Generalization of iWildCam dataset

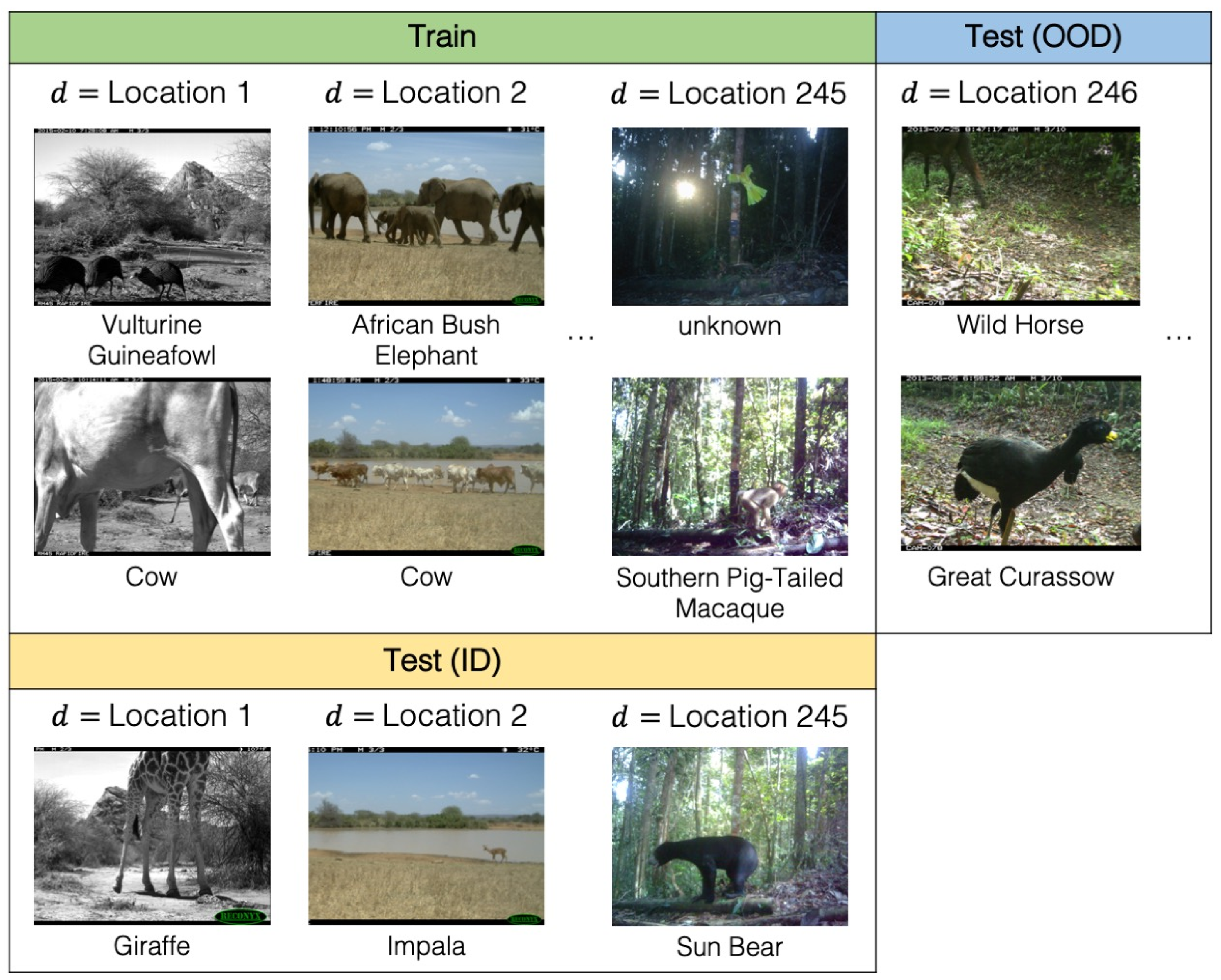
Horjea Cosmin-Marian,Vlascenco Daniel-Marian | Supervisor: Elena Burceanu

