

HiTAG: HIERARCHICAL IMAGE TAGGING WITH HYPERBOLIC VISION-LANGUAGE MODELING SUPPLEMENTARY MATERIALS

Anonymous authors

Paper under double-blind review

1 USAGE OF LARGE LANGUAGE MODELS

We leverage large language models to refine and adjust the hierarchical tag structure so that it better aligns with visual perception and ensures semantic consistency across different levels of the taxonomy. Based on this optimized hierarchy, we further employ a vision-language large model to automatically annotate tags for the images in the training sets of the benchmark. To guarantee annotation accuracy and reliability, all generated labels were subsequently manually inspected and corrected by the authors, ensuring that the final benchmark data maintains both structural consistency and high-quality annotations.

2 BENCHMARK DETAILS

2.1 HIERARCHY STRUCTURE

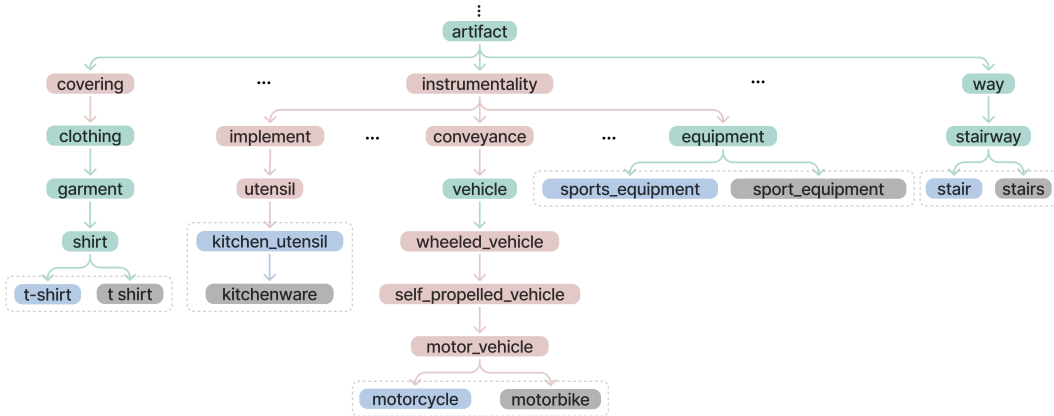


Figure 1: Step 1 of the process for constructing the hierarchical tag structure.

We show the processes of step 1 and step 2 in Figure 1 and Figure 2, respectively. In Step 1, mapping the original tag list to WordNet revealed redundant synonyms, which we removed. We then refined the list by eliminating abstract terms that are visually ambiguous. In Step 2, we identified additional common tags in WordNet, particularly under the animal branch, and integrated them into our tag tree.

We compare the coverage of general categories across hierarchical classification datasets in Table 1. SubPartImageNet contains a small number of animal and transportation categories, together with many of their subcategories. Although Mapillary Vistas 2.0 covers certain category aspects, it offers very few categories within each aspect. For example, its animal aspect includes only the categories bird and ground-animal. Our HiTag dataset encompasses a more wider range of categories.

We present the complete hierarchical tag structures of the SubPartImageNet dataset, the Mapillary Vistas 2.0 dataset and our HiTag dataset in Figure 3, Figure 4 and Figure 5, respectively. Compared to other structures, the SubPartImageNet structure is much simpler. Besides, the Mapillary Vistas 2.0

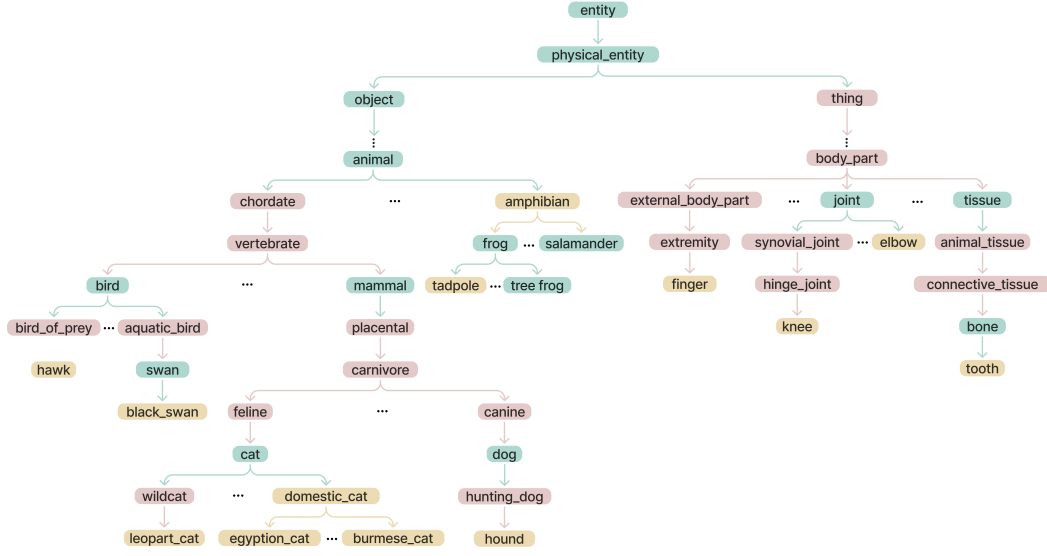


Figure 2: Step 2 of the process for constructing the hierarchical tag structure.

Table 1: Hierarchical coverage comparison of the hierarchical classification datasets.

(a) Natural object and Natural phenomenon

Dataset	Natural object							Natural phenomenon
	Plant	Animal	Person	Biological parts	Microorganism	Celestial body	Geographical feature	
SubPartImageNet	✗	✓	✗	✓	✗	✗	✗	✗
Mapillary 2.0	✗	✓	✓	✗	✓	✗	✓	✗
HiTag (ours)	✓	✓	✓	✓	✓	✓	✓	✓

(b) Event, Mythological creatures, Group, Attribute and Food

Dataset	Food	Event					Mythological creatures	Group	Attribute	Food
		Action	Competitive event	Celebratory event	Catastrophic event	Life event				
SubPartImageNet	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Mapillary 2.0	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
HiTag (ours)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

(c) Artifact

Dataset	Artifact									Communication media
	Built environment	Tool	Equipment	Furniture	Clothing	Transportation	Accessory	Decoration	Weapon	
SubPartImageNet	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
Mapillary 2.0	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
HiTag (ours)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

hierarchy exhibits extensive duplication of category names at levels 3 and 4. Compared with these structures, our architecture covers a richer set of tags and exhibits a more complete overall structure.

Since displaying the full hierarchy would require more than 60 pages, we have placed the complete hierarchical tag structure of the dataset in the anonymous repository ¹ of the provided code. It includes the name and definition of each category and is saved in two formats: hitag/-

¹Our code is available at <https://anonymous.4open.science/r/HiTag>.

