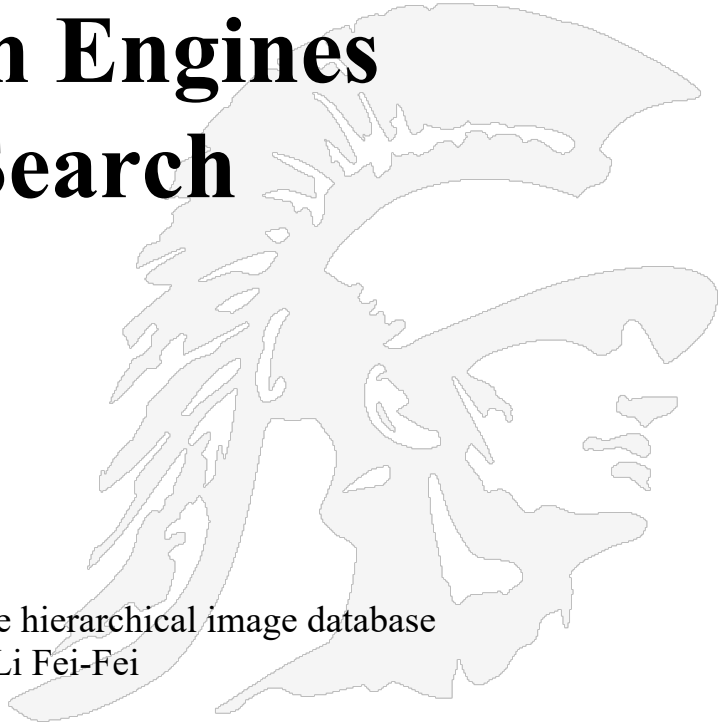
Web Search Engines
Image Search



Some slides come from ImageNet:A large-scale hierarchical image database
By Deng, Dong, Socher, Li-Jia Li, Kai Li and Li Fei-Fei

Search Engines Expand their Search to Images and Videos

- Image Search
  - Search primarily based on
    - tags (FlickR, FaceBook)
    - surrounding text
    - using image features
- Leading Video Search Engines
  - Google and YouTube are the leaders
  - other video search engines include:
    - Bing videos http://www.bing.com/videos/
    - Daily Motion
    - DuckDuckGo
    - Yahoo
    - Metacafe
    - Ask
    - Yandex
- https://www.searchenginejournal.com/best-video-search-engines/360822/#close

MediaMagic Video Production

Image and Video Searching There is a Lot of It on the Web

- **Images are returned for 19% of search queries on Google**
- **Pinterest is an American image sharing and social media service designed to enable saving and discovery of information on the internet using images, and on a smaller scale, animated GIFs and videos**
- **Visual search is a growing trend**
  - **There are over 600 million visual searches on Pinterest each month**
  - **Image-based Pinterest Ads have an 8.5% conversion rate, and Pinterest is projected to clear $1 billion a year in ad revenue by 2020"**
  - **https://www.socialmediatoday.com/news/why-visual-search-will-be-one-of-the-biggest-digital-marketing-trends-of-20/545999/**

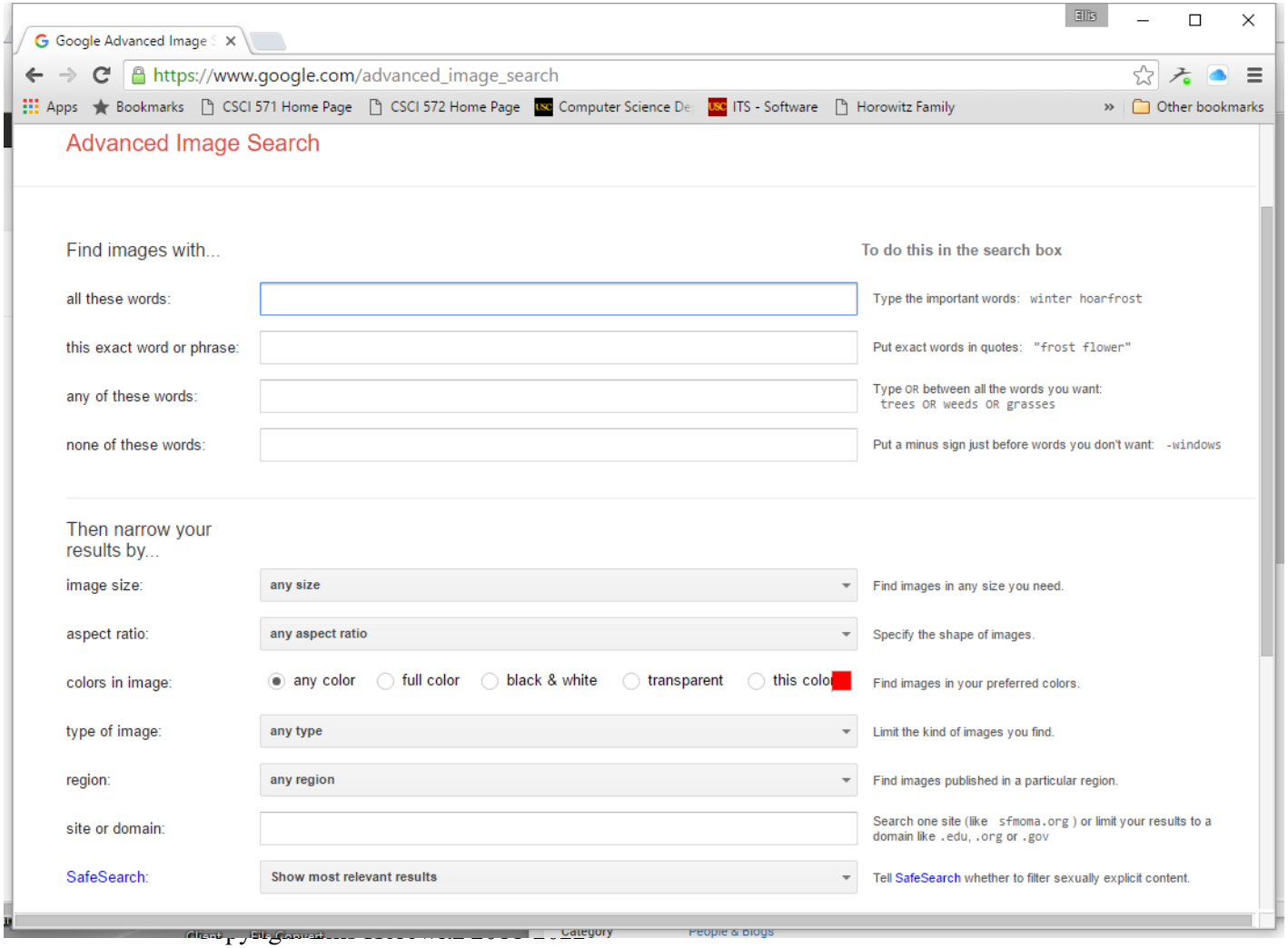
Top 100 Google searches in the U.S.

#	Keyword	Search volume
1	facebook	144,000,000
2	youtube	143,000,000
3	amazon	119,000,000
4	weather	95,300,000
5	walmart	74,400,000
6	google	70,100,000
7	wordle	61,600,000
8	gmail	59,900,000
9	target	50,400,000
10	home depot	47,000,000