Faculty of Mathematics and Informatics, University of Bucharest

Introduction

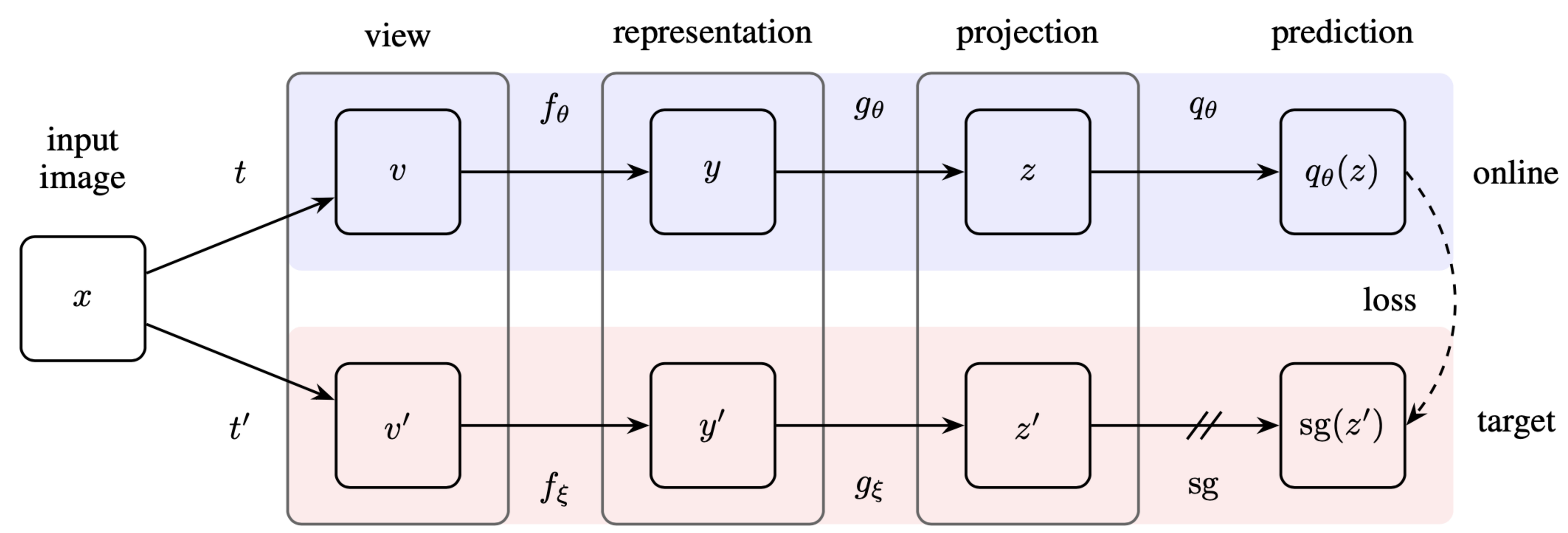
Not having a good distribution of data in a Machine Learning task can lead to our model not learning the important parts of data. Data comes from all kinds of places and it's not always as uniform as we would like it to be. We tackle this problem by using contrastive learning, more specifically we will be using BYOL to try to overcome this issue.

Dataset Description



iWildCam is a dataset part of the WILDS collection which gathered images of 182 animal species, from 313 cameras span around the globe in 20 countires. The dataset offers us both in distribution and out of distribution data splits. Moving forward we will only work with the most common 20 classes in this dataset

Method



BYOL(Bootstrap Your Own Latent) helps us deal with the generalization problem. It uses two neural networks, namely the *online* and the *target* network.They have same architecture but the target one uses different weights.For a given image x it creates two augmented views that go down the networks and at the end the *online* network it's optimized

Experiments

We first started off by training a ResNnet18 on our data. We trained it over 5 epochs using **CrossEntropyLoss** as our loss function, an **SGD** for our optimizer with a 0.01 learning rate, 0.9 momentum and 5e-4 weight decay

Model	Test-ood	Test-id
Resnet18	35.47%	53.56%

After this we decided to use the *byol-pytorch* library and try to pretrain our resnet on the ood data. After this, our accuracies look like this

Model	Test-ood	Test-id
Resnet18-pretrained	40.36%	45.27%

We see that the accuracies scores are much closer together and our model can generalize better.

Conclusions

By using a contrastive learning solution like BYOL and pretraining the model, we can decrese the impact that ood data has on our models

Future Work

iWildCam is a huge dataset, we can use the unlabeled dataset provided for our pretraining. The images are taken in sequences, which means 10 images one after the other are very similar and don't change that much since the background is the same. We can even forget about the background and only classify the animal in the picture since there is a dataset of boxes which enclose the animals in each image.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition," *https://arxiv.org/abs/1512.03385*,

[2] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, Michal Valko, "Bootstrap your own latent: A new approach to self-supervised Learning" *https://arxiv.org/abs/2006.07733*,

[3] Bootstrap Your Own Latent (BYOL), in Pytorch, https://github.com/lucidrains/byol-pytorch