A multi-domain generalizable animal detection model: exploration of potential solutions to improve the SOA

Deliverable 4

Presented by
Abdiel Fernandez - abdielfer@gmail.com
Rose Guay Hottin - guayhottin.rose@hotmail.com
Kevin Lessard - kevin.lessard@umontreal.ca
Santino Nanini - santino.nanini@umontreal.com

Department of Computer Science and Operations Research
Faculty of Arts and Science

Work presented to Alex Hernandez-Garcia
As part of the course IFT3710/6759
Projets avancés en apprentissage automatique

March 21, 2022

Fine-tuning pre-trained Faster R-CNN (Baseline)

Firstly, we adapted the pre-trained Faster R-CNN with a ResNet-50-FPN backbone pre-trained on the COCO dataset for the purpose of doing animal detection in camera traps images. We fine-tuned the last layers of the Faster R-CNN with the recommendations made in Recognition in Terra Incognita [1]. As for data augmentation, we have currently done horizontal flipping with 0.5 probability. There are 2 classes of images resulting from this model: images containing an animal or images that are just a background.

For our first run, we did 20 epoch of training while monitoring train, cis-valid and trans-valid losses. We might do more epoch while keeping the training time down. To filter out the different predicted bounding boxes of the model, we will use Non-Maximum Suppression (NMS). We will then evaluate the remaining bounding box using some metrics. A good starting option is to use mean average precision (mAP) with intersection over union (IoU) at a 0.5 threshold. After obtaining the results, we might want to try and reduce our training time to be able to have results way faster even if we lose some precision. We might try to use ResNet-34 or 18 as a backbone and reduce the size of input images. Once the results and the training time seem fine, this will be our baseline that we will try to improve with 2 methods for in-domain and out of domain distributions.

Subspace alignment based domain adaptation (Method 1)

The first method that we will implement to improve the performance of the baseline on images from camera locations unseen during training is inspired from Raj, Namboodiri and Tuytelaars (2015) [2]. This subspace alignment based domain adaptation technique was developed for R-CNN architecture but we are adapting it to work with our Faster R-CNN model. The domain adaptation is made without having labels for the target domain.

The first step is to generate feature representations of proposed detection regions for source and target (new locations with no targets) data using the fine-tuned model. Source features associated with a bounding box having an intersection over union (IoU) score higher than a certain threshold γ are stacked in the source matrix (X_{source}) and target features associated with a detection score higher than another threshold σ are stacked in the target matrix (X_{target}). Two Principal component analysis (PCA) are performed on X_{source} and X_{target} and only the first 100 dimensions are kept for both matrices to form $X_{source\_reduced}$ and $X_{target\_reduced}$. The transformation matrix M is then created, which is the matrix that minimizes the Frobenius norm of the difference between $X_{source\_reduced}M$ and $X_{target\_reduced}$. M is used to project the source data on an aligned coordinate system which is used to train the model. When testing the adapted model, target data are also projected on the aligned coordinate system.

We will compare the performance of the adapted model to the one of the baseline on locations unseen during training. Since this technique requires reducing the size of a top layer from 1024 to 100, we will also evaluate the performance of a non-adapted model with this layer size as control.

Low-light enhancement and deblurring (Method 2)

On the other hand, we will try to apply the network called LEDNet, to perform joint low-light enhancement and deblurring to normalize contrast/luminosity of the images to improve animal detection from low-light and blur pictures. The paper with our baseline will both serve as a guideline for this given goal. The point of this method is to try to see if the model generalizes well to the data we provide as input with this normalization. [3]

References

[1] Beery, Sara, Grant Van Horn, and Pietro Perona. ”Recognition in terra incognita.” Proceedings of the European conference on computer vision (ECCV). 2018.

[2] Raj, Anant, Vinay P. Namboodiri, and Tinne Tuytelaars. ”Subspace alignment based domain adaptation for rcnn detector.” arXiv preprint arXiv:1507.05578 (2015).

[3] Shangchen Zhou, Chongyi Li, Chen Change Loy : “LEDNet: Joint Low-light Enhancement and Deblurring in the Dark”, 7 Feb 2022