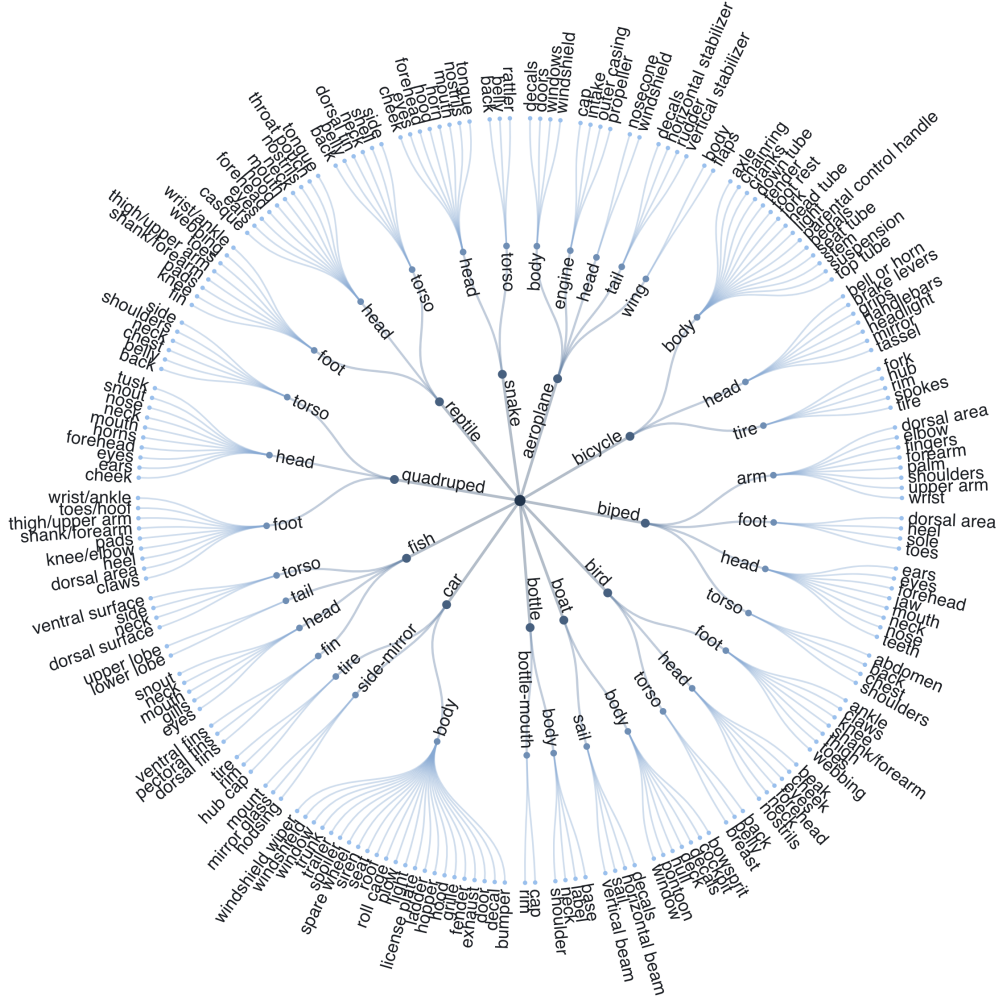
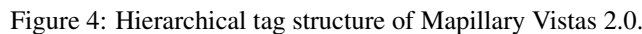


Figure 3: Hierarchical tag structure of SubPartImageNet.

data/hire_tree_file/object_structure.json and hitag/data/hire_tree_file/object_structure.txt. We show all the 3,334 tags of HiTag in 10 levels:

- Level 1 : [entity]
- Level 2 : [physical entity, natural phenomenon, mythological creatures, group, measure, attribute, event]
- Level 3 : [object, weather, snow, rain, wave, sunrise, sunset, eclipse, hot spring, flame, flood, earthquake, aurora, bolt, dew, lightning, flare, moonlight, sunshine, fire, wildfire, icicle, frost, light, shadow, darkness, sparkle, rainbow, reflect, fairy, monster, magic, ghost, queue, social group, organization, collective assembly, time, time period, score, number, numeric quantity, water level, part, scene, appearance, view, state, quality, cognition, recycling, prayer, activity, test, trade, action, competitive event, celebratory event, catastrophic event, life event, entertainment, healthcare activities, communication activities]
- Level 4 : [natural object, artifact, food, matter, hail, foggy, rainy, snowy, stormy, sunny, warm, windy, storm, snowstorm, tornado, cloudy, santa claus, cupid, angel, loki, leprechaun,



4

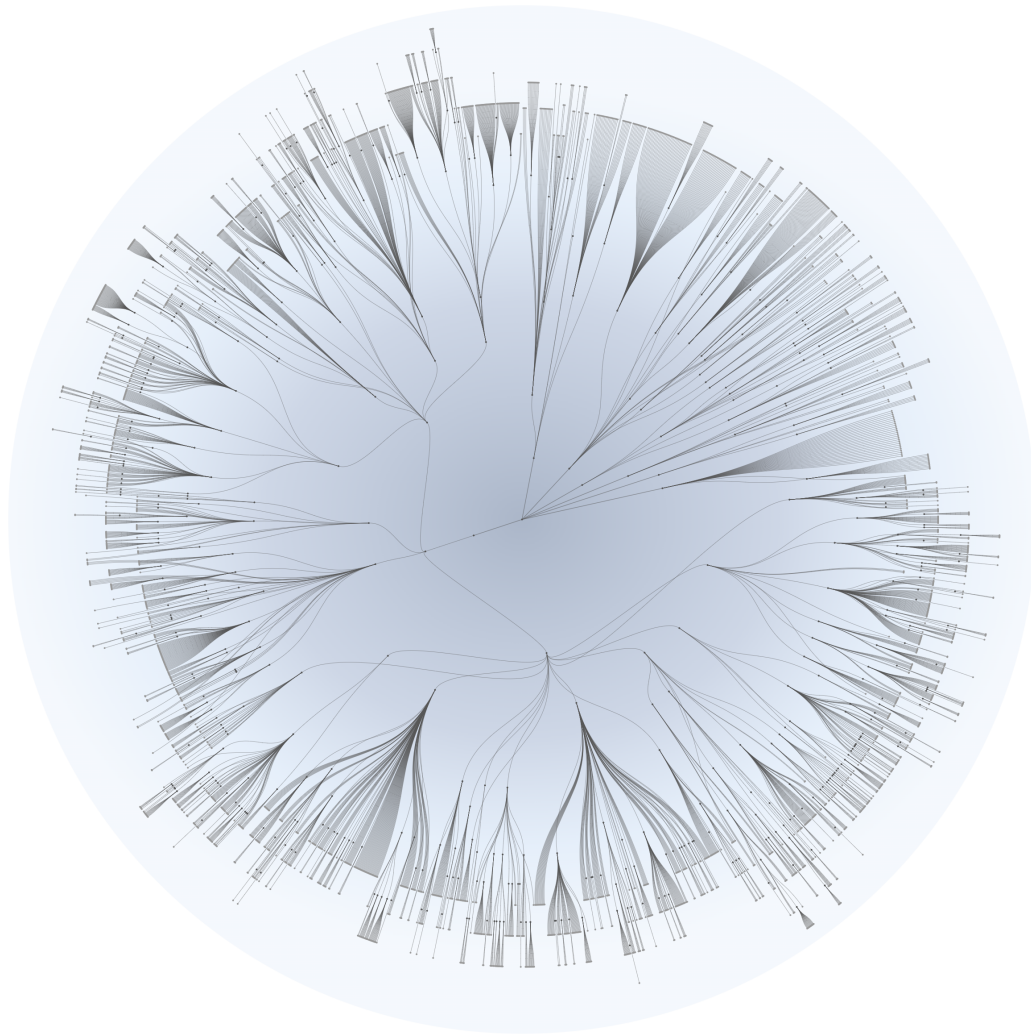


Figure 5: Visualization of the whole hierarchical tag structure which contains 3,334 tags across 10 levels.

therapy, exercise, massage, invitation, speech, interview, debate, live, call, protest, dating]

