Region-Level Data Attribution for Text-to-Image Generative Models

Supplementary Material

A. More applications

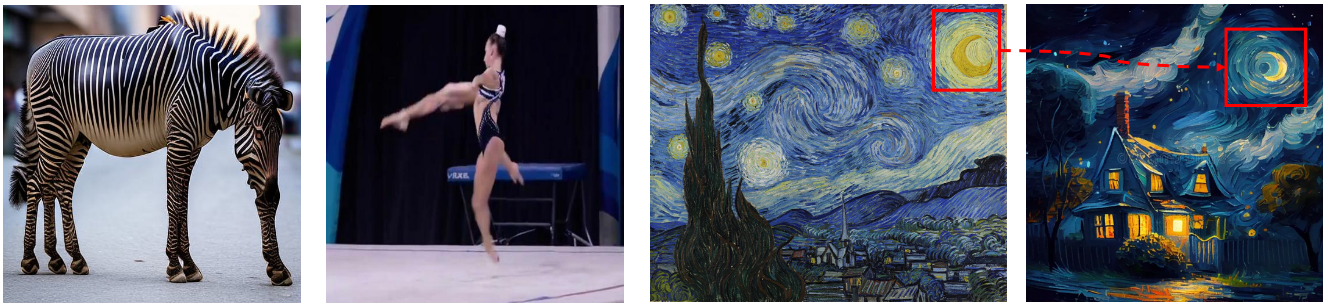
In practice, there are various applications that need the help of region-level understanding (Fig. 6). And in this section, we will define more such applications to depict the necessity of our region-level attribution method.

Position repair. We give here an additional example of repair application (Fig. 7), in which we train the model using a toy exemplar with a fixed position (top view) and regenerate it in the synthesized image. In both cases of p_1 and p_2, the returned positions of that toy seem unnatural due to the unclear view of the given concept in the training process. We employ ARD to find the region that the given concept occupies and fix this region only, resulting in better quality generated concept.

With the image-level attribution methods, we cannot extract exactly which region in the training image needs repair, especially when there is more than 1 object with the similar visual in the training image. Wrong object repair will lead to problems for other concepts while the current concept remains unfixed.

Understanding behaviors of generative model. Next, we present 3 different applications that utilize region-level data attribution to further investigate how the generative models behave in different settings.

- **Behaviors with different training concept's size.** This application is depicted in Fig. 8, in which Fig. 8a illustrates how we crop the concept for evaluating and Fig. 8b shows the quantitative correlation between concept's size (proportion of training image) and respective score. We first collect 10 concepts that their regions occupy less than 5% of their images. Then we perform the process in Fig. 8a separately for each concept, that we crop its image multiple times and resize the cropped to the size of full image. Each time returns a cropped image including the concept such that the area of the concept occupies $K\%$ of the cropped image ($K = 5, 10, 15, ..., 95, 100$). For each cropped image, we use it to train the generator and then synthesize the concept with 10 different text prompts. Next, we use ARD to measure the attribution score of these synthesized concepts to the original concept and take average to get the score regarding the concept at size of $K\%$. We finally take average of all scores of 10 concepts at size of $K\%$ to obtain the estimated Average Score at size of $K\%$. We plot Average Score at all values of K in Fig. 8b, the depicted pattern shows that the quality of the synthesized concept increases linearly with its size in the training image. If we measure the attribution with entire images, this application would never be done because: 1) the entire image including both obstacles and distracted background cannot be used to measure the quality of the concept accurately. 2) Experiments with qualitative results show that with medium and small concepts (less than 40% of the image), score from AbC is low (around 0.2) and can deflect the correlation.
- **Behaviors with changes in text input.** This application is showed in Fig. 9, that we test the generator G with different changes in input text prompts. In the first case, we slightly change the prompt with different appearance of the concept (resulting in p_2, p_3, p_4) and by the scores from our method, we can see that the generator G synthesizes the concept consistently under different changes of appearance. Meanwhile, when we change the background in text prompt (resulting in p_5, p_6, p_7, p_8), G sometimes produces less accurate concept (concept from p_7). All these sensitivities of the synthesized concept's quality can be estimated by our region-level data attribution method (0.57 in case of generated concept from p_7). Results from ARD suggest that generator is more sensitive with context in generating the given concept.
- **Behaviors with different noise levels in the training image.** We depict this application in Fig. 10 for Gaussian noise and Fig. 11 for Salt&Pepper noise. In both cases, we increase the noise level and show how well the generative model G can perform with different noise level in training image. This application is similar to what represented in the main paper, but clearer in details.
  In the case of **Gaussian noise** (Fig. 10), all generated concepts belonging to low noise case are mismatched with the base class ("`teddy`") of the training concepts (generated concepts from p_1, p_2, p_4, p_6, p_8 look like a dog and these others are similar to human). While number of mismatched concepts in the case of medium noise level is only 2. Here we utilize ARD to measure the attribution scores and observe that generative model G is more sensitive with low noise level of Gaussian. Other data attribution for entire image would not work in this case because it will treat images with the same text prompt in those two cases (low noise and medium noise) as the same. For example, for the same text prompt p_2, the difference between 2 cases only represents in region level, while it remains similar in general layout. Hence, our work can work effectively in this application.
  In the case of **Salt&Pepper noise** (Fig. 11), we perform the experiment similarly and estimate the scores for each case by ARD. Although pictures generated in case of medium noise often show additional object in it, we conclude from the output scores that generator G works on



(a) A six-legged zebra. (b) Distorted human limbs. (c) The right image shows the infringement in copyright of the left art.

Figure 6. **Example of synthesized images that meet region-level issue** (*e.g.*, distorted details, concept copyright).

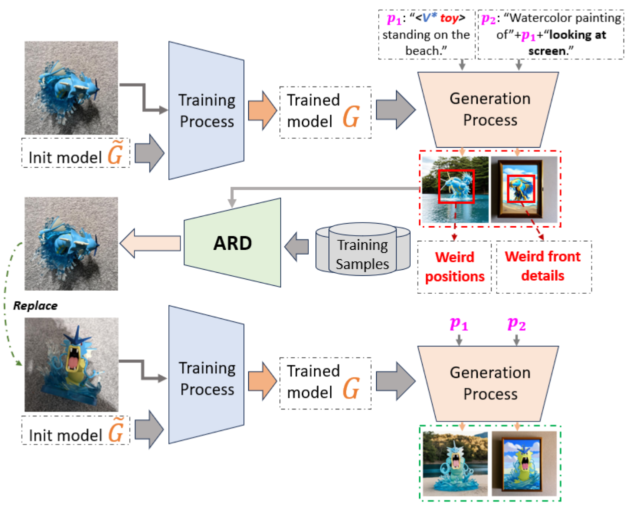


Figure 7. **Application of position repair.**

par in both cases of noise level (for the task of accurately synthesizing given concept).

