Database Overview

WebVision Database: Visual Learning and Understanding from Web Data,
Wen Li, Limin Wang, Wei Li, Erikur Agustsson, and Luc Van Gool, arXiv 1708.02862, 2017.

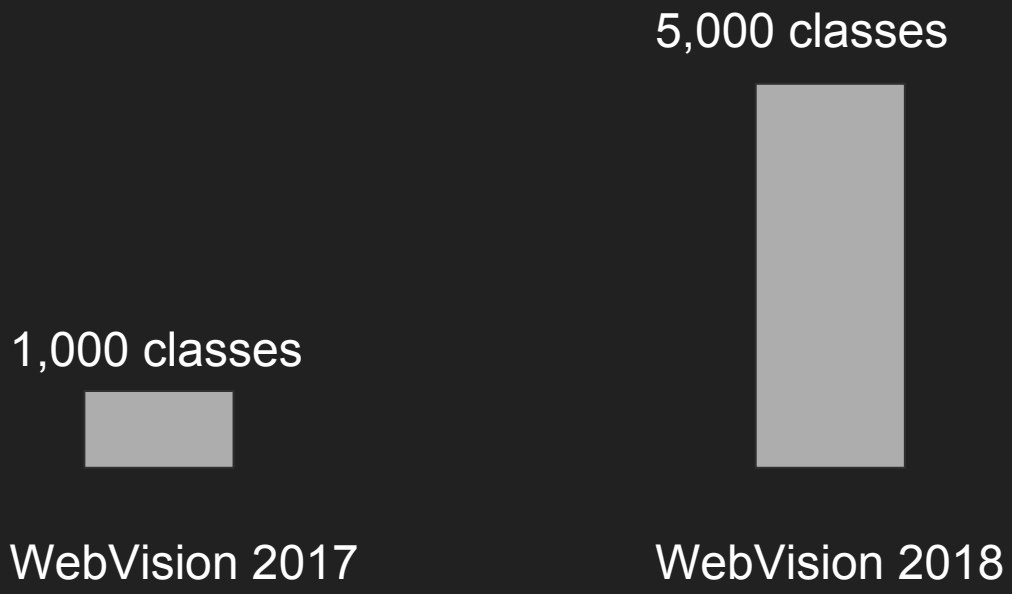
webvision

What’s new in WebVision 2018

Bigger

What’s new in WebVision 2018

Bigger number of categories



What’s new in WebVision 2018

Bigger number of images

What’s new in WebVision 2018

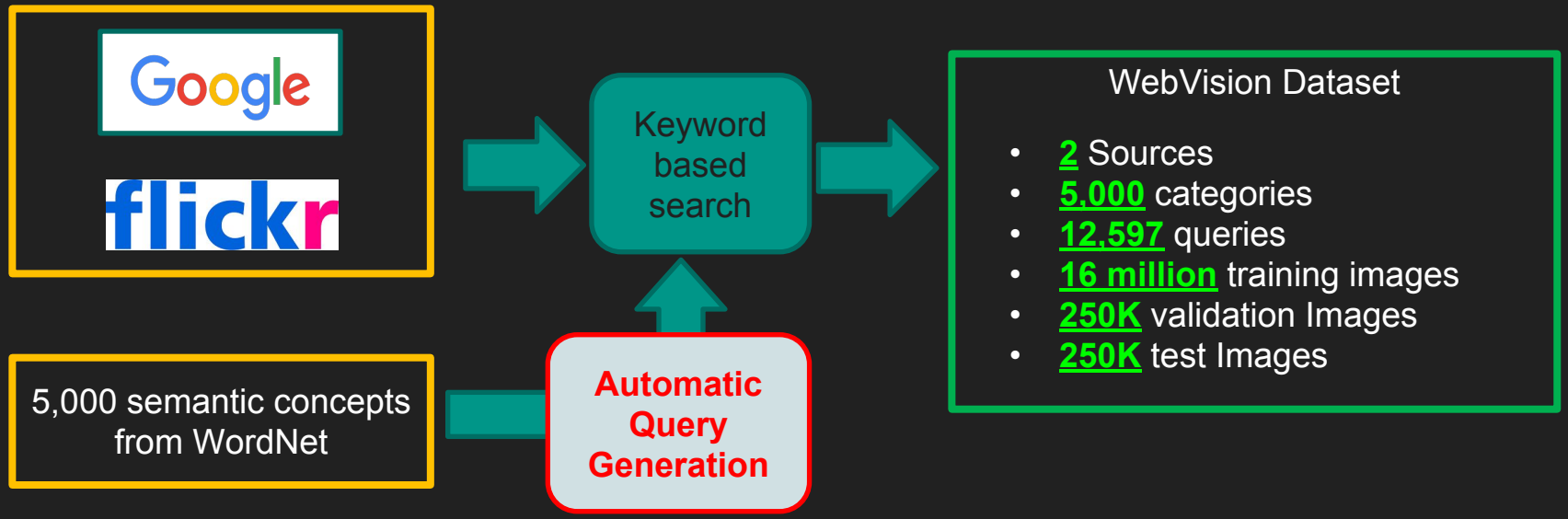
Bigger

than

Bigger

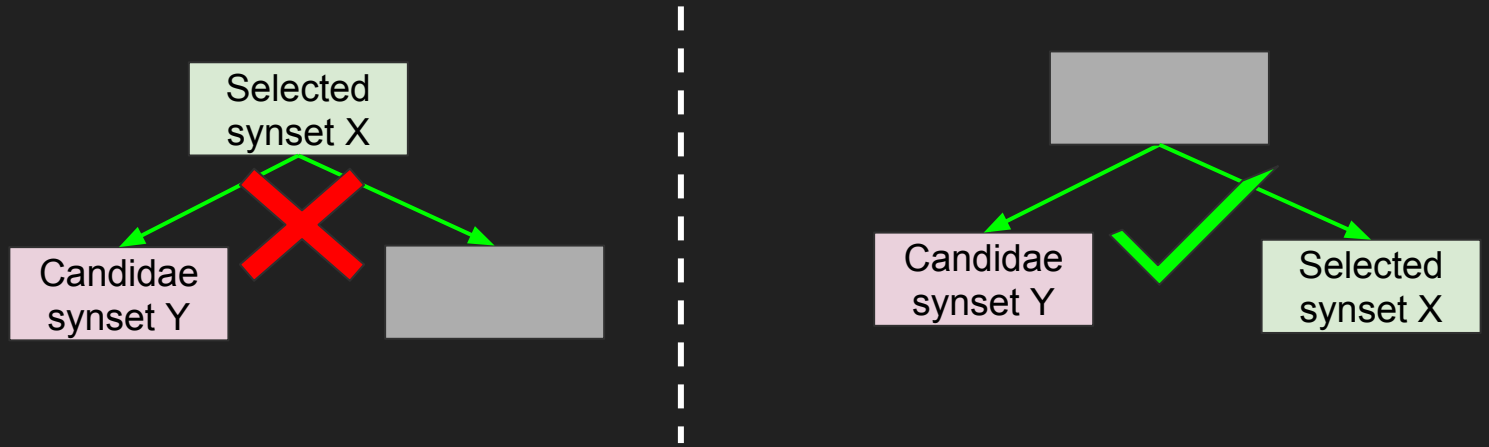
