Image Label Comparison - Google Image Dataset V5 and ResNet50

Chan Jang, Li-Chuan Lo, Samar Manjeshwar

University of Maryland, Baltimore County

Fall 2020

Table of Contents

1. Abstract 3
2. Literature Review 3
3. Methodology 4
4. Analysis 4
5. Model Evaluation 5
6. References 7

**Abstract**

For this project, our team set out to compare the image labels between the Google Image Dataset V5 and a pre-trained convolutional neural network model - ResNet50. Utilizing the University of Maryland - Baltimore County’s Big Data Cluster, our team was able to leverage the Hadoop Distributed File System to carry out this project. After initial exploratory data analysis, we realized the differences in labels between the Google Image Dataset V5 and ResNet50. This shifted the team’s data subset focus from human body parts to reptiles. Having chosen reptiles as our data subset within the Google Image Dataset V5, our analysis showed that ResNet50 goes a layer deeper than the Google Image Dataset V5 (snakes, crocodiles, lizards, but not dinosaurs). The findings were further validated after adjusting the confidence levels from 0 to 1 for the image labels. All analysis and findings can be found in the accompanying Jupyter Notebook.

**Literature Review**

As the Data Science community broadens and continues to evolve, advances in deep convolutional neural networks have also innovated to meet the dynamic needs of the field. Specifically, in the realm of image classification, deep networks have shown that the number of stacked layers (depth) can affect the efficacy of the program (Simonyan & Zisserman, 2015 as cited in He, Zhang, Ren & Sun, 2015). This discovery has created the dynamic of adding more layers to a network in order to replicate the additional efficiency in conducting classification tasks. Unfortunately, as research shows (He & Sun, 2015 as cited in He et al., 2015) that the additional layers in fact increases the training error. This can be attributed to the degradation problem of increasing depth but losing saturation. Therefore, a deep residual learning framework was introduced in an effort to tackle the degradation problem.

He et al. (2015) experimented the hypothesis that a deep residual learning framework would outperform plain networks by utilizing the ImageNet 2012 classification database. The results showed that the 34-layer ResNet had a much lower training error (25.03) when compared to a plain 34-layer network (28.54) and alleviates the degradation problem. The same incremental efficiency was shown for ResNet50 with a top-1 error rate at 22.85 (He et al., 2015, pg. 6, Table 3).

The findings from He et al. (2015) have spurred the innovation of utilizing deep residual networks in a myriad of settings. Some examples in the healthcare setting where ResNet50 has been utilized include screening for early detection of colorectal precancerous lesions (Gao, Guo, Sun & Qu, 2020) as well as lung nodule malignancy classification in chest computed tomography images (da Nobrega et al., 2020). In the current worldwide pandemic, ResNet50 has also been utilized to detect COVID-19 from CT images (Walvekar & Shinde, 2020). In each research study, the researchers utilize ResNet50’s ability to overcome the degradation problem within dense networks to accurately process images and train models within them. Our team will also be utilizing ResNet50’s superior ability to process and detect images within the Google Open Image Dataset V5.

**Methodology**

For this project our team utilized a subset of the Google Open Image Dataset V5. This subset has the bounding boxes, object segmentations and visual relationships already defined. This subset is split into 1,743,042 training images and the full validation (41,620 images) and test (125,436 images) subsets. The images are rescaled to have at most 1024 pixels on their longest side, while preserving their original aspect-ratio. The total size is 561GB plus a few GB of metadata files.

Our analysis was conducted on the UMBC Big Data Cluster which was preconfigured to have the subset of the Google Open Image Dataset. The UMBC Big Data Cluster consists of eight compute nodes (2x18 Core Xeon, 384 GB RAM, 48 TB Disk), for a total of 288 Cores, 3 TB RAM, 384 Disk on 10 GB Network.