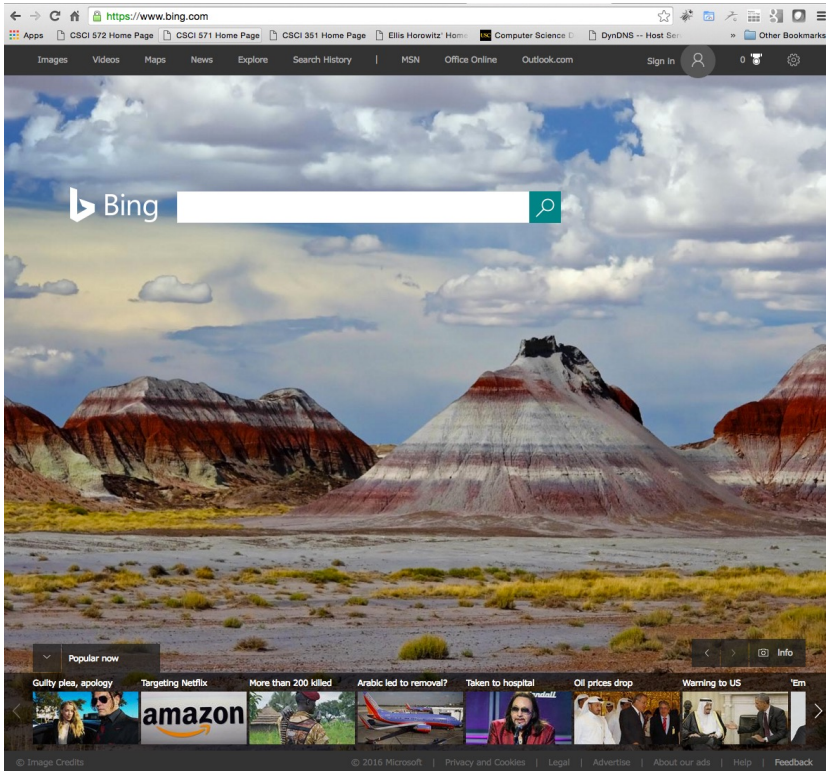
Google Advanced Image Search

text-based query criteria

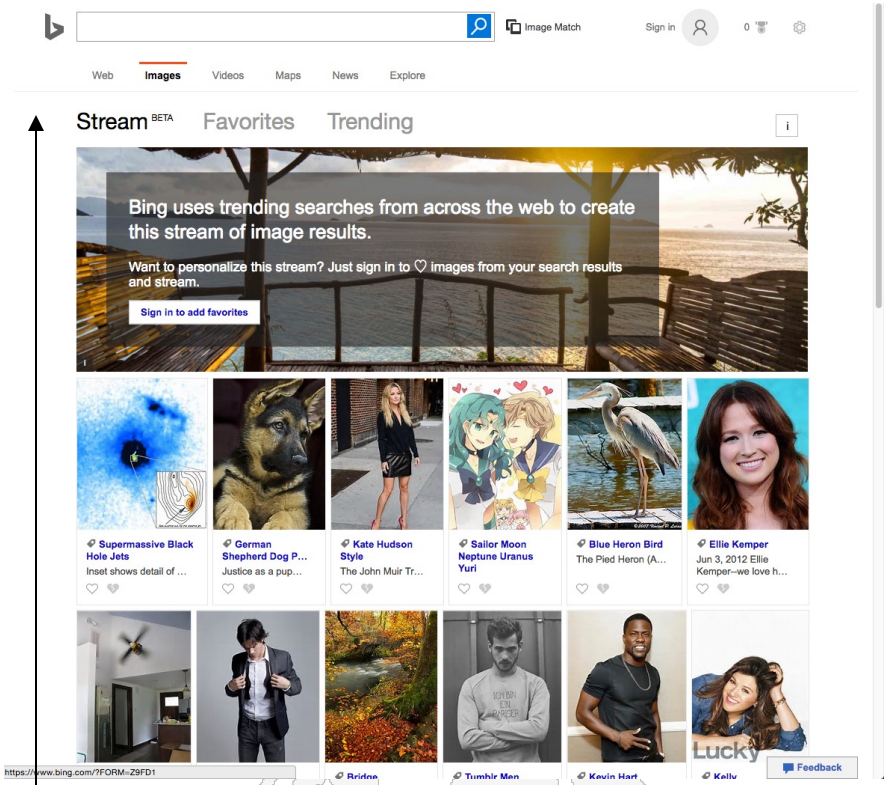
image-based query criteria



Bing Image Search



Bing Initial Page with Options: Images, Videos, Maps, News, Explore, Search History

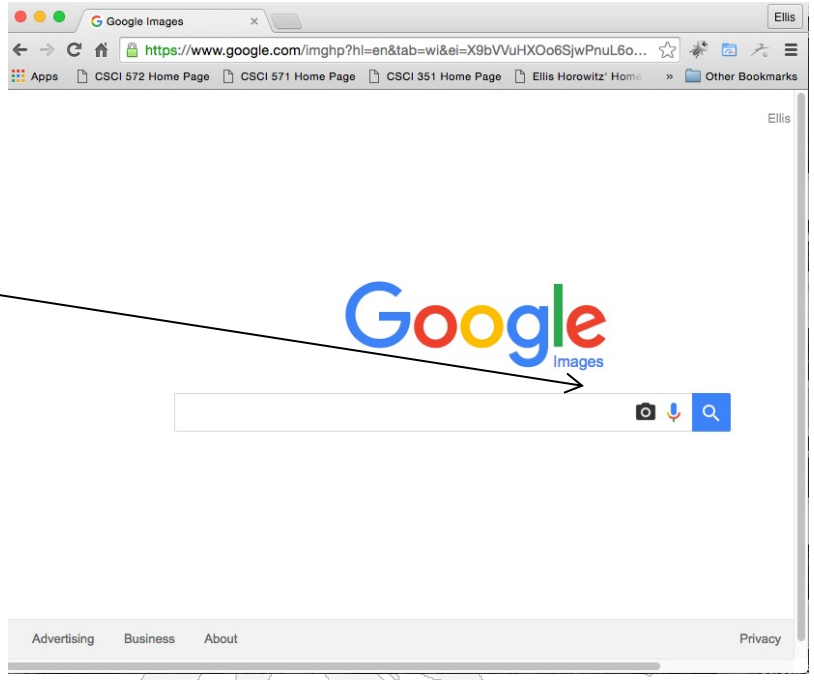


Bing Initial Images Page; Options are: Stream, Favorites, Trending

Using Google Image Search by Similarity

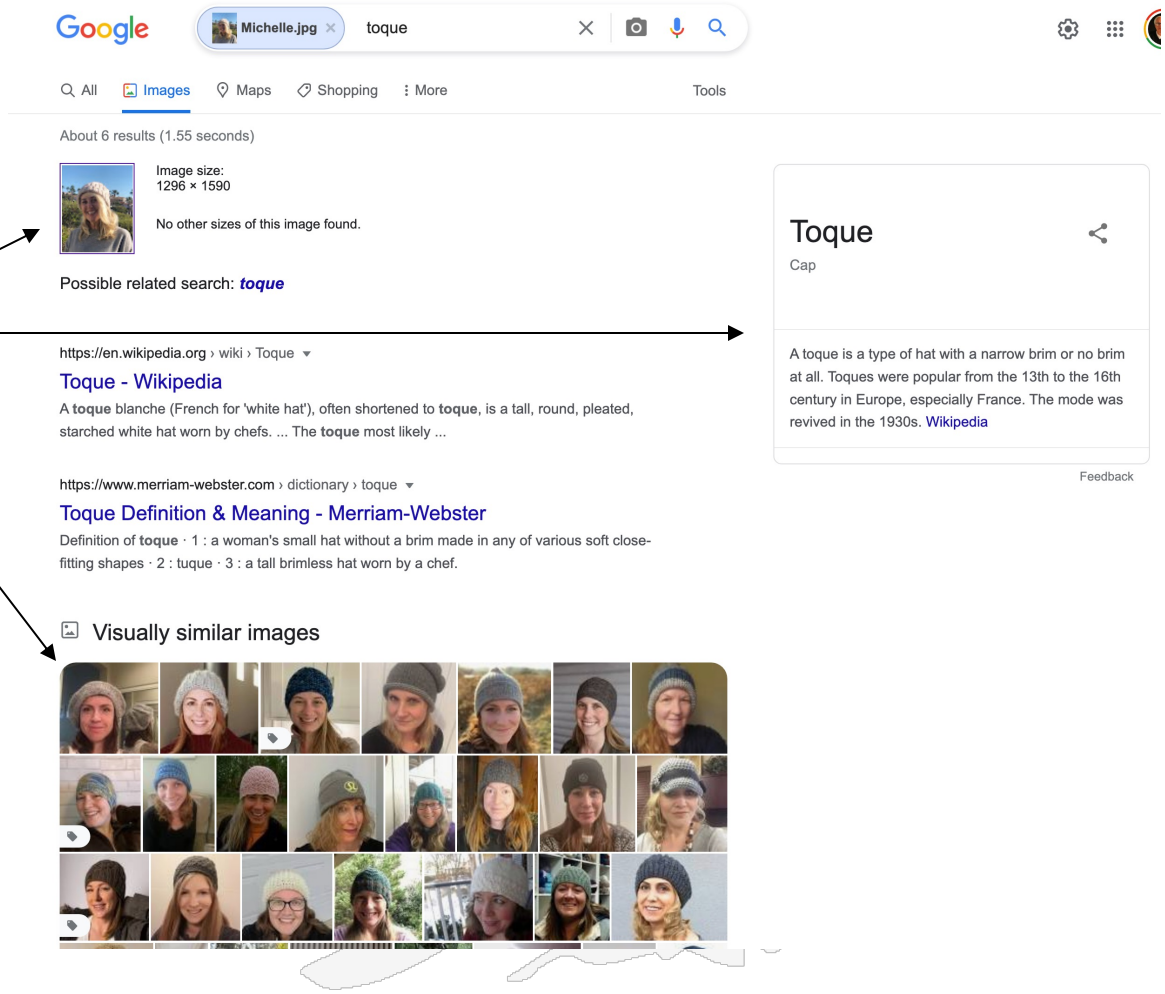
- ***Search By Image*** **is available now at images.google.com or via the "Images" tab in the right-side menu on Google.com. You should see a small camera icon on the far right side of the search bar.**
- **There are several ways to access it:**

1. **Drag and drop an image on the search bar**
2. **Click the camera icon to upload an image from your computer**
3. **Paste the URL of a photo on the web into the search bar**
4. **Use the Chrome or Firefox extensions that add a search option to your computer's contextual menu (right click)**



