COCO-AFRICA: A CURATION TOOL AND DATASET OF COMMON OBJECTS IN THE CONTEXT OF AFRICA

Victor Dibia*
IBM Research

1 ABSTRACT

This work introduces COCOA-Africa, a tool and dataset to enable the curation of image data relevant to the context of Africa. Following examples from existing datasets, this work focuses on gathering images on scenes, activities and objects that are more likely to be observed within Africa and are largely underrepresented in existing datasets. The dataset can then be used to models on tasks that can be meaningfully applied to local problems, inspire interest in AI experimentation from African students and augment existing datasets to improve representativeness for image classification, image generation and object detection tasks.

2 THE AFRICA CONTEXT

Over the last decade, the availability of large datasets has greatly enable scientific research and commercialized solutions for supervised learning tasks. Datasets such as ImageNet [1] - a dataset of over 14 million images used in the ILSVRC challenge [4, 8] - has been instrumental to achieving near human-level performance [3] on image classification tasks. Similarly, the MSCOCO dataset [6]) provides 328k images, 2.5 million labeled instances of 91 objects in their natural context and has been valuable in training models on the task of object detection, key-point detection and object segmentation[2]. While these datasets collectively provide a comprehensive pool of image data, the representation of African content and context (salient object classes and semantic meaning) is limited. For example, images of African food, clothing items, building structures and market places may visually differ from representation within these datasets. Furthermore, image captions within existing datasets may fail to adequately capture semantic meaning needed to understand activities within the context of Africa. For example, an image captioning model trained on existing datasets may capture relationships between a *river* and the activity of *fishing* but not the activity or *drawing water*, *washing clothes* as seen within the African context.

3 DATA COLLECTION

As a starting point, we begin with curating an initial dataset based on available data which will be followed by efforts to create continuous data collection infrastructure as needed.
Data Curation: We have created a web based tool (see Figure 1) to assist in a two stage data collection process - curation and labeling. The first stage focuses on the curation of candidate images for the COCO-Africa dataset. Our tool allowed a user to query multiple image search engine APIs (flickr, Google, Bing) and view the results of these queries. Queries are parameterized using a search text and additional parameters as supported by the search engine (e.g. dates, relevance etc). Hovering over each image provided additional detail on each image (title, magnified view). Annotators can quickly add a candidate image by left clicking on it or dismiss it by right clicking. The second stage focuses on high level tagging and labeling of image in which an annotator can annotate a curated image. Similar to MSCOCO [6], our tool supports the curation of object segmentation and bounding box annotations for images, and special case for face/portrait images. The tool is also designed to curate relevant datasheet [2] information such as copyright, consent etc as applicable.

*Research Staff Member at IBM T.J Watson Research Center, Yorktown Heights, NY. USA.
[1] ImageNet Datast http://www.image-net.org/
[2] Coco dataset project http://cocodataset.org

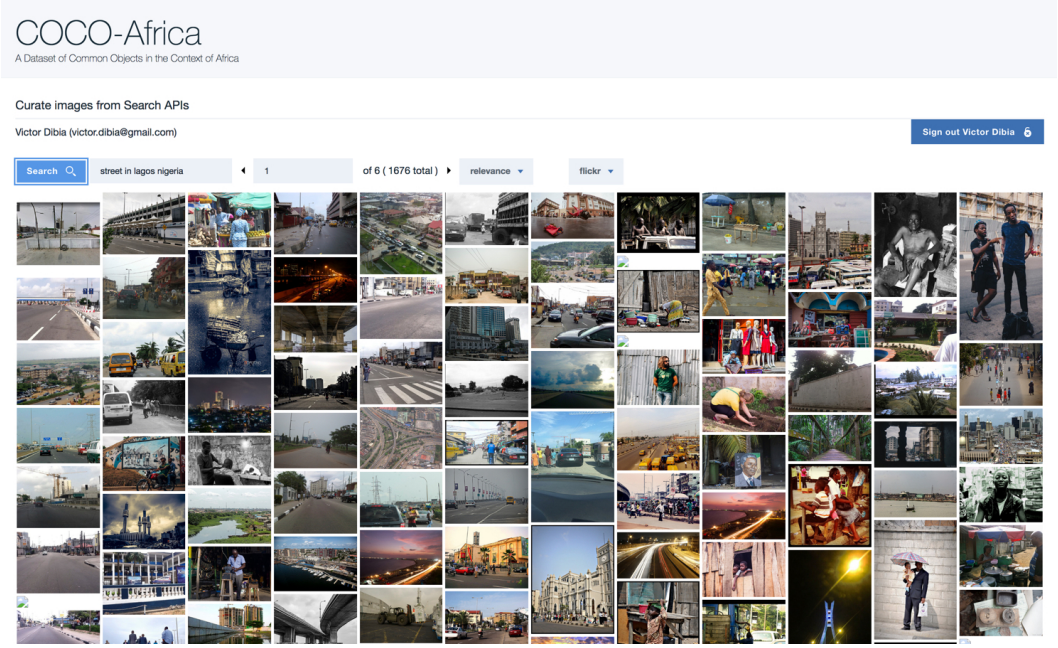

Figure 1: COCO Africa Data Curation interface. Additional information on the tool and project can be found on the Github page - https://github.com/victordibia/coco-africa

Continuous Data Collection: Based on insights on classes observed within collected data, we plan to explore the creation of a mobile app that allows users in the wild to ethically capture and contribute image data.

4 Expected Impact

The availability of datasets such as the CelebA [7] and CelebA-HQ datasets[5] have enabled research that leverage generative models in creating synthetic images. These results have profound implications for the future of digital art, creativity, film and other media industries. In addition, visual imagery specific to a region can be leveraged in predicting socio-economic markers [1] such as income and political affiliation. Assembling datasets for the Africa continent in such categories holds potential to enable similar research and application. Finally, results and examples using the COCO Africa dataset can be useful in the design of AI curriculum for students in Africa.

References

[1] Timnit Gebru, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, page 201700035, 2017.

[2] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé, and Kate Crawford. Datasheets for Datasets. 3 2018.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 11-18-Dec, pages 1026–1034, 2016.

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

[5] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *CoRR*, abs/1710.1, 10 2017.

[6] Tsung Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8693 LNCS, pages 740–755, 2014.

[7] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep Learning Face Attributes in the Wild. 11 2014.

[8] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.