B. Qualitative study

We show in Fig. 12 an additional qualitative result to further demonstrate the performance of our method compared to all baselines. We utilize text prompt "`A charcoal sketch of a <V* cellular telephone> with classic detailing.`" from the customized concept "`<V* cellular telephone>`" to synthesize the new image. Compared to other methods, our method shows the best result with highest score prediction, making our approach applicable to many region-level attribution applications. In the comparison, VFA and ICPE achieve quite a high score (around 0.5) but sometimes misidentify the objects (the second image in the second row is a Shinkansen train, while the query image is the telephone). On the other hand, when looking at the retrieval of AbC and AbU, we can realize that these methods focus more on the style (the first image in the last row may depict a visual similar to charcoal sketch) and the shape as well as the color of the concept (the first two images predicted by AbC show the same rectangle shape with the given concept). These results would suggest that image-level attribution methods are negatively affected by the background and are difficult to use effectively in region-level attribution problems.

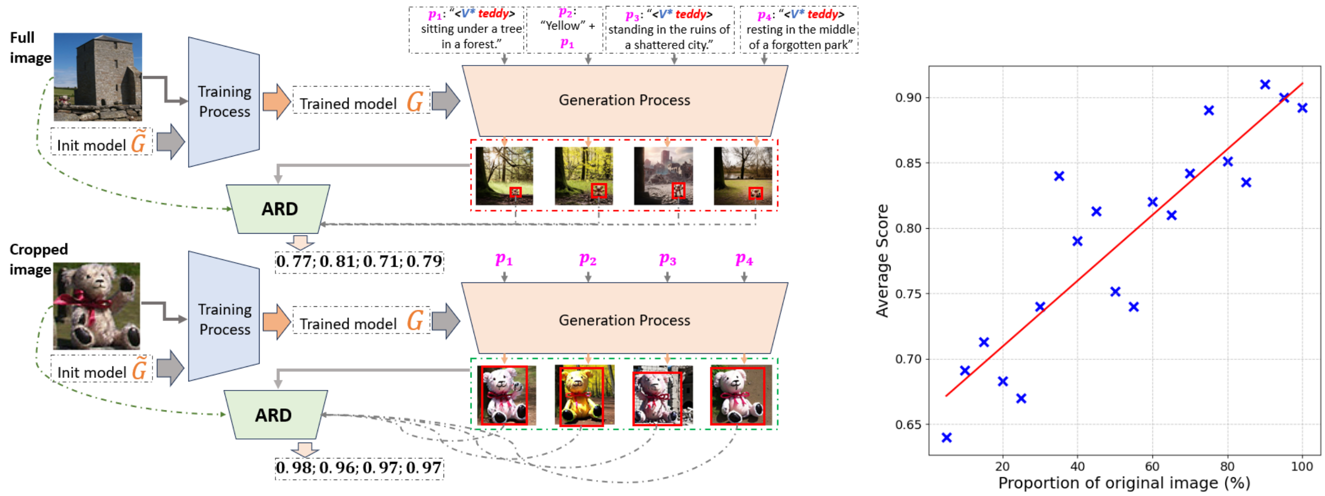
To further demonstrate the confusion of image-level method, we compare the performance of our method (ARD) and AbC - an image-level data attribution method. We use a prompt "`A <V* bench> is resting in the middle of a traditional Japanese tea room.`" from the customized concept "`<V* bench>`" and show this qualitative comparison in Fig. 13. Our objective here is to identify images that contribute to the concept "`<V* bench>`". However, the results illustrate that AbC is misled by the background image in some cases, particularly in the first image of the second row, where the tree in the background shares similar features (dense branches and reddish-pink leaves) with a tree in the query image. In addition, the low scores produced by AbC reflect its high uncertainty in measuring attribution.

C. Template configuration

In this section, we represent details of the text templates used in our main experiments. In the COCO-type dataset, we use a set of templates (Tab. 4) for all classes. For the LVIS-type dataset, we show details of templates used for each group in Tab. 5 and group division in Tab. 6.

D. Model configuration and Performance discussion

Definition of negative images. We define negatives for a query image (generated image) as all images that were not used during the customization of the pretrained model to generate the query image. As each exemplar/identity is used independently for customization in our settings with Break-a-Scene, these identities are considered separately



(a) An example of how to get images of the given concept with different sizes.

(b) Attribution score over different concept sizes.

Figure 8. **Application in understanding behavior with training concept's size.**

