හොම්බෙන් AI

HOMBEN AI

WILDLIFEDATASETS: AN OPEN-SOURCE TOOLKIT FOR ANIMAL RE-IDENTIFICATION

MegaDescriptor Methodology

Panthera
Electronic and Telecommunication Engineering
University of Moratuwa

WILDLIFE RE-IDENTIFICATION

Typically framed as a closed-set classification problem

The goal is to match new images of animals (query set) with a set of known identities (reference set).

Backbone Architecture Types

Swin - B
Transformer-based architecture

EfficientNet - B3
CNN-based architecture

Local Features Approaches

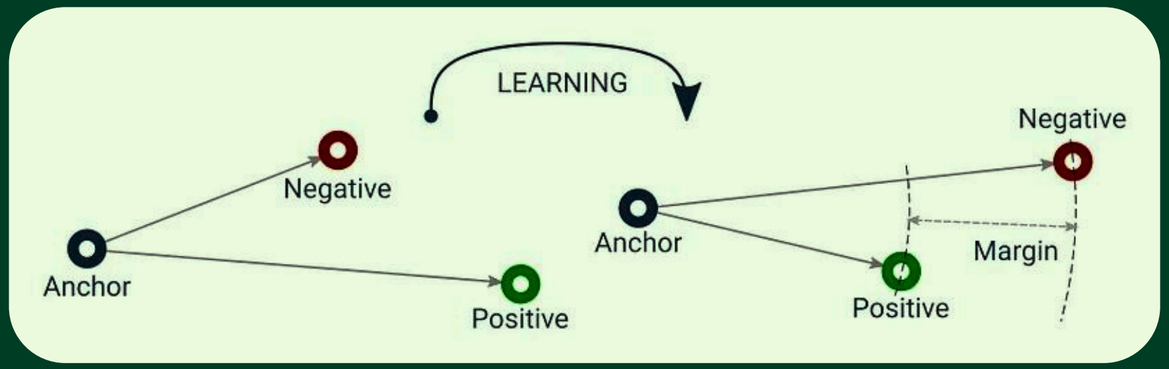
1. SIFT Discriptor
2. Superpoint Discriptor

Steps for Local Features Approaches:

- **Keypoint Extraction:** Extract keypoints (important features) and their associated descriptors (numerical representations of the keypoints) from both reference and query images.
- **Distance Calculation:** Compute the distances between all possible pairs of descriptors. Shorter distances mean the features are more similar.
- **Ratio Test:** To filter out false matches (incorrectly identified keypoints).
- **Identity Prediction:** The reference image with the highest number of matches is considered the best match for the query image.

Metric Learning Approaches

1. Triplet loss



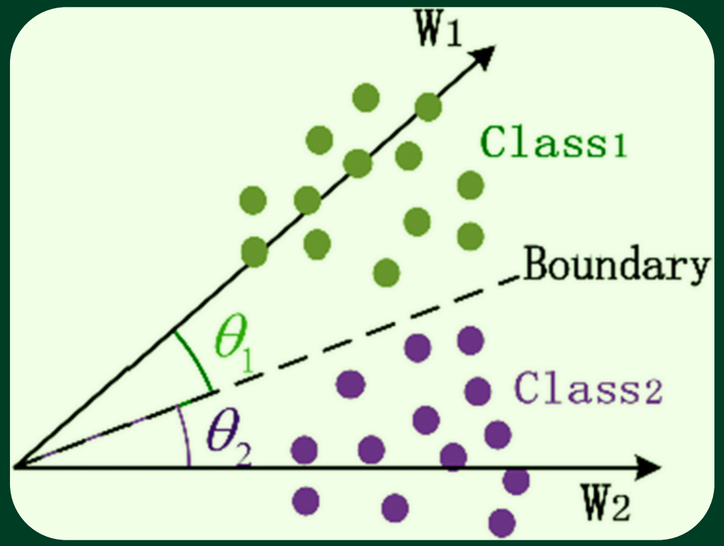
Goal is,

- Make the distance between the anchor and positive pair small.
- Make the distance between the anchor and negative pair large

There are different strategies for selecting these triplets:

All: All valid triplets in a batch.
Hard: Triplets where the negative is closer to the anchor
Semi-hard: Triplets where the positive is closer to the anchor l.