- Level 5 : [plant, animal construct, biological parts, animal, person, microorganism, celestial body, geographical feature, natural materials, atmosphere, built environment, passage, product, tool, equipment, furniture, clothing, transportation, accessories, decoration, weapon, communication media, recreational objects, money, coconut meat, chocolate, baked goods, meat, feed, drink, fast food, meal, dishes, western food, grain, dessert, sweet, condiment, dairy, egg, macaroni, spaghetti, dumpling, noodle, cereal, tofu, dough, batter, stuffing, tea bag, baby food, bento, street food, canned food, muesli, incense, amber, rust, bubble, peel, fuel, states of matter, material, youth, elderly, town, village, downtown, city, countryside, company, bank, market, school, academy, class, government, court, fire department, municipality, police, army, navy, troop, band, choir, ensemble, marching band, baseball team, basketball team, football team, ice hockey team, cricket team, art exhibition, car show, airshow, seminar, winter, summer, spring, autumn, rough, soft, smooth, modern, white, black, blue, yellow, green, pink, red, brown, purple, gray, orange color, point, round shape, figure, cylinder, curve, cube, sphere, standing, upright, industry, agriculture, shipping, jump, climb, ride, sailing, hike, chase, walk, run, flip, takeoff, surfing, taekwondo, walking, tai chi, mountain climbing, high jump, archery, swim, cycling, horse riding, ice skating, yoga, boxing, hockey, football, ping pong, rugby, bullfighting, handball, badminton, polo, karting, motorsport, rock climbing, weightlifting, mountain biking, hockey game, softball, baseball, tennis, volleyball, golf, basketball, bowling, table tennis, karate, football competition, baseball competition, tennis competition, volleyball competition, basketball competition, race, wedding party, birthday party, prom, dinner party, celebration, wedding, funeral, graduation, award ceremony, birthday, performance, juggle, light show, fashion show, talk show, stunt, karaoke, lottery, video game, billiards, card game, playing chess, playing mahjong]
- Level 6 : [vine, houseplant, algae, plant organs, vegetable, spice, tree, moss, lichen, fern, grass, cactus, bush, shrub, wheat, succulent, reed, nest, spider web, cell, animal coverings, body, muscle, wing, tail, skeleton, fin, claw, whisker, antler, penis, paw, limb stump, nipple, tooth, blood, invertebrate, amphibian, fish, reptile, bird, mammal, life stage, family relation, social role, leader, occupation, sportsman, hacker, speaker, camper, tailor, handicapped person, zombie, hiker, mold, fungi, earth, moon, sun, star, galaxy, planet, constellation, nebula, fossil, area, place, territory, surface, shoreline, harbor, rise, corner, hole, angle, habitat, wilderness, body of water, landform, world, ice cube, pebble, jade, coal, petroleum, stone, smoke, mist, fog, sky, cloud, building, infrastructure, facility, grave, estate, construction site, structure, debris, outdoor amenity, barrier, interior spaces, drain, pipe, chimney, corridor, hallway, aisle, elevator shaft, entrance, archway, goods, cargo, medicine, dye, acrylic paint, freshener, detergent, cleaner, cigar, cigarette, spray, clothespin, umbrella, hooks, tableware, container, candle, cassette deck, envelope, mp3 player, software, stamp, ink pad, paper, stationery, toiletry, flashlight, book cover, test tube, stick, personal care, charger, salt shaker, duct tape, hose, rope, tripod, easel, fishing net, crowbar, crutch, cane, ladder, connector, sharpener, brush, oar, paddle, needle, broom, pole, drill, jack, plow, rake, hoe, mower, wrench, screwdriver, saw, shovel, hammer, opener, razor, pocketknife, scissors, axe, 3d glasses, tape measure, plunger, pliers, awl, funnel, tape, ice pack, shaker, basket, bucket, pipe bowl, bin, manger, bundle, tray, ashtray, dustpan, basin, barrel, tub, jar, beaker, bottle cap, fishbowl, yoke, float, palette, ruler, pitchfork, razor blade, horseshoe, floss, rein, leash, screw, lock, clip, roller, rubber band, key, hairdryer, compass, stethoscope, syringe, thermometer, iron, magnifying glass, binoculars, handcuffs, lid, stapler, stirrup, gauge, hourglass, mallet, hanger, magnet, mousetrap, slot, thresher, copperware, match, harvester, pegboard, charcoal burner, dryer, vacuum cleaner, trimmer, concrete mixer, trouser press, appliance, musical instrument, kitchen utensil, screen, projection screen, spirit lamp, anchor, valve, pump, wiper, handle, vane, propeller, bumper, sprinkler, switch, headlight, torch, radar, lightning rod, robot, firefighting equipment, brake, laser, battery, solar panel, sports equipment, camera, kit, heater, pendulum, phone, parachute, poker chip, oil rig, life jacket, lifebuoy, plug, socket, showerhead, scale, juicer, submersible, megaphone, recording device, game controller, keyboard, computer, television, radio, calculator, atm, gear, engine, machine, loom, printer, scanner, typewriter, windmill, grinder, cotton gin, mill, slicer, lathe,

snow blower, shredder, fan, water cooler, satellite dish, cd player, amplifier, excavator, generator, record player, monitor, projector, air conditioner, fridge, sundial, microscope, lighting, telescope, remote, medical equipment, dropper, hand dryer, headphones, panel, circuit board, wire, humidifier, jukebox, electric chair, curtain, canopy, window screen, lamp, mirror, clock, watch, storage, seat, table, bedding, washstand, bathtub, toilet, fixture, textile, headdress, garment, footwear, laundry, transportation structure, vehicle, lace, hairnet, belt, broach, mask, earmuff, wristband, hairgrip, hairpin, belt buckle, collar, cigarette case, cuff, dog collar, button, zipper, mousepad, tie, glove, pocket, suspenders, knee pad, elbow pad, glasses, jewelry, mannequin, wallpaper, wind chime, picture frame, totem pole, celebratory items, brass plaque, tinsel, floral arrangement, tapestry, plume, tassel, lanyard, balloon arch, vase, disco ball, antique, ribbon, souvenir, birthday candle, gun, sword, spear, bomb, sling, dart, bullet, arrow, bow, tomahawk, brass knuckles, pike, flamethrower, cannon, missile, protective gear, net, web page, message, physical media, symbolic representation, visual identity, art, image, time based media, toy, board game, dartboard, frisbee, kite, snowball, seesaw, kaleidoscope, foosball, hula hoop, wand, music box, puzzle, domino, jigsaw puzzle, card, digital gaming, die, digital currency, currency, chocolate bar, white chocolate, cake, bread, zongzi, pie, baozi, eclair, breadstick, gingerbread house, chapati, pastry, shellfish, calamari, octopus tentacles, roe, chicken wing, escargot, sausage, ham, bacon, lamb chop, poultry, beef, pork, hay, pet food, soy milk, coffee, cocoa, tea, alcohol, juice, milk, drinking water, lemonade, milk tea, shake, breakfast, buffet, feast, picnic, caviar, roast squab, mashed potatoes, fries, bibimbap, foie gras, roast chicken, home fries, baked potato, taco, pizza, sashimi, casserole, tempura, paella, burrito, quesadilla, porridge, boiled egg, meatball, sushi, pasta, lasagna, salad, soup, fried egg, curry, fried rice, barbecue, stew, kebab, omelette, french toast, sandwich, potato chip, popcorn, fruit dish, grilled eel, onion ring, scrambled eggs, spring roll, croquette, fried chicken, hotpot, chow mein, fish sticks, meatloaf, flour, cornmeal, rice, paddy, corn cob, oats, wheat grain, corn, rye, egg tart, pudding, tiramisu, chocolate mousse, ice cream, ice cream cone, sugar cube, syrup, honey, candy, jam, jelly, bubble gum, nectar, topping, vanilla, salt, ginger, dip, vinegar, mustard, salsa, sauce, pickle, turmeric, garlic clove, sesame seed, peanut butter, sugar, cooking oil, peppercorn, butter, cream, yogurt, cheese, yolk, egg white, granola, spam, banana peel, orange peel, lemon peel, charcoal, coke, oil, ember, compound, mixture, fluid, solid, glass, plastic, rock, metal, mineral, building material, waste, fiber, dirt, wood, water, wax, insect repellent, dust, resin, latex, sponge, fur, leather, parchment, lava, foam, ash, business team, fish market, flower market, street market, fruit market, night market, vegetable market, art school, campus, university, fleet, triangle, oval, rectangle, cone, dot, rim, hook, arch, loop, dive, bike racing, marathon, slalom, formula 1, concert, circus, poker]

