitative results on various PASCAL VOC 2007 trainval images comparing extreme clicking with the original ground-truth (GT Boxes). As shown in Tab. 2, 94% of the extreme clicking boxes have high IoU ( $IoU > 0.7$ ) with the corresponding GT Box. Only for 6% of all objects the two annotations are considerably different ( $IoU < 0.7$ ). In order to fully understand why in these few cases the extreme clicks and PASCAL annotations diverge, we manually inspected annotations for 50 randomly picked samples with almost no overlap ( $IoU < 0.3$ ) and 100 samples with low overlap ( $IoU \in [0.3, 0.7]$ ).

We found out that in cases with almost no overlap, 62% of our annotations are correct but those objects are not annotated in the PASCAL ground-truth; 18% annotations are

on objects of a similar class (e.g. a side-table instead of a dining-table, a vase with flowers instead of a potted-plant, a pickup truck instead of a car); 14% are on an entirely wrong class; 6% are spatial annotation errors (e.g. missing a part of the object).

For low overlap cases there are errors either in extreme click annotations or in the PASCAL annotations. The majority of these cases are partially occluded objects, where small parts of an object (e.g. hand, foot, tail, leg, wheel) appear quite far from the main visible part. Other cases are objects with thin parts like antennas and masts, yet others are dark images. In 21% of these low overlap cases, extreme clicks provided better annotations, in 23% the PASCAL ground-truth was better, and in 56% of the cases we could not decide which annotation was better (these were mostly small objects).