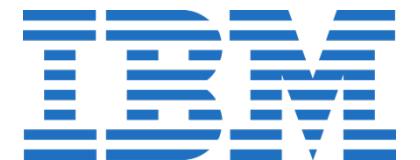


COCO-AFRICA : A Curation Tool and Dataset of Common Objects in the Context of Africa

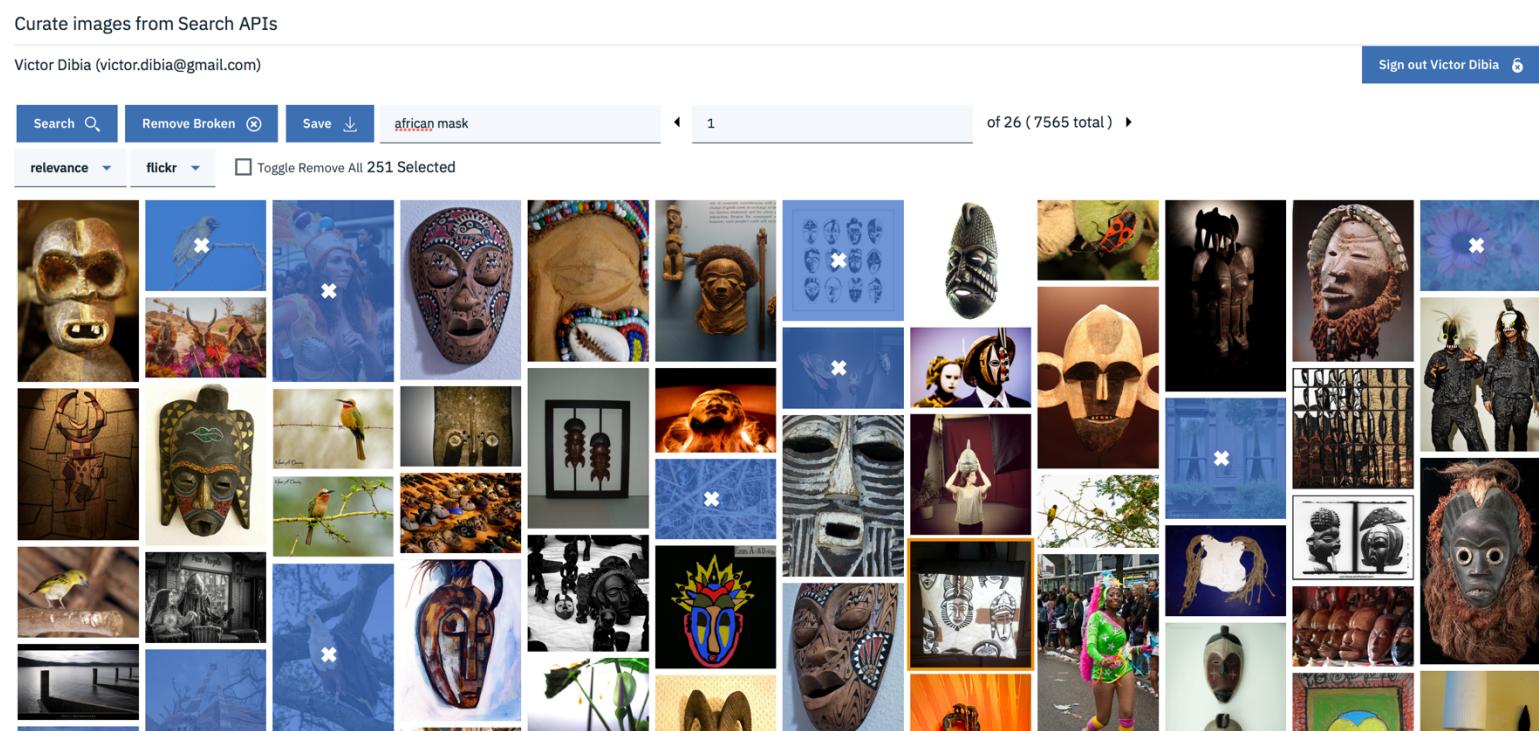


Victor Dibia | IBM TJ Watson Research Center, Yorktown Heights, NY 10598, dibiac@us.ibm.com

TLD; A data collection tool (web, mobile) and dataset for common objects in the context of Africa for teaching and experiments.

