Supplementary Material:
Webly Supervised Learning of Convolutional Networks

Anonymous ICCV submission

Paper ID 465

In the supplementary material we include:

1. Additional results on PASCAL VOC for ablation analysis.
2. Scene classification results.
3. Diagnosis results for webly supervised object detection using [4].
4. Lists of objects, scenes and attributes.

1. Additional Results on PASCAL VOC

For ablation analysis, we provide more results on the PASCAL VOC 2007 detection challenge. Please refer to Table 1 for comparison. The newly added results are shown at the bottom rows. Following the notation in the paper, “NFT” means before fine-tuning, and “FT” means after fine-tuning (100K iterations with a step size of 20K). We add three pairs of new results:

GoogleO-CI “CI” stands for “Cleaned Images”. This network is obtained by fine-tuning GoogleO on the images obtained by our object localization algorithm (described in Section 3.3). Here, the entire image is regarded as clean if at least one object is found inside it. However, the extra location information (bounding box) is not used - the input is still the full image.

GoogleO-CB “CB” stands for “Cleaned images with Bounding boxes”. Similar to GoogleO-CI, the network is fine-tuned on the cleaned images. However, instead of using the entire image, the image patches cropped by the discovered bounding boxes are fed into the network. Both GoogleO-CI and Google-CB were fine-tuned for 200K iterations, with the learning rate reduced every 40K iterations.

IN-GoogleA To see if more data can help the performance, we fine-tuned the ImageNet pretrained network [3] to all the images downloaded from Google. The fine-tuning was performed for 400K iterations, reducing learning rate every 80K iterations.

Somewhat to our surprise, we found the extra clean-up step does not help improving the detection performance. In fact, the average precision dropped for most of the categories, regardless of whether the bounding box information is used or not. We suspect one reason lies in the size of the data: after the object localization step, the number of images was cut in more than a half (~0.67M compared to ~1.5M). Also better algorithms can be devised for cleaning up web images.

On the other hand, fine-tuning ImageNet pretrained model to Google images gives slightly better result than training from scratch. However, further investigation is still needed here since IN-GoogleA has seen more images (ImageNet 1M) than GoogleA.

2. Scene Classification

To further demonstrate the usage of CNN features directly learned from the web, we also conducted a set of scene classification experiments on the MIT Indoor-67 dataset [5]. For each image, we simply computed the fc7 feature vector, which has 4096 dimensions. We did not use any data augmentation or spatial pooling technique, with the only pre-processing step normalizing the feature vector to unit \( \ell_2 \) length [6]. The default SVM parameters \((C=1)\) were fixed throughout the experiments.

Table 2 summarizes the results on the default train/test split. We can see our web based CNNs achieved very competitive performances: all the three networks achieved an accuracy at least on par with ImageNet pretrained models. Fine-tuning on hard images enhanced the features, but adding scene-related categories gave a huge boost to 66.5 (comparable to the CNN trained on Places database [9], 68.2). This indicates CNN features learned directly from the web are indeed generic.

Moreover, since we can get semantic labels (e.g. actions) other than objects or scenes from the web for free, webly supervised CNN bears a great potential to perform well on many relevant tasks - and the cost is just as low as providing a category list to query for that domain.
3. Diagnosis for Webly Supervised Object Detection

The final analysis comes from the nice tool developed in [4]. We would like to see what is the failure modes of our webly supervised object detection (when no PASCAL VOC train data is present). We took the results from our best model (Flickr-C) and fed them for diagnosis. Figure 1, 2 and 3 highlight some of the interesting observations we found. Hopefully these findings can offer more insights about how the web based detectors are performing and what are the possible directions to go for better performance.

Firstly, localization error accounts for a lot of false positives. Since Google Image Search results do not contain location information, the background is inevitably included when the detector is trained (e.g. aeroplane, dining table). Moreover, multiple objects can occur in the image, and the algorithm has no clue that they should be treated as separate pieces (e.g. bottle). Since our CNN representation is directly trained on full images, it is also learned to be invariant to such variations. All these factors caused serious localization issues.

Second, we did observe some interesting semantic drift between PASCAL categories and Google categories. For example, bicycle can also mean motorcycle on Google. Careful disambiguation of different senses for a polysemous word is needed here. Also note that our person detector is severely confused with cars, we suspect it is because “caprice” was added as a related category and it can also mean a car (chevy caprice). How to handle such issues is a fascinating future research topic by itself.

