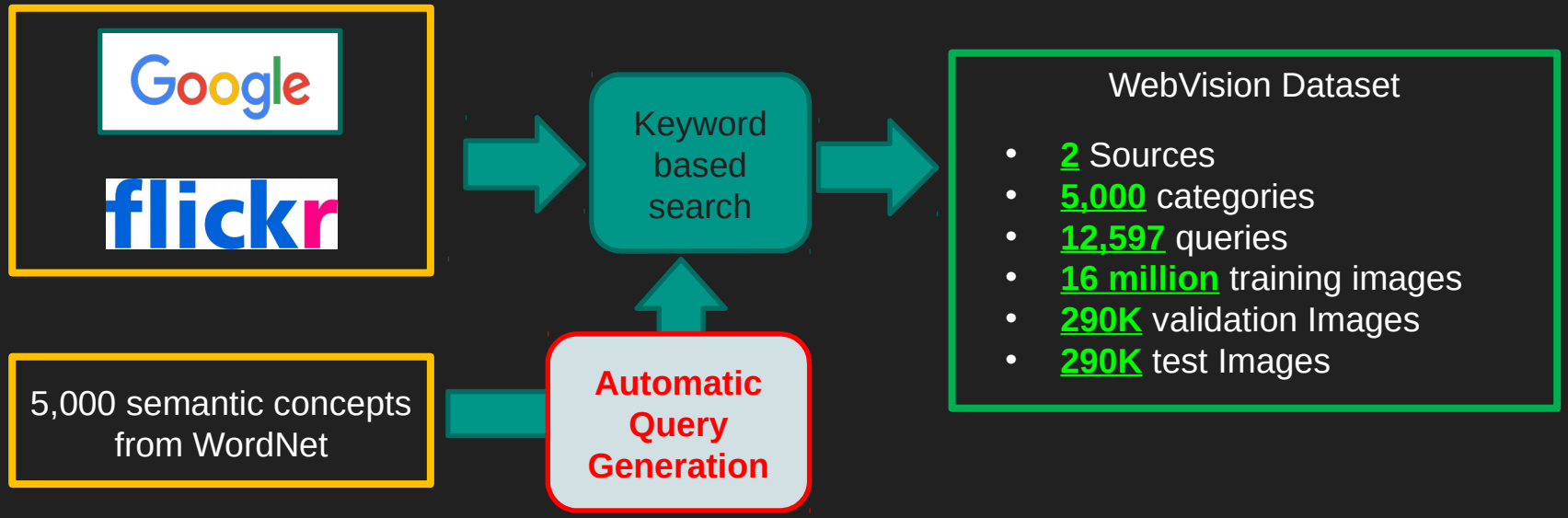
Database Overview

WebVision2.0 dataset

- **5,000** categories
- From Flickr & Google
- **16M** images
- **290K** validation images
- **290K** test images