2. ArcFace loss

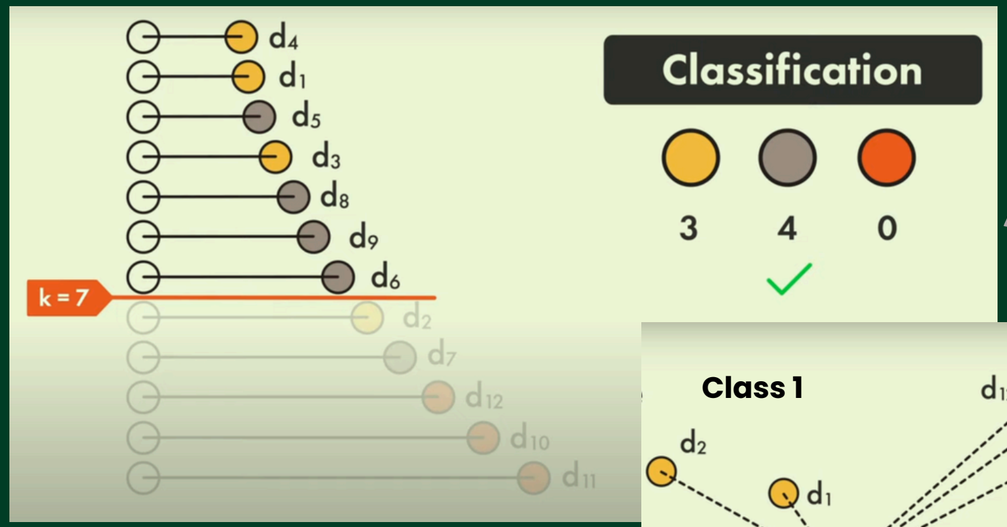


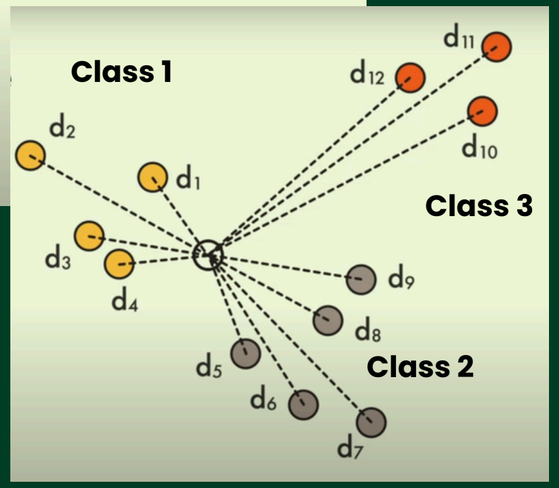
The ArcFace loss method is an enhancement of the standard classification loss (softmax), with the aim of improving the separation between different identities.

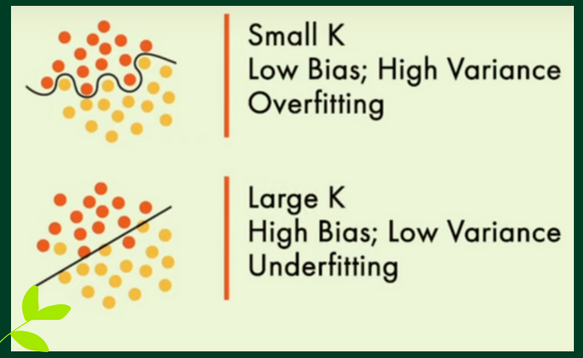
It introduces an angular margin (m) between classes