Google Image Similarity Match Example

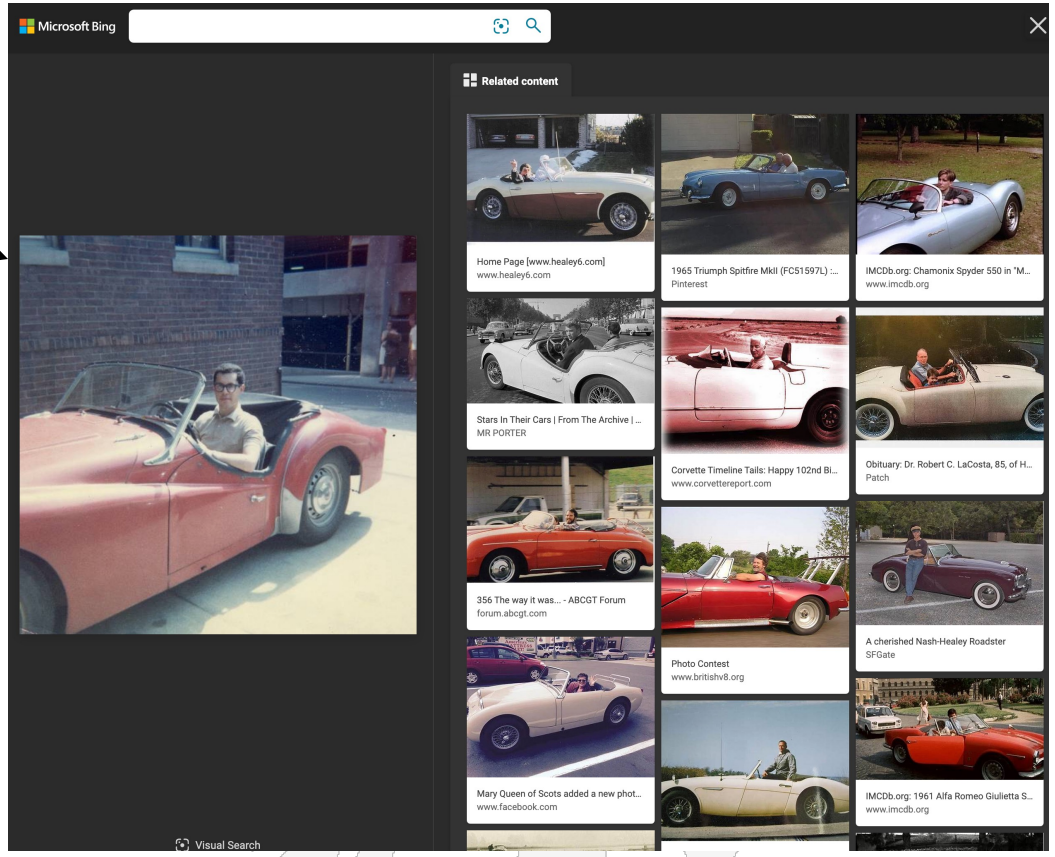
- Google will try to match an image that you drag into the search box
- The original image
- The search results



Bing Image Similarity Match Example

Original photo

Matching photos



Some Useful Links Describing Bing Image Search

https://blogs.bing.com/search-quality-insights/May-2018/Internet-Scale-Deep-Learning-for-Bing-Image-Search

http://searchengineland.com/bing-image-search-redesigned-to-add-more-image-details-to-the-results-218276

Search Engines Specializing in Images

- **TinEye** is a reverse image search engine
- It primarily uses image identification rather than keywords, metadata
- Upon submitting an image, TinEye creates a "unique and compact digital signature or fingerprint" of the image and matches it with other indexed images
- This procedure is able to match even heavily edited versions of the submitted image, but will not usually return similar images in the results.

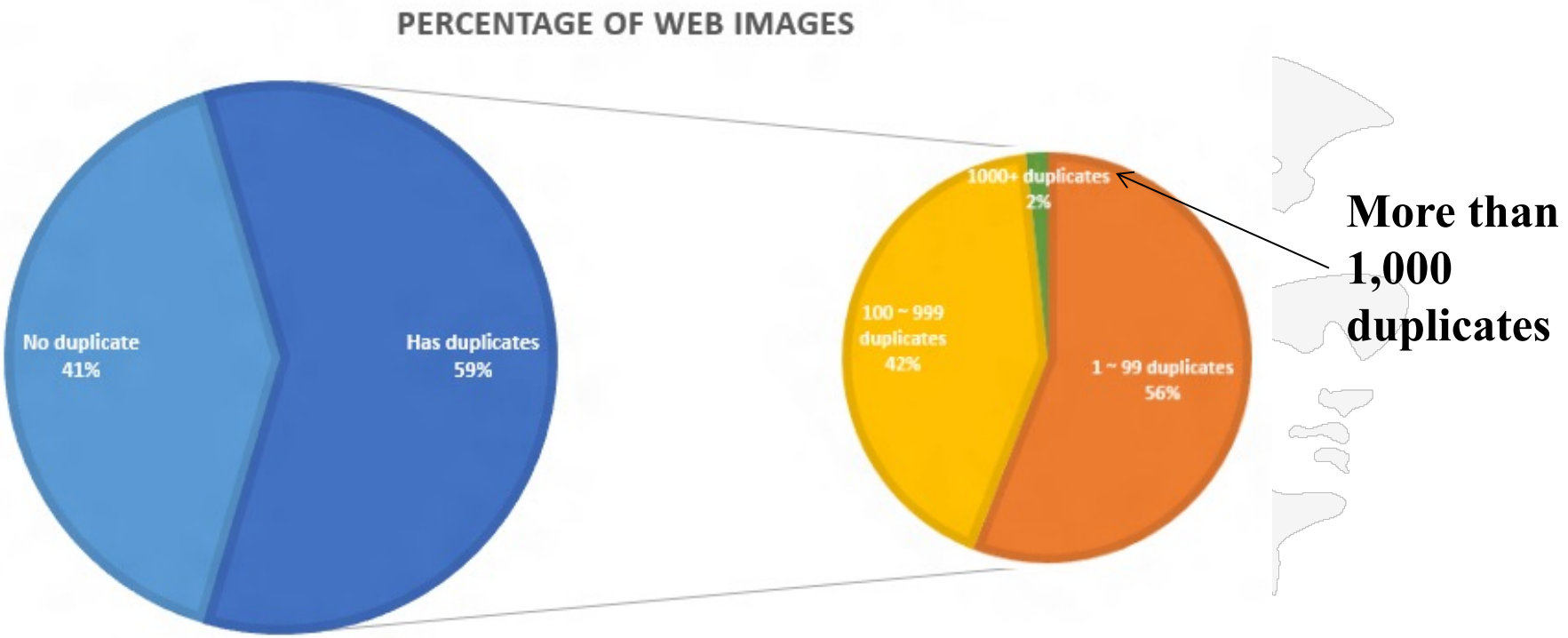
List of Image Search Engine

	Reverse Search	Price
Bing Yahoo Altavista (All 3 uses Bing algorithm)	No	Free
Ask.com (Uses Google algorithm)	No	Free
Google Image	Yes Drag and Drop Tutorial	Free
Corbis	No	Buy sell Images, some are royalty free
Imnese	No	Free
TinEye	Yes	Free
GazoPa	No	For Business
PicSearch	No	Free

https://www.searchenginejournal.com/best-image-search-engines/299963/#close

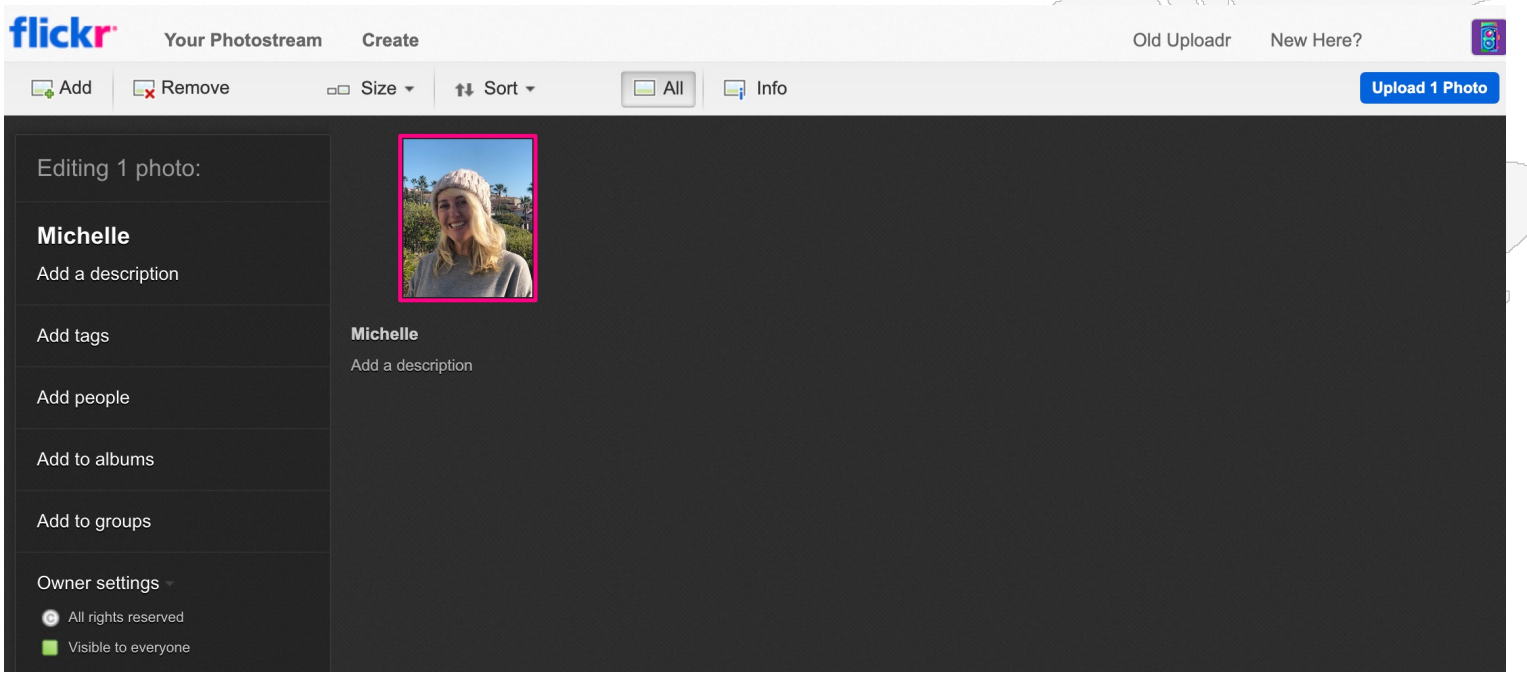
Bing on Image Duplication

- Bing Data on Image Duplication across the Web-there is a lot of it!
- 59% have at least one duplicate; many images have many more duplicates