- Level 7 : [seaweed, stem, stump, root, leaf, fruit, flower, tree root, onion, cane sugar, pumpkin, mushroom, zucchini, yam, lettuce, broccoli, cauliflower, brussels sprouts, daikon, turnip, spinach, cucumber, squash, truffle, bean, pea, beet, potato, cabbage, sprout, tomato, eggplant, asparagus, carrot, sweet potato, bell pepper, okra, celery, mint, basil, parsley, thyme, dill, rosemary, clover, geranium, pepper, butterfly pea, lavender, aloe vera, pitcher plant, artichoke, tobacco, garlic, birch, willow, bonsai, beech tree, palm tree, oak, magnolia, mangrove, maple, holly, eucalyptus tree, fig, rowan, cashew, macadamia, hazel shrub, cypress, pine, fir, cedar, ginkgo tree, cypress tree, elm, hawthorn, banana plant, apple tree, cherry tree, albizia, orange tree, olive tree, mulberry, gliricidia, guava tree, lilac, bird nest, hair, shell, feather, skin, ivory, shoulder, belly, chest, joint, head, neck, face, waist, hand, limb, foot, buttock, brain, bone, hip, heart, worm, starfish, sea cucumber, sea urchin, insect, centipede, jellyfish, coral, scorpion, spider, snail, abalone, conch, huitre, clam, tick, isopod, octopus, crab, lobster, squid, termite, sea anemone, millipede, sea spider, horseshoe crab, slug, frog, salamander, snappers, mackerels, porgies, grouper, catfish, tuna, pufferfish, trout, shark, flatfish, swordfish, seahorse, carp, goldfish, clownfish, seabass, snake, turtle, crocodile, dinosaur, lizard, platypus, crow, parrot, swan, seabird, water bird, hummingbird, ostrich, sparrow, hornbill, turkey, chicken, woodpecker, toucan, eagle, owl, vulture, heron, stork, flamingo, swallow, starling, pigeon, osprey, falcon, bluebird, robin, canary, house finch, blackbird, jay, raven, magpie, pheasant, cuckoo, emu, guinea fowl, downy woodpecker, bantam, hawk, crane, spoonbill, roadrunner, wren, mockingbird, titmouse, hoopoe, kookaburra, marine mammal, panda, bear, raccoon, fox, coyote, wolf,

lion, serval, margay, lynx, tiger, caracal, cheetah, jaguar, jaguarundi, hyena, leopard, pet, otter, meerkat, rodent, elephant, armadillo, hedgehog, lemur, monkey, ape, pig, warthog, giraffe, deer, alpaca, donkey, sheep, bison, goat, antelope, cattle, water buffalo, yak, sloth, anteater, skunk, mink, mole, bat, kangaroo, rhinoceros, camel, hippo, koala, horse, buffalo, ferret, badger, opossum, cub, baby, child, teenager, adult, boy, girl, sibling, cousin, partner, grandfather, mother, father, daughter, son, customer, patient, client, passenger, bridesmaid, host, hero, flower girl, commuter, pedestrian, tourist, protester, player, king, queen, director, shepherd, princess, prince, president, entertainer, athlete, warrior, matador, boxer, scientist, engineer, writer, farmer, artist, craftsman, driver, astronaut, student, fisherman, airman, sailor, lifeguard, flight attendant, waiter, maid, soldier, commander, lumberjack, miner, cowboy, servant, quarry worker, coachman, policeman, volunteer, bartender, judge, fireman, guard, detective, bodyguard, businessperson, referee, gardener, street vendor, bullfighter, chef, samurai, educator, nurse, doctor, trainer, architect, butcher, rescuer, pirate, fashion model, lawyer, coach, author, secretary, nanny, worker, cook, hay worker, electrician, reporter, DJ, motorcyclist, rider, snowboarder, surfer, biker, martial artist, archer, wrestler, ground, water surface, dock, leak, sea, lake, waterfall, gulf, bay, stream, river, strait, pool, waterway, pond, fjord, creek, canyon, pasture, hayfield, swamp, hillside, plateau, volcano, snow mountain, ravine, rock arch, ice cave, islet, archipelago, forest floor, continent, oasis, peninsula, woodland, plain, flat, island, hill, mountain, ridge, cliff, forest, beach, cave, shore, wetland, ski slope, headland, valley, iceberg, glacier, ice floe, dune, reef, peak, grassland, desert, quarry, mine, field, orchard, excavation, rainforest, jungle, blue sky, evening sky, night sky, platform, castle, hospital, clinic, pyramid, auto showroom, bell tower, indoor, monument, mausoleum, tower, obelisk, ruins, skyscraper, roof, stair, home, outbuildings, religious buildings, fortification, white house, prison, hydrant, lighthouse, bridge, dam, tunnel, road, port, water tower, landfill, plaza, telegraph pole, foundation, assembly line, heliport, airfield, station, terminal, wind farm, phone box, factory, workshop, mall, supermarket, office, ticket booth, car wash, spa, ktv, bar, restaurant, casino, sports facilities, dormitory, observatory, library, hotel, resort, theater, office building, rink, cinema, kindergarten, post office, zoo, botanical garden, raceway, golf course, swimming pool, gym, football stadium, post, fort, airport, museum, archive, army base, warehouse, payphone, telephone booth, flagpole, playing field, racetrack, street light, train track, well, mailbox, exit, depository, water tank, exhaust hood, lift, faucet, farm, junkyard, pharmacy, stall, store, laboratory, workplace, studio, laundromat, winery, ball pit, sandbox, fountain, playgrounds, campgrounds, maze, park, amusement park, picnic area, garden, campfire, fire pit, backyard, lawn, swing, window, hedge, paling, wall, fence, barricade, gate, door, trench, hurdle, fender, balustrade, graveyard, parking lot, cage, fireplace, aquarium, corral, forge, balcony, ceiling, sink, cockpit, cabin, basement, loft, urinal, floor, room, shower, office cubicle, capsule, soap, shampoo, toothpaste, cutlery, ice cream scoop, plate, bowl, platter, cup, tea set, chopsticks, decanter, gravy boat, napkin, tablecloth, flower pot, cooler, box, bag, bottle, can, wrapping paper, cardboard, newspaper, graph paper, tissue, notepad, blackboard, pen, crayon, pencil, ink, eraser, binder, bookmark, folder, notebook, whiteboard, clipboard, stylus, toilet paper, handkerchief, towel, toothbrush, perfume, makeup, makeup tool, comb, hairspray, shaving cream, nail polish, mascara, eyeliner, lipstick, face powder, eye shadow, mouthwash, condom, earplug, curling iron, ironing board, sanitary pad, chain, hinge, clasp, knot, pin, scrubbing brush, hairbrush, paintbrush, stilt, ski pole, mast, fishing rod, chainsaw, handsaw, bottle opener, can opener, shaver, shopping basket, flower basket, combination lock, latch, padlock, mail slot, stringed instrument, wind instrument, keyboard instrument, drumstick, drum, maraca, plectrum, barrel organ, stove, colander, peeler, masher, rolling pin, grater, whisk, blender, pan, pot, coffeepot, kettle, grill, reamer, spatula, pizza cutter, measuring cup, ladle, pressure cooker, teapot, hotplate, oven, cutting board, rice cooker, toaster, tongs, ice maker, dishwasher, skewer, electric range, waffle iron, microwave, computer monitor, knob, door handle, fire alarm, fire hydrant, fire hose, fire extinguisher, snowboard, surfboard, yoga mat, bridle, horse blanket, saddle, saddlebag, treadmill, ski, hockey stick, basketball hoop, baseball bat, baseball glove, baseball base, racket, diving board, skate, trampoline, high bar, rings, balance beam, wicket, dumbbell, barbell, weight, javelin, mat, golf equipment, tackle, boxing glove, bodyboard, tennis net, roller skates, basketball backboard, punching bag, parasail, finish line, sideline, target, gymnastics apparatus, cricket bat, snowshoe, battle rope, ball, camera lens, shutter, first-aid kit, sewing kit, carpenter's kit, smartphone, corded phone, microphone, desktop computer, laptop, tablet, server, abacus, steam engine, washing