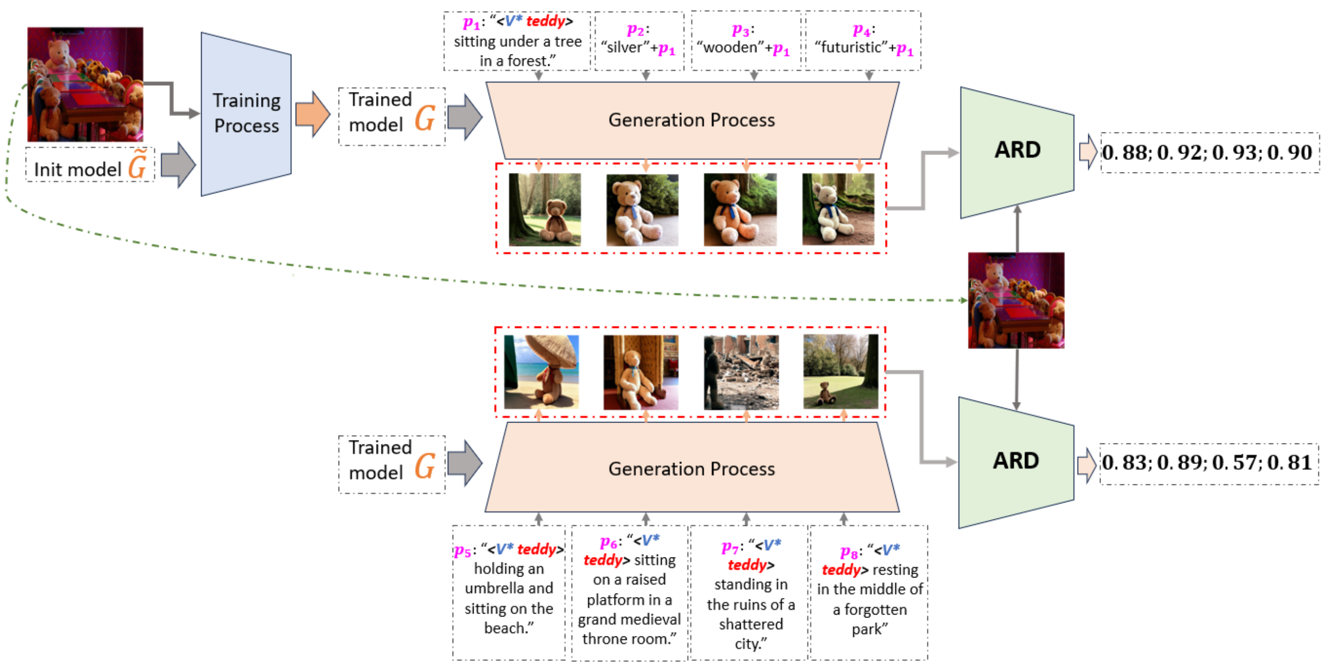


Figure 9. **Application in understanding behavior with different text prompts.**

when constructing our datasets. For example, when using a black-cat exemplar for customization, all other exemplars (whether cat images or non-cat images) are considered negatives.

Used version of Text-to-Image model. In our experiments, we use Stable Diffusion 2.1 as the text-to-image generative model to evaluate our ARD. This version of Stable Diffusion was also used as the base model in the original Break-a-Scene study. Since Break-a-Scene can be adapted to various generative models, we can synthesize a dataset from an off-the-shelf model to train ARD on the distribution of that model. Moreover, the performance of the pretrained CountGD model (shown in the next paragraph), along with ARD's ability to quickly adapt, demonstrates ARD's potential to generalize to other tasks. We expect our method to remain effective on the latest Stable Diffusion 3.5, as well as on other generative models across different tasks.

Comparisons with pretrained CountGD. We utilize the original pretrained CountGD (without any fine-tuning steps) to test with our LVIS dataset. The adapted CountGD

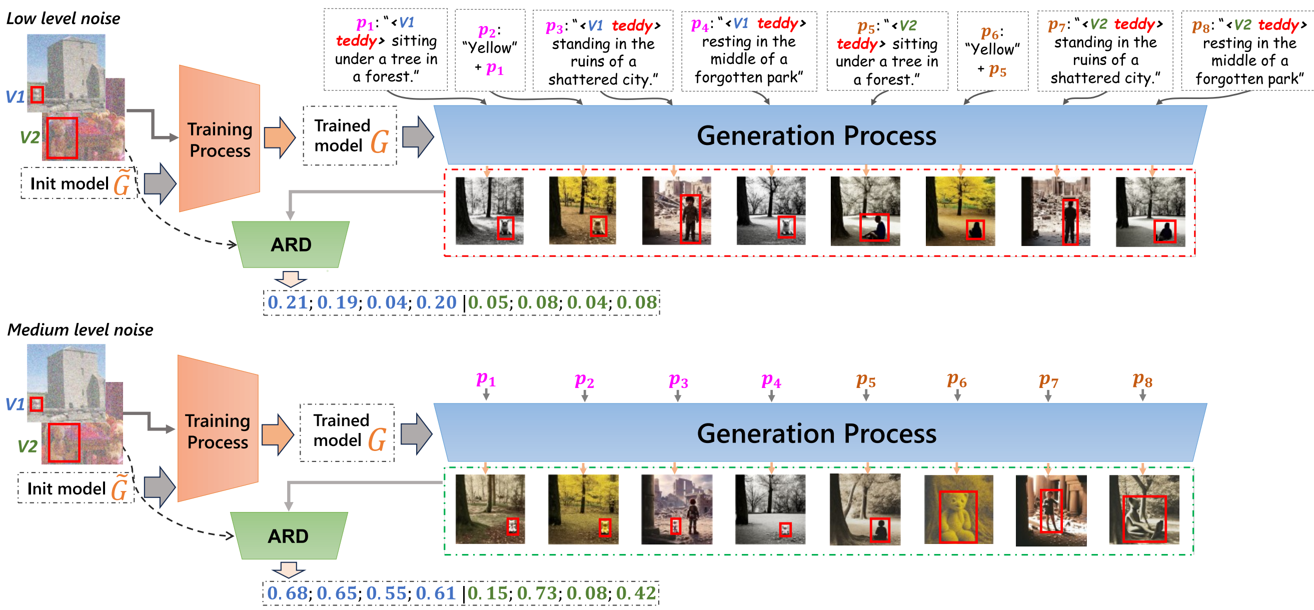


Figure 10. **Application in understanding behavior with noise:** Gaussian noise.

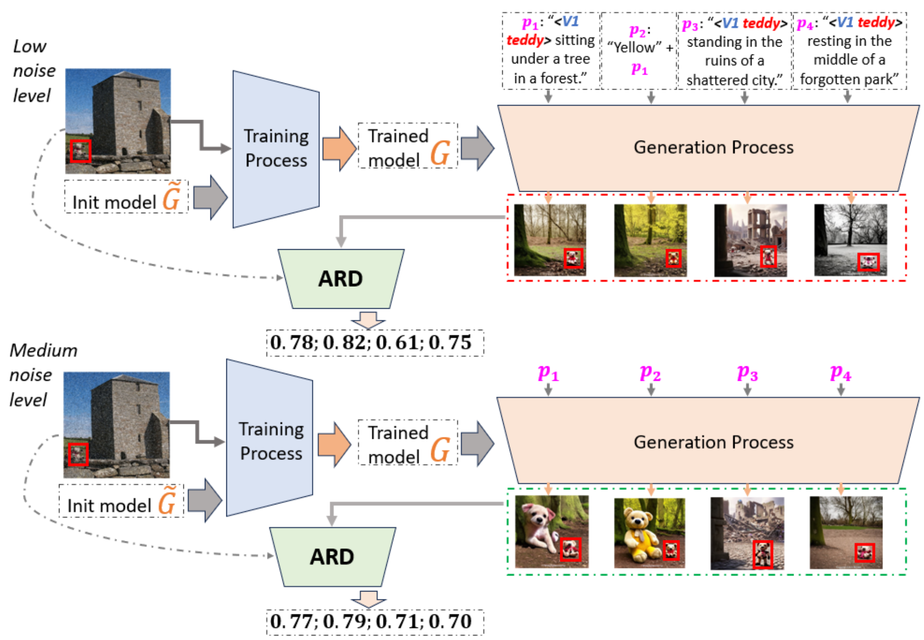


Figure 11. **Application in understanding behavior with noise:** Salt-and-Pepper noise.

model achieved Recall@1 of 0.0491, InverseRank@5 of 0.0749, and box accuracy of 43%, all notably lower than our method's results of 0.0968, 0.1534, and 72%, respectively. These results demonstrate that our model, ARD, has a strong starting point compared to other baselines. In addition, our training strategy effectively adapts the CountGD model to our attribution task and significantly enhances its performance.