COCO-Africa

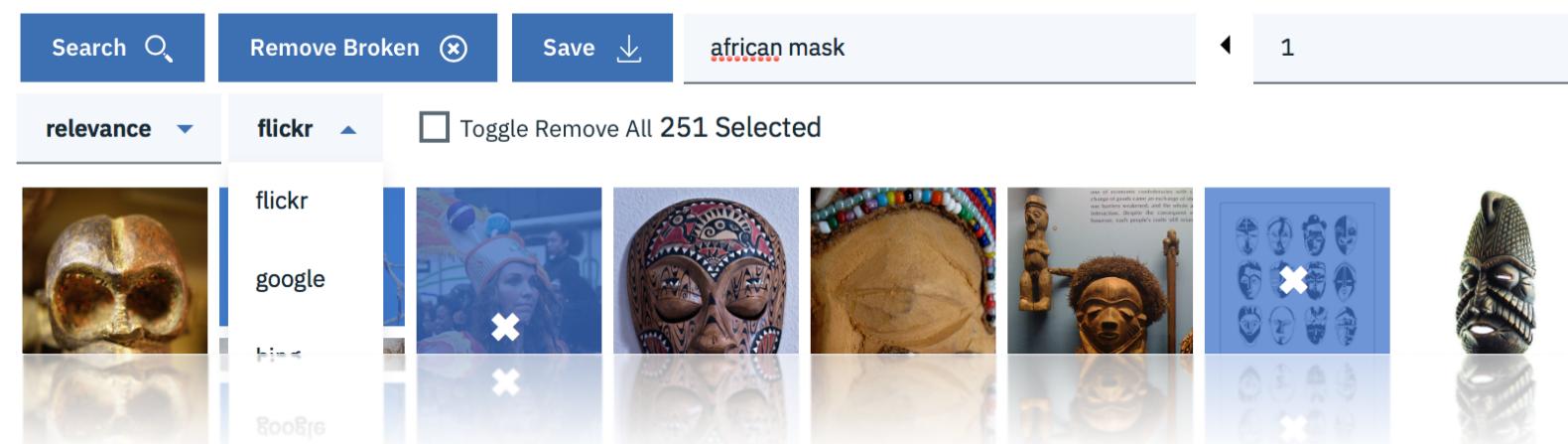
A Dataset of Common Objects in the Context of Africa



- Most available datasets today are limited in their reflection of objects and contexts within Africa.
- This work introduces COCO-Africa: a data collection tool and associated datasets focusing on objects and context of Africa.

