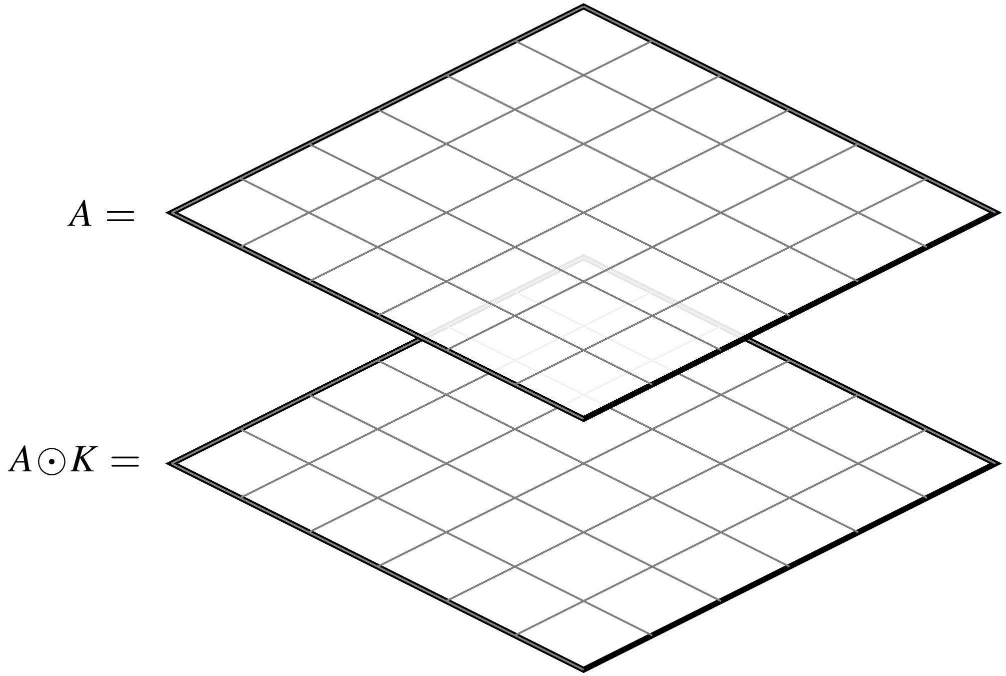
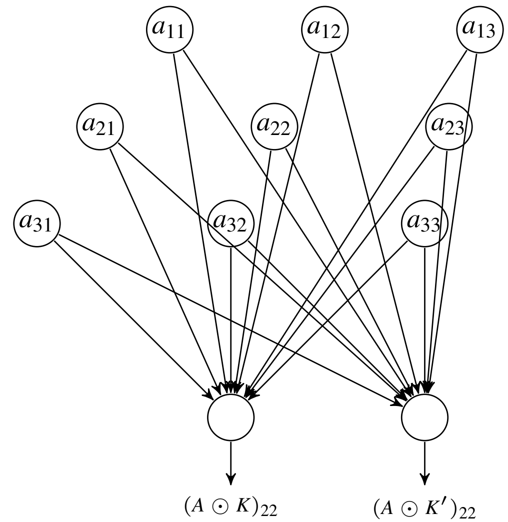


- Here, the input layer representing A and the consecutive layer representing $A \odot K$ are visualized as sheets.
- The layer that computes $A \odot K$ is often called a **convolutional layer**, although *cross-correlation layer* would be more accurate. (There is another operation called a convolution that is similar to cross-correlation, but not identical.)
- Neural networks that contain convolutional layers are called **convolutional neural networks**, even if not every layer is a convolution. Typically, the first hidden layer performs a convolution.
- Almost all networks used for image processing and computer vision problems are convolutional neural networks.