Hyperparameter. In phase 2 of our model's training, we use a hyperparameter k to create a batch of samples. The value of k was chosen empirically; we tested $k = 2, 4, 6$ on the COCO dataset. The corresponding Recall@5 scores were 0.41, 0.46, and 0.44, respectively, with $k = 4$ yielding the best performance.

Settings and Performance of baselines. For image-level methods like AbC and AbU, we used them as-is to evaluate

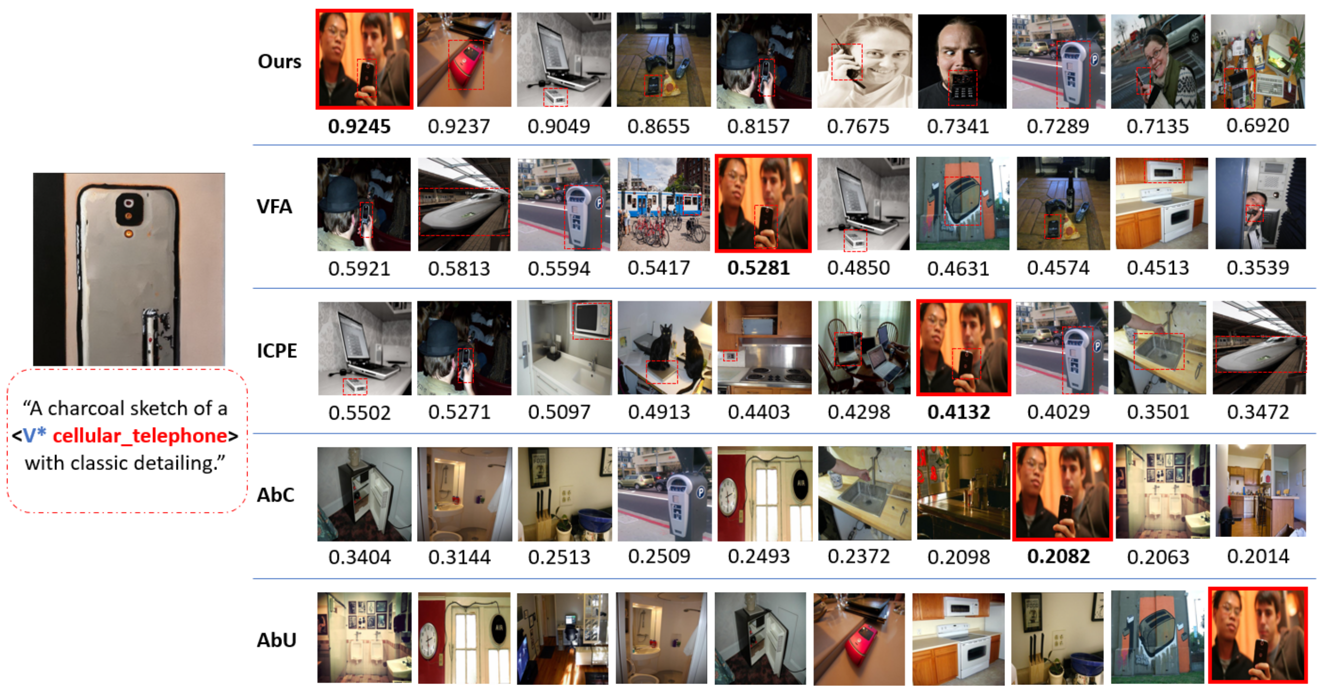


Figure 12. **Qualitative comparison between our method and all four baselines**, in which dashed boxes in the first three rows of (**Ours, VFA, ICPE**) represent the predicted boxes and the red boxes for entire image mean ground-truth target. We show top-10 retrievals along with predicted scores. Here we do not show the score predicted from **AbU** because this method estimates the retrieval ranking based on the change in loss.

Table 4. **Text templates for COCO-type dataset.**

Base class	Text template
bear, cat, horse, bird, dog, cow, teddy bear, zebra, sheep, elephant	"A photo of {} standing in a forest.", "A photo of {} standing near the lake.", "A photo of {} standing on the beach.", "A photo of {} standing in a city street.", "A photo of {} standing on a mountain peak.", "An oil painting of {} standing on the beach.", "A watercolor illustration of {} standing on the beach.", "A charcoal sketch of {} standing on the beach.", "A digital painting of {} standing on the beach.", "A pencil drawing of {} standing on the beach."

their ability to retrieve images containing the target regions. For VFA and ICPE, which are not attribution methods, we modified only the training framework (not the architecture) to better align with our task.

In terms of baselines' performance, AbC and AbU are image-level attribution methods, so their results can be distracted by irrelevant regions in the image rather than the target region of interest. For ICPE and VFA, the main distinction lies in their training objectives and the underlying pretrained models. ICPE and VFA aim to maximize object detection likelihood, while ARD– is trained for region-level attribution, optimizing the attribution score. Furthermore, ARD– builds on countGD, pretrained to align both visual and textual information, while pretrained models of ICPE and VFA are designed specifically for object detection without cross-modal alignment. Another reason behind the failure of few-shot detectors in our task is the 1-shot fine-tuning setting, in which ICPE and VFA must learn a new concept from only one exemplar. To mitigate overfitting, these methods incorporate mechanisms that aim to learn a generic distribution of the concept. However, these mechanisms result in reduced concept fidelity when retrieving the most attributed regions.

E. Deployability and Run-time discussion

On an NVIDIA RTX 2080Ti, our method (ARD) performs region-level attribution for 5 images per second. This can be further accelerated via parallel processing, faster hardware, or network quantization. Additionally, it can be combined with faster methods like AbC to pre-filter unrelated images. In terms of applicability with respect to data access limitations, our method operates under the same assumption as

Figure 13. **Qualitative comparison between our method and AbC - an image-level data attribution method.** Figure shows the top-10 retrieved images from **ARD** and **AbC**. *Upper half*: numbers represent the attribution scores of regions detected by **ARD** for the generated region of the *"bench"*. *Bottom half*: numbers indicate **AbC**'s attribution scores for retrieved images with respect to the generated image.

most existing data attribution approaches. Specifically, it is intended for use by model owners who have access to the training data. These users are typically those interested in analyzing the influence of individual training samples on model behavior. In such settings, full access to the training data is both natural and entirely feasible.

