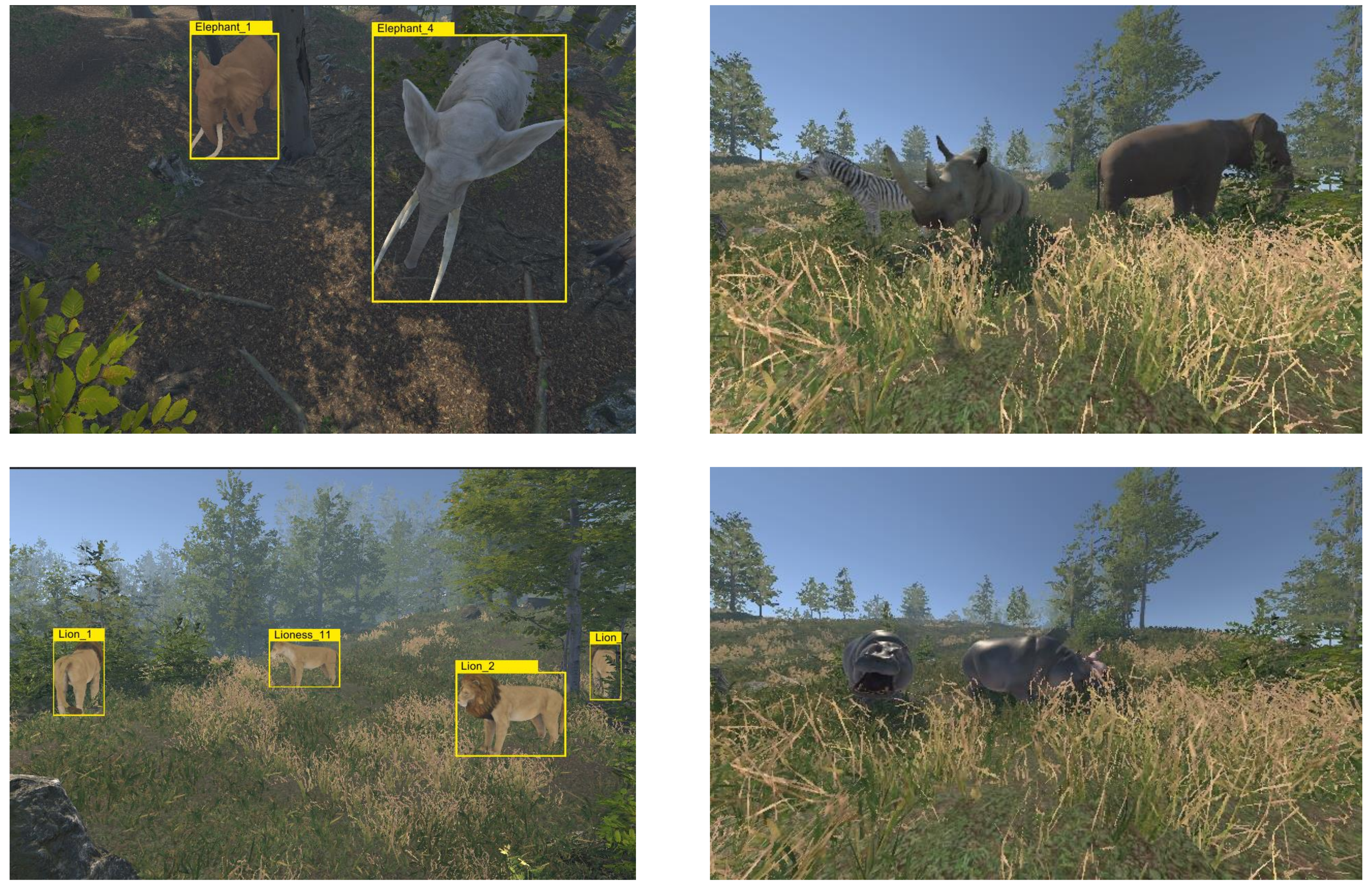
Sim-to-Real Transfer for Object Detection and Localization on Animals

Syed Talal Wasim[1], Syed N Hasany[2], Kainat Abbasi[3], Huda Feroz[3], Anisa A Ahmed[3], Mudasir H Shaikh[3], Muhammad Farhan[3]
[1]Université de Bordeaux, [2]University of Cassino, [3]Habib University

