Few-Shot Object Detection with Attention-RPN and Multi-Relation Detector

Qi Fan* HKUST qfanaa@cse.ust.hk

Wei Zhuo* Tencent wei.zhuowx@gmail.com

Chi-Keung Tang HKUST cktang@cse.ust.hk

Yu-Wing Tai Tencent yuwingtai@tencent.com

Appendix A: Implementation Details of Multi-Relation Detector

Given the support feature f_s and query proposal feature f_q with the size of $7 \times 7 \times C$, our multi-relation detector is implemented as follows. We use the sum of all matching scores from the three heads as the final matching scores.

Global-Relation Head We concatenate f_s and f_q to the concatenated feature f_c with the size of $7 \times 7 \times 2C$. Then we average pool f_c to a $1 \times 1 \times 2C$ vector. We then use an MLP with two fully connected (fc) layers with ReLU and a final fc layer to process f_c and generate matching scores.

Local-Relation Head We first use a weight-shared $1 \times 1 \times C$ convolution to process f_s and f_q separately. Then we calculate the depth-wise similarity using the equation in Section 4.2.1 of the main paper with $S = H = W = 7$. Then we use a fc layer to generate matching scores.

Patch-Relation Head We first concatenate f_s and f_q to the concatenated feature f_c with the size of $7 \times 7 \times 2C$. Then f_c is fed into the patch-relation module, whose structure is shown in Table 1. All the convolution layers followed by ReLU and pooling layers in this module have zero padding to reduce the feature map size from 7×7 to 1×1. Then we use a fc layer to generate matching scores and a separate fc layer to generate bounding box predictions.

Appendix B: More Implementation Details

B.1. Training and Fine-tuning details

Here we show more details for the experiments in Section 5.2 of the main paper.

In Section 5.2, we follow other methods to train our model on MS COCO dataset [1] and fine-tune on the target datasets. When we train our model on MS COCO, we remove the images with boxes smaller than the size of 32×32. Those boxes are usually in bad visual quality and hurt the training when they serve as support examples. When we fine-tune our model on the target datasets, we follow the same setting of other methods[2, 3, 4, 5] for fair comparison. Specifically, LSTD [2] and RepMet [3] use 5 support images per category where each image contains one or more object instances, and the Feature Reweighting [4] and Meta R-CNN [5] use a strict rule to adopt 10 object instances per category for fine-tuning.

Type	Filter Shape	Stride/Padding
Avg Pool	3x3x4096	s1/p0
Conv	1x1x512	s1/p0
Conv	3x3x512	s1/p0
Conv	1x1x2048	s1/p0
Avg Pool	3x3x2048	s1/p0

Table 1. Architecture of the patch-relation module.

B.2. Evaluation details

There are two evaluation settings in the main paper. **Evaluation setting 1:** The ablation experiments adopt the episode-based evaluation protocol defined in RepMet [3], where the setting is borrowed from the few-shot classification task [6, 7]. There are 600 random evaluation episodes in total, which guarantee every image in the test set can be evaluated in a high probability. In each episode, for N-way K-shot evaluation, there are K support images for each of the N categories, and there are 10 query images for each category where each query image containing at least one instance belonging to this category. So there are $K \times N$ supports and $10 \times N$ query images in each episode. Note that all these categories and images are randomly chosen in each episode. **Evaluation setting 2:** Other comparison experiments with baselines adopt the standard object detection evaluation protocol, which is a full-way, N-shot evaluation. During evaluation, the support branches in our model can be discarded once the support features are attained, then the support features serve as model weights for the forward process.

Appendix C: Qualitative Visualization Results

FSOD Dataset Here we show qualitative results for the ablation study in Fig. 1. From the visualization results, we can see the advantage of our model over baselines and the advantage of our dataset over benchmark dataset, *e.g.* MS COCO. Here, we present results on three categories, *i.e.* blackboard, segway and turban, which are very distinct with the training samples in appearance. From Fig. 1, we can

*Both authors contributed equally.

This research is supported in part by Tencent and the Research Grant Council of the Hong Kong SAR under grant no. 1620818.

see that both the novel attention RPN and multi-relation detector of our model play an important role in accurate detection. For example, in (1-2)(b) the Patch-Relation head fails to detect some targets while our model with only the multi-relation detector can detect more targets and suppress backgrounds. In (2-3)(d), our full model further correct the detection and pull up the scores on true targets by adding the Attention RPN, especially in the instance of (3)(d). Comparing results of our model trained on our dataset with that on MS COCO dataset, shown in Fig. 1 (1-3)(a), our model can rarely learn a matching between the support and query target when we train it on MS COCO, but it can capture the similarity between the image pair when it trains on our dataset. More detection of our full model on the 1-shot evaluation setting is shown in Fig. 2 and Fig. 3. Our model performs well even when there are objects of non-support categories in the query image.