Computing Several Cross-Correlations in Parralel

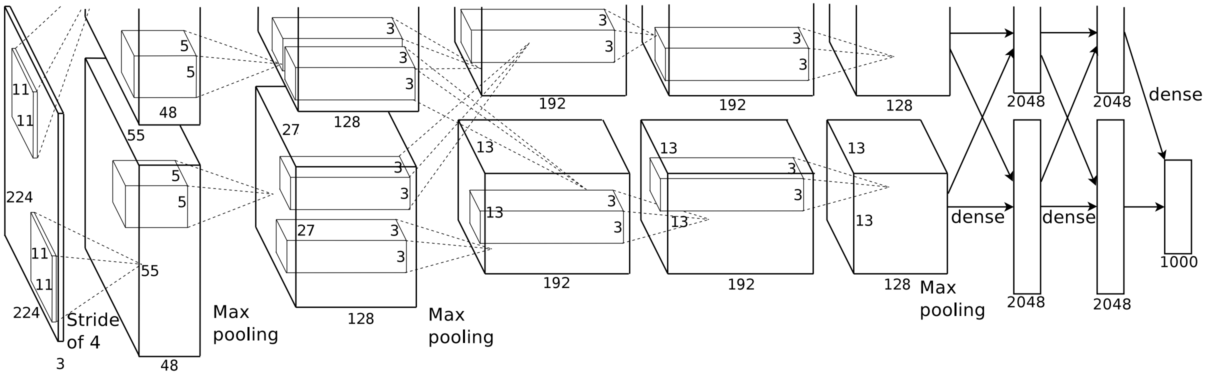


- We start with the same network as before that computes $(A \odot K)_{22}$.
- For each input vertex, we add a second connection and collect all of these in a second (linear) unit. That is, the second layer now has two units.
- The connections to the first node on the second layer still use the weights given by K. (The weights are omitted above since the figure would get too crowded.)
- Now specify a second 3×3 matrix K'. Use its entries as weights for the additional connections, collected by the second linear unit.
- The network now computes $(A \odot K)_{22}$ (as output of one unit in the second layer) and $(A \odot K')_{22}$ (as output of the other one).

OBJECT RECOGNITION TASKS

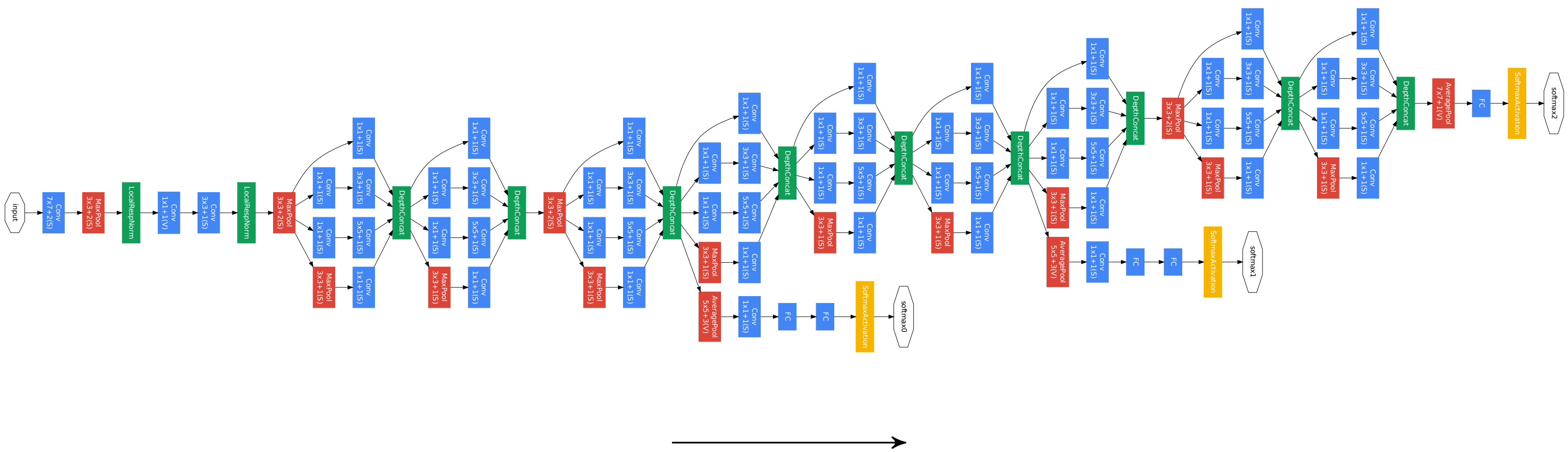
- An important benchmark problem is object recognition.
- The task is, roughly: An image is fed into a multiclass classifier, and the classifier should output the label of a/the “dominant” object in the image.
- For a picture of a car with background, the label would be “car”, possibly plus a specific type or model.
- The current state of the art for this problem are (convolutional) neural networks whose input is the entire image (i.e. there is no prior feature extraction step).
- The next two slides illustrate models that performed best in comparisons organized as a contest in 2012 and 2014.