How Search Engines do Image Indexing – First Use Tags

- Search over tags associated with images
  - Users manually add tags to images
  - Find images with tags that match the query keyword
    - FlickR, Facebook, and many other image hosting sites ask users to tag their uploaded images
- Limitations
  - Tags require human effort to create
  - Tags may be wrong, misleading, or just too general to be useful

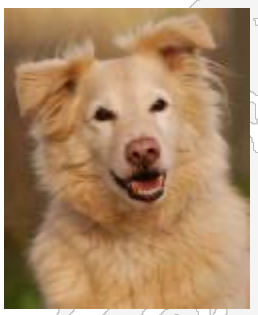


How Search Engines do Image Indexing – Next Use Surrounding Text

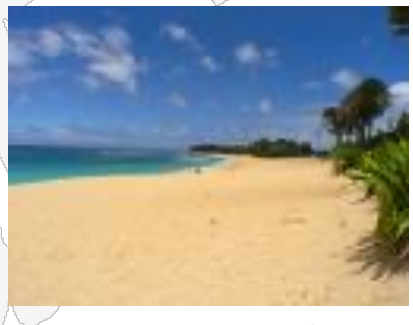
- **Use text associated with images for indexing**
  - **Search web for images, <img src=…>**

  and then use:
    - **Text in URL for image filename**
    - **Text in HTML on page**
- **Example: Google Image Search for "Sunset" gives**
  - **Sunset at Rocky Point in Australia**
  - **Sunset Beach, Oahu**
  - **Frank Smiles at Sunset**
- **A single keyword like "Sunset" produces a diverse set of results, no surprise**

Sunset at Rocky Point



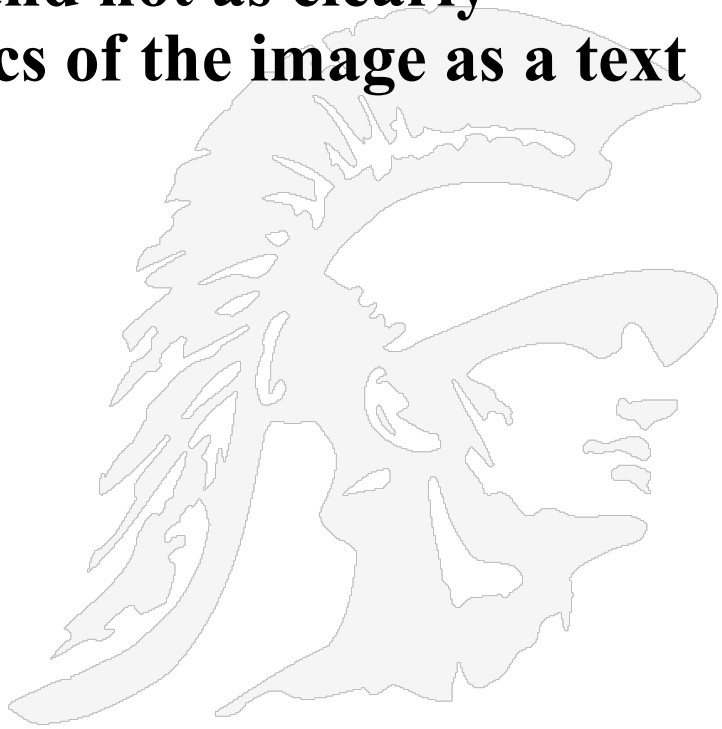
Frank Smiles at Sunset



Sunset Beach, Oahu

How Search Engines do Image Indexing - Finally Use Feature Extraction

- Feature extraction from images is far more difficult than identifying surrounding text
- Features may be low-level and not as clearly associated with the semantics of the image as a text description



Feature Extraction
3 Types of Image Features

- **Typical examined features are those related to color, texture, and shape**

1. Primitive features

- **Mean color (RGB)**
- **Color Histogram**
  - **A representation of the distribution of colors in an image**

2. Semantic features

- **Color Layout, texture etc…**

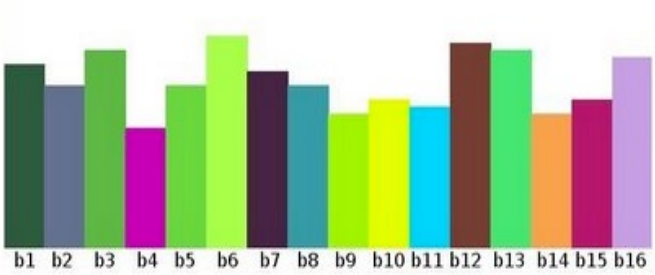
3. Domain specific features

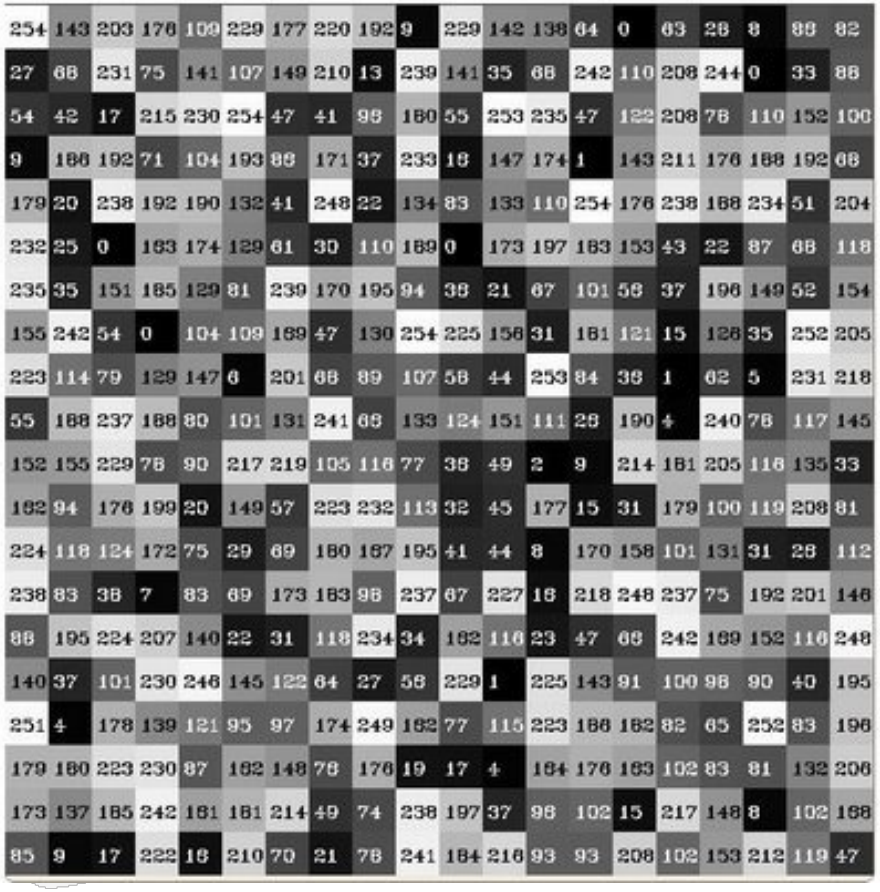
- **Face recognition, fingerprint matching etc…**

General features

Feature Extraction
Color Histograms

- Histograms are collected counts of data organized into a set of predefined bins
- A color histogram measures the intensity of the color of every pixel
- To the right is a matrix containing the intensity of an image, a value between 0 and 255
- We can divide the values in the range [0, 255] into 16 bins: [0,15]U[16,31]U. . . U[240,255] and keep count of the number of pixels that fall in each bin; we get

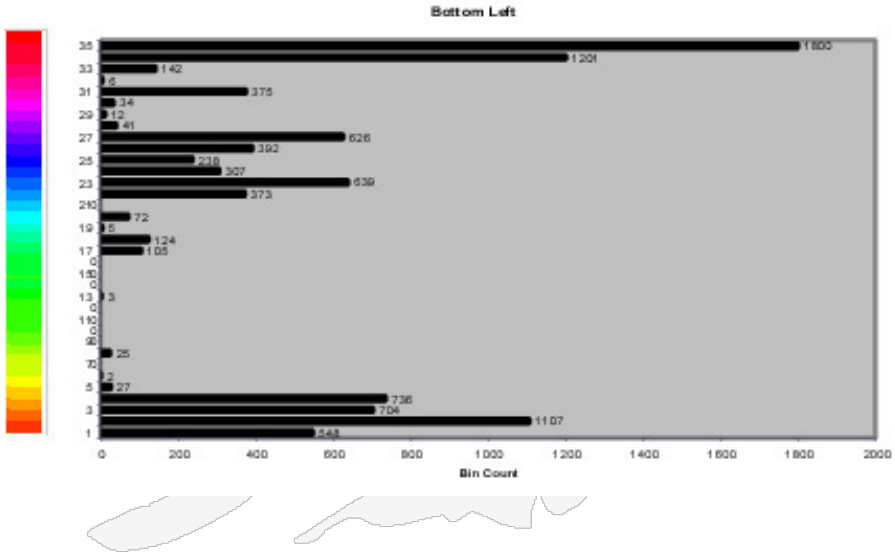




Feature Extraction Similar Color Content

- A **correlogram** for an image is a table indexed by color pairs, where the *d-th* entry for row *(i,j)* specifies the probability of finding a pixel of color *j* at a distance *d* from a pixel of color *i* in this image. Here *d* is chosen from a set of distance values D.