Cityscapes and KITTI We apply our method on benchmark datasets, such as Cityscapes and KITTI, with five shots which are randomly selected from the test dataset. Results are shown in Fig. 4 and 5 respectively. It shows that our model performs well on these datasets, even though our model is trained in a very different domain.

Appendix D: FSOD Dataset Class Split

Here we describe the training/testing class split in our proposed FSOD Dataset. This split was used in our experiments.

Training Class Split

lipstick, sandal, crocodile, football helmet, umbrella, houseplant, antelope, woodpecker, palm tree, box, swan, miniskirt, monkey, cookie, scissors, snowboard, hedgehog, penguin, barrel, wall clock, strawberry, window blind, butterfly, television, cake, punching bag, picture frame, face powder, jaguar, tomato, isopod, balloon, vase, shirt, waffle, carrot, candle, flute, bagel, orange, wheelchair, golf ball, unicycle, surfboard, cattle, parachute, candy, turkey, pillow, jacket, dumbbell, dagger, wine glass, guitar, shrimp, worm, hamburger, cucumber, radish, alpaca, bicycle wheel, shelf, pancake, helicopter, perfume, sword, ipod, goose, pretzel, coin, broccoli, mule, cabbage, sheep, apple, flag, horse, duck, salad, lemon, handgun, backpack, printer, mug, snowmobile, boot, bowl, book, tin can, football, human leg, countertop, elephant, ladybug, curtain, wine, van, envelope, pen, doll, bus, flying disc, microwave oven, stethoscope, burrito, mushroom, teddy bear, nail, bottle, raccoon, rifle, peach, laptop, centipede, tiger, watch, cat, ladder, sparrow, coffee table, plastic bag, brown bear, frog, jeans, harp, accordion, pig, porcupine, dolphin, owl, flowerpot, motorcycle, calculator, tap, kangaroo, lavender, tennis ball, jellyfish, bust, dice, wok, roller skates, mango, bread, computer monitor, sombrero, desk, cheetah, ice cream, tart, doughnut, grapefruit, paddle, pear, kite, eagle, towel, coffee, deer, whale, cello, lion, taxi, shark, human arm, trumpet, french fries, syringe, lobster, rose, human hand, lamp, bat, ostrich, trombone, swim cap, human beard, hot dog, chicken, leopard, alarm clock, drum, taco, digital clock, starfish, train, belt, refrigerator, dog bed, bell pepper, loveseat, infant bed, training bench, milk, mixing bowl, knife, cutting board, ring binder, studio couch, filing cabinet, bee, caterpillar, sofa bed, violin, traffic light, airplane, closet, canary, toilet paper, canoe, spoon, fox, tennis racket, red panda, cannon, stool, zucchini, rugby ball, polar bear, bench, pizza, fork, barge, bow and arrow, kettle, goldfish, mirror, snail, poster, drill, tie, gondola, scale, falcon, bull, remote control, horn, hamster, volleyball, stationary bicycle, dishwasher, limousine, shorts, toothbrush, bookcase, baseball glove, computer mouse, otter, computer keyboard, shower, teapot, human foot, parking meter, ski, beaker, castle, mobile phone, suitcase, sock, cupboard, crab, common fig, missile, swimwear, saucer, popcorn, coat, plate, stairs, pineapple, parrot, fountain, binoculars, tent, pencil case, mouse, sewing machine, magpie, handbag, saxophone, panda, flashlight, baseball bat, golf cart, banana, billiard table, tower, washing machine, lizard, brassiere, ant, crown, oven, sea lion, pitcher, chest of drawers, crutch, hippopotamus, artichoke, seat belt, microphone, lynx, camel, rabbit, rocket, toilet, spider, camera, pomegranate, bathtub, jug, goat, cowboy hat, wrench, stretcher, balance beam, necklace, scoreboard, horizontal bar, stop sign, sushi, gas stove, tank, armadillo, snake, tripod, cocktail, zebra, toaster, frying pan, pasta, truck, blue jay, sink, lighthouse, skateboard, cricket ball, dragonfly, snowplow, screwdriver, organ, giraffe, submarine, scorpion, honeycomb, cream, cart, koala, guacamole, raven, drawer, diaper, fire hydrant, potato, porch, banjo, hammer, paper towel, wardrobe, soap dispenser, asparagus, skunk, chainsaw, spatula, ambulance, submarine sandwich, axe, ruler, measuring cup, scarf, squirrel, tea, whisk, food processor, tick, stapler, oboehartebeest, modem, shower cap, mask, handkerchief, falafel, clipper, croquette, house finch, butterfly fish, lesser scaup, barbell, hair slide, arabian camel, pill bottle, springbok, camper, basketball player, bumper car, wisent, hip, wicket, medicine ball, sweet orange, snowshoe, column, king charles spaniel, crane, scoter, slide rule, steel drum, sports car, go kart, gearing, tostada, french loaf, granny smith, sorrel, ibex, rain barrel, quail, rhodesian ridgeback, mongoose, red backed sandpiper, penlight, samoyed, pay phone, barber chair, wool, ballplayer, malamute, reel, mountain goat, tusker, longwool, shopping cart, marble, shuttlecock, red breasted merganser, shutter, stamp, letter opener, canopic jar, warthog, oil filter, petri dish, bubble, african crocodile, bikini, brambling, siamang, bison, snorkel, loafer, kite balloon, wallet, laundry cart, sausage dog, king penguin, diver, rake, drake, bald eagle, retriever, slot, switchblade, orangutan, chacma, guenon, car wheel, dandie dinmont, guanaco, corn, hen, african hunting dog, pajama, hay,