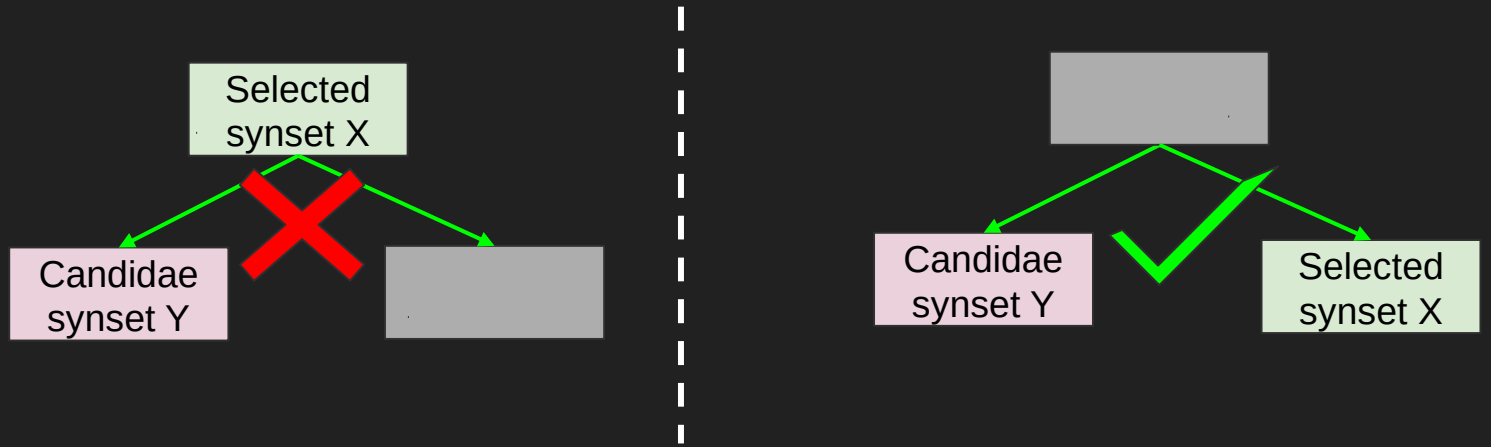
Dataset Construction

Automatic query generation instead of manual way



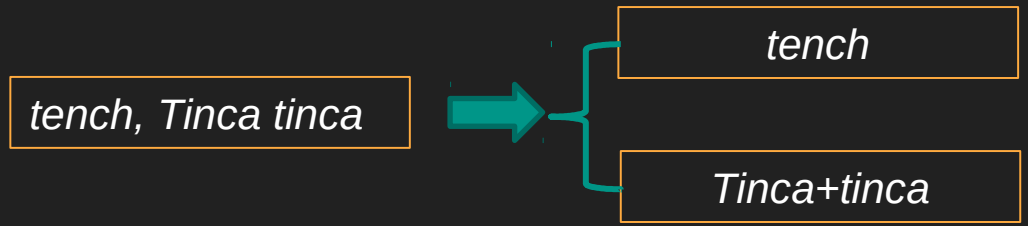
5000 Synsets

- 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

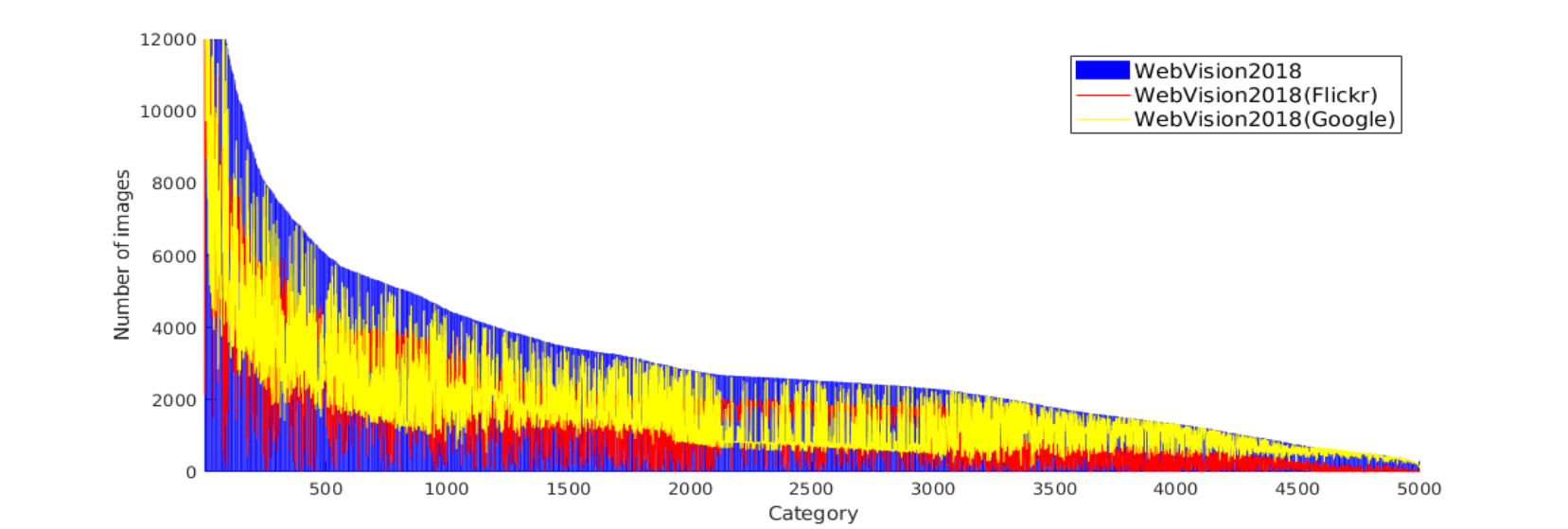
- Synsets are processed in order
- Each synset is splitted into multiple words, and each word is a query
- If a query is overlapped with existing queries, it will be discarded
- If no query is valid for a synset, we combine each word with each word in its parental node to get extended queries.
- If none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



