Fine Grained Image Classification for Wildlife

Srishti Yadav

Datasets, Papers and Ideas

1 Datasets mentioned in papers

1. iNat2017
   - **Data:** https://github.com/visipedia/inat_comp/tree/master/2017
2. iNat2018 and iNat2019
   - **Data:** https://github.com/visipedia/inat_comp/blob/master/2018/README.md
   - **Data:** https://github.com/visipedia/inat_comp
   - **Details:** The dataset is similar to iNat2017 with small differences, which are mentioned in the website.
3. Herbarium Dataset:
   - **Paper:** https://drive.google.com/file/d/1HPyY82IwGkKlp3ow13JCDtn1G-s_mgGJ/view
4. Cassava (leaves) images:
   - **Paper:** https://drive.google.com/file/d/1GW0Ak_fS0ZMXcy89B7di1xNF1MBIga_4/view
5. Birds:
   - **Paper:** https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Horn_Building_a_Bird_2015_CVPR_paper.pdf
6. Animal Species (camera trap)
   - **Paper:** https://arxiv.org/pdf/2004.10340.pdf
7. UCSD Birds 200
   - **Data:** http://www.vision.caltech.edu/visipedia/CUB-200.html
   - **Paper:** https://authors.library.caltech.edu/27452/1/CUB_200_2011.pdf
8. Birdsnap Large-scale Fine-grained Visual Categorization of Birds
   - **Data:** Link doesn't work anymore
   - **Paper:** https://openaccess.thecvf.com/content_cvpr_2014/papers/Berg_Birdsnap_Large-scale_Fine-grained_2014_CVPR_paper.pdf
9. Stanford Dogs
   - **Data:** http://vision.stanford.edu/aditya86/ImageNetDogs/
   - **Paper:** https://people.csail.mit.edu/khosla/papers/fgvc2011.pdf
10. Oxford Dogs

  - **Data:** https://www.robots.ox.ac.uk/~vgg/data/pets/
  - **Paper:** https://www.robots.ox.ac.uk/~vgg/publications/2012/parkhi12a/parkhi12a.pdf

11. Flowers
  - **Data:** Link mentioned in the paper doesn't work. Couldn't find the datasest
  - **Paper:** https://www.ics.uci.edu/~welling/teaching/273ASpring09/nilsback06.pdf

2 Papers

2.1 Presence-Only Geographical Priors for Fine-Grained Image Classification [1]

Objective:

- Explores how can we use additional meta-data available to make better classification (in this case animal species).
- Explores how to make best use of additional meta data which comes with most images today.

Summary:

- Knowing where a given image was taken can provide a strong prior for what objects it may contain.
- paper provide a novel training loss to capture these relationship
- The data they assemble can have unrelated image and location dataset as long as both contain the same categories.
- At test time, given an image and where and when it was taken, they aim to estimate which category it contains.
- Location information is incorporated as bayesian spatio-temporal prior. Also, during the modelling, spatio temporal (longitude, latitude, time) are independent from the image classifier
- It is difficult and time consuming to have information on where and when a given category has been a.) observed to be present and b.) observed to be absent. Hence, the paper explores presence-only setting (***novelty***)

2.2 Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection [3]

Objective:

- Focuses on birds recognition.

Summary:

- Beginner friendly paper which helps understands who are citizen scientist, why are they better than regular human/mechanical turk. It even compares quality of existing datasets and annotators to give the rationale on what/who are better and why.
- Helps understand the need of dataset in the field of wildlife and conservation despite having datsets like ImageNet and CUB-200-2011.
- Classifier they tested on was based on CNN-fc6 features and pose-normalized deep CNN (a past work by same authors).

2.3 The iNaturalist species classification and detection dataset [4]

Objective:

- Focuses on species of plants and animals captured in wide variety of situations, different camera types, varying image quality, feature large class imbalance and verified by citizen scientists.

Summary

- **Details:** There are a total of 5,089 categories in the dataset, with 579,184 training images and 95,986 validation images. For the training set, the distribution of images per category follows the observation frequency of that category by the iNaturalist community. Therefore, there is a non-uniform distribution of images per category.
- **Experiments:** Classification experiments were done using ResNet, Inception V3, Inception ResNet V2 and MobileNet
- **Known issues:** a.) Doesn't contains additional annotations such as sex and life stage attributes, habitat tags, and pixel level labels for the four super-classes that were challenging to annotate. b.) Need of an efficient algorithm that works when the test set contains classes that were never seen during training.

2.4 Analytical guidelines to increase the value of citizen science data: using eBird data to estimate species occurrence [5]

The paper dealt with what are the best approaches/checklist to make sure that the data collected by citizen scientist can be utilized in a proper way. It wasn't much about data itself or the methodologies where such data have been used. Good paper to understand about what makes good data collection checklist when working with citizen scientists.