- An **autocorrelogram** captures *spatial correlation between identical colors only*. This information is a subset of the correlogram and consists of rows of the form *(i,j)* only.



Feature Extraction
Color Layout

- **Need for Color Layout**
  - ***Global*** **color features give too many false positives**
- **How color layout works**
  - **Divide the whole image into sub-blocks**
  - **Extract features from each sub-block**
- **One step further**
  - **Divide into regions based on color feature concentration**
  - **This process is called *segmentation*.**

Example: Color layout

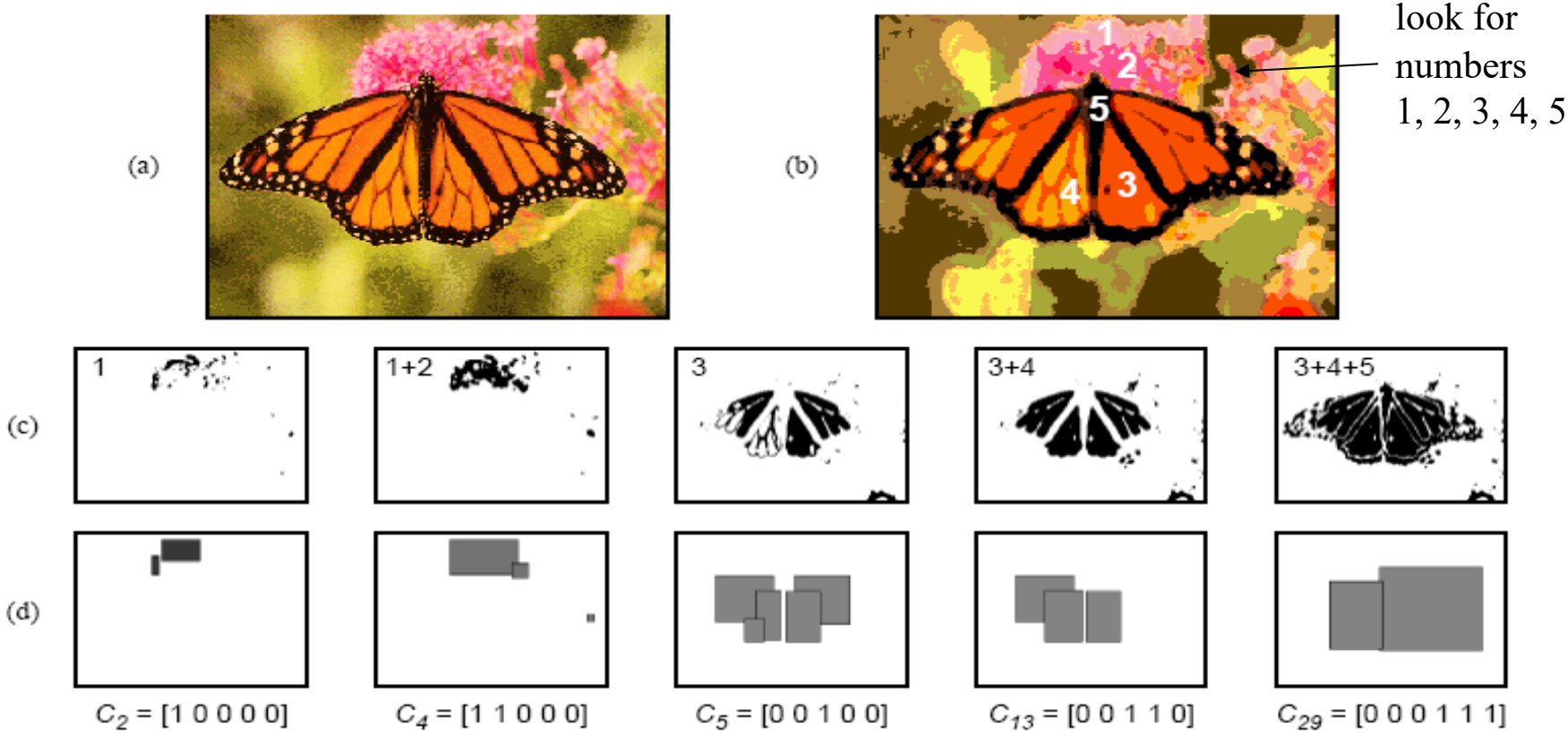


FIGURE 2. (a) *Butterfly* color image, (b) processed color image with 30 colors, (c) pixels from image (b) belonging to color set C_i, (d) minimum bounding rectangles (MBRs) for extracted regions used to index the image collection.

**** Image adapted from Smith and Chang : Single Color Extraction and Image Query**

Example: Texture and Shape

- **Texture – an innate property of all surfaces**
  - **e.g. clouds, trees, bricks, hair etc…**
  - **Refers to visual patterns of homogeneity**
  - **Texture is spatial arrangement of gray levels in the image**
- **Shape features describe the form of object boundaries and edges**
- **Examples:**

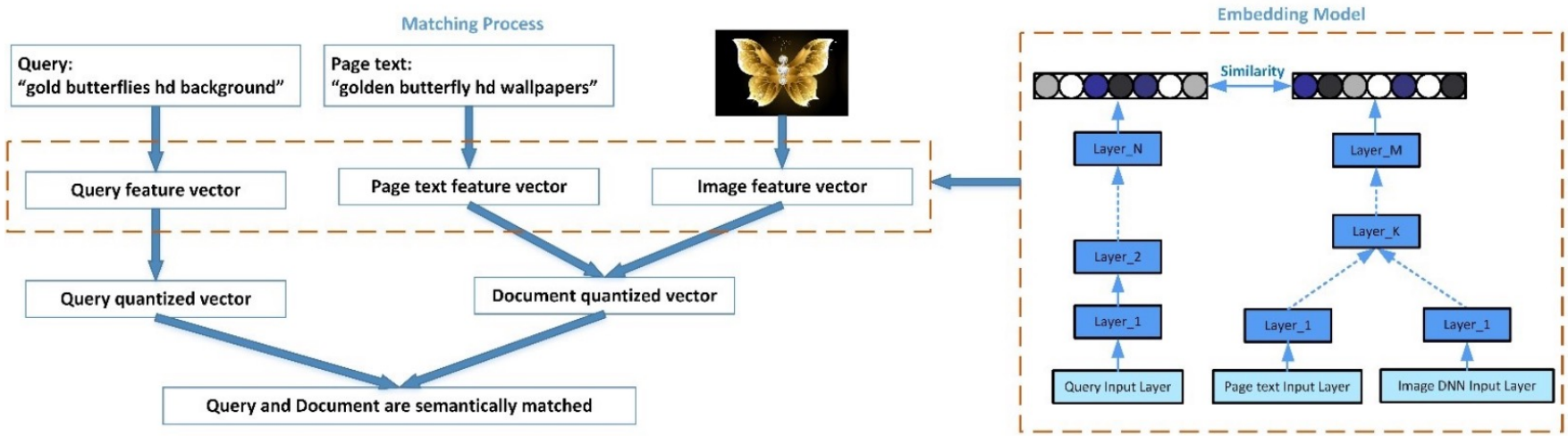
shapes

How Bing Does Deep Learning of Images

- Deep learning maps an image to a vector through a process called *image embedding*
- A model is first trained using deep learning techniques to map (embed) a query and an image to a high dimensional vector space in such a way that the corresponding vectors are similar if the image is semantically relevant to the query, and further apart otherwise
- Top level image matching algorithm involves 3 steps
  1. A matching stage - to select candidate images from a huge index of images
  2. Multiple ranking stages - to use computationally expensive methods to score each candidate image independently and rank all images
  3. Multiple set ranking stages - to re-rank previous candidates lists considering information from the entire candidate set, not just independent images