CVPR VIRTUAL JUNE 19-25

Motivation

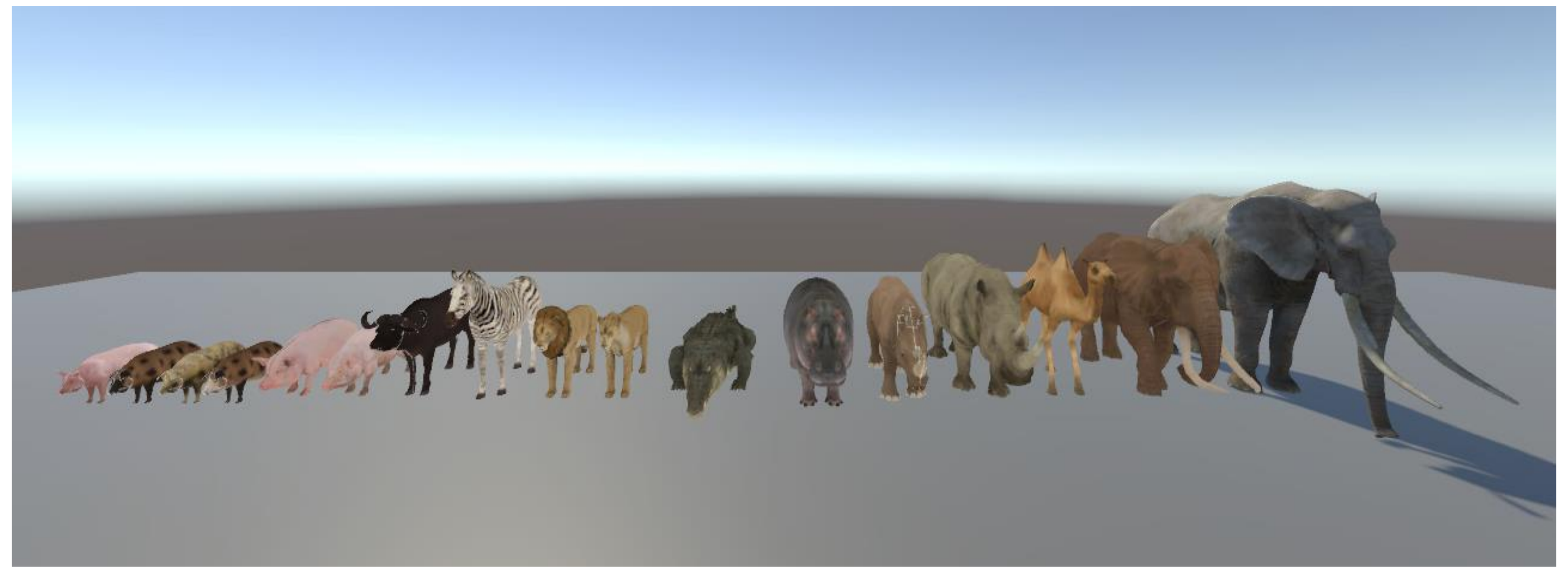
- **Animal Detection and Localization:** The task of detection and localization is central to many further tasks like tracking.
- **Lack of Hand-Labelled Animal Data:** Large-Scale Hand-Labelled data for animals is time consuming and expensive to produce
- **Simulated Data for Pretraining:** Pretraining with simulated data can improve accuracy on task with limited data
- **Simulated Data Allows More Variability:** Can create a variety of data covering many possible environments, randomizations, poses. We are in control to produce millions of optimal training samples.



Images on Left: Sample Images from Simulated Dataset 1 (Top view and flat area scene)
Images on Right: Sample Images from Simulated Dataset 2

Methodology

- Tested on the task of object detection and localization
- Datasets generated using Unity3D.
- 9 common categories of animals were chosen; Buffalo, Elephant, Rhinoceros, Zebra, Pig, Crocodile, Hippopotamus, Lion, and Camel.



- Real and Simulated (two versions) datasets are built for evaluation purposes
- Real data was gathered using images from various large scale datasets
- For Simulated Dataset 1, animals were placed in four different background scenes with each image having only one category of animals.
- For Simulated Dataset 2, the images were a more zoomed version of animals with a mixture of different categories of animals in the same background scenes as Simulated Dataset 1.
- All results were generated on YOLOv3 model using a train/validation/test split of 0.7/0.2/0.1.

Results

- Pretraining with simulated data can give performance comparable to pretraining with large scale general datasets like MS COCO across most categories
- Simulated data pretraining performs much worse on specific categories where 3D models do not capture the variability of the category (Elephant, Zebra and Buffalo).

Class	Test for Real Data		Test for Simulated Data 1		Test for Simulated Data 2	
	recall	ap	recall	ap	recall	ap
Zebra	0.933	0.898	0.372	0.160	0.319	0.127
Elephant	0.917	0.859	0.427	0.209	0.420	0.210
Lion	0.831	0.771	0.840	0.731	0.826	0.683
Crocodile	0.920	0.878	0.874	0.779	0.895	0.837
Buffalo	0.954	0.902	0.459	0.291	0.445	0.297
Hippopotamus	0.977	0.923	0.842	0.754	0.941	0.829
Rhino	0.946	0.938	0.519	0.447	0.526	0.458
Camel	0.912	0.855	0.912	0.857	0.883	0.838
Pig	0.686	0.591	0.879	0.664	0.745	0.626
mAP@50IoU		0.846		0.543		0.545

Challenges

- Background scenes lack variability; unable to capture natural habitat of different animals.
- Current 3D models of animals not diverse enough; unable to capture variety in terms of age, size, growth, intra-class features.
- Foreground objects' randomization sometimes led to the overlapping and occlusion of 3D models.
- For some 3D models, the color of foreground and background terrain is similar.