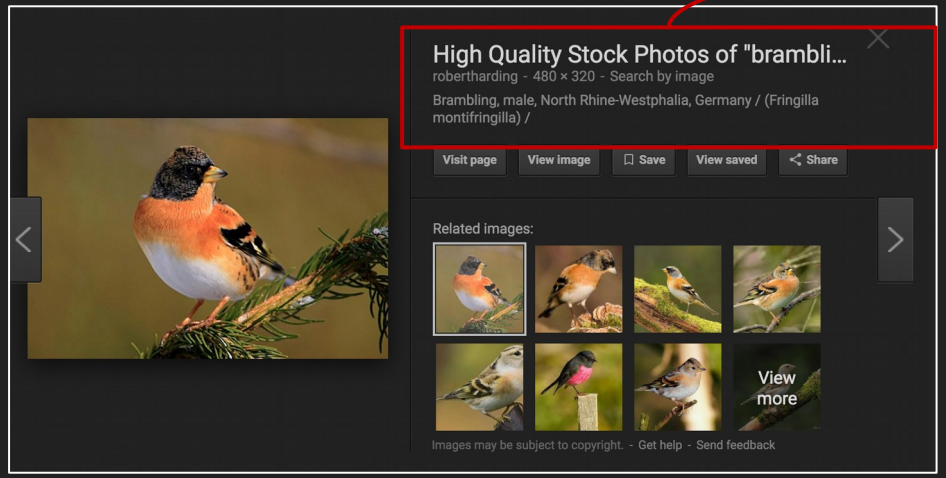
Class distribution

Highly imbalanced

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

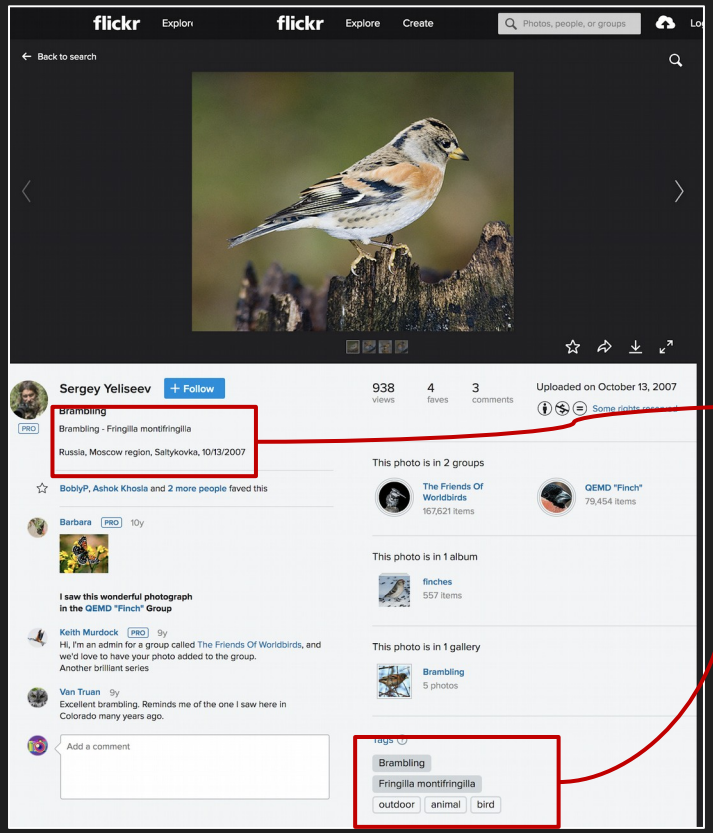


Meta Information - Google Images



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

Meta Information - Flickr Images



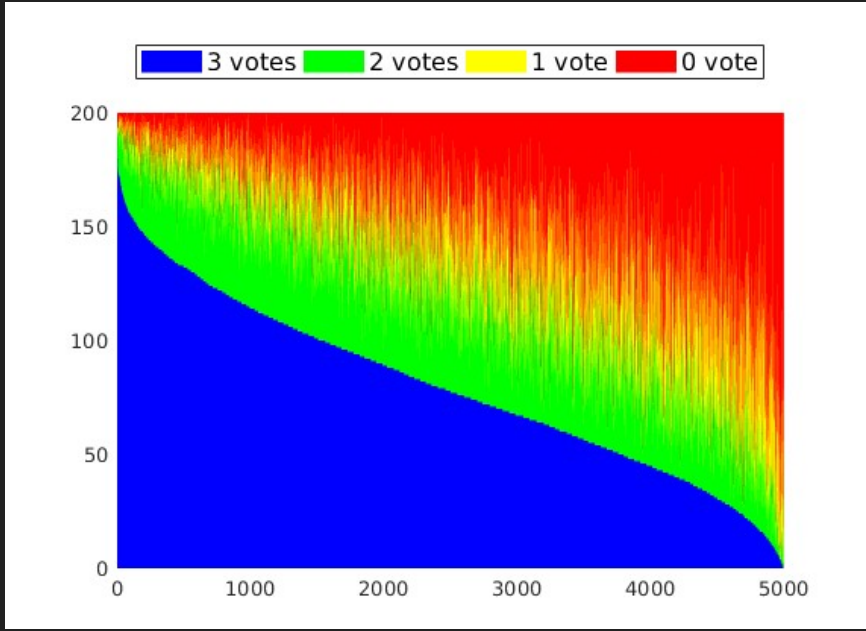
- Title: ``*Brambling*'';
- Description:``*Brambling - Fringilla montifringilla Russia, Moscow region, Saltykovka, 10/13/2007*'';
- Tags: "*Brambling*", "*Fringilla montifringilla*";

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).



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 large scale 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

Challenge Overview

Challenge Task

WebVision Image Classification Task

- Learn models on the WebVision train set and evaluate on the val and test set

Challenge Platform: CodaLab

webvision

WebVision Challenge 2020

Organized by sumansaha - Current server time: June 13, 2020, 1:30 p.m. UTC

First phase	End
Development	Competition Ends
March 28, 2020, midnight UTC	June 7, 2020, 11:59 p.m. UTC

Learn the Details | Phases | Participate | Results

- Overview
- Evaluation
- Terms and Conditions
- get_starting_kit

Challenge