Finally, there is massive confusion within animal categories (e.g. horse vs. cow). We believe it is mainly caused by the domain shifting between Google Images and PASCAL images. Adding Flickr data can benefit this case (only Google images are used to train our detectors).

Table 1. Additional Results on VOC-2007 (PASCAL data used).

| In-GoogleA-NFT | 55.0 60.2 39.5 33.4 23.8 58.7 60.6 48.6 23.2 44.9 35.8 44.1 50.3 60.1 42.9 23.9 49.0 39.8 49.7 54.4 44.9 |
| IN-GoogeA-FT | 64.6 69.7 44.1 34.8 28.4 66.1 72.1 56.2 30.7 57.0 45.8 50.4 61.9 66.5 54.8 31.7 58.0 43.0 56.8 62.9 52.8 |

Table 2. Scene Classification Results on MIT Indoor-67 Dataset.

<table>
<thead>
<tr>
<th>Indoor-67</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ImageNet [9]</td>
<td>56.8</td>
</tr>
<tr>
<td>OverFeat [6]</td>
<td>58.4</td>
</tr>
<tr>
<td>GoogleO</td>
<td>58.1</td>
</tr>
<tr>
<td>GoogleA</td>
<td>66.5</td>
</tr>
<tr>
<td>Flickr</td>
<td>59.2</td>
</tr>
</tbody>
</table>
Figure 2. Top false positives for selected categories on PASCAL VOC 2007 detection with Flickr-C. From top down: aeroplane, bicycle, bottle, cat.

4. Lists of Concepts

4.1. Lists for Network Training

We obtained three lists of categories from ImageNet Challenge [7], SUN database [8] and NEIL knowledge base [2].

ImageNet syn-sets are transformed to its surface forms by just taking the first explanation. Most of them are focusing on object categories. To better assist querying and reducing noise, we remove the suffix (usually correspond to attributes, e.g. indoor/outdoor) of the SUN categories. These two lists are augmented by the categories listed on NEIL knowledge base, which are better suited as query words. We also used NEIL’s list of attributes.