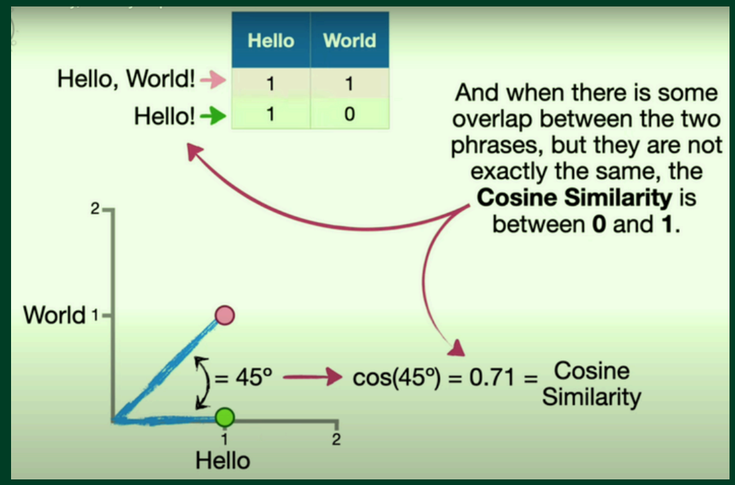
MATCHING STRATEGY

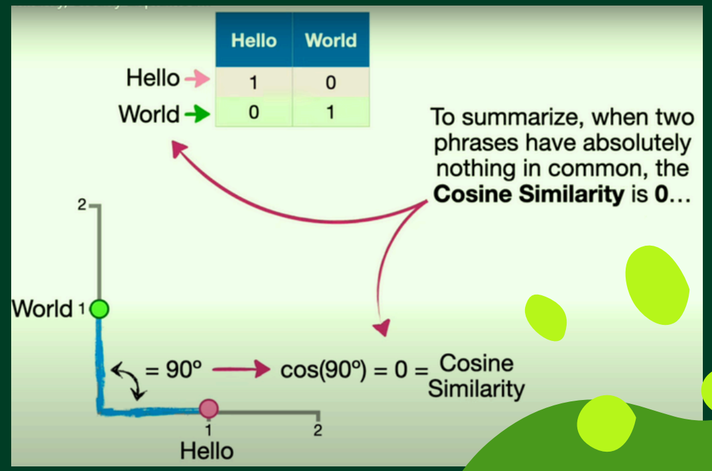
We essentially create a **1-nearest-neighbor classifier** using **cosine similarity**











TRAINING STRATEGY

- Datasets, split **80% for training** (reference set) and **20% for testing** (query set).

- A **closed-set problem** : All identities in the query set are also in the reference set.

- The models are optimized using the **SGD optimizer** with momentum for **100 epochs**

HYPERPARAMETER TUNNING

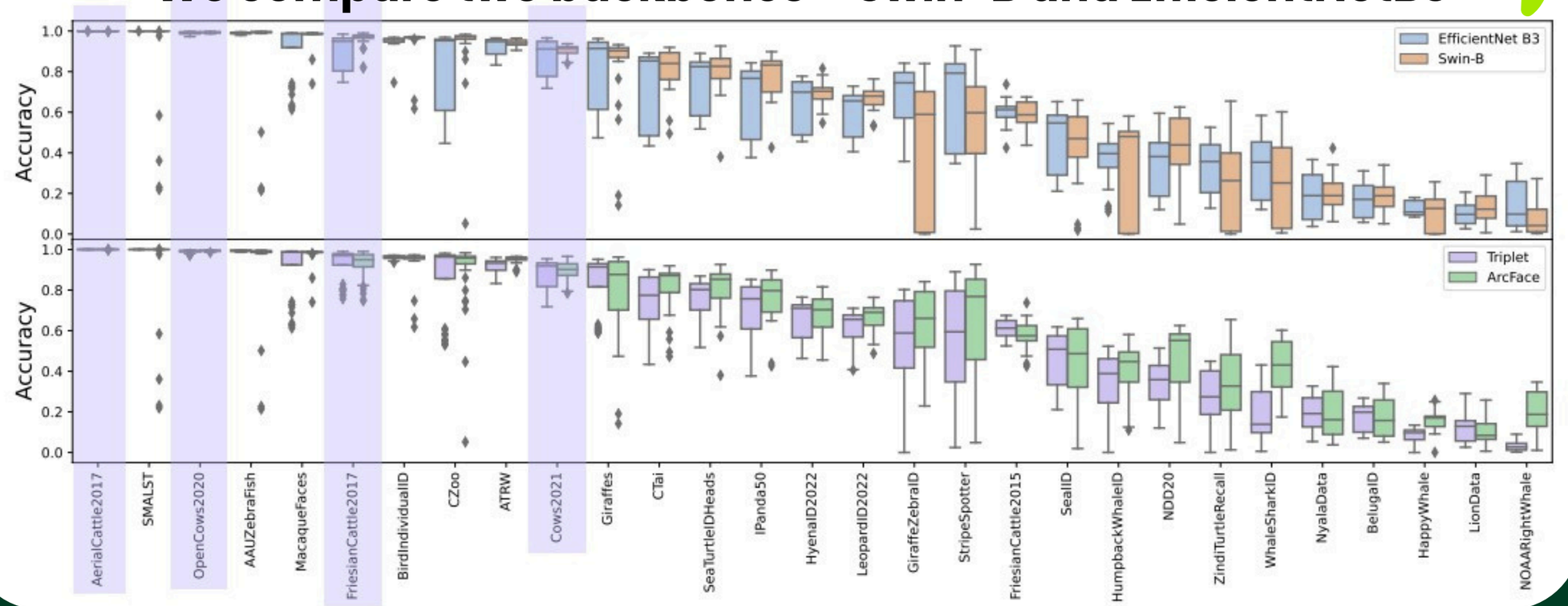
The performance of the metric learning approaches is usually highly dependent on training data and optimization hyperparameters