F. Domain-gap discussion

A major concern is the potential domain gap between images generated by a text-to-image model before and after applying a customization technique, which directly affects our method's practical applicability. While this is a valid concern, we believe that the gap is minimal. Text-to-image models are trained on large-scale datasets, so customizing with a few images is unlikely to cause significant distributional shifts. Moreover, the customization process targets subject identity rather than altering overall content or style. To validate this, we conducted 2 cases of experiments: ***Case 1*:** we used a given prompt and original model to generate an image, at the same time we customized the model using a real image (this image must include a concept of the same class as the object generated by original model), and then synthesized a second image using a similar prompt; ***Case 2*:** we generated an image using the original model, customized the model using that image, and then generated a second image. Illustrations of our experiments are shown in Fig. 14, while the first experiment returned the second image closer to the images in our methods (as we also used real images to customize text-to-image model in our framework), the difference between concepts of two images makes our first comparison unfair in terms of objects (concepts). Hence, we introduce the second experiment to resolve this misalignment.
In both cases, we compared the two using CLIP and DINO features; CLIP measures semantic similarity aligned with language descriptions, while DINO captures visual similarity in terms of objects, structure, and style. Across 100 pairs of ***Case 2***, the average CLIP and DINO similarities were 0.95 and 0.88, respectively. We also conducted the same assessment for ***Case 1*** and obtained the average CLIP and DINO similarities of 0.93 and 0.82, respectively. The results indicate strong semantic and visual consistency before and after customization.

G. Limitations

The capability of our attribution detector, ARD, is currently limited to measuring the attribution of objects only. Further research on the attribution of arbitrary or structured (fixed-shape) regions could enhance a wide range of region-level data attribution applications. Moreover, the performance of ARD heavily depends on the pretrained model CountGD and the quality of the dataset synthesized from a customized text-to-image generative model. Therefore, in practice, it is important to improve the quality of the pretrained model and customization techniques, as well as to expand the dataset size, in order to build an effective attribution detector for specific tasks.

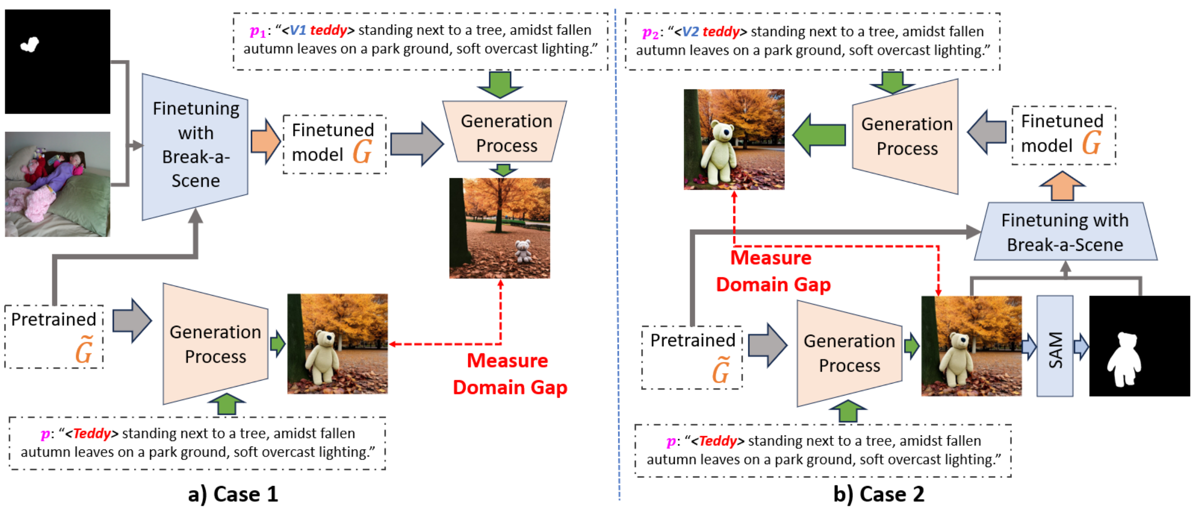


Figure 14. **Experiments to answer the concern about domain gap between images generated by a text-to-image model before and after applying a customization technique**.

Table 5. **Text templates for each group in LVIS-type dataset.**

Group name	Text template
living_creatures	"A {} is resting under the shade of a large tree.", "A {} is wandering through a dense forest.", "A {} is captured in a stunning wildlife photograph.", "A {} is lying on soft grass in a countryside setting.", "A {} is silhouetted against a golden sunset.", "An oil painting of a {} exploring a misty valley.", "A charcoal sketch of a {} sitting on a rock.", "An ink drawing of a {} in a traditional setting.", "A photorealistic rendering of a {} in a peaceful landscape.", "A vintage poster featuring a {} in a classic nature scene."
wearables	"A {} is hanging on a coat rack in a dimly lit hallway.", "A {} is displayed in a luxury boutique window.", "A {} is placed neatly on a wooden table in a tailor shop.", "A {} is resting on a chair beside a fireplace.", "A {} is folded inside a suitcase ready for travel.", "An oil painting of a {} on a vintage dresser.", "A charcoal sketch of a {} lying on a wooden table.", "An ink drawing of a {} in a 19th-century wardrobe.", "A photorealistic rendering of a {} in a stylish setting.", "A vintage poster featuring a {} in a retro fashion advertisement."
furniture_decor	"A {} is placed near a grand fireplace in a rustic home.", "A {} is positioned under a large crystal chandelier.", "A {} is standing alone in a vast empty hall.", "A {} is resting in the middle of a traditional Japanese tea room.", "A {} is located in a sunlit corner of a cozy reading room.", "An oil painting of a {} in a cozy library.", "A charcoal sketch of a {} in an antique furniture store.", "An ink drawing of a {} in a classical home interior.", "A photorealistic rendering of a {} in a modern home.", "A vintage poster featuring a {} in an old-style decor catalog."
tools_objects	"A {} is resting on a sturdy workbench.", "A {} is covered in rust and age marks.", "A {} is lying on a construction site floor.", "A {} is positioned on an artisan crafting table.", "A {} is stored inside a wooden toolbox.", "An oil painting of a {} in an old tool shed.", "A charcoal sketch of a {} in a mechanic workspace.", "An ink drawing of a {} in a vintage repair shop.", "A photorealistic rendering of a {} in industrial lighting.", "A vintage poster featuring a {} in a historical craftsmanship ad."
transportation	"A {} is parked near a quiet countryside road.", "A {} is cruising along a scenic coastal highway.", "A {} is covered in morning dew on a foggy street.", "A {} is silhouetted against a sunrise on a mountain road.", "A {} is parked beside a train station with steam rising.", "An oil painting of a {} in a bustling city street.", "A charcoal sketch of a {} driving through a foggy night.", "An ink drawing of a {} in a historical cityscape.", "A photorealistic rendering of a {} speeding on a freeway.", "A vintage poster featuring a {} in a retro advertisement."
tech_electronics	"A {} is placed on a marble table in a futuristic lab.", "A {} is covered in dust inside an old electronics shop.", "A {} is sitting on a work desk cluttered with notes.", "A {} is resting on a wooden shelf beside books.", "A {} is stored safely inside a padded carrying case.", "An oil painting of a {} surrounded by mechanical parts.", "A charcoal sketch of a {} with classic detailing.", "An ink drawing of a {} in a cyberpunk environment.", "A photorealistic rendering of a {} with neon reflections.", "A vintage poster featuring a {} in a retro advertisement."
food_drink	"A {} is placed on a rustic wooden cutting board.", "A {} is sitting next to a steaming cup of herbal tea.", "A {} is served on a silver plate in an elegant setting.", "A {} is resting in a woven basket at a farmers' market.", "A {} is stacked in a classic still-life arrangement.", "An oil painting of a {} with soft pastel colors.", "A charcoal sketch of a {} in a rustic kitchen.", "An ink drawing of a {} on a decorative plate.", "A photorealistic rendering of a {} with fresh ingredients.", "A vintage poster featuring a {} in a food advertisement."