In total we have 2,240 objects and 89 attributes to train the GoogleNet focusing on objects.
Figure 3. Top false positives for selected categories on PASCAL VOC 2007 detection with Flickr-C. From top down: dining table, horse, person, tv monitor.
cap, shower curtain, shrubbery, shubunkin, shuttlecock, siamang, siamese cat, siamese husky, siamese tiger, side-winder, sienna cathedral, sieve, sild, silky terrier, silver, silver seattle terrier, sink, skateboard, skeleton, ski, skil, ski mask, skin texture, skoda forman, sks, skunk, sky, sleeping bag, slide rule, sliding door, slimy sculpin, slipper, slot, sloth bear, slr, slug, sma c31, smallie, smart car, smoke, snail, snailfish, snake, snakehead, snorkel, snow leopard, snowmobile, snowplow, snowy weather, soap dispenser, soccer ball, soccer net, sock, sofa, softball bat, soft coated wheaton terrier, solid dish, soldering iron, solo oobe, sobrembo, sonata, sony cybershot, sorrel, soup, soup bowl, space bar, space heater, space rover, space shuttle, spadeshift, spaghetti bolognaisne, spaghetti squash, sparring, sparrow, spatula, speckled, speedboat, sphinx, spider, spiderman, spider monkey, spider web, spigola, spindle, spiny lobster, sponge, spoon, spoonbill, sports car, sportsperson, sportster, spotlight, spotted salamander, splinker, square shape, squash ball, squash casserole, squirrel, squirrel monkey, ss 90, ssangyong roadius stivic, staffordshire bullterrier, stage, stair, standard poodle, standard schnauzer, stapler, star, starsfish, starfish, statue, statue of liberty, steam engine, steam locomotive, steam whistle, steel arch bridge, steel drum, step, stephansdom, stereo, sterling rover, stethoscope, steve jobs, stick, still water, stingray, stinkhorn, stockpot, stole, stone, stonge-hen, stone wall, stadium, stop, stopwatch, street, trailer, streetcar, stripes, stripe texture, straganooff, strumstick, studio couch, stupa, sturgeon, suburub 1000, suburub liberty, suburub outback, suburub trezia, submarine,suburban, subway, subway train, sugar, suit, sukiyaki, sulphur butterfly, sulphur crested cockatoo, sun, sunbeam alpine, sun bear, sun-dial, sunflower, sunglass, sunglasses, sunny weather, sun perch, sunscreen, supercar, superman, supernova, supersonic jet, surgoenfish, suspension bridge, sussex spaniel, suv, suzuki forenza, suzuki forenza wagon, suzuki freewind, suzuki fronte, swab, swan, sweatshirt, swimming trunks, swing, switch, sword, sxt neon, syclone, symmetrical, syringe, tabby, table, table lamp, tabor drum, tahoe, tailed frog, Taj mahal, tam tam, tank, tape measure, tape player, taphon, tarabuka, tarantula, tart, tartan, tarte tatin, tawnywine, tasmanian devil, tata nano, tea, teacup, teapot, measure, tape player, taphon, tarabuka, tarantula, tart, tartan, tarte tatin, tawnywine, tasmanian devil, tata nano, tea, teacup, teapot, vegas, vellum, velvet, vending machine, venus, vera zvonarea, verruphon, vertical, vertical cylinder, vertical lines, vestment, viaduct, vice tool, vicuna, vijay singh, vijn, vine snake, vintage guitar, violet, violin, viper, viper snake, vizsla, voca, volcano, volkswagen, volkswagen beetle, volkswagen polo, volleyball, volvo bus, vor indicator, vulture, waffle iron, wagon, walker hound, walking stick, wall, wallaby, wall clock, wallet, wardrobe, warm scenery, warplane, warthog, washbasin, washer, washing machine, watch, water bottle, waterbuck, water buffalo, water jug, water ouzel, water snake, water tower, wave, waverunner, weakfish, weasel, web site, weevil, weimaraner, welsh springer spaniel, west highland white terrier, whale, wheel, whippet, whiptail, whisk, whiskey jug, whistle, white, whitefish, white stork, white wolf, whitung, wide, wig, wild boar, window, window screen, window shade, windshield, windsor tie, wine, wine bottle, wing, wire, wire haired fox terrier, wiry, wobegong, wok, wolf, wolf spider, wolfpenterger, woman, wombat, wooden legs, wooden spoon, wood rabbit, wool, worm fence, wre, xindi, yamaha, yawl, yellow, yellow clorox bottle, yellowfin bream, yellow lady slipper, yellow tail, yellow tail fish, yoda, yorkshire terrier, yurt, zebra, zebra shark, zia syed, zinedine zidane, zofp, zucchini.

Later we found around 1/4 of the 2000+ categories are related to vehicles (aeroplane, bicycle, boat, bus, car, motorcycle, train), with a lot of them being specific car models. Therefore, this list is heavily skewed to vehicles.

To train the GoogleA net, the following list of scene categories are also included: abbey, access road, afghanistan, agra, air base, airfield, aircraft, airline, airline cabin, airport, airport entrance, airport terminal, airport ticket counter, alaska, alcove, alley, alliance bank stadium, american football field, amphitheater, amsterdam, amusement arcade, amusement park, anechoic chamber, angel stadium, apartment building, apple store, aspe, aquarium, aquatic theater, aqueduct, arcade, arch, archaeological excavation, archive, argentina, arlington stadium, armory, army base, arrival gate, art gallery, artificial lake, artists loft, art school, art studio, ashbourne, assembly line, astrodome, athens, athletic field, atlanta county stadium, atium home, atium public, attic, att park, auditorium, auto factory, auto mechanics, auto racing paddock, auto showroom, autozone park, aylesbury, backstage, backstory, badlands, badminton court, baggage claim, bakery kitchen, bakery shop, balcony, barnard, park, ballpark, ball field, ballroom, bamboo forest, bank, bank vault, banquet hall, baptistery, bar, barbershop, barn, barndoor, barnyard, barack, baseball diamond, baseball field, baseball stadium, base- ment, basilica, basketball arena, basketball court, bathhouse, bathroom, batters box, batting cage, battenage, bayou, bazaar, beach, beach house, beauty salon, bedchamber, bedroom, beer garden, beer hall, beijing, belfry, bell foundry, bern, berth, berth deck, betting shop, bicycle racks, bindery, biology laboratory, bistro, bleachers, bloomingdale, boardwalk, boat cargo deck, boat deck, boathouse, bog, bolivia, bollywood, bologna, bomb shelter, bookbindery, bookstore, booth, botanical garden, bowling alley, bow window, boxing ring, box seat, brave field, brazil, breakroom, brewery, brickyard, bridge, building complex, building facade, bullpen, bullring, burial chamber, busch memorial stadium, bus depot, bus interior, bus shelter, bus station, butchers shop, butte, cabana, cabin, cafeteria, call center, camden stadium, camden yards, campsite, campus, canal natural, canal urban, candlestick park, candy store, canteen, canyon, caravansary, car...
4.2. List for Category Expansion