Figure 1. Qualitative results for ablation study. (a) Our full model trained on MS COCO; (b) Our model only with Patch Relation head trained on FSOD; (c) Our full model without attention RPN trained on FSOD; (d) Our full model trained on FSOD. We visualize the boxes with scores larger than 0.8 or the box with highest score in (3)(c).

dingo, meat loaf, kid, whistle, tank car, dungeness crab, pop bottle, oar, yellow lady's slipper, mountain sheep, zebu, crossword puzzle, daisy, kimono, basenji, solar dish, bell, gazelle, agaric, meatball, patas, swing, dutch oven, military uniform, vestment, cavy, mustang, standard poodle, chesapeake bay retriever, coffee mug, gorilla, bearskin, safety pin, sulphur crested cockatoo, flamingo, eider, picket fence, dhole, spaghetti squash, african elephant, coral fungus, pelican, anchovy pear, oystercatcher, gyromitra, african grey, knee pad, hatchet, elk, squash racket, mallet, greyhound, ram, racer, morel, drumstick, bovine, bullet train, bernese mountain dog, motor scooter, vervet, quince, blenheim spaniel, snipe, marmoset, dodo, cowboy boot, buckeye, prairie chicken, siberian husky, ballpoint, mountain tent, jockey, border collie, ice skate, button, stuffed tomato, lovebird, jinrikisha, pony, killer whale, indian elephant, acorn squash, macaw, bolete, fiddler crab, mobile home, dressing table, chimpanzee, jack o' lantern, toast, nipple, entlebucher, groom, sarong, cauliflower, apiary, english foxhound, deck chair, car door, labrador retriever, wallaby, acorn, short pants, standard schnauzer, lampshade, hog, male horse, martin, loudspeaker, plum, bale, partridge, water jug, shoji, shield, american lobster, nailfile, poodle, jackfruit, heifer, whippet, mitten, eggnog, weimaraner, twin bed, english springer, dowitcher, rhesus, norwich terrier, sail, custard apple, wassail, bib, bullet, bartlett, brace, pick, carthorse, ruminant, clog, screw, burro, mountain bike, sunscreen, packet, madagascar cat, radio telescope, wild sheep, stuffed peppers, okapi, bighorn, grizzly, jar, rambutan, mortarboard, raspberry, gar, andiron, paintbrush, running shoe, turnstile, leonberg, red wine, open face sandwich, metal screw, west highland white terrier, boxer, lorikeet, interceptor, ruddy turnstone, colobus, pan, white stork, stinkhorn, american coot, trailer truck, bride, afghan hound, motorboat, bassoon, quesadilla, goblet, llama, folding chair, spoonbill, workhorse, pimento, anemone fish, ewe, megalith, pool ball, macaque, kit fox, oryx, sleeve, plug, battery, black stork, saluki, bath towel, bee eater, baboon, dairy cattle, sleeping bag, panpipe, gemsbok, albatross, comb, snow goose, cetacean, bucket, packhorse, palm, vending machine, butternut squash, loupe, ox, celandine, appenzeller, vulture, crampon, backboard, european gallinule, parsnip, jersey, slide, guava, cardoon, scuba diver, broom, giant schnauzer, gordon setter, staffordshire bullterrier, conch, cherry, jam, salmon, matchstick, black swan, sailboat, assault rifle, thatch, hook, wild boar, ski pole, armchair, lab coat, goldfinch, guinea pig, pinwheel, water buffalo, chain, ocarina, impala, swallow, mailbox, langur, cock, hyena, marimba, hound, knot, saw, eskimo dog, pembroke, sealyham terrier, italian greyhound, shih tzu, scotch terrier, yawl, lighter, dung beetle, dugong, academic gown, blanket, timber wolf, minibus, joystick, speedboat, flagpole, honey, chessman, club sandwich, gown, crate, peg, aquarium, whooping crane, headboard, okra, trench coat, avocado, cayuse, large yellow lady's slipper, ski mask, dough, bassarisk, bridal gown, terrapin, yacht, saddle, redbone, shower curtain, jennet, school bus, otterhound, irish terrier, carton, abaya, window shade, wooden spoon, yurt, flat coated retriever, bull mastiff, cardigan, river boat, irish wolfhound, oxygen mask, propeller, earthstar, black footed ferret, rocking chair, beach wagon, litchi, pigeon.