Dataset Construction

Automatic query generation instead of manual way



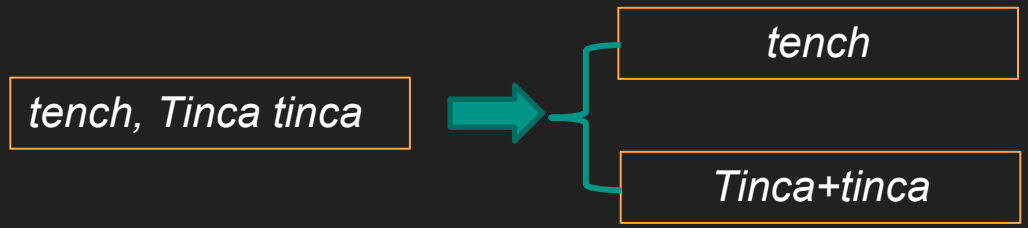
5000 Synsets

- Synsets from ILSVRC2012 dataset are the first 1,000 synsets
- The other 4,000 synsets are selected as follows
  - Sort the remaining synsets in WordNet in descending order according to popularity (the number of images in ImageNet)
  - A synset is valid if and only if it does not cause semantic overlap, i.e., there is not selected synset that is the ancestor node or child node of this synset in WordNet.



Synset to Queries

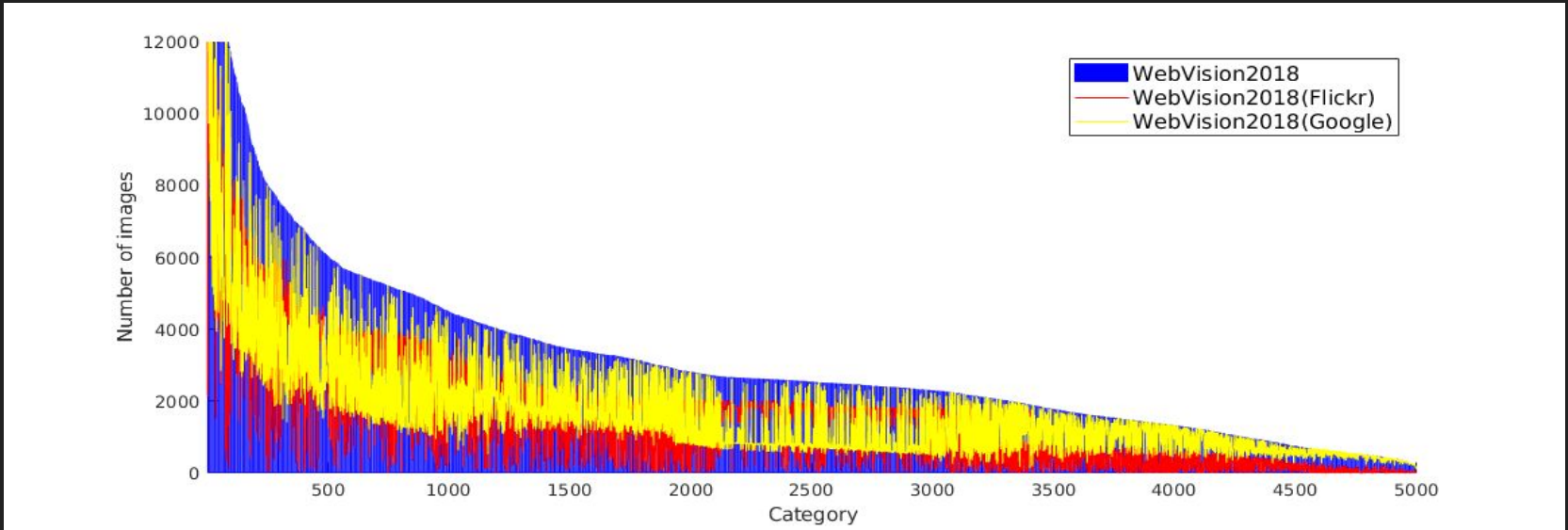
- Synsets are processed in order
- Each synset is splitted into multiple words, and each word is a query
- If a query is overlapped with existing queries, it will be discarded
- If no query is valid for a synset, we combine each word with each word in its parental node to get extended queries.
- If none of those extended query is valid, we discard this synset.
- In total, we get 12,597 queries for 5,000 synsets