The goal of this challenge is to advance the area of learning knowledge and representation from web data. The web data not only contains huge numbers of visual images, but also rich meta information concerning these visual data, which could be exploited to learn good representations and models. In 2020, we organize two tracks for WebVision challenge: (1) Track 1: WebVision Image Classification Task (2) Track 2: WebVision Task on Mining YouTube Dataset.

Challenge Schedule

Development

Start: March 28, 2020, midnight

Description: The Development Leaderboard is based on a fixed random subset of 50% of the test images. To submit, upload a .zip file containing a predictions.txt file with the prediction in the format used in the dev kit. An example submission file can be found at: https://www.vision.ee.ethz.ch/webvision/2019/files/example_submission.zip

Testing

Start: June 2, 2020, midnight

Description: To submit, upload a .zip file containing a predictions1.txt, ..., predictions5.txt file with the prediction in the format used in the dev kit. The file with the best top-5 accuracy will be used to determine the winner. Please also include a readme.txt file with a description for your entry. An example submission file can be found at: https://www.vision.ee.ethz.ch/webvision/2019/files/example_submission_testphase.zip

Competition Ends

June 7, 2020, 11:59 p.m.

Submission Policies

- Each participant may have maximum 10 submissions during development phase.
- Each team may have 1 submissions (containing 5 predictions) during test phase.
- Learn vision models from noisy data (WebVision dataset).
- No extra data is allowed to use.