STATE-OF-THE-ART IN 2012



- This is an illustration (taken from the research article) of the convolutional network that first demonstrated enourmous improvements in computer vision benchmark tasks.
- "Stride of 4" refers to a convolutional layer that applies 96 kernels in parallel.
- Each of the big blocks in the figure represents a convolutional layer.
- In between the convolutional layers, additional operations are performed ("pooling" and a form of normalization).
- "Pooling" refers to operations that collect outputs from a rectangular patch adjacent units and summarize them in a single unit. That reduces layer size.
- "Dense" refers to a layer that is fully connected (all possible edges from one layer to the next are present). These are located at towards the output end of the network, where layer size has already been reduced.

STATE-OF-THE-ART IN 2014



- Layers: Convolution (blue), pooling (red), various others.
- This network was designed by Google (the one of the previous page in academia).

Szegedy et al: "Going deeper with convolutions"

How do we evaluate which methods work?

- The basic evaluation of data mining/machine learning methods is conducted by individual research groups and reported in scientific articles.
- These results use different data sets, different cross-validation setups, etc. That makes them hard to compare.
- It is easy to cheat, too. That is not in anyone's long-term interest as a researcher, but it happens.
- It is easy to make mistakes, e.g. by getting your cross-validation wrong.

Benchmark data sets

- Benchmark data sets are sets of labelled data used by many researchers to make results more comparable.
- Early examples in computer vision are the *Berkeley Segmentation Dataset and Benchmark* (2001, for image segmentation) and the *Caltech 101* dataset (2004, for object categorization).

Challenges

- To make evaluation (not just data) comparable, some research groups organize competitions (often called “challenges” in computer vision and machine learning).
- The organizers specify a task (e.g. a classification problem) and a performance goal (e.g. “achieve minimal classification error on the test data”).
- Research groups can sign up to participate.
- A set of labelled data is made available to participants, for use as training data.
- The organizers hold out a test data set (which is kept secret). At the end of the competition, all participating groups submit their final trained model, the organizers run it on the test data, and report the results.

ILSVRC

- The best-known example is the *ImageNet Large-Scale Visual Recognition Challenge* (or ILSVRC). which evaluates how well an algorithm can perform certain vision tasks, like classifiying and locating objects in images.
- In 2012, a “deep” neural network drastically improved on previous ILSVRC results. That was one of the triggers for the current interest of the tech industry in machine learning. The network is the one picture on slide 323.