Backbone	$\{\texttt{Swin}-\texttt{B}, \texttt{EfficientNet}-\texttt{B3}\}$
Learning rate	$\{0.01, 0.001\}$
ArcFace margin	$\{0.25, 0.5, 0.75\}$
ArcFace scale	$\{32, 64, 128\}$
Triplet mining	$\{\texttt{all}, \texttt{semi}, \texttt{hard}\}$
Triplet margin	$\{0.1, 0.2, 0.3\}$

All relevant hyperparameters and their appropriate values are listed here.

ABLATION STUDIES

We compare two backbones – Swin-B and EfficientNetB3



We compare Triplet / ArcFace methods

In most cases, the Swin-B with ArcFace showing better performance than EfficientNet-B3 and Triplet.

In cattle datasets, swin-B shows better mean accuracy and lower variability. So swin-B architecture is best for our case.

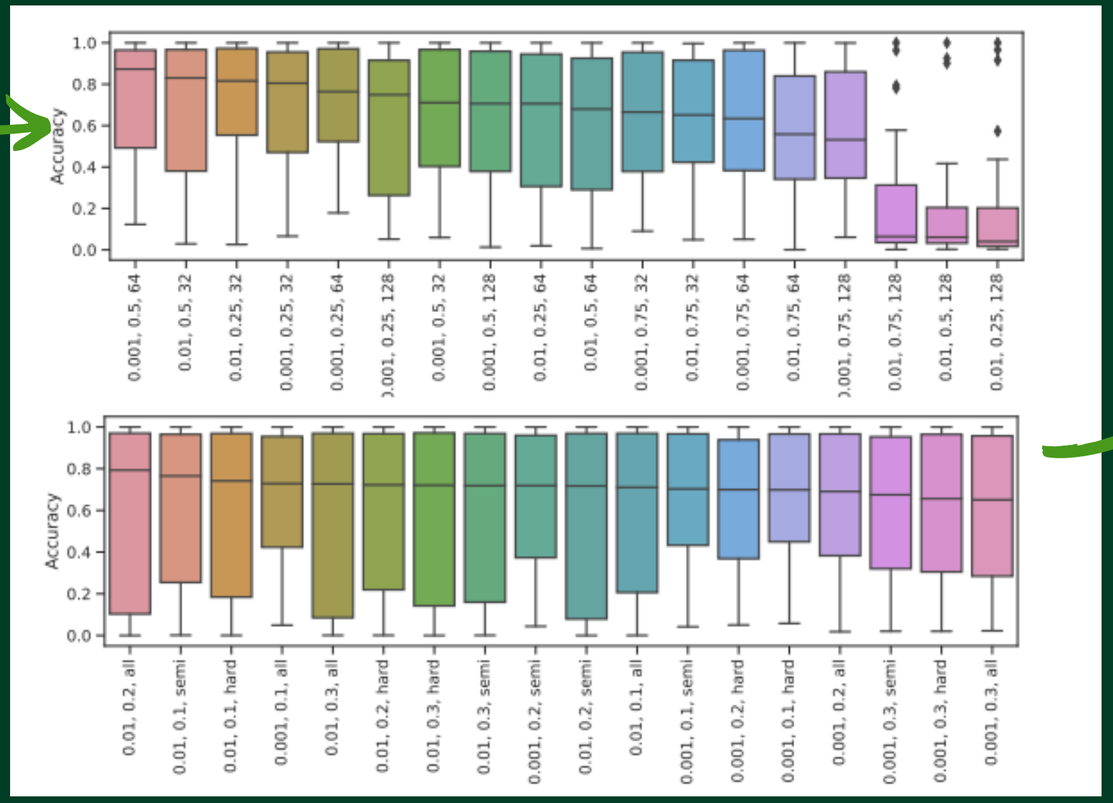
ABLATION OF HYPERPARAMETERS

ArcFace Loss

Best Setting

Learning Rate (lr) = 0.001
Margin (m) = 0.5
Scale (s) = 64

With best settings we can get best median accuracy.