For category expansion (Section 3.3), we used more categories for each of the 20 PASCAL VOC classes. Since our list is skewed toward vehicles, some categories (e.g., bottle, horse) did not receive any augmented category after semantic verification.

- aeroplane: aircraft, fighter plane, cessna aircraft, plane, airplane, airliner, supersonic jet, warplane, f16, boeing dreamliner, airbus 340, biplane, lufthansa plane, airbus a320, airbus a380, airbus a330, airbus 330, delta airlines, american airlines.
- bicycle: hero cycles, cannondale, bicycle built for two, bmx, mountain bike, giant defy.
- bird: loriikeet, indigo bunting, goldfinch, bee eater, hummingbird, toucan, broadbill, macaw, sparrow, chickadee, red backed sandpiper, magpie, baya weaver, pigeon, dove, house finch, brambling, eagle, seagull, seagull, spoonbill, bald eagle, water ouzel, plover, jacamar, peacock, junco, hornbill, vulture, coucal, white stork, swan, african grey, quail, partridge, bittern.
- boat: ship, container ship, schooner, sailboat, navy destroyer, catamaran, aircraft carrier, titanic, speedboat, trawmoran, motorboat, barge, fireboat, lifeboat, gondola, canoe, yawl, kayak.
- bottle: NA
- bus: school bus, double decker, articulated bus, coach bus, volvo bus, trolleybus, minibus, greyhound bus.
- car: sports car, supercar, lamborghini, mazda, smart car, ferrari, hatchback, roadster, corvette, toyota prius, tesla, convertible, camaro, porche carrera, hyundai, saleen s7, bugy, lincoln car, nissan auto, coupe, dodge, audi a8, fusion hybrid, sedan, viper, mitsubishi, Plymouth, scion, 1950 car, infiniti, ac car, honda hsc, honda nxs, mazda senku, rover, c max, nissan versa, acura, cabriolet, toyota supra, jaguar car, bmw 320, grigio telesto, chevrolet panther, citroen projet, fiat abarth, audi a6, vauxhall corsa, porche boxtor, toyota spyder, plymouth challenger, gumpert apollo, suv, mazdaspeed, crossover, honda incentives, fiat, kia venga, citroen, lancia, datsun.
- cat: tabby, persian cat, angora, siamese cat, tiger cat, egyptian cat, lynx.
- chair: armchair, rocking chair, throne, bench, stool, folding chair, office chair, aeron, barber chair, dentist chair.
- cow: bull, oxen, ox, water buffalo.
- dining table: table, desk.
- dog: golden retriever, labrador retriever, german shepherd, eskimo dog, maltese dog, beagle, pug, pomeranian, samoyed, siberian husky, staffordshire bullterrier, english foxhound, malinois, chihuahua, malamute, toy terrier, vizsla, shetland sheepdog, american staffordshire terrier, border collie, west highland white terrier, collie, boxer, kelpie, dalmation, rottweiler, saint bernard, papillon, flat coated retriever, walker hound, brittany spaniel, greater swiss mountain dog, bull mastiff, chow, miniature poodle, cocker spaniel, toy poodle, rhodesian ridgeback, saluki, basset, greyhound, great pyrenees, welsh springer spaniel, lakeland terrier, basjeni, miniature pinscher, japanese spaniel, kuvasz.
- horse: NA
- motorcycle: motor scooter, yamaha, tmx, bsa, m50, dirt bike, honda vtr, honda cr85r, honda cross tourer, sportster, honda cbr.
- person: woman, actor, actress, man, girl, face, boy, baseball, katy perry, andre agassi, barack obama, andrew garfield, baby, caprice, ender wiggin, sportsperson.
- potted plant: fern, parsley.
- sheep: goat.
- sofa: couch, studio couch.
- train: bullet train, electric locomotive, locomotive, steam locomotive, steam engine, tv, subway train, passenger car, tram.
- tv monitor: tv, television, monitor, screen, imac.

References