ILSVRC TASKS

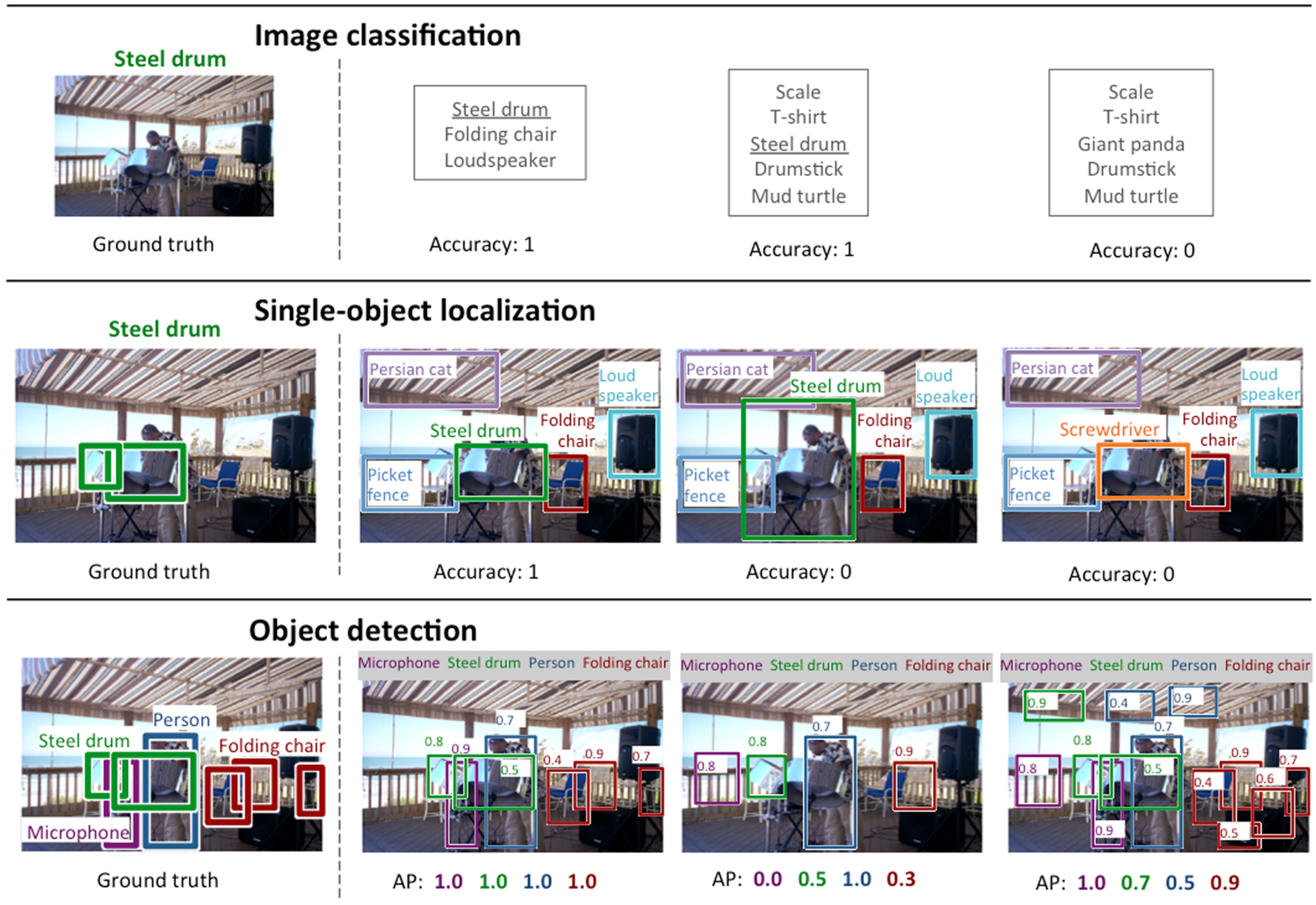
(1) *Image classification* (2010–2014): Algorithms produce a list of object categories present in the image.

(2) *Single-object localization* (2011–2014): Algorithms produce a list of object categories present in the image, along with an axis-aligned bounding box indicating the position and scale of *one* instance of each object category.

(3) *Object detection* (2013–2014): Algorithms produce a list of object categories present in the image along with an axis-aligned bounding box indicating the position and scale of *every* instance of each object category.

from Russakovsky et al: "ImageNet Large Scale Visual Recognition Challenge"

ILSVRC Tasks



from Russakovsky et al: "ImageNet Large Scale Visual Recognition Challenge"

Cross Validation Setup

Year	Train images (per class)	Val images (per class)	Test images (per class)
Image classification annotations (1000 object classes)			
ILSVRC2010	1,261,406 (668–3047)	50,000 (50)	150,000 (150)
ILSVRC2011	1,229,413 (384–1300)	50,000 (50)	100,000 (100)
ILSVRC2012-14	1,281,167 (732–1300)	50,000 (50)	100,000 (100)

size of smallest class

size of largest class

- The data is split into a large training set, plus a validation and a test set.
- Research groups download the training set.
- The validation data sits on a server on which research groups can upload their trained models. The server runs the model on the validation data and reports the accuracy estimate to the researchers, who can use this feedback to improve their model.
- The test data is withheld. After a submission deadline, all submitted models are run on the test data to produce an "official" result.

from Russakovsky et al: "ImageNet Large Scale Visual Recognition Challenge"