FAQ Webpage

Frequently Asked Questions

- **Can I use the ImageNet images or the ImageNet pretrained models?**
  No. The main target of WebVision challenge is to push the envolope of learning visual representation without human annotations. So human annotated data is strictly prohibited to be used (Text data will be an exception). Therefore, ImageNet images or ImageNet pretrained models are not allowed to be used in any form.

- **Can I use external images without human annotations?**
  No. For fairness, we restrict the challenge to use only WebVision training images. You are not allowed to use other web image datasets like YFCC100. You are not allowed to crawl web images by yourself, too.

- **Can I use the text data (tags, description, caption) in the WebVision dataset?**
  Yes, and we encourage you to do so. It has shown in the literature that such textual information could provide useful supervision for training models.

- **Can I use external text data, or models pretrained with external text data, with or without human annotation?**
  Yes, and we also encourage you to do so. This does not conflict with our target of learning visual representation without human annotationos. Therefore, WordNet, Knowledge Graph, etc. can be used. Models trained using external text data are also allowed, such as Word2Vec, BERT models, and so on. **Note that the text data or models should be pulicly available. You should explicitly state in your final submission that what text datasets/models are used.**

- **Can I crawl text data according to WebVision concepts by myself, and use it as training data?**
  Yes. There is no restriction on non-visual data except the data should be publicly available. So people could reproduce the results. If you crawl text data by yourself, please clearly state it in your submission, and make it available to public before the final submission deadline. An URL should be provided in the method description part of your submission.

If you have other questions, please drop an email to webvisionworkshop AT gmail.com

Provided Tools https://github.com/qinenergy/webvision-2020-public

WebVision 2020: The 4th Challenge on Visual Understanding by Learning from Web Data

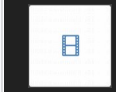
Development kit for the 2020 Visual Understanding by Learning from Web Data (WebVision) Challenge. In this repository, you can find the information you need to participate in the challenge. This includes instructions to download the data, run and eval a baseline model.

Please see the main website for competition details, rules and dates.

You can find the development kits for the two competition tracks by following these links:

Track 1: WebVision Image Classification Task

- Track website
- Submission System
- Dev Kit

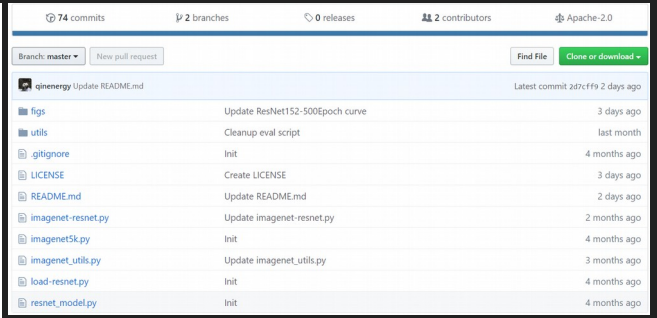


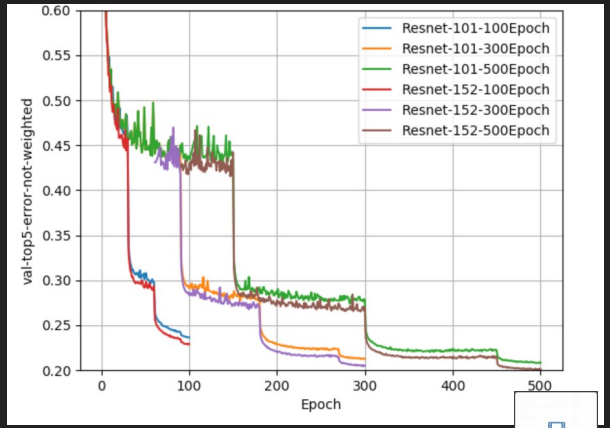
Baseline

Pretrained models

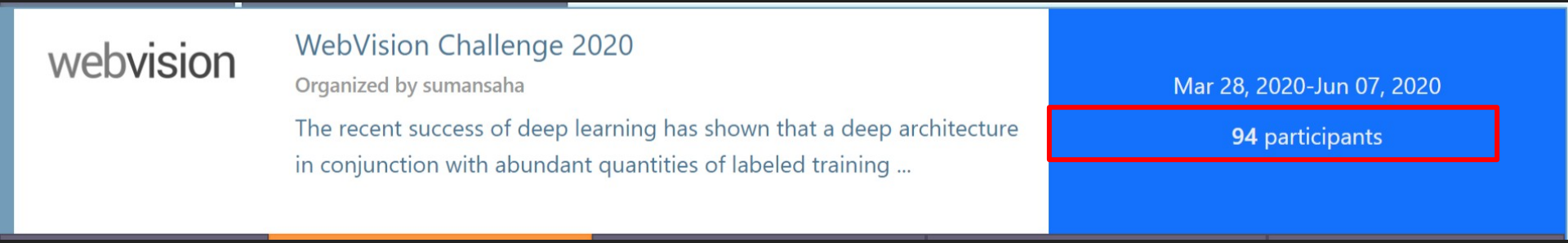
We offer several pretrained models. **Due to the class imbalance in WebVision, we duplicated the file items in train.txt such that different classes have equal number of training samples. You might want to add similar strategies in imagenet5k.py or modify your own train.txt. Check utils/upsample.py for an example.**

Model	Top1-Val-Error	Top5-Val-Error	Download
ResNet-50 (101 Epoch)	54.28%	30.69%	link
ResNet-50 (205 Epoch)	52.10%	28.51%	link
ResNet-101 (100 Epoch)	52.21%	28.62%	link
ResNet-101 (200 Epoch)	50.12%	26.78%	link
ResNet-101 (300 Epoch)	48.97%	25.74%	link
ResNet-101 (500 Epoch)	48.38%	25.21%	link
ResNeXt-101 (100 Epoch)	50.62%	27.11%	link
ResNet-152 (100 Epoch)	51.23%	27.80%	link
ResNet-152 (200 Epoch)	48.98%	25.75%	link
ResNet-152 (300 Epoch)	48.05%	24.88%	link
ResNet-152 (500 Epoch)	47.31%	24.31%	link
ResNet-152-SE (100 Epoch)	51.61%	28.02%	link





Number of participants



4 teams submitted valid results during the test phase to image classification track.

Challenge Results

Results				
Rank	Team Name	Affilication	top-5 accuracy	top-1 accuracy
1	**smart_image**	**Huawei Inc.**	**82.97 (1)**	**61.17 (1)**
2	**fISHpAM**	**Wechat AI, Tencent**	**82.01 (2)**	**59.76 (2)**
3	**pci**	**Pcitech**	**79.88 (3)**	**57.38 (3)**
4	AntVision	Unknown	77.37 (4)	53.93 (4)

Team: smart_image

Our work is implemented on Huawei ModelArts platform [1], which slightly improve accuracy while being much faster in training. As for the algorithms, the main idea is to leverage area under the margin and knowledge distillation for handling noise labels, as well as a algorithm for learning an ensemble model.

Team: fISHpAM

Modalities: Image, Query ID, text

We use pretraining and ensembling techniques to improve the performance. Using WordNet, each image can be mapped to several word tags (e.g., noun and adjective.). Then base models are pretrained with those multi-label images and different network architectures. Totally, there are 43 learned models. For ensembling, we use xgboost tool to dig the abilities of learned models with a part of training set. Other methods include large-scale finetuning, hard sampling and class-balanced sampling.

Team: PCI_AI

Modalities: Image, Query ID, meta information

Our method is based on the ResNet and ResNet variants, ResNet101、 ResNet152[1]、 ResNext101[2] and ResNest101[3]. Due to limited resources, we use fp16、 part of training samples and less training epochs to speed up. We totally trained 8 models. In the test stage, We use multi-scale、 multi-crop and multi-model fusion.

Program Schedule

9:00	Opening Remarks
9:10	Dataset/Challenge Overview
9:30	**Participant Presentation by Huawei**
9:40	**Participant Presentation by Tencent**
9:50	**Participant Presentation by Pcitech**
10:00	**Live Q&A Session**

10:15	Paper Session (ID 1-3)
10:30	Live Q&A Session
10:36	Paper Session (ID 4-6)
10:51	Live Q&A Session
11:00	Award Session & Closing Remarks