machine, slot machine, vending machine, sewing machine, arcade game, coffee maker, fruit machine, milking machine, stone mill, meat grinder, coffee grinder, water mill, flour mill, ceiling fan, floor fan, gramophone, swab, bandage, sunshade, chandelier, lampshade, oil lamp, floor lamp, droplight, wall lamp, table lamp, light bulb, rear-view mirror, hand mirror, alarm clock, wall clock, digital clock, pocket watch, cabinet, dresser, shelf, bookcase, tv cabinet, wine rack, coat rack, drawer, chair, couch, stool, bench, baby seat, stroller, billiard table, kitchen table, counter, desk, workbench, dining table, table tennis table, dressing table, conference table, operating table, glass table, nightstand, altar, bed, bedclothes, mattress, carpet, pillow, cradle, headboard, fabric, cloth, headband, cap, hat, helmet, crown, tiara, wig, sleepwear, sleeve, uniform, leotard, wet suit, swimwear, veil, leggings, vest, suit, protective suit, bib, pants, diaper, robe, shirt, sweater, scarf, cloak, coat, overall, sportswear, dress, underpants, cheongsam, hanfu, baby clothes, crop top, jumpsuit, underclothes, hoodie, wetsuit, blouse, waistband, hood, costume, coverall, apron, stocking, tights, socks, shoe, bus station, subway station, traffic cone, bus stop, railroad, spacecraft, vehicle structure, bus, train, stretcher, dolly, rocket, escalator, sleigh, construction vehicle, vessel, tank, unicycle, skateboard, bicycle, scooter, baby carriage, tricycle, railcar, cart, wagon, car, ambulance, snowplow, golf cart, motorcycle, truck, go-kart, tractor, forklift, ski lift, chairlift, plane, drone, helicopter, glider, balloon, airship, snowmobile, atv, self-balancing scooter, pickup truck, skibob, horse-drawn vehicle, trailer, cable car, elevator, subway, segway, eye mask, bolo tie, goggles, sunglasses, pendant, necklace, cufflinks, bracelet, anklet, pearl, jewel, earrings, ring, bead, brooch, fireworks, christmas decoration, jack-o'-lantern, snowman, chinese knot, gift, lantern, confetti, wreath, bouquet, rifle, handgun, machine gun, revolver, paintball gun, scimitar, broadsword, rapier, dagger, machete, crossbow, halberd, mortar, gas mask, extinguisher, armor, shield, blueprint, receipt, bulletin board, billboard, menu, gift card, recipe, placard, scoreboard, map, journal, sticker, poster, list, ticket, document, sheet music, handwriting, postcard, bible, fiction, poetry, letter, checkbook, passbook, greeting card, manuscript, publication, magazine, newspapers, book, record, album, cassette, dvd, videotape, fax, reel, sign, signal, icon, avatar, mark, award, design, identity card, credit card, keycard, business card, qr code, barcode, clip art, tattoo, ceramic, fictional character, photography, sculpture, dance, painting, street art, drawing, mosaic, portrait, medical image, photo, illustration, graph, picture, diagram, advertisement, news, press conference, notice, movie, auditory expression, drama, puppet, playhouse, lego, doll, teddy, toy car, toy plane, toy gun, piggy bank, mahjong, chess, game board, playing card, currency unit, currency form, chocolate cake, cupcake, pancake, tortilla, birthday cake, waffle, wedding cake, doughnut, brownie, fruitcake, mooncake, crouton, toast, bagel, croissant, pretzel, muffin, biscuit, baguette, steamed bread, bun, cinnamon roll, profiterole, eel, mussel, scallop, clam meat, lobster meat, oyster, shrimp, crabmeat, whelk, steak, haystack, liquor, wine, beer, cocktail, whisky, orange juice, chocolate milk, ice water, bottled water, broth, tomato soup, hamburger, hot dog, submarine sandwich, easter egg, candy bar, fudge, candy cane, lollipop, cotton candy, caramel, jelly bean, mint candy, whipped cream, frosting, tomato sauce, hot sauce, gravy, gel, gas, powder, crystal, foil, copper, silver, gold, steel, bronze, brass, turquoise, quartz, ruby, sapphire, emerald, gravel, granite, marble, plaster, beam, plank, tile, brick, concrete, cement, tarmac, block, garbage, yarn, wool, cotton, straw, flax, coir, raffia, silk, string, twine, sand, clay, mud, lumber, bamboo, circle, square]

- Level 8 : [trunk, branch, sugarcane, greenery, banana leaf, maple leaf, tobacco leaves, pod, acorn, cherry, peach, olive, pear, apple, seed, raisin, grape, lychee, passion fruit, apricot, citrus fruit, mango, melon, berry, avocado, pomegranate, banana, kiwi, hami melon, dragon fruit, starfruit, nectarine, papaya, prickly pear, mangosteen, plum, durian, pineapple, jackfruit, jujube, calabash, water lily, lotus, iris, daffodil, tulip, lily, marigold, daisy, poppy, anemone, dahlia, petunia, orchid, sunflower, chrysanthemum, peony, carnation, cherry blossom, chinese rose, clivia, hibiscus flower, blossom, rose, rhododendron, fuchsia, poinsettia, oleander, heather, hydrangea, hibiscus, blueweed, violet, thistle, pansy, morning glory, livingstone daisy, dandelion, jasmine, squill, coconut palm, hairstyle, mane, beard, lash, seashell, ankle, elbow, wrist, knee, eye, nose, ear, mouth, finger, arm, leg, lap, toe, barefoot, planarian, leech, pest, larva, beetle, flea, bee, wasp, ant, cricket, grasshopper, stick insect, praying mantis, dragonfly, moth, butterfly, earwig, vespa, hermit crab, king crab, tadpole, tree frog, sardine, whale shark, stingray, koi, python, iguana, gecko, macaw,