**Analysis**

Within the UMBC Big Data Cluster, we were able to connect to HDFS and utilize Pandas as well as a Spark Session to complete our analysis. The following code was an important part of the initial package configuration - “conf.set(‘spark.yarn.dist.files’, ‘/keras\_data/resnet50\_weights\_tf\_dim\_ordering\_tf\_kernals.h5’)” which gave access to the file and distributed the copy of file to every node in the cluster. Following this step, we were able to read in the files from the Google Image Dataset V5 and stored them in Spark dataframes. After joining the label name columns in both data frames, a frequency distribution was created to visualize the top 50 image labels. Flowers were the highest subcategory of image data within the dataset. As previously mentioned, our team decided to focus on the subset of reptiles for the scope of this project. Therefore, we created subsets of data within the reptile subcategory. This included snakes, dinosaurs, crocodiles, and lizards.

We then confirmed that all bounding boxes within these selected images were associated with at least one image. We were then able to create UDFs for each bounding box in order to identify the coordinates of edges of each bounding box. After joining the image data frame and bounding boxes for the subset of data created for reptiles, we were able to use the UDF data to extract image chips. In order to confirm that we had successfully linked the image and the corresponding bounding boxes, we displayed a random image of each subset (snakes, dinosaurs, crocodiles, and lizards). Having successfully displayed the image for each subset, we concluded by creating set labels for each subset.

**Model Evaluation**

For our chosen model, , a similar sequence was created.

[model = ResNet50(weights= f’{os.getcwd()}/resnet50\_weights\_tf\_dim\_ordering\_tf\_kernals include\_top = True)]

First, a UDF was created to evaluate the chip data of each subset of reptiles.

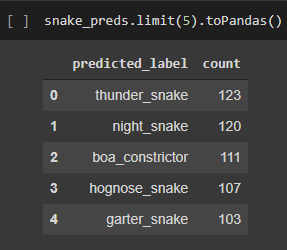
[snake\_chip\_data = snake\_subset.withColumn(“prediction”, udf\_evaluate\_chip(snake\_subset.Data))]

This was followed by prediction labels and confidence values from the returned map.

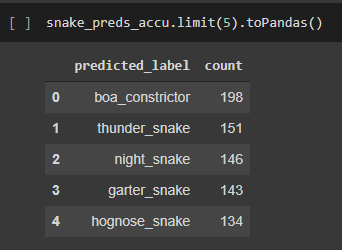
[snake\_predictions = snake\_chip\_Data.limit(6)]

[snake\_predictions = snake\_chip\_data.select(explode(col(“prediction)).alias(“predicted\_label”, “predicted\_score”]

As shown below for the snake subdata, ResNet50 identifies the type of snake in each image. This is not the case within the Google Image Dataset V5 where images did not specify the type of snake, crocodile, dinosaur, or lizard.



In an effort to increase the accuracy of the labels, we filtered the chip data subsets to include those records that had a confidence score of 1 and ran the predictions again. This process showed a significant increase in image label counts for snakes, crocodiles, and lizards but not much on dinosaurs.



With this information, our team was able to conclude that the image labels from the pre-trained model (ResNet50) showed a greater level of specificity for images within the reptile subcategory. Future research can replicate this study as well as the other dimensions of the Google Image Dataset V5. Our findings support the research done by He et al., 2015 in that ResNet50 can in fact mitigate the degradation issue of deep neural networks. The image label and prediction were highly accurate for the reptile dimension.

**References**

da Nóbrega, R. V. M., Rebouças Filho, P. P., Rodrigues, M. B., da Silva, S. P. P., Dourado

Júnior, C. M. J. M., & de Albuquerque, V. H. C. (2020). Lung nodule malignancy

classification in chest computed tomography images using transfer learning and

convolutional neural networks. *Neural Computing & Applications, 32*(15), 11065–11082.

<https://doi-org.proxy-bc.researchport.umd.edu/10.1007/s00521-018-3895-1>

Gao, J., Guo, Y., Sun, Y. & Qu, G. (2020). Application of deep learning for early screening of

colorectal precancerous lesions under white light endoscopy. *Computational and*

*Mathematical Methods in Medicine*. <https://doi.org/10.1155/2020/8374317>

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Deep residual learning for image recognition. arXiv

preprint arXiv:1512.03385.

Walvekar, S., & Shinde, D. (2020). Detection of COVID-19 from ct images using resnet50.

*International Conference on Communication and Information Processing.*