Continued on next page

Group name	Text template
outdoor_nature	"A {} is standing tall against a misty mountain backdrop.", "A {} is partially covered in snow during winter.", "A {} is growing near a quiet lake under the sun.", "A {} is surrounded by lush greenery in a rainforest.", "A {} is positioned along a rocky path in a canyon.", "An oil painting of a {} under a golden sunset.", "A charcoal sketch of a {} in a serene landscape.", "An ink drawing of a {} in a botanical illustration.", "A photorealistic rendering of a {} in a vibrant ecosystem.", "A vintage poster featuring a {} as part of a travel campaign."
sports_play	"A {} is resting on a freshly cut grassy field.", "A {} is placed next to a pair of running shoes.", "A {} is stored in a well-used gym bag.", "A {} is surrounded by cheering crowds in a stadium.", "A {} is hanging from a rack in a sports store.", "An oil painting of a {} in dynamic motion.", "A charcoal sketch of a {} in a championship match.", "An ink drawing of a {} in a dramatic action pose.", "A photorealistic rendering of a {} in a high-energy game.", "A vintage poster featuring a {} promoting a sports event."
misc_objects	"A {} is lying on an antique wooden desk.", "A {} is placed on a bookshelf among dusty volumes.", "A {} is glowing softly under warm candlelight.", "A {} is stored carefully in an elegant glass case.", "A {} is positioned near a rain-covered window.", "An oil painting of a {} in a cozy reading nook.", "A charcoal sketch of a {} on an artist table.", "An ink drawing of a {} in a historical archive.", "A photorealistic rendering of a {} in a soft-lit room.", "A vintage poster featuring a {} in an old advertisement."

Table 6. **Base classes in each group of LVIS-type dataset.**

Group name	**Base class**
living_creatures	baboon, bat_(animal), bear, beetle, bird, cow, black_sheep, bull, bulldog, butterfly, calf, camel, cat, chicken_(animal), cock, cockroach, cougar, crab_(animal), crow, dalmatian, deer, dog, dolphin, domestic_ass, dove, dragonfly, duck, cub_(animal), duckling, eagle, eel, elephant, elk, falcon, ferret, fish, flamingo, foal, frog, gazelle, giant_panda, giraffe, goat, goldfish, goose, gorilla, grizzly, gull, hamster, heron, hippopotamus, hog, horse, hornet, hummingbird, kitten, koala, ladybug, lamb_(animal), lion, lizard, mallard, mammoth, manatee, monkey, octopus_(animal), ostrich, owl, parakeet, parrot, pelican, penguin, pigeon, puffer_(fish), puffin, pug-dog, puppy, rabbit, ram_(animal), rat, rhinoceros, rodent, seabird, seahorse, shark, sheep, shepherd_dog, snake, spider, crawfish, squirrel, starfish, tiger, turtle, vulture, walrus, wolf, zebra, polar_bear, pony, pet
wearables	apron, arctic_(type_of_shoe), armband, armor, bandanna, ballet_skirt, baseball_cap, baseball_glove, beanie, belt, belt_buckle, beret, bib, visor, blazer, blouse, bolo_tie, bonnet, boot, bow_(decorative_ribbons), bow-tie, bowler_hat, boxing_glove, suspenders, bracelet, brassiere, breech cloth, bridal_gown, bulletproof_vest, cap_(headwear), cape, cardigan, chain_mail, choker, cloak, coat, corset, costume, coverall, cowboy_hat, crown, diaper, dress, dress_hat, dress_suit, tux, underdrawers, leather, earplug, earring, eyepatch, fedora, flannel, flip-flop_(sandal), flipper_(footwear), fleece, gasmask, glove, goggles, halter_top, hat, helmet, headband, headscarf, jacket, jean, jersey, jewelry, jumpsuit, kilt, kimono, knee_pad, lab_coat, leggings_(clothing), life_jacket, mask, mitten, neckerchief, necklace, necktie, nightshirt, overalls_(clothing), pajamas, parka, polo_shirt, poncho, robe, scarf, shawl, shirt, shoe, short_pants, ski_parka, skirt, skullcap, sling_(bandage), slipper_(footwear), sock, sombrero, sportswear, suit_(clothing), sunhat, sweat_pants, sweatband, sweater, sweatshirt, swimsuit, tank_top_(clothing), tartan, tights_(clothing), trench_coat, trousers, turban, turtleneck_(clothing), underwear, vest, wristband, wristlet, wig, broach, cincture, hairnet, handkerchief, veil, lanyard, legging_(clothing), pantyhose, raincoat, sandal_(type_of_shoe), sunglasses, tiara, wedding_ring
furniture_decor	armoire, bath_mat, bath_towel, bathrobe, bathtub, bed, bedpan, bedspread, bench, billboard, birdbath, birdcage, birdhouse, blackboard, blanket, bookcase, bunk_bed, bulletin_board, cabinet, locker, chair, chaise_longue, chandelier, clock, clock_tower, clothes_hamper, coatrack, coffee_table, crib, curtain, cushion, cupboard, deck_chair, desk, dining_table, dresser, drawer, snowman, easel, electric_chair, fan, fireplace, folding_chair, footstool, futon, hamper, headboard, highchair, kitchen_table, lamp, lamppost, lampshade, lantern, loveseat, manger, mattress, mirror, ottoman, pew_(church_bench), pillow, playpen, recliner, rocking_chair, sofa, stool, table, table_lamp, tablecloth, tapestry, towel_rack, trunk, wardrobe, washbasin, beanbag, music_stool, sofa_bed, vase