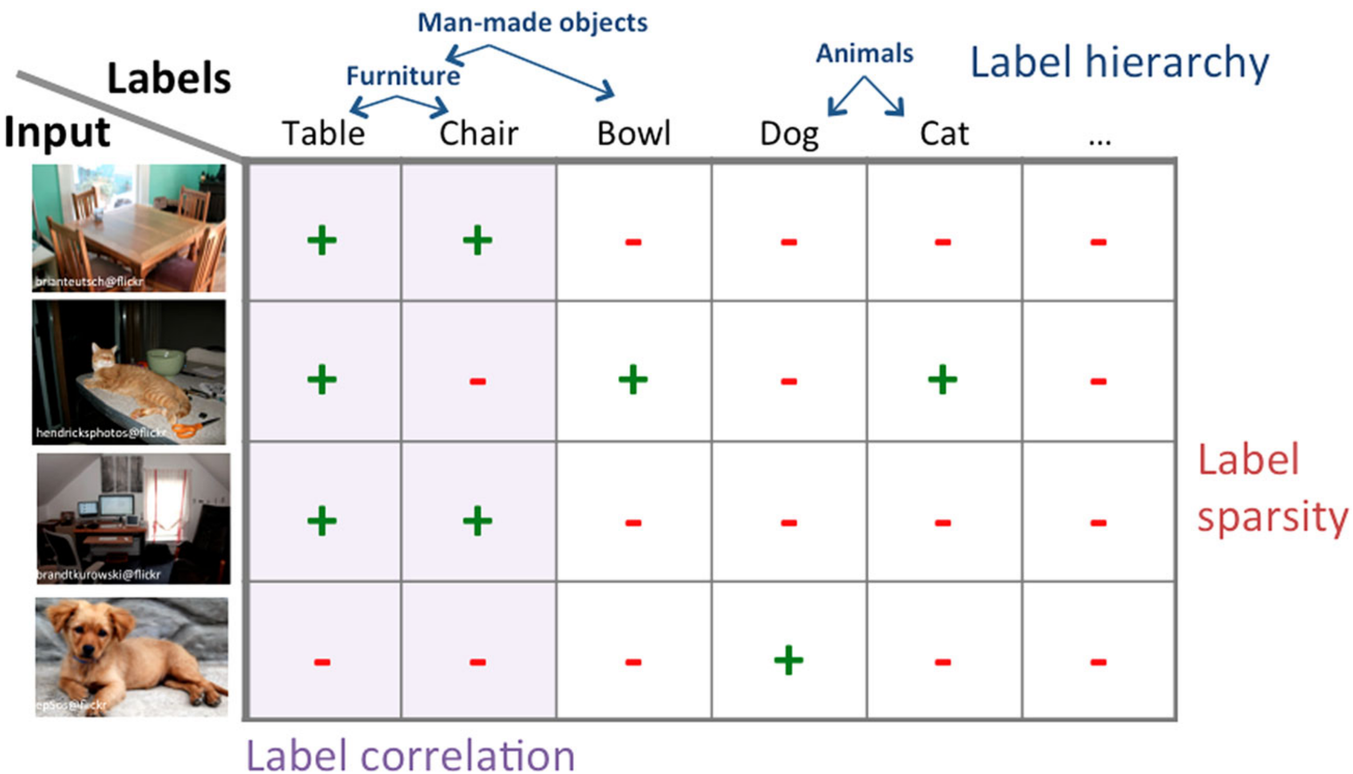
Example Rules

- Each research group is limited to two validation steps per week. (One team famously cheated its way around this rule in 2015.)
- There are separate contests that do or do not permit additional training data to be used.

Crowdsourcing

- The data is collected from image search engines.
- It does not come with reliable labels for training, validation and testing.
- The class labels are added by crowdsourcing.
- The labels are structured hierarchically, i.e. there is a meta-category “cars” which contains specific types of cars as subcategories. The challenge task is to predict the most specific labels (the leaves in the hierarchy tree).
- The next two slides illustrate how the categories are structured.

ILSVRC Data



from Russakovsky et al: "ImageNet Large Scale Visual Recognition Challenge"