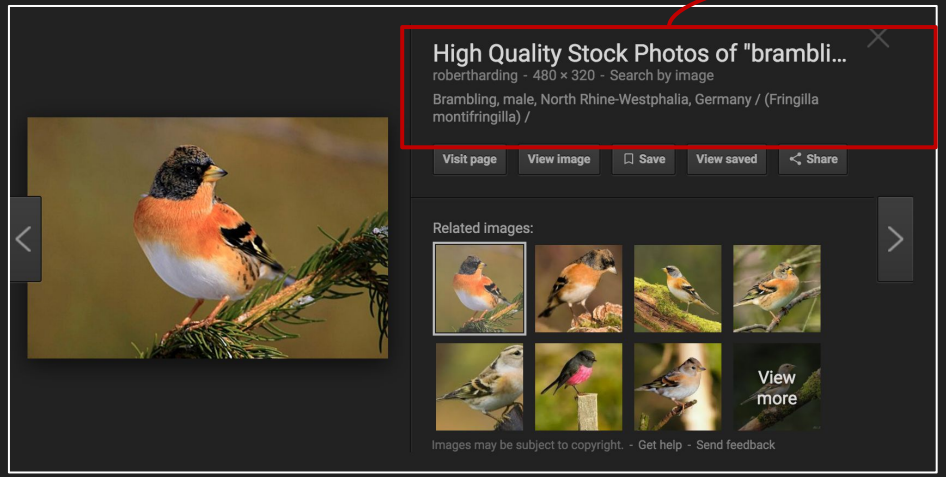
Class distribution

Highly imbalanced

#images/class varies, subject to #queries/class and the availability of images

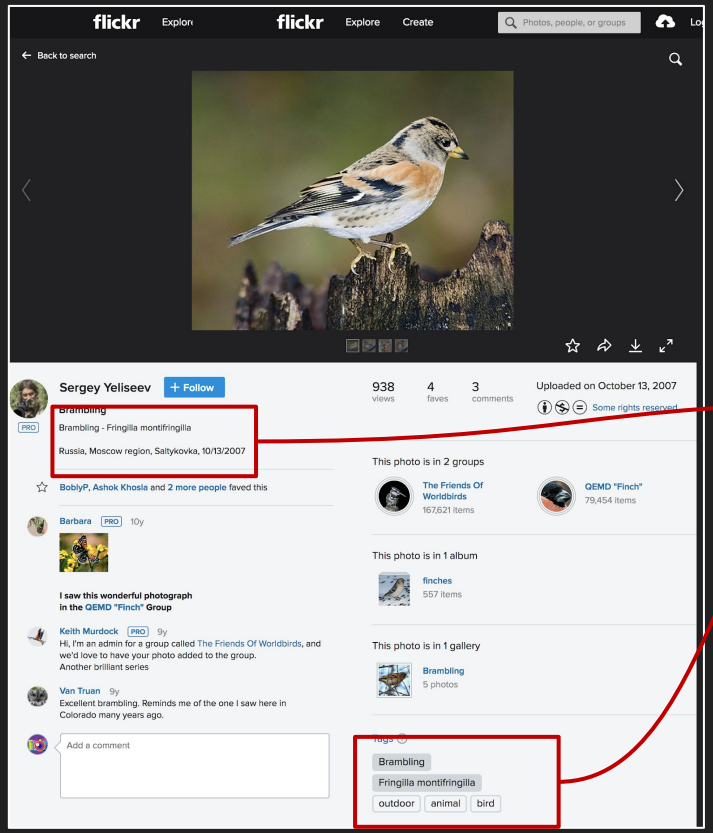


Meta Information - Google Images



- Title: ``*High Quality Stock Photos of brambling*";
- Description: ``*Brambling, male, North Rhine-Westphalia, Germany (Fringilla montifringilla)*";