2.5 The Unreasonable Effectiveness of Noisy Data for Fine-Grained Recognition[6]

Objective Leverage free, noisy data from the web to train effective models of fine-grained recognition.

Summary

- Interesting paper on using noisy data from the web.
- They sample images directly from Google search, using all returned images as images for a given category. For L-Bird and L-Butterfly, queries are for the scientific name of the category, and for L-Aircraft and L-Dog queries are simply for the category name (e.g."Boeing 737-200" or "Pembroke Welsh Corgi").
- Active learning-based approach to collect the data.
- The active learning begins by training a classifier on a seed set of input images and labels (i.e.the Stanford Dogs training set), then proceeds by iteratively picking a set of images to annotate, obtaining labels with human annotators, and re-training the classifier.
- Inception V3 is the base classifier
- To avoid images overlap between GT and web images, aggressive duplication procedure with all ground truth test sets and their corresponding web images is performed using a SOTA for learning similarity metric between images.

Questions:

- How reliable are the search results for a single category from web, otherwise there are a lot of False Positives being introduced.
- Is it limited to extracting category information or do we need extra labels like position, time etc.
- How much human annotators are required to be involved or are they even involved?
- Haven't worked much with active learning but since algo queries user to label data from time to time; does this query apply on a subset of the images to verify the annotations or create annotations.

3 Problems that can be explored:

1. Images in datasets represent usually one region. By limiting the geographic region, the flora and fauna seen across the locations remain consistent. However, this introduces the problem of algorithm not being useful for new regions.
2. How to identify species never seen before
3. Problem of long- tailed data distribution (i.e., a few classes account for most of the data, while most classes are under-represented) - Currently being solved using Low-shot learning at various places.
   - **Note:** From my personal observation, when we have higher number of samples from one class and very low samples of other class, it becomes hard to train the model to learn both classes. Deeper model sometimes don't work; not only they make computation slow, I personally didn't find that results improved significantly. I tried the approach to weight the datasets inversely proportional to their sample size. The idea was that the class with less sample will have more weightage etc. and hence I expected that model might work better. It did not. We augmented the data to increase the samples, but, it still did not help. I

couldn't come up with a better weighing strategy to balance the classes to improve (converge) loss over time. It's a small problem which can be explored for datasets like iNat etc. Works like [1] have attempted to solve this issue using joint- enbedding space.

4. Can we apply transfer learning here i.e. apply results from well-represented categories to least-represented one?
5. How to augment data from different sources. Work like [2] attempts to do the same with camera wild traps by using data from citizen scientist and remote sensing, together.
6. How to choose proper and right amount of relevant instances from a training set of geotagged observations? (Found this question one of the papers so noted here)
7. In iNat2017, bounding boxes were not collected on the Plantae,Fungi, Protozoa or Chromista super-classes because these super-classes exhibit properties that make it difficult to box the individual instances (e.g. close up of trees, bushes, kelp,etc.). An alternate form of pixel annotations need to be explored potentially from a more specialized group of crowd worker that may be more appropriate for these classes.
8. Most datasets are either from North/South America (animal species) or Africa (plants, food). Potential to explore more regions from the point of dataset, if available. In particular for camera traps, websites like Wildlife Insights show data majority for South America and Africa.
9. How can we scale our work using freely available (noisy) data? Work like [6] shows an interesting approach.
10. How to classify the images correctly when the distinct features of the species are obscured or multiple species are in the same scenes

Some recent papers like [7] discuss data unification efforts (I am assuming they mean going from production to pipeline).

References

1. Mac Aodha, O., Cole, E., & Perona, P. (2019). Presence-only geographical priors for fine-grained image classification. In Proceedings of the IEEE International Conference on Computer Vision (pp. 9596-9606).
2. Beery, S., Cole, E., & Gjoka, A. (2020). The iWildCam 2020 Competition Dataset. arXiv preprint arXiv:2004.10340.
3. Van Horn, G., Branson, S., Farrell, R., Haber, S., Barry, J., Ipeirotis, P., ... & Belongie, S. (2015). Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 595-604).
4. Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., ... & Belongie, S. (2018). The inaturalist species classification and detection dataset. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8769-8778).
5. Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Gutierrez, V. R., Robinson, O. J., Miller, E. T., ... & Fink, D. (2020). Analytical guidelines to increase the value of citizen science data: using eBird data to estimate species occurrence. bioRxiv, 574392.
6. Krause, J., Sapp, B., Howard, A., Zhou, H., Toshev, A., Duerig, T., ... & Fei-Fei, L. (2016, October). The unreasonable effectiveness of noisy data for fine-grained recognition. In European Conference on Computer Vision (pp. 301-320). Springer, Cham.

7. Kulkarni, S., Gadot, T., Luo, C., Birch, T., & Fegraus, E. (2020). Unifying data for fine-grained visual species classification. arXiv preprint arXiv:2009.11433.