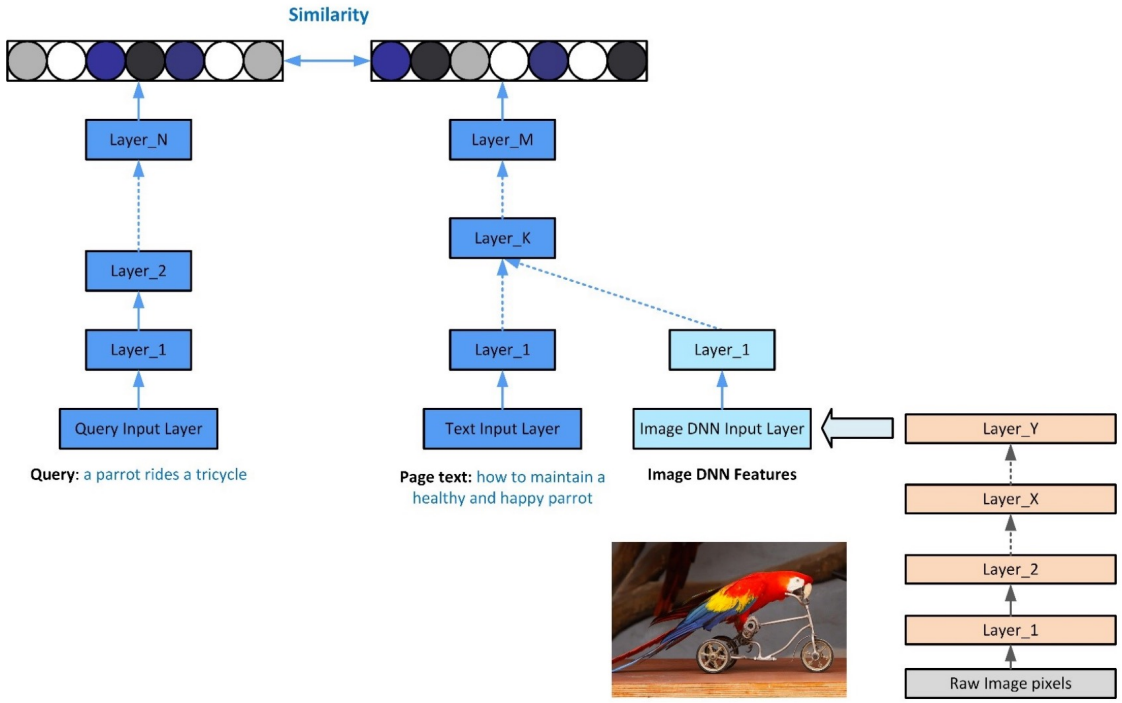
Bing Deep Learning in Matching

- The matching stage selects a candidate set of images from billions of images
- During matching, it is essential to obtain a result set with high recall and moderate precision.
- query, image and page text are embedded by a deep learning model into embedding vectors. We then use an efficient vector search algorithm to scan a billions-scale index to find the N-best document vectors for a query vector.
- below: the left part shows that even though the page text 'golden butterfly hd wallpapers' is not exactly the same as the query 'gold butterflies hd background', we are able to retrieve the relevant image based on query embedding, page text embedding and image embedding, which are generated from the deep learning model as shown in the right diagram.



Bing Deep Learning in Ranking Final Stage

- We now rank all the candidate images according to their relevance to the query
- Pruning of images has occurred - we reuse the previously computed embedding vectors but do more exact calculations of semantic distance between the vectors.
- For example, in the diagram below, the page contains the text 'how to maintain a healthy and happy parrot' however we are still able to recognize that the image matches the query 'a parrot rides a tricycle' because we are looking inside the image using image embedding.

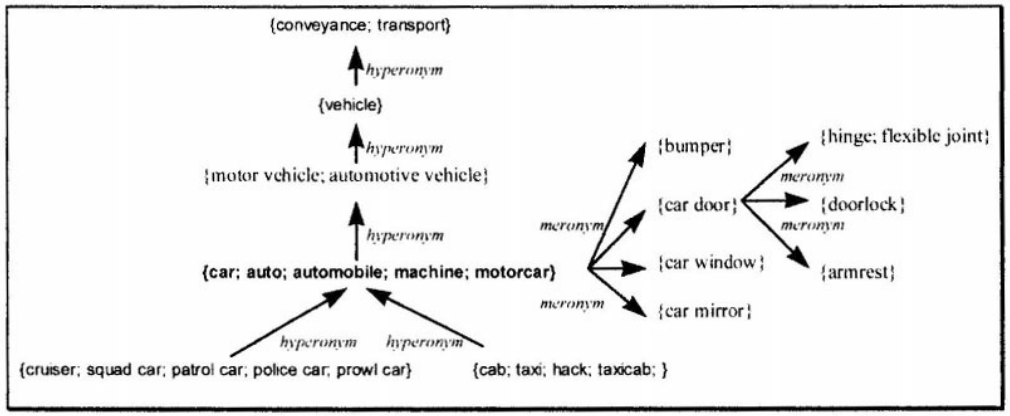


ImageNet: A Large-Scale Hierarchical Image Database

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei Dept. of Computer Science, Princeton University, USA

Recall WordNet

- We discussed WordNet in the lecture on Question/Answering;
- WordNet is a large lexical database of English.
  - Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.
- WordNet as an ontology



Imagenet.org

- **ImageNet** is an image database organized *according to the WordNet hierarchy*, in which each node of the hierarchy is depicted by hundreds and thousands of images.
- Currently there are an average of over five hundred images per node
- 14,197,122 images covering 21,841 synsets indexed

IMAGENET

14,197,122 images, 21841 synsets indexed

Home Download Challenges About

Not logged in. Login I Signup

An Update to the ImageNet Website and Dataset

March 11, 2021

We are proud to see ImageNet's wide adoption going beyond what was originally envisioned. However, the decade-old website was burdened by growing download requests. To serve the community better, we have redesigned the website and upgraded its hardware. The new website is simpler; we removed tangential or outdated functions to focus on the core use case—enabling users to download the data, including the full ImageNet dataset and the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

Meanwhile, the computer vision community has progressed, and so has ImageNet. The dataset was created to benchmark object recognition—at a time when it barely worked. The problem then was how to collect labeled images at a sufficiently large scale to be able to train complex models in laboratories. Today, computer vision is in real-world systems impacting people's Internet experience and daily lives. An emerging problem now is how to make sure computer vision is fair and preserves people's privacy. We are continually evolving ImageNet to address these emerging needs.

In a FAT* 2020 paper, we filtered 2,702 synsets in the "person" subtree that may cause problematic behaviors of the model. We have updated the full ImageNet data on the website to remove these synsets. The update does not affect the 1,000 categories in ILSVRC.

In a more recent paper, we investigate privacy issues in ILSVRC. 997 out of 1000 categories in ILSVRC are not people categories; nevertheless, many incidental people are in the images, whose privacy is a concern. We first annotated faces in the images and then constructed a face-blurred version of ILSVRC. Experiments show that one can use the face-blurred version for benchmarking object recognition and for transfer learning with only marginal loss of accuracy. We release our face annotations to facilitate further research on privacy-aware visual recognition.

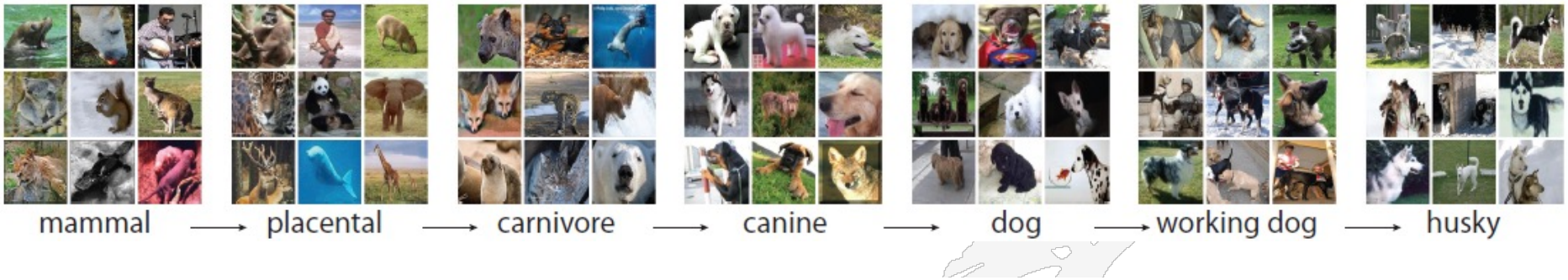
Team members working on these new improvements: Kaiyu Yang (Princeton), Jacqueline Yau (Stanford), Li Fei-Fei (Stanford), Jia Deng (Princeton), Olga Russakovsky (Princeton).

© 2020 Stanford Vision Lab, Stanford University, Princeton University imagenet.help.desk@gmail.com Copyright infringement

https://www.image-net.org/

ImageNet is a Knowledge Ontology

- **Below is the taxonomy for a Husky dog from WordNet and the corresponding ImageNet taxonomy of images**



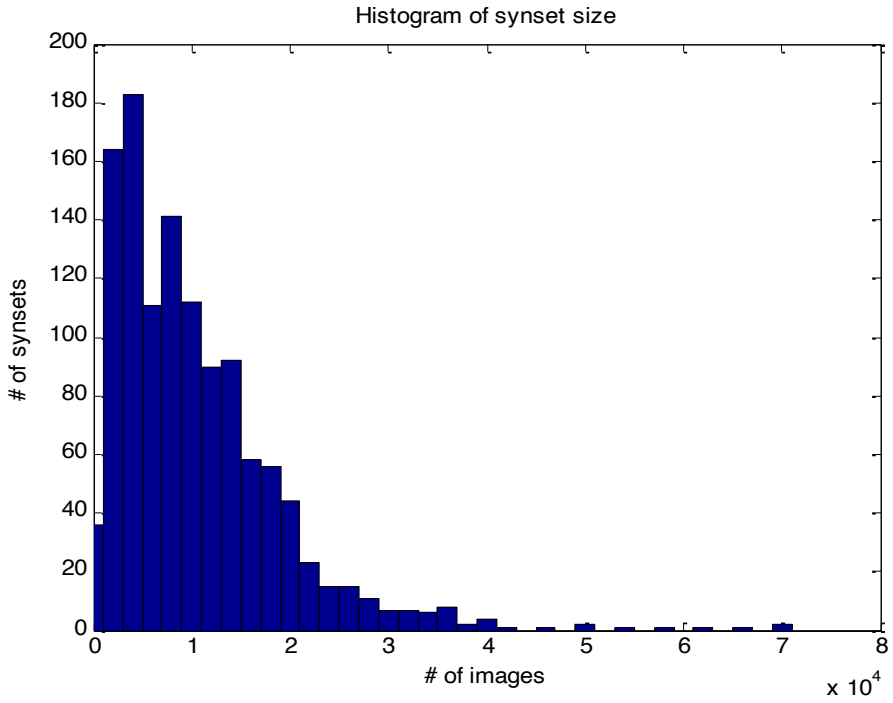
- S: (n) Eskimo dog, **husky** (breed of heavy-coated Arctic sled dog)
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) working dog (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
      - S: (n) dog, domestic dog, Canis familiaris (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) *"the dog barked all night"*
        - S: (n) canine, canid (any of various fissiped mammals with nonretractile claws and typically long muzzles)
          - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have four or five clawed digits on each limb"*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a placenta; all mammals except monotremes and marsupials)
              - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
                - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
                  - S: (n) chordate (any animal of the phylum Chordata having a notochord or spinal column)
                    - S: (n) animal, animate being, beast, brute, creature, fauna (a living organism characterized by voluntary movement)
                      - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
                        - S: (n) living thing, animate thing (a living (or once living) entity)
                          - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                            - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                              - S: (n) physical entity (an entity that has physical existence)
                                - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