cockatoo, African grey parrot, budgie, cob, black swan, pelican, cormorant, penguin,
 gull, auk, mollymawk, goose, duck, hen, bald eagle, barn owl, egret, quail, partridge,
 peacock, ruffed grouse, seal, walrus, whale, manatee, bear cub, polar bear, brown bear,
 cougar, mountain lion, snow leopard, cat, dog, rat, squirrel, hamster, mouse, beaver, guinea
 pig, rabbit, mammoth, baby elephant, squirrel monkey, macaque, baboon, chimpanzee,
 orangutan, gorilla, piglet, fawn, roe deer, reindeer, moose, lamb, wildebeest, calf, bull, cow,
 zebra, pony, mule, man, woman, elder, twin, newlywed, wife, husband, magician, musician,
 joker, dancer, actor, juggler, snake charmer, strongman, climber, hockey player, football
 player, skater, skateboarder, runner, swimmer, skier, basketball player, baseball player,
 gymnast, tennis player, rock climber, goalkeeper, motorcycle racer, rugby player, soccer
 goalkeeper, soccer player, weightlifter, poet, journalist, beekeeper, photographer, sculptor,
 painter, conductor, makeup artist, welder, construction worker, plumber, barber, carpenter,
 blacksmith, mechanic, builder, stonecutter, truck driver, bus driver, taxi driver, racing driver,
 pilot, police officer, teacher, professor, dentist, surgeon, vet, snowfield, lakeshore, coast,
 river bank, wheat field, corn field, flower field, rice field, sunflower field, farmland, podium,
 catwalk, scaffold, deck, stage, boxing ring, gravestone, chinese tower, eiffel tower, dome,
 tile roof, manor house, log cabin, yurt, house, parking space, beehive, tree house, flower bed,
 shelter, hangar, courtyard, terrace, pavilion, stable, chicken coop, barn, garage, outhouse,
 conservatory, shed, dog house, shrine, monastery, church tower, chapel, mosque, temple,
 church, pulpit, suspension bridge, tower bridge, asphalt road, promenade, sidewalk, curb,
 street, highway, driveway, railway line, crosswalk, intersection, canal, trail, alley, bike
 lane, path, crossroad, police station, power station, fire station, gas station, broadcasting
 station, bowling alley, stadium, bullring, sports field, airport runway, fire escape, silo,
 treasury, storehouse, repair shop, salon, newsstand, pet store, flower shop, bakery, chemistry
 lab, music studio, recording studio, art studio, slide, amusement ride, water park, display
 window, skylight, floor window, city wall, glass wall, climbing wall, seawall, doorplate,
 screen door, car door, elevator door, glass door, glass floor, tile flooring, wood floor, suite,
 hotel room, storage room, bedroom, dining room, classroom, hospital room, jail cell,
 hall, recreation room, gallery, bathroom, engine room, dressing room, kitchen, restroom,
 sauna, locker, computer room, meeting room, dance studio, hotel lobby, laundry room,
 toilet bowl, fork, spoon, knife, paper plate, glass plate, glass bowl, drinking glass, teacup,
 chalice, paper cup, mug, shipping container, carton, pencil case, crate, safe, toolbox, gift
 box, computer case, briefcase, luggage, shopping bag, backpack, tote bag, sleeping bag,
 sack, duffel, handbag, shoulder bag, clutch bag, gift bag, paper bag, jug, flask, thermos,
 gourd, glass bottle, baby bottle, watering can, spray can, fountain pen, sharpie, board eraser,
 makeup brush, makeup mirror, hair roller, thumbtack, zither, harp, guitar, violin, cello, erhu,
 recorder, sax, panpipes, flute, bugle, trombone, tuba, whistle, kazoo, trumpet, mouth organ,
 musical keyboard, piano, harpsichord, accordion, cymbal, snare drum, tambourine, gas
 stove, wood-burning stove, frying pan, wok, roaster, saucepot, dutch oven, saucepan, water
 ski, badminton racket, tennis racket, ping pong paddle, roller skate, golf ball, golf club,
 rugby ball, cricket ball, bowling ball, basketball ball, soccer ball, baseball ball, tennis ball,
 volleyball ball, billiard ball, adhesive bandage, bookshelf, supermarket shelf, folding chair,
 rocking chair, swivel chair, feeding chair, wheelchair, throne, barber chair, office chair,
 sofa, daybed, footrest, folding stool, bar stool, hassock, pew, window seat, park bench,
 training bench, bunk bed, berth, hammock, hospital bed, dog bed, double bed, single bed,
 waterbed, canopy bed, crib, quilt, blanket, bedcover, sheet, throw pillow, patch, shower
 cap, beret, skullcap, swim cap, visor, straw hat, dress hat, witch hat, sun hat, cowboy hat,
 fedora, baseball cap, party hat, top hat, christmas hat, football helmet, bicycle helmet, safety
 helmet, racing helmet, military uniform, school uniform, bikini, safety vest, jeans, shorts,
 sweatpants, kimono, bathrobe, dress shirt, polo shirt, t-shirt, sweatshirt, jersey, cardigan,
 turtleneck, headscarf, poncho, shawl, jacket, raincoat, overcoat, lab coat, fur coat, sport coat,
 trench coat, skirt, sari, gown, kilt, lace dress, evening dress, maxi dress, bandeau, corset,
 cosplay, halloween costume, boot, sandal, running shoe, high heel, skiing shoes, leather
 shoe, slipper, flip-flop, satellite, space station, space shuttle, windshield, taillight, wheel, tire,
 steering wheel, seat belt, dashboard, car seat, car trunk, school bus, double-decker bus, train
 car, locomotive, steam train, raft, yacht, boat, sailboat, passenger ship, warship, pirate ship,
 cargo ship, mountain bike, exercise bike, caboose, freight car, passenger car, shopping cart,
 wheelbarrow, hand truck, sports car, taxi, racing car, limo, convertible, jeep, sedan, police
 car, lorry, tow truck, fire truck, garbage truck, food truck, transporter, police van, bulldozer,

biplane, jet, airliner, seaplane, cargo aircraft, bomber, hot air balloon, elevator car, diamond, glass bead, sparkler, firecracker, christmas light, christmas tree, christmas ball, paper lantern, halloween pumpkin, chinese lantern, bulletproof vest, chainmail, checklist, movie ticket, certificate, passport, license, contract, bill, work card, comic book, paperback book, hardback book, atlas, textbook, dictionary, file, register, note, cd, symbol, alarm, armband, air sock, brake light, traffic light, buoy, milestone, label, earmark, watermark, medal, trophy, architecture, pattern, emblem, flag, cross, logo, license plate, letter logo, badge, pottery, porcelain, cartoon character, filming, face close-up, figurine, bronze sculpture, ice sculpture, sand sculpture, statue, ballet, folk dance, mural, oil painting, watercolor painting, portrait painting, monochrome, doodle, sketch, manga, id photo, wedding photo, family photo, selfie, cartoon illustration, fashion illustration, cartoon, documentary, horror film, science fiction film, animated film, music, audio, laugh, conversation, argument, opera, comedy, play, musical, barbie, water gun, chessboard, dollar, pound sterling, rmb, coin, paper money, check, champagne, sake, steam, ice, snowflake, amethyst, citrine]