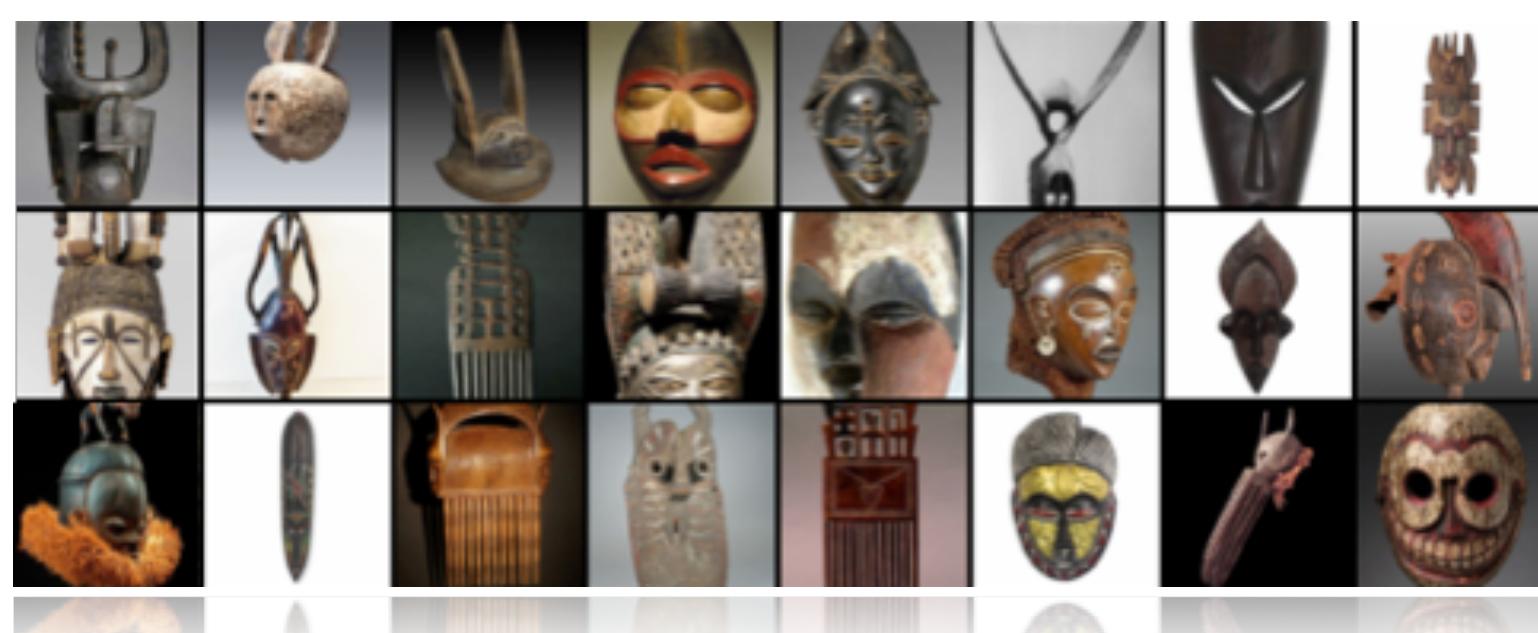
Data Collection Tool

- Paginated **web search** (support for Flickr api), **removal of broken links**, **curation of results**.
- **Ethical considerations:** The tool is being designed to preserve information on the source of each image as possible. All images remain the copyright of their respective owners.



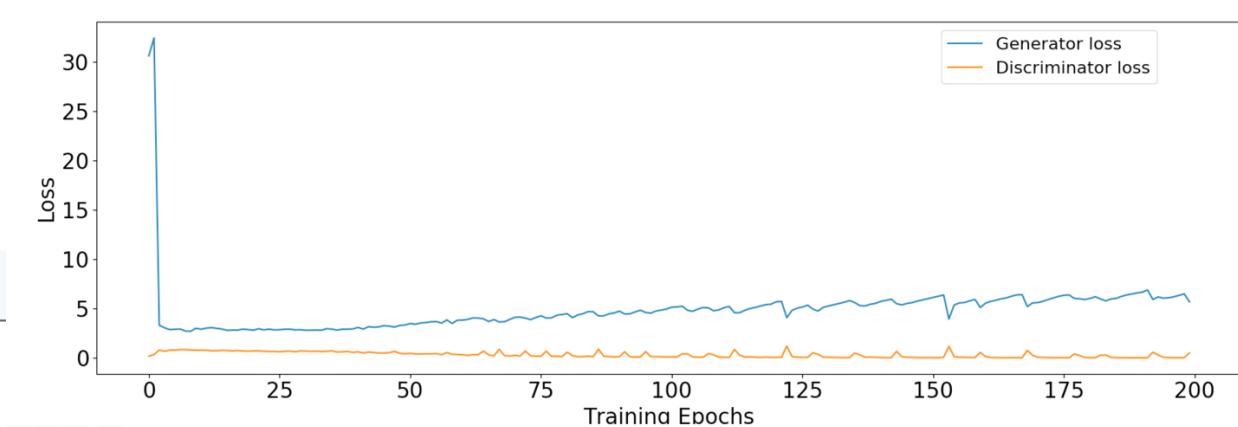
Current Dataset: COCO Africa Masks

- Manually curated dataset of 10,000 African Masks. Collected to foster artistic exploration of African art and AI.
- Introduces an additional (complex) dataset for evaluating the performance of various GAN architectures.



Learning to Generate African Masks

Baseline results using an unconditional GAN (DCGAN [1]) to generate African mask images using the our dataset.



Training parameters: BatchSize=64, beta1=0.5, imageSize=64, lr=0.0002, ndf=64, ngf=64, niter=200, nz=100. V100 GPU.

Next Steps

- **Category selection study.**
Study to finetune the list of objects and contexts which further efforts will focus on.
- **Data Collection Tools**
 - Improve web UI (support for other search sources).
 - Incentivized (mobile) data collection effort to collect images (and other data types) within the selected list of objects and contexts.
- **Baseline Models and Demonstrations**
Provide baselines for tasks associated with each dataset and trained models where possible.
- **Data cleaning and release.**
Data release in compliance with ethical guidelines for community use.

<https://github.com/victordibia/coco-africa>

Sample code, trained model, sample images.