Testing Class Split

beer, musical keyboard, maple, christmas tree, hiking equipment, bicycle helmet, goggles, tortoise, whiteboard, lantern, convenience store, lifejacket, squid, watermelon, sunflower, muffin, mixer, bronze sculpture, skyscraper, drinking straw, segway, sun hat, harbor seal, cat furniture, fedora, kitchen knife, hand dryer, tree house, earrings, power plugs and sockets, waste container, blender, briefcase, street light, shotgun, sports uniform, wood burning stove, billboard, vehicle registration plate, ceiling fan, cassette deck, table tennis racket, bidet, pumpkin, tablet computer, rhinoceros, cheese, jacuzzi, door handle, swimming pool, rays and skates, chopsticks, oyster, office building, ratchet, salt and pepper shakers, juice, bowling equipment, skull, nightstand, light bulb, high heels, picnic basket, platter, cantaloupe, croissant, dinosaur, adhesive tape, mechanical fan, winter melon, egg, beehive, lily, cake stand, treadmill, kitchen & dining room table, headphones, wine rack, harpsichord, corded phone, snowman, jet ski, fireplace, spice rack, coconut, coffeemaker, seahorse, tiara, light switch, serving tray, bathroom cabinet, slow cooker jalapeno, cartwheel, laelia, cattleya, bran muffin, caribou, buskin, turban, chalk, cider vinegar, bannock, persimmon, wing tip, shin guard, baby shoe, euphonium, popover, pulley, walking shoe, fancy dress, clam, mozzarella, peccary, spinning rod, khimar, soap dish, hot air balloon, windmill, manometer, gnu, earphone, double hung window, conserve, claymore, scone, bouquet, ski boot, welsh poppy, puffball, sambuca, truffle, calla lily, hard hat, elephant seal, peanut, hind, jelly fungus, pirogi, recycling bin, in line skate, bialy, shelf bracket, bowling shoe, ferris wheel, stanhopea, cowrie, adjustable wrench, date bread, o ring, caryatid, leaf spring, french bread, sergeant major, daiquiri, sweet roll, polypore, face veil, support hose, chinese lantern, triangle, mulberry, quick bread, optical disk, egg yolk, shallot, strawflower, cue, blue columbine, silo, mascara, cherry tomato, box wrench, flipper, bathrobe, gill fungus, blackboard, thumbtack, longhorn, pacific walrus, streptocarpus, addax, fly orchid, blackberry, kob, car tire, sassaby, fishing rod, baguet, trowel, cornbread, disa, tuning fork, virginia spring beauty, samosa, chigetai, blue poppy, scimitar, shirt button.

References

[1] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014.

[2] Hao Chen, Yali Wang, Guoyou Wang, and Yu Qiao. Lstd: A low-shot transfer detector for object detection. In *AAAI*, 2018.

[3] Leonid Karlinsky, Joseph Shtok, Sivan Harary, Eli Schwartz, Amit Aides, Rogerio Feris, Raja Giryes, and Alex M Bronstein. Repmet: Representative-based metric learning for classification and few-shot object detection. In *CVPR*, 2019.