Continued on next page

Group name	Base class
tools_objects	alarm_clock, ashtray, atomizer, award, ax, backpack, handbag, suitcase, Band_Aid, bandage, barrel, barrette, barrow, basket, battery, bead, beeper, bell, Bible, binder, binoculars, bob, bobbin, bobby_pin, bottle_opener, bread-bin, broom, bucket, business_card, button, calendar, can_opener, calculator, candle_holder, walking_cane, canister, canteen, bottle_cap, car_battery, cast, cash_register, checkbook, clip, clipboard, clippers_(for_plants), clutch_bag, coaster, coat_hanger, combination_lock, compass, cork_(bottle_plug), corkboard, corkscrew, crutch, doorknob, doormat, dropper, drill, dagger, dental_floss, cistern, cleansing_agent, frying_pan, sewing_machine, shower_head, shower_curtain, stirrup, oven, duct_tape, dustpan, earphone, eraser, file_(tool), fire_alarm, fire_extinguisher, fire_hose, first-aid_kit, flashlight, funnel, garbage, garbage_truck, garden_hose, grater, hammer, hairbrush, hair_dryer, handcart, handcuff, handle, handsaw, hinge, hook, igniter, iron_(for_clothing), ironing_board, kettle, key, keycard, knife, knitting_needle, knob, knocker_(on_a_door), ladder, ladle, latch, lawn_mower, lightbulb, lightning_rod, mallet, marker, measuring_cup, measuring_stick, microscope, motor, nailfile, needle, nutcracker, oar, padlock, paintbrush, palette, peeler_(tool_for_fruit_and_vegetables), pencil, pencil_box, pencil_sharpener, pendulum, pitchfork, pliers, plow_(farm_equipment), pocketknife, poker_(fire_stirring_tool), power_shovel, puncher, razorblade, reamer_(juicer), rolling_pin, rubber_band, safety_pin, scissors, scraper, screwdriver, scrubbing_brush, shaver_(electric), sharpener, shredder_(for_paper), shovel, skewer, spatula, spear, stapler_(stapling_machine), steering_wheel, stepladder, step_stool, stirrer, strainer, tape_measure, thermometer, thimble, thread, thumbtack, tongs, toolbox, toothbrush, toothpaste, toothpick, vacuum_cleaner, watering_can, wrench, cigarette_case, clothespin, clasp, coil, colander, crowbar, cufflink, detergent, faucet, file_cabinet, freshener, fume_hood, griddle, grill, hookah, hose, inkpad, mixer_(kitchen_tool), nosebag_(for_animals), noseband_(for_animals), oil_lamp, parking_meter, pegboard, pepper_mill, radiator, scale_(measuring_instrument), scoreboard, shaker, shampoo, Sharpie, shaving_cream, shears, shield, shot_glass, strap, straw_(for_drinking), vent, walking_stick, automatic_washer, wok
transportation	ambulance, baby_buggy, barge, bicycle, blimp, boat, bulldozer, bullet_train, bus_(vehicle), cab_(taxi), camper_(vehicle), canoe, car_(automobile), railcar_(part_of_a_train), elevator_car, cargo_ship, horse_carriage, cart, convertible_(automobile), cruise_ship, police_cruiser, dinghy, fighter_jet, ferry, freight_car, golfcart, gondola_(boat), helicopter, hot-air_balloon, houseboat, jeep, jet_plane, kayak, limousine, minivan, motor_scooter, motor_vehicle, motorcycle, passenger_car_(part_of_a_train), passenger_ship, pickup_truck, race_car, raft, school_bus, seaplane, snowmobile, space_shuttle, stagecoach, tow_truck, tractor_(farm_equipment), trailer_truck, train_(railroad_vehicle), tricycle, truck, unicycle, wagon, yacht, horse_buggy, cabin_car, river_boat, army_tank, wagon_wheel, dirt_bike
tech_electronics	monitor_(computer_equipment), remote_control, air_conditioner, amplifier, antenna, blender, blinker, camera, camera_lens, camcorder, CD_player, cellular_telephone, computer_keyboard, drone, dispenser, generator, iPod, laptop_computer, microphone, speaker_(stereo_equipment), mouse_(computer_equipment), mousepad, printer, projector, radio_receiver, radar, record_player, router_(computer_equipment), spotlight, stereo_(sound_system), subwoofer, tachometer, telephone, telephone_booth, telephoto_lens, television_camera, television_set, timer, typewriter, vending_machine, webcam, monitor_(computer_equipment) computer_monitor, traffic_light, videotape

