Natural Adversarial Examples Supplementary Materials

Dan Hendrycks
UC Berkeley

Kevin Zhao*
University of Washington

Steven Basart*
UChicago

Jacob Steinhardt, Dawn Song
UC Berkeley

1. Appendix

2. Expanded Results

2.1. Full Architecture Results

Full results with various architectures are in Table 1.

2.2. More OOD Detection Results and Background

Works in out-of-distribution detection frequently use the maximum softmax baseline to detect out-of-distribution examples [4]. Before neural networks, using the reject option or a $k+1$st class was somewhat common [1], but with neural networks it requires auxiliary anomalous training data. New neural methods that utilize auxiliary anomalous training data, such as Outlier Exposure [5], do not use the reject option and still utilize the maximum softmax probability. We do not use Outlier Exposure since that paper's authors were unable to get their technique to work on ImageNet-1K with 224×224 images, though they were able to get it work on Tiny ImageNet which has 64×64 images. We do not use ODIN since it requires tuning hyperparameters directly using out-of-distribution data, a criticized practice [5].

We evaluate three additional out-of-distribution detection methods, though none substantially improve performance. We evaluate method of [2], which trains an auxiliary branch to represent the model confidence. Using a ResNet trained from scratch, we find this gets a 14.3% AUPR, around 2% less than the MSP baseline. Next we use the recent Maximum Logit detector [3]. With DenseNet-121 the AUPR decreases from 16.1% (MSP) to 15.8% (Max Logit), while with ResNeXt-101 (32×8d) the AUPR of 20.5% increases to 20.6%. Across over 10 models we found the MaxLogit technique to be slightly worse. Finally, we evaluate the utility of self-supervised auxiliary objectives for OOD detection. The rotation prediction anomaly detector [6] was shown to help improve detection performance for near-distribution yet still out-of-class examples, and with this auxiliary objective the AUPR for ResNet-50 does not change; it is 16.2% with the rotation prediction and 16.2% with the MSP. Note this method requires training the network and does not work out-of-the-box.

*Equal Contribution.

Figure 1: A demonstration of color sensitivity. While the leftmost image is classified as "banana" with high confidence, the images with modified color are correctly classified. Not only would we like models to be more accurate, we would like them to be calibrated if they wrong.

2.3. Calibration

In this section we show IMAGENET-A calibration results.

Uncertainty Metrics. The ℓ_2 *Calibration Error* is how we measure miscalibration. We would like classifiers that can reliably forecast their accuracy. Concretely, we want classifiers which give examples 60% confidence to be correct 60% of the time. We judge a classifier's miscalibration with the ℓ_2 Calibration Error [7].

Our second uncertainty estimation metric is the *Area Under the Response Rate Accuracy Curve (AURRA).* Responding only when confident is often preferable to predicting falsely. In these experiments, we allow classifiers to respond to a subset of the test set and abstain from predicting the rest. Classifiers with quality uncertainty estimates should be capable identifying examples it is likely to predict falsely and abstain. If a classifier is required to abstain from predicting on 90% of the test set, or equivalently respond to the remaining 10% of the test set, then we should like the classifier's uncertainty estimates to separate correctly and falsely classified examples and have high accuracy on the selected 10%. At a fixed response

	ImageNet-A (Acc %)	ImageNet-O (AUPR %)
AlexNet	1.77	15.44
SqueezeNet1.1	1.12	15.31
VGG16	2.63	16.58
VGG19	2.11	16.80
VGG19+BN	2.95	16.57
DenseNet121	2.16	16.11
ResNet-18	1.15	15.23
ResNet-34	1.87	16.00
ResNet-50	2.17	16.20
ResNet-101	4.72	17.20
ResNet-152	6.05	18.00
ResNet-50+Squeeze-and-Excite	6.17	17.52
ResNet-101+Squeeze-and-Excite	8.55	17.91
ResNet-152+Squeeze-and-Excite	9.35	18.65
ResNet-50+DeVries Confidence Branch	0.35	14.34
ResNet-50+Rotation Prediction Branch	2.17	16.20
Res2Net-50 (v1b)	14.59	19.50
Res2Net-101 (v1b)	21.84	22.69
Res2Net-152 (v1b)	22.4	23.90
ResNeXt-50 (32 $\times$ 4d)	4.81	17.60
ResNeXt-101 (32 $\times$ 4d)	5.85	19.60
ResNeXt-101 (32 $\times$ 8d)	10.2	20.51
DPN 68	3.53	17.78
DPN 98	9.15	21.10
DeiT-tiny	7.25	17.4
DeiT-small	19.1	20.9
DeiT-base	28.2	24.8