ImageNet is a Knowledge Ontology

- S: (n) **car**, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) *"he needs a car to get to work"*
  - *direct hyponym / full hyponym*
  - ***part meronym***
    - S: (n) accelerator, accelerator pedal, gas pedal, gas, throttle, gun (a pedal that controls the throttle valve) *"he stepped on the gas"*
    - S: (n) air bag (a safety restraint in an automobile; the bag inflates on collision and prevents the driver or passenger from being thrown forward)
    - S: (n) auto accessory (an accessory for an automobile)
    - S: (n) automobile engine (the engine that propels an automobile)
    - S: (n) automobile horn, car horn, motor horn, horn, hooter (a device on an automobile for making a warning noise)
    - S: (n) buffer, fender (a cushion-like device that reduces shock due to an impact)
    - S: (n) bumper (a mechanical device consisting of bars at either end of a vehicle to absorb shock and prevent serious damage)
    - S: (n) car door (the door of a car)
    - S: (n) car mirror (a mirror that the driver of a car can use)
    - S: (n) car seat (a seat in a car)
    - S: (n) car window (a window in a car)
    - S: (n) fender, wing (a barrier that surrounds the wheels of a vehicle to block splashing water or mud) *"in Britain they call a fender a wing"*
    - S: (n) first gear, first, low gear, low (the lowest forward gear ratio in the gear box of a motor vehicle; used to start a car moving)
    - S: (n) floorboard (the floor of an automobile)
    - S: (n) gasoline engine, petrol engine (an internal-combustion engine that burns gasoline; most automobiles are driven by gasoline engines)
    - S: (n) glove compartment (compartment on the dashboard of a car)
    - S: (n) grille, radiator grille (grating that admits cooling air to car's radiator)
    - S: (n) high gear, high (a forward gear with a gear ratio that gives the greatest vehicle velocity for a given engine speed)
    - S: (n) hood, bonnet, cowl, cowling (protective covering consisting of a metal part that covers the engine) *"there are powerful engines under the hoods o cowling in order to repair the plane's engine"*
    - S: (n) luggage compartment, automobile trunk, trunk (compartment in an automobile that carries luggage or shopping or tools) *"he put his golf bag in the*
    - S: (n) rear window (car window that allows vision out of the back of the car)
    - S: (n) reverse, reverse gear (the gears by which the motion of a machine can be reversed)
    - S: (n) roof (protective covering on top of a motor vehicle)
    - S: (n) running board (a narrow footboard serving as a step beneath the doors of some old cars)
    - S: (n) stabilizer bar, anti-sway bar (a rigid metal bar between the front suspensions and between the rear suspensions of cars and trucks; serves to stabilize the chassis)
    - S: (n) sunroof, sunshine-roof (an automobile roof having a sliding or raisable panel) *"'sunshine-roof' is a British term for 'sunroof'"*
    - S: (n) tail fin, tailfin, fin (one of a pair of decorations projecting above the rear fenders of an automobile)
    - S: (n) third gear, third (the third from the lowest forward ratio gear in the gear box of a motor vehicle) *"you shouldn't try to start in third gear"*
    - S: (n) window (a transparent opening in a vehicle that allow vision out of the sides or back; usually is capable of being opened)

Some Synsets Have Very Few Images While Some Have Many

- "Mammal" subtree (1180 synsets)
  - Average # of images per synset: 10.5K



Most populated	Least populated
Humankind (118.5k)	Algeripithecus minutus (90)
Kitty, kitty-cat (69k)	Striped muishond (107)
Cattle, cows (65k)	Mylodonitid (127)
Pooch, doggie (62k)	Greater pichiciego (128)
Cougar, puma (57k)	Damaraland mole rat (188)
Frog, toad (53k)	Western pipistrel (196)
Hack, jade, nag (50k)	Muishond (215)

Constructing ImageNet

- 2-step process

Step 1 :
Collect candidate images Via the Internet

→

Step 2 :
Clean up the candidate Images by humans

Step 1:Collect Candidate Images from the Internet

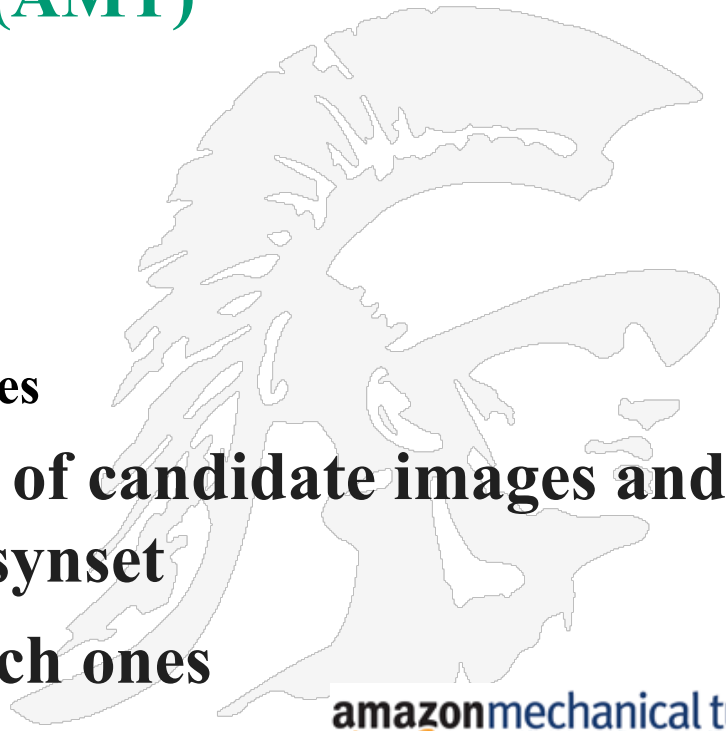
- For each synset, the queries are the set of WordNet synonyms
- Accuracy of Internet Image search results: 10 %
  - For 500-1000 clean images, needs 10K images
- Query expansion
  - Synonyms: German police dog, German shepherd dog
  - Appending words form ancestors: sheepdog, dog
- Multiple Languages
  - Italian, Dutch, Spanish, Chinese　e.g. 德国牧羊犬，pastore tedesco
- More engines: Yahoo! , flickr, Google
- Parallel downloading

Step 2:Clean Up the Candidate Images by Humans

- Rely on humans to verify each candidate image collected for a given synset
- Amazon Mechanical Turk (AMT)

- used for labeling vision data
- 300 images: 0.02 dollar
- 14,197,122 images: 946 dollars
- 10 repetition: 9460 dollars
- Jul 2008 -Apr 2010:11 million images

- Present the users with a set of candidate images and the definition of the target synset
- let users select the best match ones



amazonmechanical turk
Artificial Artificial Intelligence

A Task on AMT

(amazon mechanical turk)

Workers do annotation on AMT
-Multiple annotations for each images

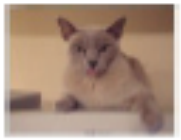
Annotation Results
- An average of > 97% accuracy

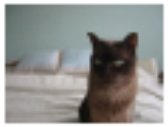
Ensure Accuracy

- Users Enhancement
- Provide wiki and Google links for definitions
- Make sure workers read the definition
  - Definition quiz
- Allow more feedback. E.g. “unimaginable synset” expert opinion

Ensure Accuracy

- Human users make mistakes
- Not all users follow the instructions
- Users do not always agree with each other
  - Subtle or confusing synsets, e.g. Burmese cat
- Quality Control System





User 1	Y	Y	Y
User 2	N	Y	Y
User 3	N	Y	Y
User 4	Y	N	Y
User 5	Y	Y	Y
User 6	N	N	Y

#Y	#N	Conf Cat	Conf BCat
0	1	0.07	0.23
1	0	0.85	0.69
1	1	0.46	0.49
2	0	0.97	0.83
0	2	0.02	0.12
3	0	0.99	0.90
2	1	0.85	0.68

Quality Control System

- randomly sample an initial subset of images to users
  - Have multiple users independently label same image
- obtain a confidence score table, indicating the probability of an image being a good image given the user votes
  - Different categories requires different levels of consensus
- Proceed until a pre-determined confidence score threshold reached