[4] Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, and Trevor Darrell. Few-shot object detection via feature reweighting. In *ICCV*, 2019.

[5] Xiaopeng Yan, Ziliang Chen, Anni Xu, Xiaoxi Wang, Xiaodan Liang, and Liang Lin. Meta r-cnn : Towards general solver for instance-level low-shot learning. In *ICCV*, 2019.

[6] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *NeurIPS*, 2016.

[7] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *ICLR*, 2017.

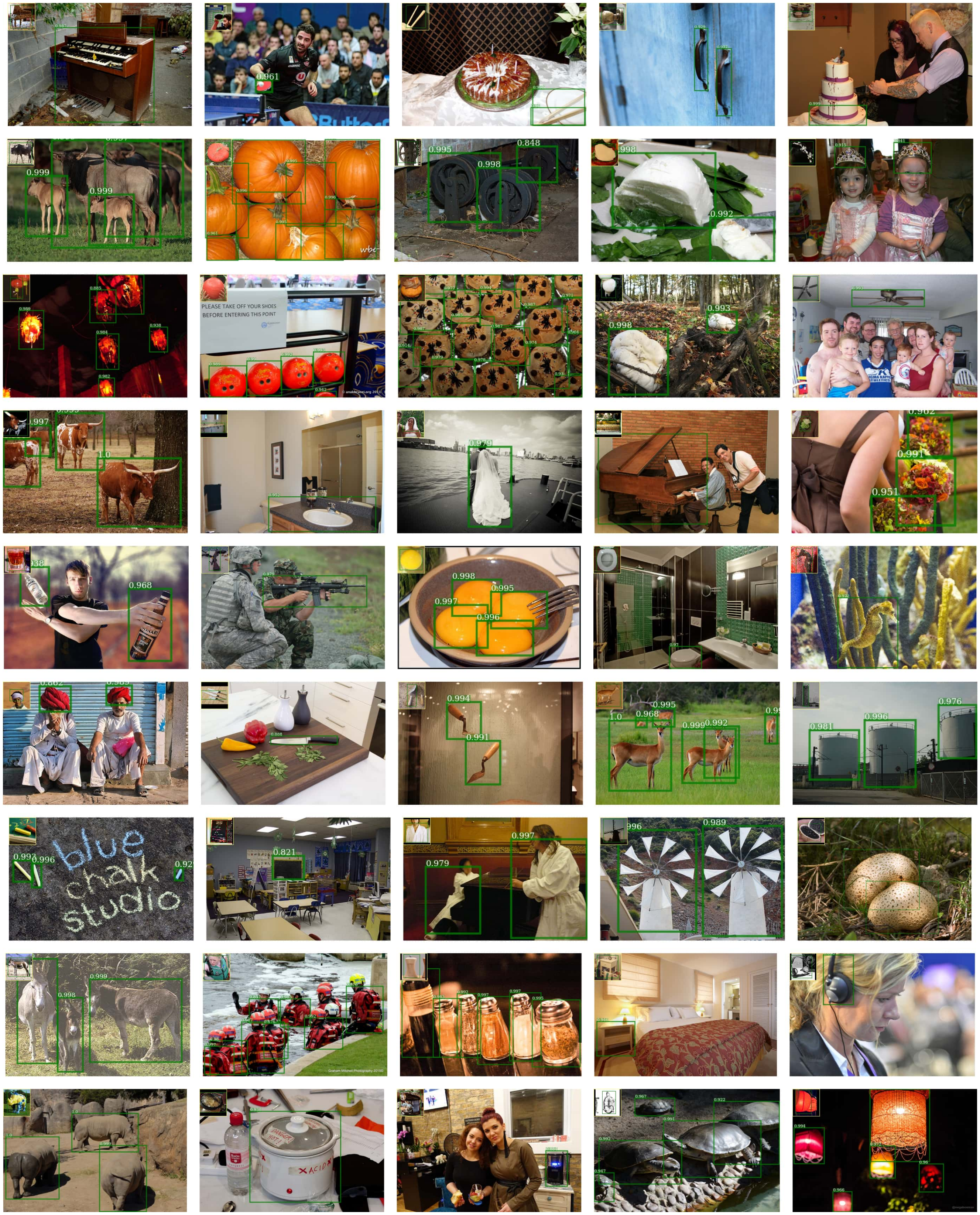

Figure 2. Qualitative 1-shot object detection results on our test set. We visualize the bounding boxes with score larger than 0.8.

