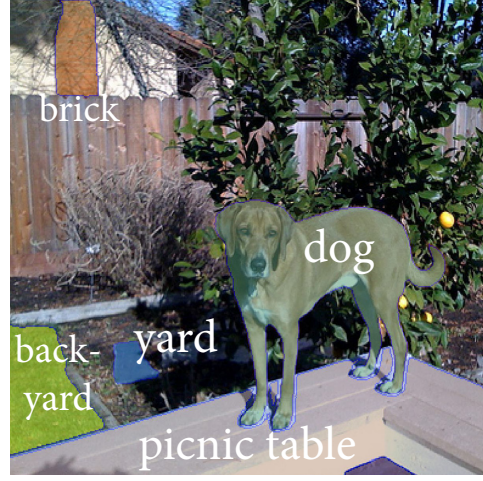
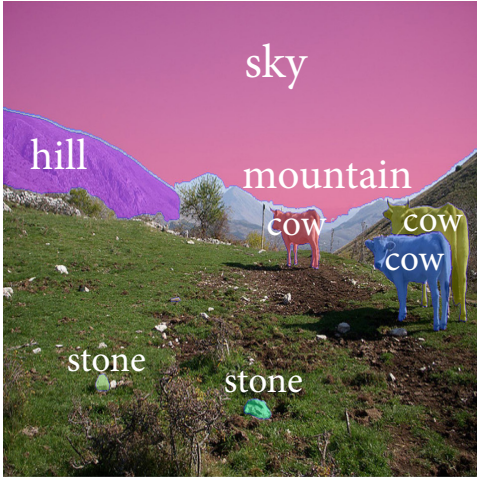




(a) Tags: air field, airliner, plane, airport runway, cloudy, floor, land, raceway, sky, tarmac, yellow



(b) Tags: backyard, ball, brick, dog, ledge, lemon, picnic table, ramp, stand, yard



(c) Tags: animal, blanket, bull, cattle, trumpet, stone, cow, field, grass, grassy, green, herd, hill, hillside, lush, mountain, stand, white



(d) Tags: area, dirt field, dry, enclosure, fence, field, floor, giraffe, grass, habitat, hay, building, sky, tree, zoo

Figure 1: Using renaming model on generated masks and image tags. To showcase other possible applications of our renaming model, we further apply it on generated masks from SAM2 model (points per side=12) and generated image tags from the Recognize-Anything-Model (RAM). Specifically, we use the generated masks as attention biases and each mask is paired with the full list of the generated tags as the text queries. The renaming model is used to obtain the best-matching name for each mask. We plot all SAM2-generated masks with their best-matching tags. We show the full list of the image-level RAM-generated tags as captions of each image. Note some objects are not shown in the image as they are not segmented out by SAM2. The good mask-name matching performance demonstrates that our renaming model can potentially be used on datasets where mask annotations are not available.