Table 1: Expanded IMAGENET-A and IMAGENET-O architecture results. Note IMAGENET-O performance is improving more slowly.

rate, we should like the accuracy to be as high as possible. At a 100% response rate, the classifier accuracy is the usual test set accuracy. We vary the response rates and compute the corresponding accuracies to obtain the Response Rate Accuracy (RRA) curve. The area under the Response Rate Accuracy curve is the AURRA. To compute the AURRA in this paper, we use the maximum softmax probability. For response rate p, we take the p fraction of examples with highest maximum softmax probability. If the response rate is 10%, we select the top 10% of examples with the highest confidence and compute the accuracy on these examples. An example RRA curve is in Figure 2 .

3. IMAGENET-A Classes

The 200 ImageNet classes that we selected for IMAGENET-A are as follows. goldfish, great white shark, hammerhead, stingray, hen, ostrich, goldfinch, junco, bald eagle, vulture, newt, axolotl, tree frog, iguana, African chameleon, cobra, scorpion, tarantula, centipede, peacock, lorikeet, hummingbird, toucan, duck, goose, black swan, koala, jellyfish, snail, lobster, hermit crab, flamingo, american egret, pelican, king penguin, grey whale, killer whale, sea lion, chihuahua, shih tzu, afghan hound, basset hound, beagle, bloodhound, italian greyhound, whippet, weimaraner, yorkshire terrier, boston terrier, scottish terrier, west highland white terrier, golden retriever, labrador retriever, cocker spaniels, collie, border collie, rottweiler, german shepherd dog, boxer, french bulldog, saint bernard, husky, dalmatian, pug, pomeranian, chow chow, pembroke welsh corgi, toy poodle, standard poodle, timber wolf, hyena, red fox, tabby cat, leopard, snow leopard, lion, tiger, chee-

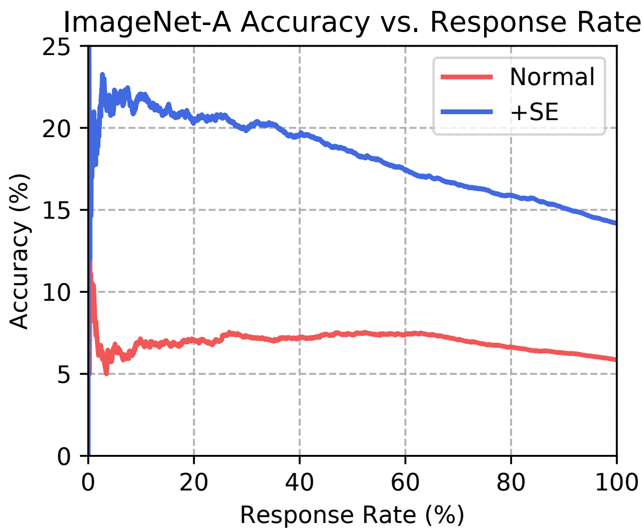

Figure 2: The Response Rate Accuracy curve for a ResNeXt-101 (32×4d) with and without Squeeze-and-Excitation (SE). The Response Rate is the percent classified. The accuracy at a $n\%$ response rate is the accuracy on the $n\%$ of examples where the classifier is most confident.

tah, polar bear, meerkat, ladybug, fly, bee, ant, grasshopper, cockroach, mantis, dragonfly, monarch butterfly, starfish, wood rabbit, porcupine, fox squirrel, beaver, guinea pig, zebra, pig, hippopotamus, bison, gazelle, llama, skunk, badger, orangutan, gorilla, chimpanzee, gibbon, baboon, panda, eel, clown fish, puffer fish, accordion, ambulance, assault rifle, backpack, barn, wheelbarrow, basketball, bathtub, lighthouse, beer glass, binoculars, birdhouse, bow tie, broom, bucket, cauldron, candle, cannon, canoe, carousel, castle, mobile phone, cowboy hat, electric guitar, fire engine, flute, gasmask, grand piano, guillotine, hammer, harmonica, harp, hatchet, jeep, joystick, lab coat, lawn mower, lipstick, mailbox, missile, mitten, parachute, pickup truck, pirate ship, revolver, rugby ball, sandal, saxophone, school bus, schooner, shield, soccer ball, space shuttle, spider web, steam locomotive, scarf, submarine, tank, tennis ball, tractor, trombone, vase, violin, military aircraft, wine bottle, ice cream, bagel, pretzel, cheeseburger, hotdog, cabbage, broccoli, cucumber, bell pepper, mushroom, Granny Smith, strawberry, lemon, pineapple, banana, pomegranate, pizza, burrito, espresso, volcano, baseball player, scuba diver, acorn,