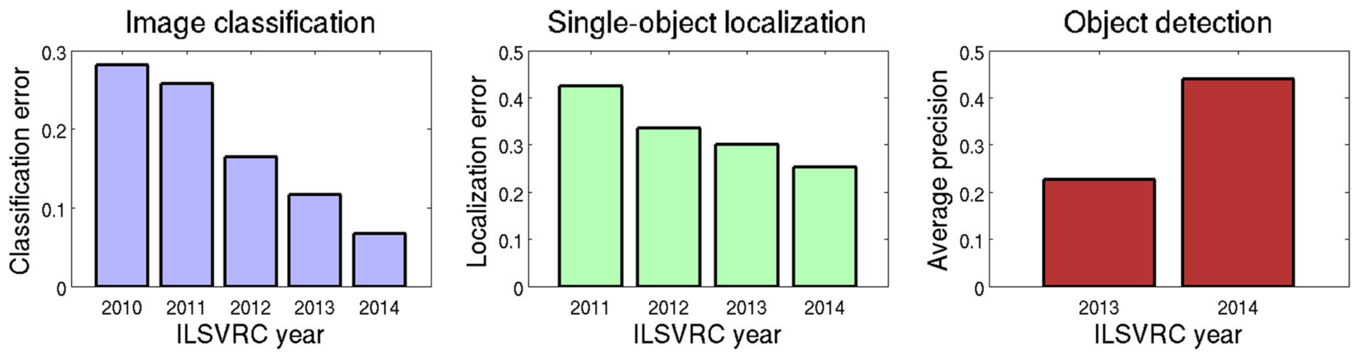
- wind instrument: a musical instrument in which the sound is produced by an enclosed column of air that is moved by the breath (such as trumpet, french horn, harmonica, flute, etc)
  - (17) trumpet: a brass musical instrument with a narrow tube and a flared bell, which is played by means of valves. often has 3 keys on top
  - (18) french horn: a brass musical instrument consisting of a conical tube that is coiled into a spiral, with a flared bell at the end
  - (19) trombone: a brass instrument consisting of a long tube whose length can be varied by a u-shaped slide
  - (20) harmonica
  - (21) flute: a high-pitched musical instrument that looks like a straight tube and is usually played sideways (please do not confuse with oboes, which have a distinctive straw-like mouth piece and a slightly flared end)
  - (22) oboe: a slender musical instrument roughly 65cm long with metal keys, a distinctive straw-like mouthpiece and often a slightly flared end (please do not confuse with flutes)
  - (23) saxophone: a musical instrument consisting of a brass conical tube, often with a u-bend at the end

- food: something you can eat or drink (includes growing fruit, vegetables and mushrooms, but does not include living animals)
  - food with bread or crust: pretzel, bagel, pizza, hotdog, hamburgers, etc
    - (24) pretzel
    - (25) bagel, beigel
    - (26) pizza, pizza pie
    - (27) hotdog, hot dog, red hot
    - (28) hamburger, beefburger, burger
  - (29) guacamole
  - (30) burrito
  - (31) popsicle (ice cream or water ice on a small wooden stick)
  - fruit
    - (32) fig
    - (33) pineapple, ananas
    - (34) banana
    - (35) pomegranate
    - (36) apple
    - (37) strawberry
    - (38) orange
    - (39) lemon
  - vegetables
    - (40) cucumber, cuke
    - (41) artichoke, globe artichoke
    - (42) bell pepper
    - (43) head cabbage
  - (44) mushroom
- items that run on electricity (plugged in or using batteries); including clocks, microphones, traffic lights, computers, etc
  - (45) remote control, remote
  - electronics that blow air
    - (46) hair dryer, blow dryer
    - (47) electric fan: a device for creating a current of air by movement of a surface or surfaces (please do not consider hair dryers)
  - electronics that can play music or amplify sound
    - (48) tape player
    - (49) iPod
  - (50) microphone, mike
  - computer and computer peripherals: mouse, laptop, printer, keyboard, etc
    - (51) computer mouse
    - (52) laptop, laptop computer
    - (53) printer (please do not consider typewriters to be printers)
    - (54) computer keyboard
  - (55) lamp
  - electric cooking appliance (an appliance which generates heat to cook food or boil water)
    - (56) microwave, microwave oven
    - (57) toaster
    - (58) waffle iron
    - (59) coffee maker: a kitchen appliance used for brewing coffee automatically
  - (60) vacuum, vacuum cleaner
  - (61) dishwasher, dish washer, dishwashing machine
  - (62) washer, washing machine: an electric appliance for washing clothes
  - (63) traffic light, traffic signal, stoplight
  - (64) tv or monitor: an electronic device that represents information in visual form
  - (65) digital clock: a clock that displays the time of day digitally
- kitchen items: tools,utensils and appliances usually found in the kitchen
  - electric cooking appliance (an appliance which generates heat to cook food or boil water)
    - (56) microwave, microwave oven
    - (57) toaster
    - (58) waffle iron
    - (59) coffee maker: a kitchen appliance used for brewing coffee automatically
  - (61) dishwasher, dish washer, dishwashing machine
  - (66) stove
  - things used to open cans/bottles: can opener or corkscrew
    - (67) can opener (tin opener)
    - (68) corkscrew
  - (69) cocktail shaker
  - non-electric item commonly found in the kitchen: pot, pan, utensil, bowl, etc