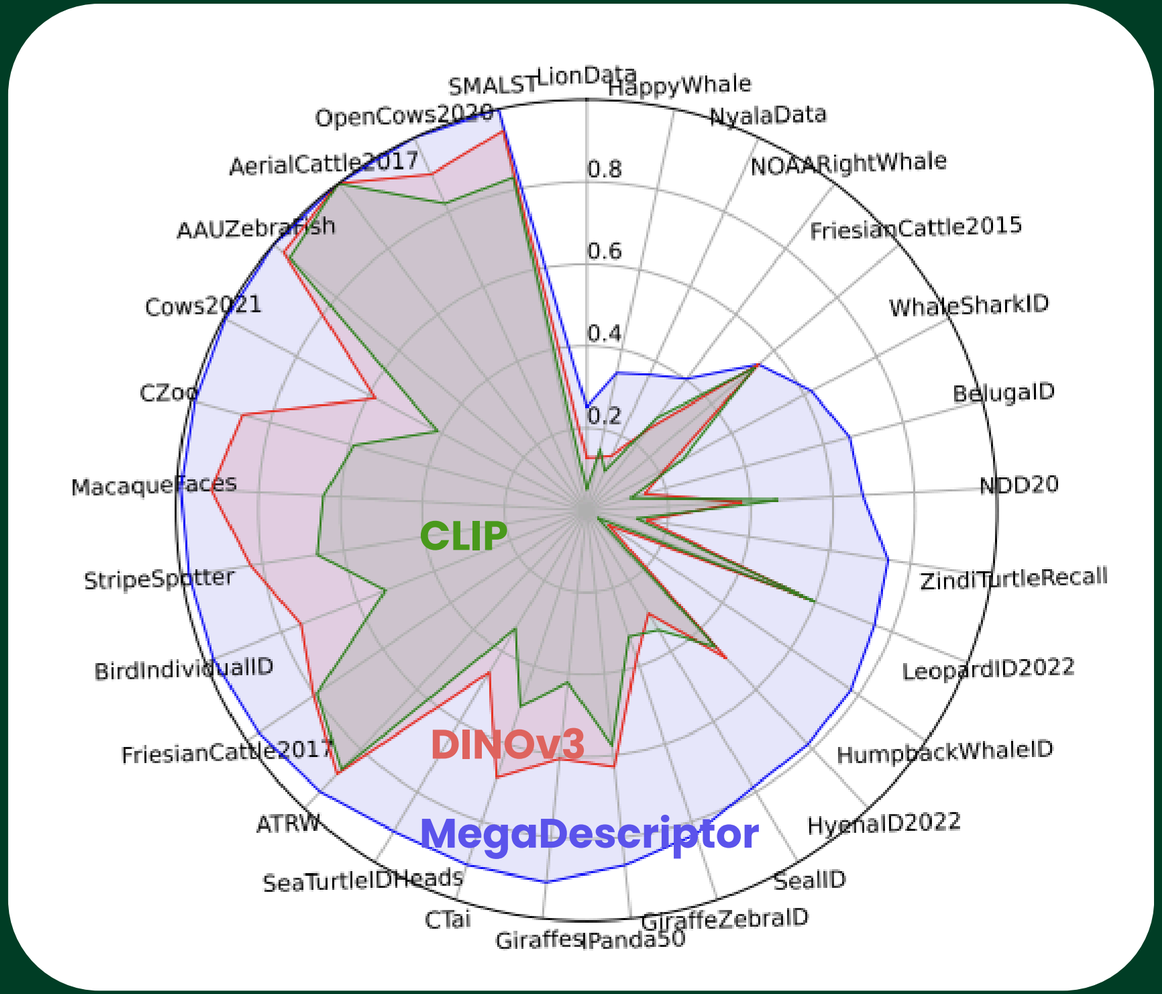


Triplet Loss

Large variations in acuracy in triplet loss.
Therefore more sensitive to hyperparameters.

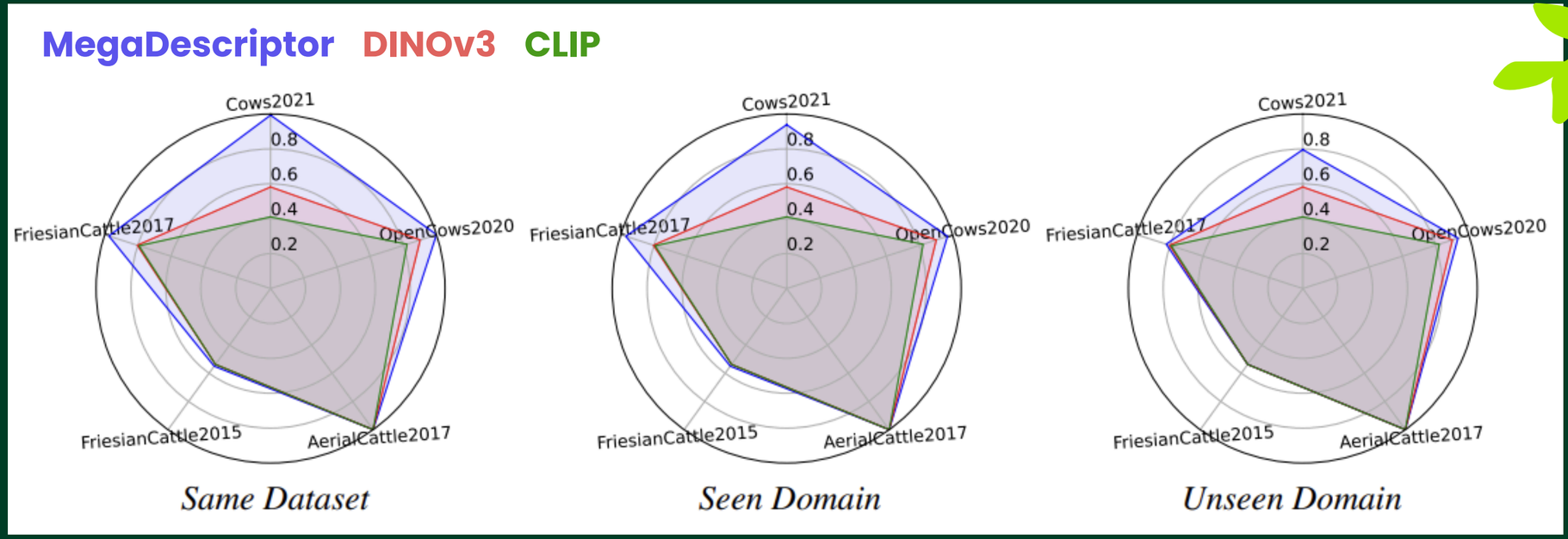
Performance for all setting combinations. (swin-B backbone)

PRE-TRAINED MODELS PERFORMANCE EVALUATION



- We can see MegaDescriptor performs better than DINOv3 and CLIP for all data sets.
- MegaDescriptor is the best way for animal re-identiffication.

These models perform in seen domain and unseen domain as follows,



MegaDescriptor performs better than others in all domains.

THANK YOU

FOR WATCHING

Panthera

Electronic and Telecommunication Engineering

University of Moratuwa