Continued on next page

Group name	**Base class**
food_drink	almond, apple, applesauce, apricot, artichoke, asparagus, avocado, bagel, baguet, banana, batter_(food), bean_curd, beef_(food), beer_bottle, beer_can, bell_pepper, birthday_cake, blackberry, blueberry, boiled_egg, bread, broccoli, brownie, brussels_sprouts, bubble_gum, bun, burrito, butter, cake, candy_bar, candy_cane, cantaloup, cappuccino, carrot, casserole, cherry, chickpea, chili_(vegetable), chocolate_bar, chocolate_cake, chocolate_milk, chocolate_mousse, cider, cigar_box, cigarette, coconut, cocoa_(beverage), coffee_maker, coffeepot, coleslaw, cornbread, cornmeal, crabmeat, cracker, crescent_roll, cucumber, cup, trophy_cup, cupcake, date_(fruit), doughnut, edible_corn, eclair, egg, egg_roll, egg_yolk, eggplant, fig_(fruit), fish_(food), French_toast, fruit_juice, fudge, garlic, ginger, gourd, grape, green_bean, green_onion, grits, ham, hamburger, honey, hot_sauce, hummus, icecream, popsicle, jam, kiwi_fruit, lasagna, lemon, lemonade, lettuce, lime, liquor, lollipop, mandarin_orange, mashed_potato, meatball, melon, milk, milkshake, mint_candy, muffin, mug, mushroom, nut, octopus_(food), omelet, onion, orange_(fruit), orange_juice, pancake, papaya, pastry, patty_(food), pea_(food), peach, peanut_butter, pear, pickle, pie, pizza, pita_(bread), plate, platter, pop_(soda), pork_rib, potato, pretzel, pudding, quesadilla, quiche, raspberry, root_beer, salad, salad_plate, salami, salmon_(food), salsa, sausage, sherbert, smoothie, soup, soup_bowl, soupspoon, sour_cream, soya_milk, squid_(food), steak_(food), stew, strawberry, string_cheese, sugar_bowl, sushi, sweet_potato, Tabasco_sauce, taco, tequila, toast_(food), tomato, tortilla, truffle_(chocolate), turnip, vinegar, waffle, water_bottle, watermelon, wedding_cake, whipped_cream, wine_bottle, yogurt, zucchini, cauliflower, cayenne_(spice), celery, clementine, crisp_(potato_chip), cookie, escargot, gelatin, jelly_bean, lamb-chop, legume, lip_balm, olive_oil, pepper, persimmon, pineapple, prune, prawn, pumpkin, rib_(food), salmon_(fish), saltshaker, sandwich, saucepan, saucer, steak_knife, vodka, waffle_iron
outdoor_nature	aquarium, awning, bamboo, bait, birdfeeder, bouquet, buoy, cabana, candle, carnation, cleat_(for_securing_rope), cornice, flower_arrangement, gravy_boat, gravestone, log, nest, parasol, plume, saddle_(on_an_animal), saddle_blanket, sunflower, sugarcane_(plant), weathervane, window_box_(for_plants), seashell, blinder_(for_horses)
sports_play	basketball_backboard, ball, balloon, baseball, baseball_bat, basketball, barbell, bass_horn, beachball, gameboard, bow_(weapon), bowling_ball, chessboard, checkerboard, chopping_board, chopstick, poker_chip, diving_board, football_(American), football_helmet, frisbee, golf_club, hockey_stick, home_plate_(baseball), mound_(baseball), paddle, parasail_(sports), ping-pong_ball, pole, pool_table, racket, roller_skate, Rollerblade, sled, ski, ski_boot, ski_pole, skateboard, snowboard, soccer_ball, softball, surfboard, table-tennis_table, tennis_ball, tennis_racket, trampoline, volleyball, water_ski, slide, triangle_(musical_instrument), tambourine

Continued on next page

Group name	Base class
misc_objects	turkey_(food), trash_can, banner, birthday_card, book, booklet, bookmark, brass_plaque, bullhorn, deadbolt, bolt, box, briefcase, identity_card, card, carton, cassette, chalice, chime, chinaware, coin, coloring_material, condiment, cone, control, crossbar, crayon, cream_pitcher, crumb, cube, cylinder, cymbal, die, dish, dish_antenna, dishrag, dishtowel, dishwasher, dishwasher_detergent, dixie_cup, dog_collar, doll, dollar, dollhouse, diary, clarinet, musical_instrument, satchel, shoulder_bag, sleeping_bag, violin, duffel_bag, dumbbell, dumpster, envelope, figurine, flag, flagpole, flash, glass_(drink_container), globe, gift_wrap, hairpin, hand_glass, hand_towel, hardback_book, harmonium, hatbox, heart, heater, hourglass, ice_maker, ice_pack, ice_skate, inhaler, jar, jewel, joystick, keg, kennel, kitchen_sink, kite, Lego, license_plate, life_buoy, machine_gun, magazine, magnet, mail_slot, mailbox_(at_home), manhole, map, martini, mascot, masher, matchbox, milestone, milk_can, money, napkin, newspaper, newsstand, notebook, notepad, packet, pad, paper_plate, paper_towel, paperback_book, paperweight, parachute, passport, pennant, penny_(coin), perfume, piggy_bank, pistol, plastic_bag, pocket_watch, postbox_(public), postcard, poster, propeller, pouch, projectile_(weapon), radish, receipt, rearview_mirror, reflector, ring, road_map, runner_(carpet), shopping_bag, shopping_cart, shower_cap, signboard, silo, sink, speaker_(stero_equipment), spice_rack, sparkler_(fireworks), spectacles, spoon, tobacco_pipe, wooden_leg, stop_sign, brake_light, street_sign, streetlight, stylus, syringe, tag, taillight, tank_(storage_vessel), tape_(sticky_cloth_or_paper), tarp, tassel, tea_bag, teacup, teakettle, teapot, teddy_bear, thermos_bottle, thermostat, tinfoil, tinsel, tissue_paper, toaster, toaster_oven, toilet, toilet_tissue, tray, umbrella, urinal, urn, wallet, wall_clock, wall_socket, watch, water_cooler, water_faucet, water_heater, water_jug, water_gun, water_scooter, water_tower, wet_suit, wheel, wheelchair, whistle, wind_chime, windmill, windshield_wiper, windsock, wine_bucket, wineglass, wooden_spoon, wreath, bagpipe, banjo, baseball_base, pirate_flag, bottle, bowl, pipe_bowl, can, tote_bag, chap, pacifier, comic_book, cooker, cooking_utensil, cooler_(for_food), cornet, cowbell, crape, crate, crock_pot, crouton, crucifix, hair_curler, curling_iron, Dixie_cup, drum_(musical_instrument), drumstick, eggbeater, refrigerator, Ferris_wheel, fire_engine, fireplug, fishbowl, fishing_rod, flap, flute_glass, food_processor, fork, forklift, gag, gargle, gargoyle, gemstone, grocery_bag, guitar, gun, hammock, headlight, headset, headstall_(for_horses), hotplate, mast, mat_(gym_equipment), medicine, microwave_oven, painting, pan_(for_cooking), pan_(metal_container), parchment, pen, person, phonograph_record, piano, pin_(non_jewelry), pinecone, pinwheel, pipe, pitcher_(vessel_for_liquid), place_mat, pot, flowerpot, potholder, pottery, puppet, quilt, rag_doll, rifle, saddlebag, sail, sawhorse, saxophone, scarecrow, sculpture, soap, solar_array, sponge, statue_(sculpture), stove, mop, sword, telephone_pole, cover, towel, toy, tripod, vat, yoke_(animal_equipment)