    - (90) tie: a long piece of cloth worn for decorative purposes around the neck or shoulders, resting under the shirt collar and knotted at the throat (NOT a bow tie)
  - headdress, headgear: clothing for the head (hats, helmets, bathing caps, etc)
    - (87) bathing cap, swimming cap: a cap worn to keep hair dry while swimming or showering
    - (91) hat with a wide brim
    - (92) helmet: protective headgear made of hard material to resist blows
  - (93) miniskirt, mini: a very short skirt
  - (94) brassiere, bra: an undergarment worn by women to support their breasts
  - (95) sunglasses

- living organism (other than people): dogs, snakes, fish, insects, sea urchins, starfish, etc.
  - living organism which can fly
    - (96) bee
    - (97) dragonfly
    - (98) ladybug
    - (99) butterfly
    - (100) bird
  - living organism which cannot fly (please don't include humans)
    - living organism with 2 or 4 legs (please don't include humans):
      - mammals (but please do not include humans)
        - feline (cat-like) animal: cat, tiger or lion
          - (101) domestic cat
          - (102) tiger
          - (103) lion
        - canine (dog-like animal): dog, hyena, fox or wolf
          - (104) dog, domestic dog, canis familiaris
          - (105) fox: wild carnivorous mammal with pointed muzzle and ears and a bushy tail (please do not confuse with dogs)
        - animals with hooves: camels, elephants, hippos, pigs, sheep, etc
          - (106) elephant
          - (107) hippopotamus, hippo
          - (108) camel
          - (109) swine: pig or boar
          - (110) sheep: woolly animal, males have large spiraling horns (please do not confuse with antelope which have long legs)
          - (111) cattle: cows or oxen (domestic bovine animals)
          - (112) zebra
          - (113) horse
          - (114) antelope: a graceful animal with long legs and horns directed upward and backward
        - (115) squirrel
        - (116) hamster: short-tailed burrowing rodent with large cheek pouches
        - (117) otter
        - (118) monkey
        - (119) koala bear
        - (120) bear (other than pandas)
        - (121) skunk (mammal known for its ability fo spray a liquid with a strong odor; they may have a single thick stripe across back and tail, two thinner stripes, or a series of white spots and broken stripes
        - (122) rabbit
        - (123) giant panda: an animal characterized by its distinct black and white markings
        - (124) red panda: Reddish-brown Old World raccoon-like carnivore
      - (125) frog, toad
      - (126) lizard: please do not confuse with snake (lizards have legs)
      - (127) turtle
      - (128) armadillo
      - (129) porcupine, hedgehog
    - living organism with 6 or more legs: lobster, scorpion, insects, etc.
      - (130) lobster: large marine crustaceans with long bodies and muscular tails; three of their five pairs of legs have claws
      - (131) scorpion
      - (132) centipede: an arthropod having a flattened body of 15 to 173 segments each with a pair of legs, the foremost pair being modified as prehensors
      - (133) tick (a small creature with 4 pairs of legs which lives on the blood of mammals and birds)
      - (134) isopod: a small crustacean with seven pairs of legs adapted for crawling
      - (135) ant
    - living organism without legs: fish, snake, seal, etc. (please don't include plants)
      - living organism that lives in water: seal, whale, fish, sea cucumber, etc.
        - (136) jellyfish
        - (137) starfish, sea star
        - (138) seal
        - (139) whale
        - (140) ray: a marine animal with a horizontally flattened body and enlarged winglike pectoral fins with gills on the underside
        - (141) goldfish: small golden or orange-red fishes
      - living organism that slides on land: worm, snail, snake
        - (142) snail
        - (143) snake: please do not confuse with lizard (snakes do not have legs)
- vehicle: any object used to move people or objects from place to place
  - a vehicle with wheels
    - (144) golfcart, golf cart
    - (145) snowplow: a vehicle used to push snow from roads

ILSVRC RESULTS



from Russakovsky et al: “ImageNet Large Scale Visual Recognition Challenge”