n01443537, n01484850, n01494475, n01498041, n01514859, n01518878, n01531178,

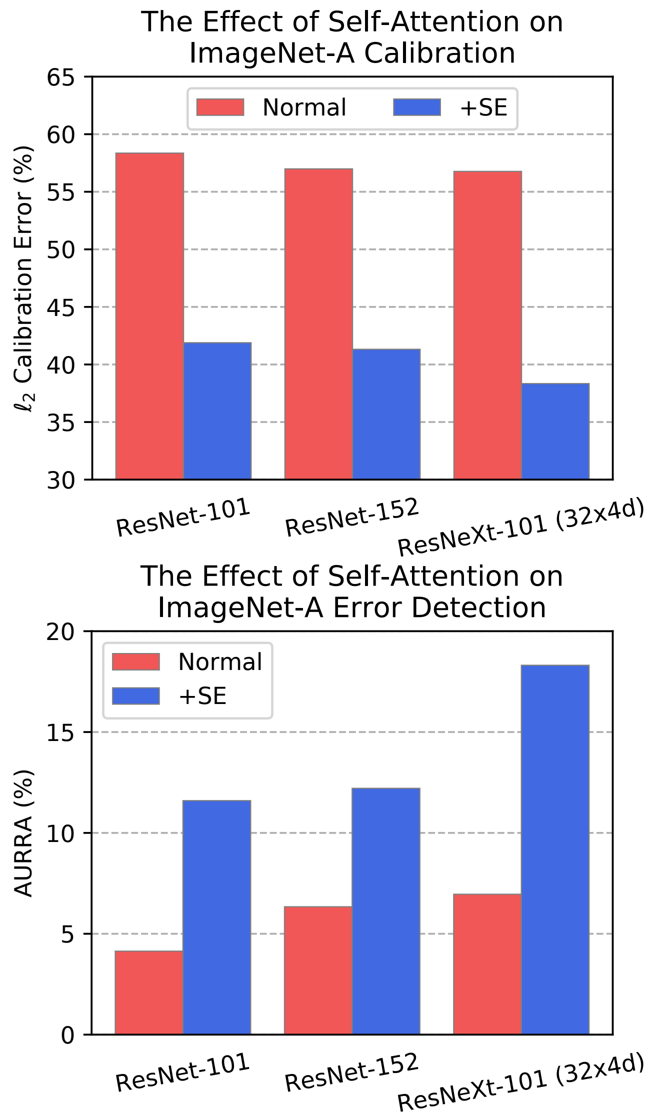

Figure 3: Self-attention's influence on IMAGENET-A ℓ_2 calibration and error detection.

n01534433, n01614925, n01616318, n01630670, n01632777, n01644373, n01677366, n01694178, n01748264, n01770393, n01774750, n01784675, n01806143, n01820546, n01833805, n01843383, n01847000, n01855672, n01860187, n01882714, n01910747, n01944390, n01983481, n01986214, n02007558, n02009912, n02051845, n02056570, n02066245, n02071294, n02077923, n02085620, n02086240, n02088094, n02088238, n02088364, n02088466, n02091032, n02091134, n02092339, n02094433, n02096585, n02097298, n02098286, n02099601, n02099712, n02102318, n02106030, n02106166, n02106550, n02106662, n02108089, n02108915, n02109525, n02110185, n02110341,