- Level 9 : [nut, coffee bean, pumpkin seed, sunflower seed, orange, lime, grapefruit, lemon, watermelon, winter melon, muskmelon, cantaloupe, persimmon, strawberry, blueberry, blackberry, bayberry, raspberry, cranberry, bud, bald, hair color, blonde hair, brown hair, gray hair, red hair, white hair, black hair, curly hair, purple hair, blue hair, green hair, orange hair, pink hair, yellow hair, straight hair, braided hair, ponytail, long hair, short hair, braid, eyebrow, eyelash, tear, beak, fly, mosquito, roach, locust, caterpillar, maggot, mealworm, nymph, ladybird, duckling, wood duck, sea lion, dolphin, killer whale, humpback whale, egyptian cat, kitten, american shorthair, persian cat, british shorthair, burmese cat, ocelot, domestic cat, siamese, abyssinian, leopard cat, puppy, german shepherd, labrador, poodle, chihuahua, dachshund, rottweiler, doberman pinscher, husky, beagle, boxer dog, bulldog, german shorthair pointer, great dane, afghan hound, schnauzer, rhodesian ridgeback, sheepdog, gordon setter, setter, shih-tzu, weimaraner, tibetan mastiff, maltese, basset, malamute, dalmatian, samoyed, ibizan hound, shiba inu, golden retriever, chow chow, bichon frise, corgi, foxhound, hound, saint bernard, bernese mountain dog, pit bull, bride, bassist, guitarist, rocker, violinist, pianist, saxophonist, violist, singer, drummer, ballet dancer, child actor, ice hockey player, diver, catcher, backdrop, apartment, cottage, farmhouse, villa, palace, tent, hut, igloo, zebra crossing, television studio, football field, basketball court, volleyball court, tennis court, granary, ammunition storage, carousel, ferris wheel, pantry, vault, wine cellar, closet, operating room, concert hall, banquet hall, conference hall, auditorium, art gallery, soup spoon, champagne flute, guzheng, denim jacket, leather jacket, miniskirt, ballet skirt, pencil skirt, wedding dress, fishing boat, tugboat, motorboat, ferry, canoe, rowboat, lifeboat, inflatable boat, swan boat, sail, cruise ship, aircraft carrier, cruiser, battleship, submarine, container ship, fighter jet, diploma, traffic sign, stop sign, speed limit sign, parking sign, road sign, warning sign, welcome sign, written symbol, bronze medal, gold medal, silver medal, stripe, banner, pirate flag, national flag, white flag, red flag, pop, jazz, disco]

- Level 10 : [pine cone, peanut, macadamia nut, hazelnut, coconut, chestnut, walnut, rapper, lead singer, ballerina, font, alphabet, check mark, dollar sign, character, percent sign, tricolor, german flag, south korean flag, italian flag, japanese flag, canada flag, brazilian flag, chinese flag, indian flag, mexican flag, australian flag, russian flag, south african flag, argentine flag, union jack, american flag]

2.2 METRICS

We employed four metrics to evaluate the predicted hierarchical structures.

- **Tree Edit Distance** *TED*: Measures the minimum number of edit operations (insertion, deletion, etc.) required to transform the predicted tag tree into the ground-truth tag tree.

- **Jaccard Similarity** J : Treats the set of nodes in the predicted and ground-truth trees as two separate sets, and computes the ratio of their intersection to their union:

$$J = \frac{|\text{Predicted Nodes} \cap \text{Ground-truth Nodes}|}{|\text{Predicted Nodes} \cup \text{Ground-truth Nodes}|}. \quad (1)$$

- **Hierarchical Precision** P_H : Represents the proportion of correctly predicted nodes among all predicted nodes:

$$P_H = \frac{|\text{Predicted Nodes} \cap \text{Ground-truth Nodes}|}{|\text{Predicted Nodes}|}. \quad (2)$$

- **Hierarchical Recall** R_H : Represents the proportion of correctly predicted nodes among all ground-truth nodes:

$$R_H = \frac{|\text{Predicted Nodes} \cap \text{Ground-truth Nodes}|}{|\text{Ground-truth Nodes}|}. \quad (3)$$

2.3 DESCRIPTION GENERATION

To enrich each tag’s semantics, we provide **6** concise visual descriptions:

- **1 CLIP-style description**: the fixed template a photo of a {tag}.
- **5 LLM descriptions**: We prepend the tag’s curated definition so the model understands its intended meaning. The prompt details are as follows.

```
llm_prompts = [
    f"Describe concisely what a(n) {tag} looks like:",
    f"How can you identify a(n) {tag} concisely?",
    f"What does a(n) {tag} look like concisely?",
    f"What are the identifying characteristics of a(n) {tag}:",
    f>Please provide a concise description of the visual
      characteristics of {tag}:"
]
...
user_prompt = (
    f"Here is the intended meaning of '{tag}':
    {definition}\n\n"
    f"{llm_prompt}"
)
```

2.4 DATA ANNOTATION

We use the following prompts to annotate image with corresponding tags:

```
system_prompt :
    "You are an image classifier."# Output only a valid Python
    list of recognized category names, and nothing else."
user_prompt:
    "You need to find which categories exist in this image and
    only output a plain list of category names for the current
    image based on following rules:"
    "1. You need to find which categories exist in this image
    based on the categories and their corresponding definitions in
    the dict {cate_dict}."
    "2. Categories should strictly reflect visible content
    without interpretation or assumption. Choose the most
    appropriate categories for current image from the list and no
    more than 30 categories."
    "3. Only output a list of category names for the image
    without any extra text (e.g. explanations, reasoning). For
    example, the final output should look like:[class1, class2, ...]"
```

3 EXPERIMENT

3.1 QUALITY COMPARISON OF DIFFERENT IMAGE TAGGING METHODS

In Figure 6, we present a comparison between our method HiTag and other image tagging approaches. The results show that HiTag can recognize a richer set of hierarchical tags.



HiTag (ours): level 0: entity; level 1: physical entity | natural phenomenon | measure | attribute; level 2: object | snow | organization | collective assembly | time | appearance | view; level 3: natural object | artifact | matter | business group | night view; level 4: plant | person | atmosphere | built environment | tool | equipment | decoration | communication media | states of matter | market | winter | orange color; level 5: tree | occupation | sky | cloud | building | infrastructure | facility | outdoor amenity | barrier | container | canopy | lamp | storage | celebratory items | visual identity | street market; level 6: blue sky | evening sky | home | outbuildings | road | plaza | stall | christmas decoration | crystal; level 7: sidewalk | street | display window

RAM++: blanket | building | christmas market | city | floor | person | market | market square | mill | plaza | snow | stall | town | town square | walk | winter

RAM: Christmas market | Christmas tree | stall | market square | snow | people | stroll | town | building



HiTag (ours): level 0: entity; level 1: physical entity | natural phenomenon | measure | attribute; level 2: object | time | appearance | view | state; level 3: natural object | artifact | matter | season | color | horizon | outdoor | rural | calm; level 4: plant | geographical feature | natural materials | atmosphere | built environment | material | yellow; level 5: plant organs | tree | grass | bush | shrub | surface | shoreline | body of water | landform | sky | cloud | infrastructure; level 6: stem | leaf | flower | water surface | sea | lake | hillside | snow mountain | hill | mountain | road | gravel; level 7: trunk | greenery | path

RAM++: blanket | bush | path | hill | hillside | lake | lake district | lush | mountain | mountain path | road | trail | tree | water | yellow

RAM: brush | dirt road | flower | path | hillside | lake | lead to | mountain | mountain path | road | trail | tree | water | yellow



