Google Landmark Recognition Challenge

1. Abstract

We present the proposal of building a convolutional network system in attempt to recognize famous landmarks around the world.

2. Instroduction

While ImageNet attracts a lot of attention and a lot of models achieve high accuracy, the computer vision area lacks models for recognizing landmarks. We will be building a convolutional network model using Google-landmarks dataset since it is the largest landmark dataset available. This posts a huge challenge due to the smaller size of the dataset compare to ImageNet[1], and also the shared features between landmarks that were built in the same era and architectural style.

3. Related Work

There have been numerous innovations in the architectures of convolutional neural networks that drastically boosted the accuracy of some complicated image classification tasks. ResNet [2], for example, successfully obtained a Top-5 classification accuracy of 96.53% on ImageNet. Later innovations including VGG, Inception and different variations of ResNet have also improved the accuracy and efficiency in the image classification field. By using the combinations of these architectures, the previous competitors in the Kaggle Google Landmark Recognition Challenge [3] were able to achieve fairly good results. The solution given by the first-place group used ResNet-101, ResNeXt-101, SE-ResNet-10, SE-ResNeXt-101 and SENet-154 as their backbone networks [4]. The third-place solution used FishNet-150, ResNet-101 and SE-ResNeXt-101 as backbones. The evaluation metrics for the final test is the Global Average Precision (GAP) [3] Both groups achieved around 0.3 GAP score.

4. Technical overview

4.1. Data Cleaning

The dataset, Google-Landmarks-v2, provided by Google, contains 5 million images of more than 200,000 different landmarks. The images were collected from photographers around the world who labeled their photos and supplemented them with historical and lesser-known images from Wikimedia Commons [5].

For the data cleaning stage, we are considering adopting similar strategies mentioned by Ozaki et al [6]. The first step is to remove all classes with no more than 3 training samples (53,435 classes in total). Then by applying spatial verification to the filtered images by k nearest neighbor search, we expect the cleaned dataset contains around 2 million images with roughly 100,000 labels.

4.2. Modeling

We Will be doing our training on AWS instance due to more powerful machine and will be using TensorFlow to construct our CNN and experiment with different hyperparamters including number of layers, learning rate, epoch and batch size. Based on the experience of previous competitors, the variations of ResNet, ResNet-101 and SE-ResNeXt-101, achieved fairly good scores. We are considering using these two achitectures as backbones trained with cosine-softmax or softmax based losses.

5. Expected Outcome

After applying the data cleaning method, the training data should not contain classes with the size of the training samples smaller than three. Visually unrelated images within the same class should also be discarded. After the meticulous construction of models and tuning of the parameters, we expect our model to achieve around 0.25 in GAP score.

References

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015. 1

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2016.90 1

[3] "Kaggle Google Landmark Recognition Challenge 2019 second google landmark recognition challenge," https://www.google.com/search?q=kaggle+google+landmark, accessed: 2019-10-15. 1

[4] Y. Gu and C. Li, "Team jl solution to google landmark recognition 2019," 2019. 1

[5] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, "Large-scale image retrieval with attentive deep local features," *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. [Online]. Available: http://dx.doi.org/10.1109/iccv.2017.374 1

[6] K. Ozaki and S. Yokoo, "Large-scale landmark retrieval/recognition under a noisy and diverse dataset," *ArXiv*, vol. abs/1906.04087, 2019. 1