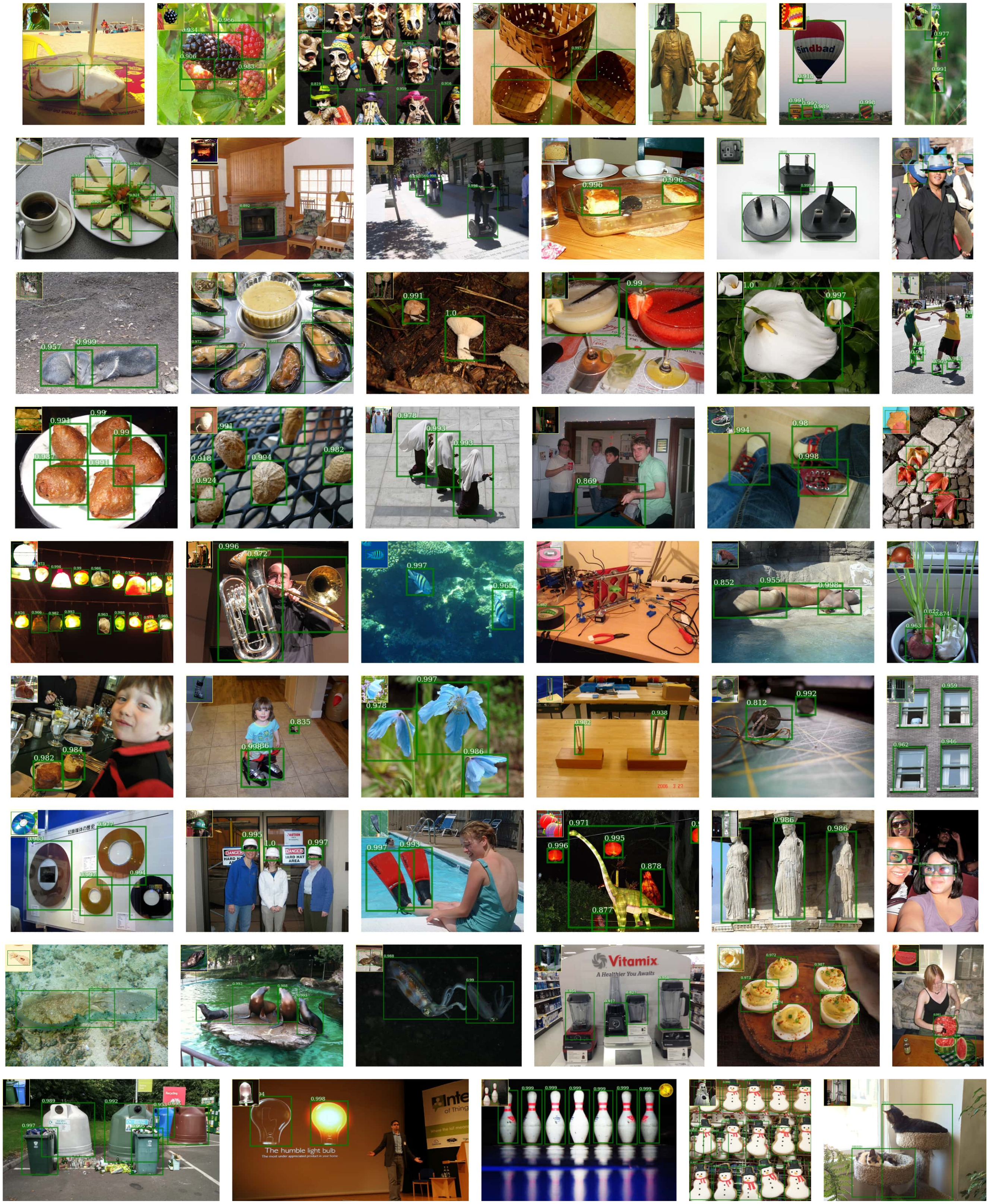

Figure 3. Qualitative results of our 1-shot object detection on test set. We visualize the bounding boxes with score larger than 0.8.

Figure 4. Qualitative results of our 5-shot car detection on Cityscapes. We visualize the bounding boxes with score larger than 0.8. The first image is a training example.

Figure 5. Qualitative results of our 5-shot car detection on KITTI. We visualize the bounding boxes with score larger than 0.8.