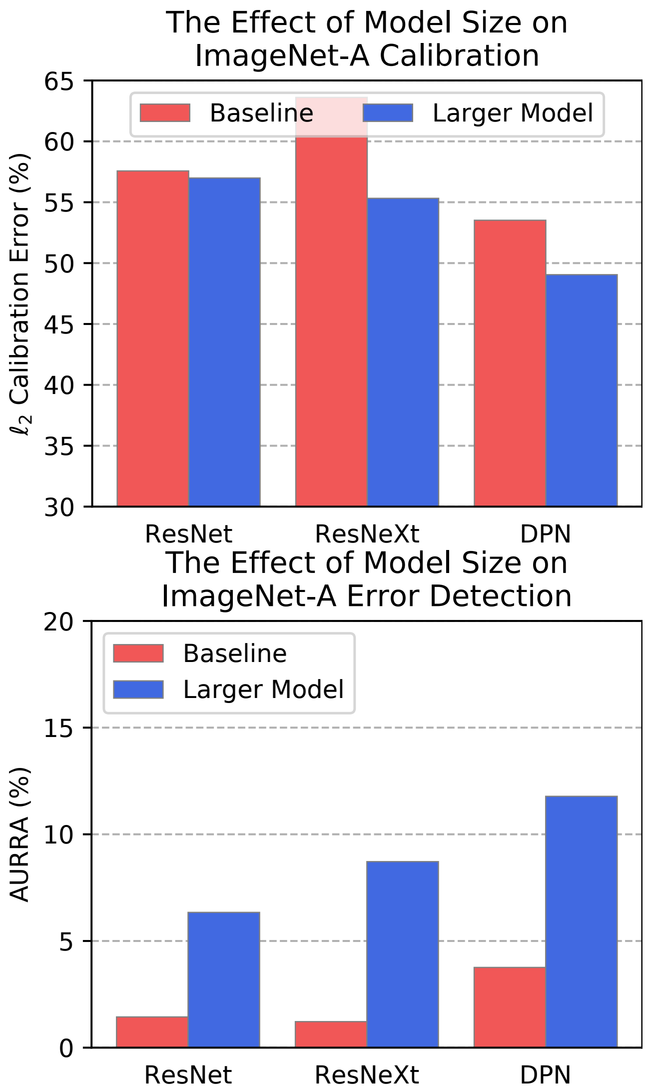

Figure 4: Model size's influence on IMAGENET-A ℓ_2 calibration and error detection.

n02110958, n02112018, n02112137, n02113023, n02113624, n02113799, n02114367, n02117135, n02119022, n02123045, n02128385, n02128757, n02129165, n02129604, n02130308, n02134084, n02138441, n02165456, n02190166, n02206856, n02219486, n02226429, n02233338, n02236044, n02268443, n02279972, n02317335, n02325366, n02346627, n02356798, n02363005, n02364673, n02391049, n02395406, n02398521, n02410509, n02423022, n02437616, n02445715, n02447366, n02480495, n02480855, n02481823, n02483362, n02486410, n02510455, n02526121, n02607072, n02655020, n02672831, n02701002, n02749479, n02769748, n02793495, n02797295, n02802426, n02808440, n02814860, n02823750, n02841315, n02843684, n02883205, n02906734, n02909870, n02939185, n02948072, n02950826, n02951358, n02966193, n02980441, n02992529, n03124170, n03272010, n03345487, n03372029, n03424325, n03452741, n03467068, n03481172, n03494278, n03495258, n03498962, n03594945, n03602883, n03630383, n03649909, n03676483, n03710193, n03773504, n03775071, n03888257, n03930630, n03947888, n04086273, n04118538, n04133789, n04141076, n04146614, n04147183, n04192698, n04254680, n04266014, n04275548, n04310018, n04325704, n04347754, n04389033, n04409515, n04465501, n04487394, n04522168, n04536866, n04552348, n04591713, n07614500, n07693725, n07695742, n07697313, n07697537, n07714571, n07714990, n07718472, n07720875, n07734744, n07742313, n07745940, n07749582, n07753275, n07753592, n07768694, n07873807, n07880968, n07920052, n09472597, n09835506, n10565667, n12267677,