HiTag (ours): level 0: entity; level 1: physical entity | attribute; level 2: object | view; level 3: natural object | artifact | matter | outdoor | pull; level 4: plant | biological parts | animal | tool | material; level 5: plant organs | tree | tail | whisker | paw | mammal | rope | connector | fur; level 6: leaf | hair | head | face | limb | foot | lemur | knot | twine; level 7: greenery | eye | nose | ear | mouth

RAM++: balance | rope | knot | lemur | leprechaun | rope bridge | sit | stand | tail | tree

RAM: attach | rope | knot | lemur | rope bridge | sit | stand | tail | tree



HiTag (ours): level 0: entity; level 1: physical entity | natural phenomenon | measure | attribute; level 2: object | time | appearance | view | state; level 3: natural object | artifact | matter | season | color | horizon | outdoor | rural | calm; level 4: plant | geographical feature | natural materials | atmosphere | built environment | material | yellow; level 5: plant organs | tree | grass | bush | shrub | surface | shoreline | body of water | landform | sky | cloud | infrastructure; level 6: stem | leaf | flower | water surface | sea | lake | hillside | snow mountain | hill | mountain | road | gravel; level 7: trunk | greenery | path

RAM++: blanket | bush | path | hill | hillside | lake | lake district | lush | mountain | mountain path | road | trail | tree | water | yellow

RAM: brush | dirt road | flower | path | hillside | lake | lead to | mountain | mountain path | road | trail | tree | water | yellow

Figure 6: Comparison of Tag Recognition Performance with Image Tagging Methods

3.2 ABLATION STUDY ON ENTAILMENT WEIGHT

We compared the impact of different entailment weights on performance. Based on the results of Figure 7, we conclude that 10 is an appropriate value for entailment weight.

3.3 MODEL PERFORMANCE WITH DOMAIN TRANSFER

In Table2, we present a comparison between our model and the baseline model RAM++ after domain transfer on the COCO dataset.

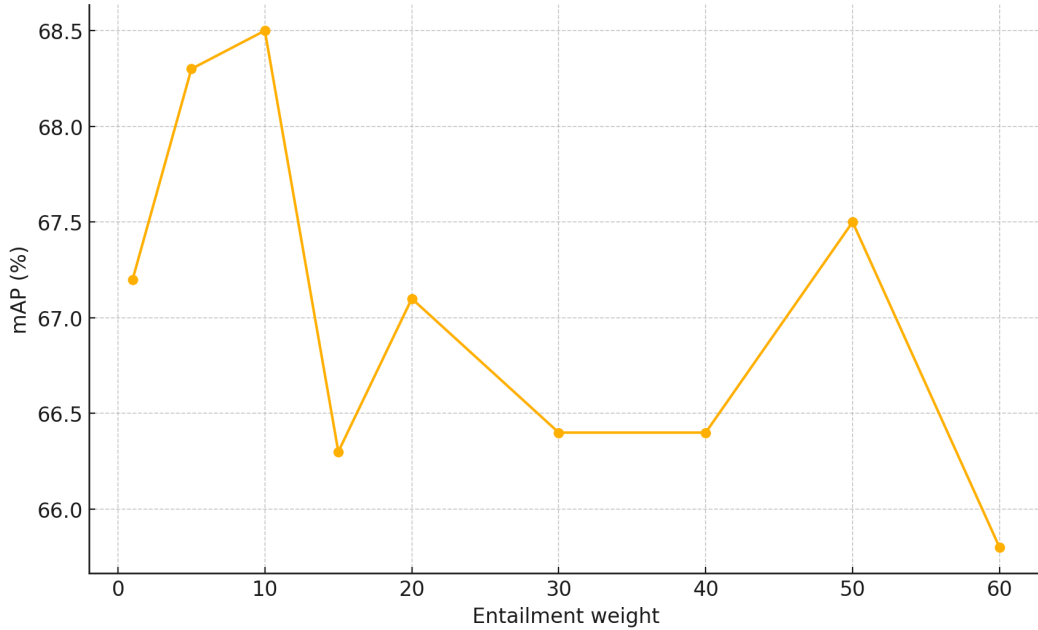


Figure 7: Performance on OpenImages-Uncommon at Different Entailment Weights

Table 2: **Comparison of Model Performance With Domain Transfer.** The following datasets are all evaluated under the open-vocabulary setting.

Methods	Training Datasets	Fine-tuning Datasets	COCO
RAM++	CC3M	COCO	75.5
HiTag (ours)	CC3M	COCO	79.5

4 METHOD DETAILS

4.1 HYPERBOLIC SPACE

Hyperbolic space’s geodesics exhibit tree-like branching, making it well-suited for hierarchical structures that expand exponentially at each level. Moreover, even low-dimensional hyperbolic spaces can simultaneously capture hierarchical relationships and node similarities, providing a natural and efficient geometric paradigm for representing and learning from hierarchical data. In practice, the Poincaré ball model and the Lorentz model are two commonly used isometric models for hyperbolic spaces. Because the Poincaré ball is bounded by the unit sphere, gradients tend to vanish near the boundary (i.e., as $r \rightarrow 1$), which affects training stability. In contrast, the Lorentz model is unbounded, and its exponential/logarithmic maps are more numerically stable, making it more suitable for deep learning.

In the Lorentz model, geodesics are the shortest paths between points. For any $\mathbf{x}, \mathbf{y} \in L^n$, the two-dimensional plane containing $\{\mathbf{x}, \mathbf{y}\}$ and the origin intersects L^n along a unique geodesic. The Lorentz distance is the arc length of this geodesic:

$$d_L(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{k}} \operatorname{arcosh}(-k \langle \mathbf{x}, \mathbf{y} \rangle_L). \quad (4)$$

4.2 IMAGE-TEXT ALIGNMENT

We define a tag-image attention to model the interaction between tags and images. Specifically, let Q denote the tag features $\mathbf{x}^{(\text{tag})}$, and let K and V represent the image features. The tag-image attention

aggregates relevant evidence vectors from the image for each category, which can be formulated as:

$$F_{i \rightarrow tag} = \text{softmax}\left(\frac{\text{Log}(Q) \text{Log}(K)^T}{\sqrt{d}}\right) \text{Log}(V), \quad (5)$$

where $F_{i \rightarrow tag}$ is the tag-image similarity feature encoding the interaction between the tag and the image, and d is a normalization constant. For computational simplicity, we map the features into Euclidean space via the $\text{Log}_o(\cdot)$ function according to Eq equation ???. Furthermore, $F_{i \rightarrow tag}$ is passed through a linear layer to obtain the score of each tag, followed by a sigmoid function to obtain the final prediction probability p_{tag} :

$$p_{tag} = \sigma(\mathbf{W} F_{i \rightarrow tag} + b), \quad (6)$$

where \mathbf{W} and b are the weight matrix and bias term of the linear layer, respectively, and $\sigma(\cdot)$ denotes the sigmoid function.

Following Ben-Baruch et al. (2020), we employ Asymmetric Loss to compute the tag-image alignment loss $\mathcal{L}_{tagalign}$. Likewise, the caption-image alignment loss $\mathcal{L}_{capalign}$ is computed in the same manner. Therefore, the overall image-text alignment loss \mathcal{L}_{align} is defined as:

$$\mathcal{L}_{align} = \mathcal{L}_{tagalign} + \mathcal{L}_{capalign} \quad (7)$$

REFERENCES

Emanuel Ben-Baruch, Tal Ridnik, Nadav Zamir, Asaf Noy, Itamar Friedman, Matan Protter, and Lihi Zelnik-Manor. Asymmetric loss for multi-label classification, 2020.