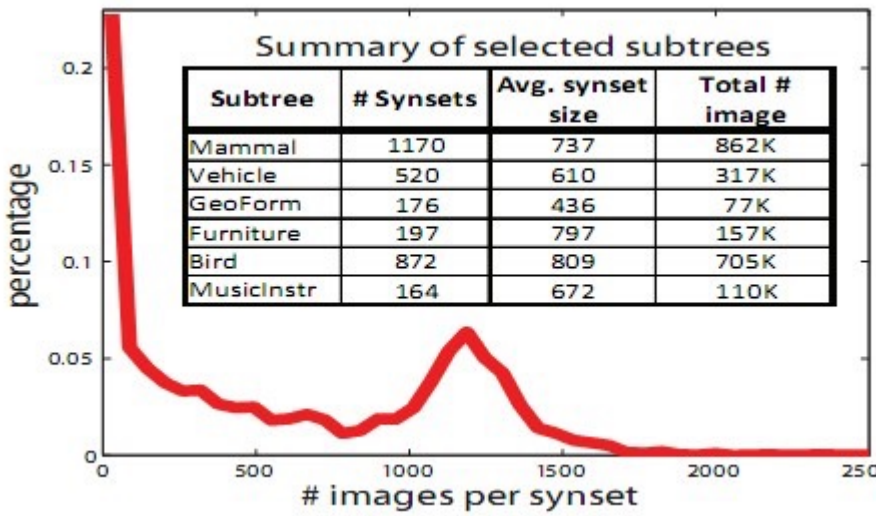
Properties of ImageNet

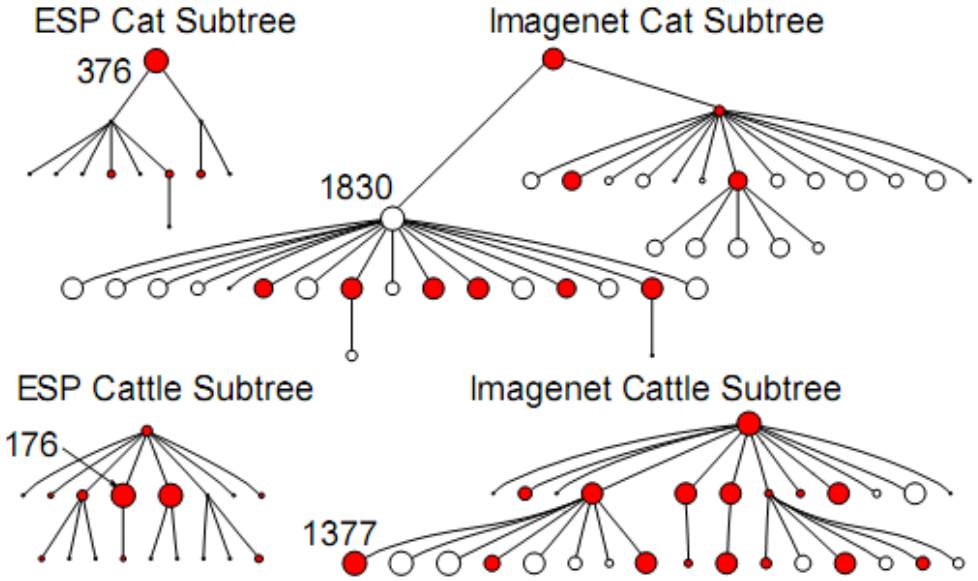
- **Scale**: 12 subtrees, 3.2 million images, 5247 categories

- **Hierarchy**: densely populated semantic hierarchy, based on WordNet

Summary of selected subtrees

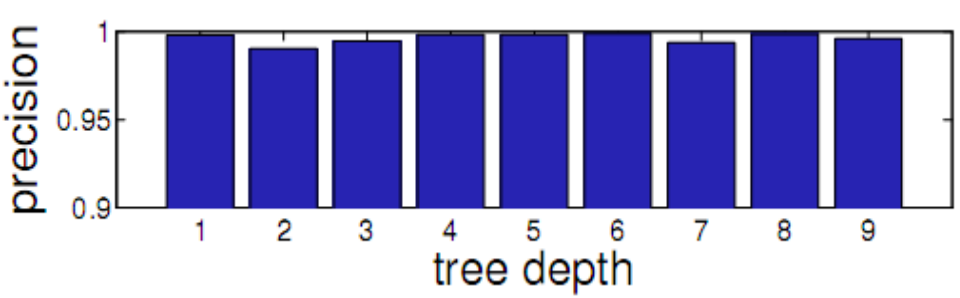
Subtree	# Synsets	Avg. synset size	Total # image
Mammal	1170	737	862K
Vehicle	520	610	317K
GeoForm	176	436	77K
Furniture	197	797	157K
Bird	872	809	705K
MusicInstr	164	672	110K



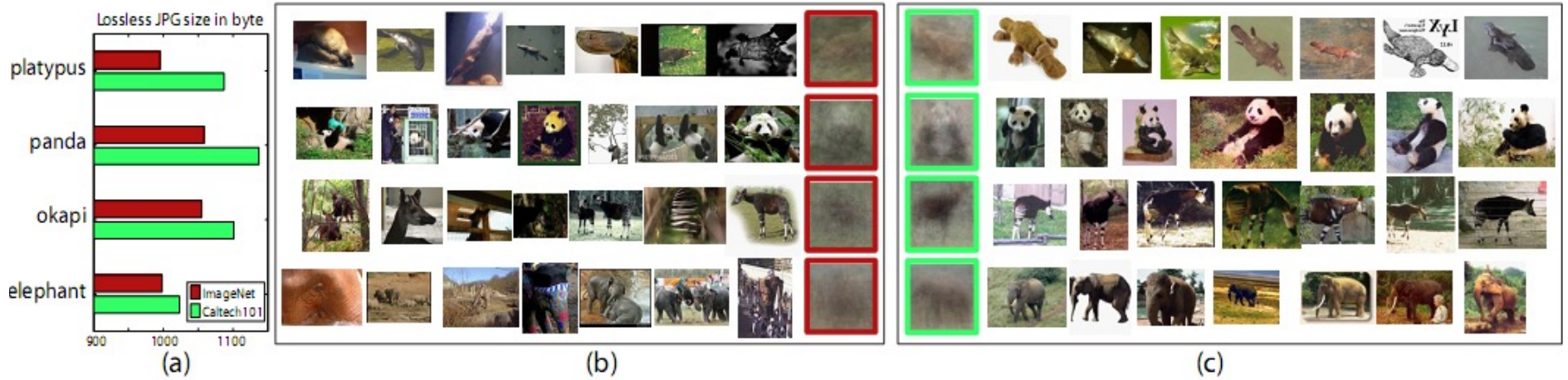


Properties of ImageNet

- **Accuracy**: clean dataset at all level



- Diversity: variable appearances, positions, view points, poses, background clutter, occlusions.



ImageNet Applications

- ImageNet is useful for many computer vision applications such as object recognition, image classification and object localization.
- Images for ImageNet were collected from various online sources. ImageNet doesn't own the copyright for any of the images.
  - For public access, ImageNet provides image thumbnails and URLs from where the original images were downloaded. Researchers can use these URLs to download the original images.
  - Those who wish to use the images for non-commercial or educational purpose, can create an account on ImageNet and request access. This will allow direct download of images from ImageNet. This is useful when the original sources of images are no longer available.
- Non-parametric Object Recognition
- Tree Based Image Classification
- Automatic Object Localization

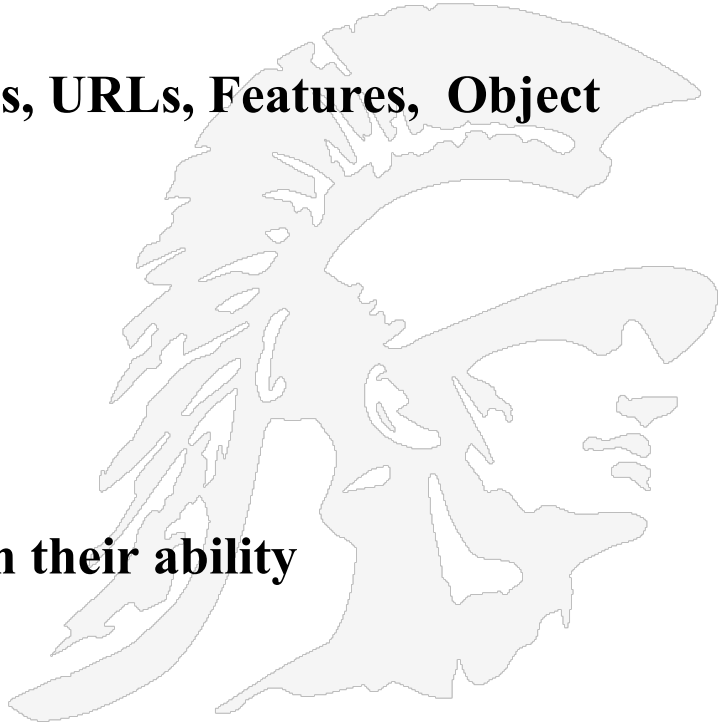
Pros and Cons

Pros

1. Crowdsourcing
2. Benchmarking
3. Open: Download Original Images, URLs, Features, Object Attributes, API

Cons

1. Improve algorithm: PageRank
2. AMT: hierarchical users based on their ability
3. Only one tag per image



Summary by the ImageNet Developers

- **ImageNet is intended to serve as**
  - A dataset
  - A knowledge ontology
- **Construction of large-scale image dataset is a relatively new research area**
  - Crowdsourcing might be the future of many such tasks
  - see an ImageNet developer's TED talk, https://www.youtube.com/watch?v=40riCqvRoMs
- **Benchmarking: what does classifying 10k+ image categories tell us?**
  - Computation matters
  - Size matters
  - Density matters
  - Hierarchy matters