'Stingray;' 'goldfinch, Carduelis carduelis;' 'junco, snowbird;' 'robin, American robin, Turdus migratorius;' 'jay;' 'bald eagle, American eagle, Haliaeetus leucocephalus;' 'vulture;' 'eft;' 'bullfrog, Rana catesbeiana;' 'box turtle, box tortoise;' 'common iguana, iguana, Iguana iguana;' 'agama;' 'African chameleon, Chamaeleo chamaeleon;' 'American alligator, Alligator mississipiensis;' 'garter snake, grass snake;' 'harvestman, daddy longlegs, Phalangium opilio;' 'scorpion;' 'tarantula;' 'centipede;' 'sulphur-crested cockatoo, Kakatoe galerita, Cacatua galerita;' 'lorikeet;' 'hummingbird;' 'toucan;' 'drake;' 'goose;' 'koala, koala bear, kangaroo bear, native bear, Phascolarctos cinereus;' 'jellyfish;' 'sea anemone, anemone;' 'flatworm, platyhelminth;' 'snail;' 'crayfish, crawfish, crawdad, crawdaddy;' 'hermit crab;' 'flamingo;' 'American egret, great white heron, Egretta albus;' 'oystercatcher, oyster catcher;' 'pelican;' 'sea lion;' 'Chihuahua;' 'golden retriever;' 'Rottweiler;' 'German shepherd, German shepherd dog, German police dog, alsatian;' 'pug, pug-dog;' 'red fox, Vulpes vulpes;' 'Persian cat;' 'lynx, catamount;' 'lion, king of beasts, Panthera leo;' 'American black bear, black bear, Ursus americanus, Euarctos americanus;' 'mongoose;' 'ladybug, ladybeetle, lady beetle, ladybird, ladybird beetle;' 'rhinoceros beetle;' 'weevil;' 'fly;' 'bee;' 'ant, emmet, pismire;' 'grasshopper, hopper;' 'walking stick, walkingstick, stick insect;' 'cockroach, roach;' 'mantis, mantid;' 'leafhopper;' 'dragonfly, darning needle, devil's darning needle, sewing needle, snake feeder, snake doctor, mosquito hawk, skeeter hawk;' 'monarch, monarch butterfly, milkweed butterfly, Danaus plexippus;' 'cabbage butterfly;' 'lycaenid, lycaenid butterfly;' 'starfish, sea star;' 'wood rabbit, cottontail, cottontail rabbit;' 'porcupine, hedgehog;' 'fox squirrel, eastern fox squirrel, Sciurus niger;' 'marmot;' 'bison;' 'skunk, polecat, wood pussy;' 'armadillo;' 'baboon;' 'capuchin, ringtail, Cebus capucinus;' 'African elephant, Loxodonta africana;' 'puffer, pufferfish, blowfish, globefish;' 'academic gown, academic robe, judge's robe;' 'accordion, piano accordion, squeeze box;' 'acoustic guitar;' 'airliner;' 'ambulance;' 'apron;' 'balance beam, beam;' 'balloon;' 'banjo;' 'barn;' 'barrow, garden cart, lawn cart, wheelbarrow;' 'basketball;' 'beacon, lighthouse, beacon light, pharos;' 'beaker;' 'bikini, two-piece;' 'bow;' 'bow tie, bow-tie, bowtie;' 'breastplate, aegis, egis;' 'broom;' 'candle, taper, wax light;' 'canoe;' 'castle;' 'cello, violoncello;' 'chain;' 'chest;' 'Christmas stocking;' 'cowboy boot;' 'cradle;' 'dial telephone, dial phone;' 'digital clock;' 'doormat, welcome mat;' 'drumstick;' 'dumbbell;' 'envelope;' 'feather boa, boa;' 'flagpole, flagstaff;' 'forklift;' 'fountain;' 'garbage truck, dustcart;' 'goblet;' 'go-kart;' 'golfcart, golf cart;' 'grand piano, grand;' 'hand blower, blow dryer, blow drier, hair dryer, hair drier;' 'iron, smoothing iron;' 'jack-o'-lantern;' 'jeep, landrover;' 'kimono;' 'lighter, light, igniter, ignitor;' 'limousine, limo;' 'manhole cover;' 'maraca;' 'marimba, xylophone;' 'mask;' 'mitten;' 'mosque;' 'nail;' 'obelisk;' 'ocarina, sweet potato;' 'organ, pipe organ;' 'parachute, chute;' 'parking meter;' 'piggy bank, penny bank;' 'pool table, billiard table, snooker table;' 'puck, hockey puck;' 'quill, quill pen;' 'racket, racquet;' 'reel;' 'revolver, six-gun, six-shooter;' 'rocking chair, rocker;' 'rugby ball;' 'saltshaker, salt shaker;' 'sandal;' 'sax, saxophone;' 'school bus;' 'schooner;' 'sewing machine;' 'shovel;' 'sleeping bag;' 'snowmobile;' 'snowplow, snowplough;' 'soap dispenser;' 'spatula;' 'spider web, spider's web;' 'steam locomotive;' 'stethoscope;' 'studio couch, day bed;' 'submarine, pigboat, sub, U-boat;' 'sundial;' 'suspension bridge;' 'syringe;' 'tank, army tank, armored combat vehicle, armoured combat vehicle;' 'teddy, teddy bear;' 'toaster;' 'torch;' 'tricycle, trike, velocipede;' 'umbrella;' 'unicycle, monocycle;' 'viaduct;' 'volleyball;' 'washer, automatic washer, washing machine;' 'water tower;' 'wine bottle;' 'wreck;' 'guacamole;' 'pretzel;' 'cheeseburger;' 'hotdog, hot dog, red hot;' 'broccoli;' 'cucumber, cuke;' 'bell pepper;' 'mushroom;' 'lemon;' 'banana;' 'custard apple;' 'pomegranate;' 'carbonara;' 'bubble;' 'cliff, drop, drop-off;' 'volcano;' 'ballplayer, baseball player;' 'rapeseed;' 'yellow lady's slipper, yellow lady-slipper, Cypripedium calceolus, Cypripedium parviflorum;' 'corn;' 'acorn.'

