Statistical ML Project Proposal: Kaggle Landmarks Challenges

Anirudha Sundaresan, Maxence Dutreix, Nathan Hatch, Raghav Kuppan,
Pranav Shenoy Kasargod Pattanashetty

April 27, 2018

Project Summary

Deep learning has considerably improved the state of the art in image classification. Using convolutional neural network (CNN) architectures like ResNet [He et al., 2016], it is possible to achieve a top-5 classification accuracy of 96.53% on ImageNet [Deng et al., 2009]. However, there are extensions and variations of these classification problems that have not been entirely solved:

1. While large, ImageNet contains only 1000 classes. We would like computer vision systems to be able to handle the much wider variety of classes that occur in the natural world.

2. Deep learning still has trouble when the training datasets are very small. For instance, we would like a computer vision system to be able to recognize a new species of bird given only one or two training examples.

3. A related but less well-studied problem in computer vision is *retrieval*. Given an image and a database, we would like to retrieve the images from the database which are most similar to the query image.

In order to stimulate progress on these problems, Noh et al. [2017] recently released the "Google Landmarks" dataset. The dataset contains ∼1M images of ∼13K landmarks around the globe, with a very skewed distribution: some landmarks have thousands of examples, but many have just one. Associated with this dataset are two current challenges on Kaggle. For the Landmark Recognition Challenge, the goal is to identify the landmark in each query image (i.e. classification). For the Landmark Retrieval Challenge, the goal is to indicate which database images contain the same landmark as the query image (i.e. retrieval).

For this project, we propose to enter both competitions. As a baseline for the Retrieval Challenge, we will implement DELF [Noh et al., 2017]. This algorithm extracts features for the database images using a fine-tuned CNN, reduces their dimensionality using principal component analysis (PCA), and embeds the result in an indexing system for approximate nearest-neighbors search. Because DELF is itself a first attempt at large-scale image retrieval, we hope to be able to improve upon it to achieve competitive results in the Retrieval Challenge. For instance:

- DELF restricts its attention weights to be positive. That is nice for visualization, but perhaps allowing negative attention could improve the feature embeddings.

- On smaller datasets like Oxford5k [Philbin et al., 2007] and Paris6k [Philbin et al., 2008], the results are better when DELF is combined with Deep Image Retrieval [Gordo et al., 2016]. Analyzing the differences between these algorithms could suggest ways to improve DELF's performance on the larger dataset.

As a baseline for the Recognition Challenge, we will fine-tune a pretrained CNN on the Google Landmarks dataset. Novel ideas for improvements are as yet unknown. To the group members, another reason that this is interesting is that for several of us, it will be our first time doing deep learning.

Detailed Project Description

See project summary.

Collaboration Plan

References

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

Albert Gordo, Jon Almazán, Jerome Revaud, and Diane Larlus. Deep image retrieval: Learning global representations for image search. In *European Conference on Computer Vision*, pages 241–257. Springer, 2016.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Hyeonwoo Noh, Andre Araujo, Jack Sim, Tobias Weyand, and Bohyung Han. Large-scale image retrieval with attentive deep local features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3456–3465, 2017.

James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*. IEEE Computer Society, 2007. ISBN 1-4244-1179-3. URL http://dblp.uni-trier.de/db/conf/cvpr/cvpr2007.html#PhilbinCISZ07.

James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.