Meta Information - Flickr Images



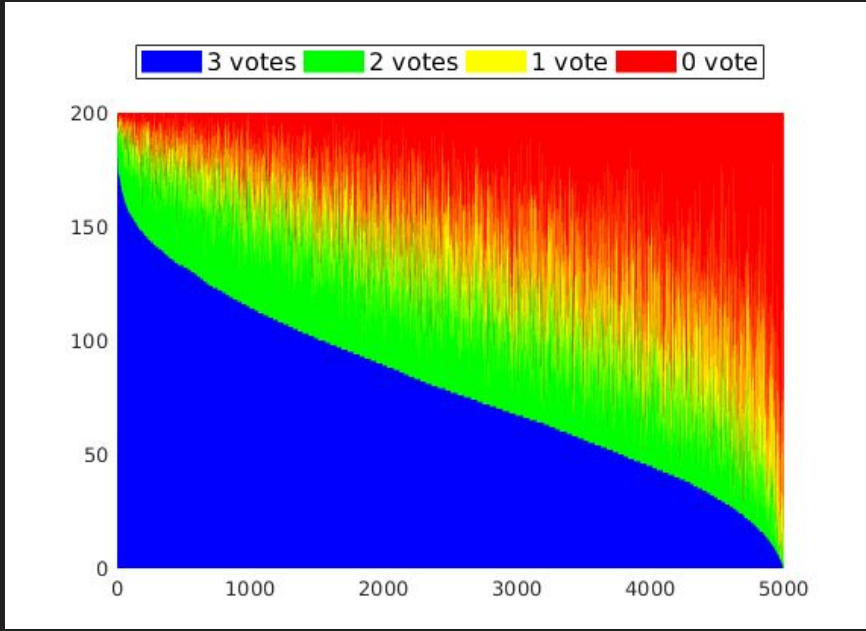
- Title: ``*Brambling*";
- Description:``*Brambling - Fringilla montifringilla Russia, Moscow region, Saltykovka, 10/13/2007*";
- Tags: "*Brambling*", "*Fringilla montifringilla*";

Noise

Ask users if the image is correctly labeled or not.

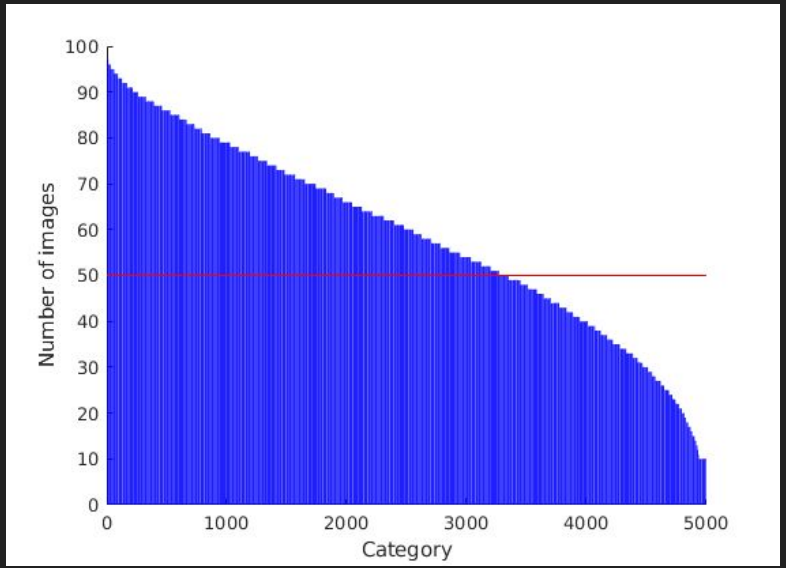
Each Image is annotated by three users.

About 59% images are inliers (with at least 2 votes).



Validation and Test Sets

- Inlier images are highly imbalanced among different classes.
- We preserve this natural imbalance in web images.
- Evenly splitting inlier images into two sets, leading to our validation and test sets.



Evaluation Metric

Due to the imbalance in number of images per class in the val/test set, we use the mean of per class top-5 accuracy as the evaluation metric,

$$ACC = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c}\sum_{i=1}^{N_c} acc(\mathbf{p}_i, y_i)$$

Summary

- A bigger web image dataset with 16M images from 5,000 categories.
- Automatic query generation from WordNet synset
- Preserve the nature of images in the wild:
  - Noisy labels,
  - imbalanced training data
  - imbalanced validation/test data
- Meta information is available