Their WordNet IDs are as follows.

n01498041, n01531178, n01534433, n01558993, n01580077, n01614925, n01616318, n01631663, n01641577, n01669191, n01677366, n01687978, n01694178, n01698640, n01735189, n01770081, n01770393, n01774750, n01784675, n01819313, n01820546, n01833805, n01843383, n01847000, n01855672, n01882714, n01910747, n01914609, n01924916, n01944390, n01985128, n01986214, n02007558, n02009912, n02037110, n02051845, n02077923, n02085620, n02099601, n02106550, n02106662, n02110958, n02119022, n02123394, n02127052, n02129165, n02133161, n02137549, n02165456, n02174001, n02177972, n02190166, n02206856, n02219486, n02226429, n02231487, n02233338, n02236044, n02259212, n02268443, n02279972, n02280649, n02281787, n02317335, n02325366, n02346627, n02356798, n02361337, n02410509, n02445715, n02454379, n02486410, n02492035, n02504458, n02655020, n02669723, n02672831, n02676566, n02690373, n02701002, n02730930, n02777292, n02782093, n02787622, n02793495, n02797295, n02802426, n02814860, n02815834, n02837789, n02879718, n02883205, n02895154, n02906734, n02948072, n02951358, n02980441, n02992211, n02999410, n03014705, n03026506, n03124043, n03125729, n03187595, n03196217, n03223299, n03250847, n03255030, n03291819, n03325584, n03355925, n03384352, n03388043, n03417042, n03443371, n03444034, n03445924, n03452741, n03483316, n03584829, n03590841, n03594945, n03617480, n03666591, n03670208, n03717622, n03720891, n03721384, n03724870, n03775071, n03788195, n03804744, n03837869, n03840681, n03854065, n03888257, n03891332, n03935335, n03982430, n04019541, n04033901, n04039381, n04067472, n04086273, n04099969, n04118538, n04131690, n04133789, n04141076, n04146614, n04147183, n04179913, n04208210, n04235860, n04252077, n04252225, n04254120, n04270147, n04275548, n04310018, n04317175, n04344873, n04347754, n04355338, n04366367, n04376876, n04389033, n04399382, n04442312, n04456115, n04482393, n04507155, n04509417, n04532670, n04540053, n04554684, n04562935, n04591713, n04606251, n07583066, n07695742, n07697313, n07697537, n07714990, n07718472, n07720875, n07734744, n07749582, n07753592, n07760859, n07768694, n07831146, n09229709, n09246464, n09472597, n09835506, n11879895, n12057211, n12144580, n12267677.

4. IMAGENET-O Classes

The 200 ImageNet classes that we selected for IMAGENET-O are as follows.

'goldfish, Carassius auratus;' 'triceratops;' 'harvestman, daddy longlegs, Phalangium opilio;' 'centipede;' 'sulphur-crested cockatoo, Kakatoe galerita, Cacatua galerita;' 'lorikeet;' 'jellyfish;' 'brain coral;' 'chambered nautilus, pearly nautilus, nautilus;' 'dugong, Dugong dugon;' 'starfish, sea star;' 'sea urchin;' 'hog, pig, grunter, squealer, Sus scrofa;' 'armadillo;' 'rock beauty, Holocanthus tricolor;' 'puffer, pufferfish, blowfish, globefish;' 'abacus;' 'accordion, piano accordion, squeeze box;' 'apron;' 'balance beam, beam;' 'ballpoint, ballpoint pen, ballpen, Biro;' 'Band Aid;' 'banjo;' 'barbershop;' 'bath towel;' 'bearskin, busby, shako;' 'binoculars, field glasses, opera glasses;' 'bolo tie, bolo, bola tie, bola;' 'bottlecap;' 'brassiere, bra, bandeau;' 'broom;' 'buckle;' 'bulletproof vest;' 'candle, taper, wax light;' 'car mirror;' 'chainlink fence;' 'chain saw, chainsaw;' 'chime, bell, gong;' 'Christmas stocking;' 'cinema, movie theater, movie theatre, movie house, picture palace;' 'combination lock;' 'corkscrew, bottle screw;' 'crane;' 'croquet ball;' 'dam, dike, dyke;' 'digital clock;' 'dishrag, dishcloth;' 'dogsled, dog sled, dog sleigh;' 'doormat, welcome mat;' 'drilling platform, offshore rig;' 'electric fan, blower;' 'envelope;' 'espresso maker;' 'face powder;' 'feather boa, boa;' 'fireboat;' 'fire screen, fireguard;' 'flute, transverse flute;' 'folding chair;' 'fountain;' 'fountain pen;' 'frying pan, frypan, skillet;' 'golf ball;' 'greenhouse, nursery, glasshouse;' 'guillotine;' 'hamper;' 'hand blower, blow dryer, blow drier, hair dryer, hair drier;' 'harmonica, mouth organ, harp, mouth harp;' 'honeycomb;' 'hourglass;' 'iron, smoothing iron;' 'jack-o'-lantern;' 'jigsaw puzzle;' 'joystick;' 'lawn mower, mower;' 'library;' 'lighter, light, igniter, ignitor;' 'lipstick, lip rouge;' 'loupe, jeweler's loupe;' 'magnetic compass;' 'manhole cover;' 'maraca;' 'marimba, xylophone;' 'mask;' 'matchstick;' 'maypole;' 'maze, labyrinth;' 'medicine chest, medicine cabinet;' 'mortar;' 'mosquito net;' 'mousetrap;' 'nail;' 'neck brace;' 'necklace;' 'nipple;' 'ocarina, sweet potato;' 'oil filter;' 'organ, pipe organ;' 'oscilloscope, scope, cathode-ray oscilloscope, CRO;' 'oxygen mask;' 'paddlewheel, paddle wheel;' 'panpipe, pandean pipe, syrinx;' 'park bench;' 'pencil sharpener;' 'Petri dish;' 'pick, plectrum, plectron;' 'picket fence, paling;' 'pill bottle;' 'ping-pong ball;' 'pinwheel;' 'plate rack;' 'plunger, plumber's helper;' 'pool table, billiard table, snooker table;' 'pot, flowerpot;' 'power drill;' 'prayer rug, prayer mat;' 'prison, prison house;' 'punching bag, punch bag, punching ball, punchball;' 'quill, quill pen;' 'radiator;' 'reel;' 'remote control, remote;' 'rubber eraser, rubber, pencil eraser;' 'rule, ruler;' 'safe;' 'safety pin;' 'saltshaker, salt shaker;' 'scale, weighing machine;' 'screw;' 'screwdriver;' 'shoji;' 'shopping cart;' 'shower cap;' 'shower curtain;' 'ski;' 'sleeping bag;' 'slot, one-armed bandit;' 'snowmobile;' 'soap dispenser;' 'solar dish, solar collector, solar furnace;' 'space heater;' 'spatula;' 'spider web, spider's web;' 'stove;' 'strainer;' 'stretcher;' 'submarine, pigboat, sub, U-boat;' 'swimming trunks, bathing trunks;' 'swing;' 'switch, electric switch, electrical switch;' 'syringe;' 'tennis ball;' 'thatch, thatched roof;' 'theater curtain, theatre curtain;' 'thimble;' 'throne;' 'tile roof;' 'toaster;' 'tricycle, trike, velocipede;' 'turnstile;' 'umbrella;' 'vending machine;' 'waffle iron;' 'washer, automatic washer, washing machine;' 'water bottle;' 'water tower;' 'whistle;' 'Windsor tie;' 'wooden spoon;' 'wool, woolen, woollen;' 'crossword puzzle, crossword;' 'traffic light, traffic signal, stoplight;' 'ice lolly, lolly, lollipop, popsicle;' 'bagel, beigel;' 'pretzel;' 'hotdog, hot dog, red hot;' 'mashed potato;' 'broccoli;' 'cauliflower;' 'zucchini, courgette;' 'acorn squash;' 'cucumber, cuke;' 'bell pepper;' 'Granny Smith;' 'strawberry;' 'orange;' 'lemon;' 'pineapple, ananas;' 'banana;' 'jackfruit, jak, jack;' 'pomegranate;' 'chocolate sauce, chocolate syrup;' 'meat loaf, meatloaf;' 'pizza, pizza pie;' 'burrito;' 'bubble;' 'volcano;' 'corn;' 'acorn;' 'hen-of-the-woods, hen of the woods, Polyporus frondosus, Grifola frondosa.'

Their WordNet IDs are as follows.

n01443537, n01704323, n01770081, n01784675, n01819313, n01820546, n01910747, n01917289, n01968897, n02074367, n02317335, n02319095, n02395406, n02454379, n02606052, n02655020, n02666196, n02672831, n02730930, n02777292, n02783161, n02786058, n02787622, n02791270, n02808304, n02817516, n02841315, n02865351, n02877765, n02892767, n02906734, n02910353, n02916936, n02948072, n02965783, n03000134, n03000684, n03017168, n03026506, n03032252, n03075370, n03109150, n03126707, n03134739, n03160309, n03196217, n03207743, n03218198, n03223299, n03240683, n03271574, n03291819, n03297495, n03314780, n03325584, n03344393, n03347037, n03372029, n03376595, n03388043, n03388183, n03400231, n03445777, n03457902, n03467068, n03482405, n03483316, n03494278, n03530642, n03544143, n03584829, n03590841, n03598930, n03602883, n03649909, n03661043, n03666591, n03676483, n03692522, n03706229, n03717622, n03720891, n03721384, n03724870, n03729826, n03733131, n03733281, n03742115, n03786901, n03788365, n03794056, n03804744, n03814639, n03814906, n03825788, n03840681, n03843555, n03854065, n03857828, n03868863, n03874293, n03884397, n03891251, n03908714, n03920288, n03929660, n03930313, n03937543, n03942813, n03944341, n03961711, n03970156, n03982430, n03991062, n03995372, n03998194, n04005630, n04023962, n04033901, n04040759, n04067472, n04074963, n04116512, n04118776, n04125021, n04127249, n04131690, n04141975, n04153751, n04154565, n04201297, n04204347, n04209133, n04209239, n04228054, n04235860, n04243546, n04252077, n04254120, n04258138, n04265275, n04270147, n04275548, n04330267, n04332243, n04336792, n04347754, n04371430, n04371774, n04372370, n04376876, n04409515, n04417672, n04418357, n04423845, n04429376, n04435653, n04442312, n04482393, n04501370, n04507155, n04525305, n04542943, n04554684, n04557648, n04562935, n04579432, n04591157, n04597913, n04599235, n06785654, n06874185, n07615774, n07693725, n07695742, n07697537, n07711569, n07714990, n07715103, n07716358, n07717410, n07718472, n07720875, n07742313, n07745940, n07747607, n07749582, n07753275, n07753592, n07754684, n07768694, n07836838, n07871810, n07873807, n07880968, n09229709, n09472597, n12144580, n12267677, n13052670.

References

[1] P. Bartlett and M. Wegkamp. Classification with a reject option using a hinge loss. *J. Mach. Learn. Res.*, 9:1823–1840, 2008.

[2] Terrance Devries and Graham W. Taylor. Learning confidence for out-of-distribution detection in neural networks. *ArXiv*, abs/1802.04865, 2018.

[3] Dan Hendrycks, Steven Basart, Mantas Mazeika, Mohammadreza Mostajabi, J. Steinhardt, and D. Song. Scaling out-of-distribution detection for real-world settings. *arXiv: 1911.11132*, 2020.

[4] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *ICLR*, 2017.

[5] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. *ICLR*, 2019.

[6] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and D. Song. Using self-supervised learning can improve model robustness and uncertainty. In *NeurIPS*, 2019.

[7] A. Kumar, P. Liang, and T. Ma. Verified uncertainty calibration. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.