Re-Identification of Individual Animals Using ArcFace Architectures

Connor Kanally
Department of Integrated Systems Engineering
The Ohio State University
Columbus, OH
kannally.4@osu.edu

Sal Hargis
Department of Integrated Systems Engineering
The Ohio State University
Columbus, OH
hargis.29@osu.edu

Abstract

Individual animal re-identification (Re-ID) is a critical task for wildlife conservation, enabling population monitoring without invasive tagging. However, traditional computer vision models often struggle with the fine-grained visual similarity between individuals of the same species and significant variations in viewpoint and background. In this work, we explore a two-stage pipeline to address these challenges with the goal of increasing the accuracy of individual animal classification. First, we employ backbone models, including ResNet-50 and EfficientNet-B3, trained with ArcFace loss on a unified multi-species dataset to perform coarse global retrieval for the top 100 candidates for each individual . Then, we implement LightGlue, a deep geometric matcher, to re-rank the top candidates based on local feature consistency . The top R-1 and R-20 accuracies are compared from the baseline models to the pipeline using LightGlue as a fine-grained individual classifier . The results show that while LightGlue improves accuracy for lower-performing baselines, it can struggle to identify individuals when backgrounds and viewpoints change, as it may focus on background features rather than the individual.

1 Introduction

Kaggle is the home to many online data sets and challenges available to challenge the community to develop advanced machine learning algorithms for many different applications. One of the more recent public challenges is AnimalCLEF25, an animal re-identification challenge in which participants are invited to develop machine learning approaches for animal re-identification (re-ID) [1]. Animal re-ID is a challenge of classifying individual animals of different species (rather than just identifying a species) based on images of the individual. The data set provided for the AnimalCLEF challenge is the WildlifeReID-10k dataset [2].

To address the limitations and challenges inherent to the Re-ID problem, we chose to explore a coarse-to-fine re-identification pipeline. We utilized backbone models optimized with ArcFace loss to learn a vectorized embedding space so that we could leverage similarity scores for each individual. ArcFace was selected because it creates an embedding space where angle is equivalent to similarity, providing a mechanism to group related images together, which Softmax is incapable of performing. The output of the first step was the top 100 best candidates for each test individual. To mitigate the "semantic confusion" problem found with global embedding approaches, we then utilized a

second-stage feature matching step using LightGlue. This deep matcher treats the animal's surface as a geometric landscape, aligning local keypoints to verify identity independent of global appearance. As this technique is computationally expensive, we limited its use to the top 100 candidates identified with the coarse-grained classification of the first stage. We evaluate this approach on the WildlifeReID-10k dataset as part of the AnimalCLEF25 Kaggle challenge, demonstrating how this coarse-fine pipeline offers increased accuracy as opposed to global embedding approaches, while still considering computational efficiency.

2 Motivation

Animal re-identification is an important aspect of modern ecological research, facilitating the tracking of population dynamics, migration patterns, and behavioral studies. Unlike species classification, which categorizes animals into broad biological groups, re-identification requires distinguishing unique individuals. For example, distinguishing "Zebra A" from "Zebra B" based on subtle traits and attributes of individual animals. Although existing technologies and algorithms are able to distinguish species from one another (e.g., cats vs dogs), there is also a need for approaches enabling the identification of different individuals in the same species. This transforms the classic classification problem into a much more difficult problem, as the model needs to be sensitive enough to fine-grained feature differences. This also poses new challenges, as being overly sensitive to features could also decrease robustness when animals are viewed from different perspectives or angles, or when their background changes.

As a result of these challenges, current Re-ID models face significant limitations. Standard Convolutional Neural Networks (CNNs) trained with global embedding losses often overfit to background noise or fail when presented with significant viewpoint variations found in real-world data. For example, a model may correctly identify a whale from its blowhole but fail to recognize the same individual from a dorsal view, as the key features of the same animal may change depending on this view. Furthermore, different individuals of the same species may often be indistinguishable from the perspective of global feature extractors, potentially leading to high false-positive rates (over-generalizing individuals as part of a species).

3 Approach

The approach was divided into two sub-tasks: (1) to develop an initial advanced model to perform closed-set recognition and then (2) modify such model to perform open-set recognition. Here, closed-set recognition means to test a machine learning model with the same individuals in the training set (although the training and testing include different images), while open set-recognition requires the model to be tested on individuals not in the training set and to to output 'new individual' when necessary. The approach to (1) is first presented followed by the approach to (2).

3.1 Baseline Approach

Before building more advanced approaches to the AnimalCLEF25 [1] challenge, an initial baseline approach was applied to understand the nature of the animal re-ID task. The standard ResNet-50 model was fine-tuned through training on the provided data set to determine the initial accuracy for both species classification and individual animal identification.

Next, focusing on the task of individual identification, and two models were fine tuned to and tested to determine the front-end of the model architecture. The front end of the pipeline was varied between EfficientNet-B3 [5] and ResNet-50 [3]. EfficientNet-B3 and ResNet-50 are both convolution based networks with some key differences. ResNet-50 uses residual layers and skip connections [3], while EfficientNet-B3 uses compound scaling methods [5]. Both were specifically fine-tuned to the Wildlife-10k dataset from the AnimalCLEF25 challenge. Both models were also optimized with ArcFace Loss with an embedding size of 512 dimensions.

3.2 Advanced Approach

The second stage of the pipeline re-evaluates the top 100 candidates to correct ranking errors from the coarse grained rankings. The top 100 candidates for each individual are inputted into LightGlue,

a deep Graph Neural Network (GNN) that performs sparse feature matching [4]. LightGlue employs self-attention and cross-attention mechanisms to determine point correspondences while adaptively pruning unmatchable points (outliers). For each candidate pair, a geometric score is calculated based on the number of valid inlier matches. The candidate list is then re-ranked in descending order of this geometric score. As even the LightGlue approach is computationally expensive, it was never run on the entire dataset, but always a subset of the top 100 from the initial stage.

To perform open set-recognition, the better of the two (ResNet-50 with LightGlue and EfficientNet-B3 with LightGlue) was adapted to begin to explore open set recognition. The number of features identified by LightGlue in the query image relative to the number of matching points in the candidate images was computeted to determine the threshold at which to differentiate between existing individuals and new individuals (e.g., individuals the model was not trained on).

3.3 Libraries and Computational Resources

Python libraries including seaborn, numpy, pandas were used for data manipulation and plotting, while python machine learning libraries including pytorch, torch vision, and sk-learn were used for the machine learning tasks. All code was run using the Ohio Supercomputer Center pitzer cluster.

4 Experiments, Evaluation, and Validation

4.1 Data

We evaluated our pipeline on the unified WildlifeReID-10k dataset, which aggregates images from 36 distinct wildlife datasets. This dataset includes 10k individuals across 140k images that span a wide array of sea, land, and air species. The backgrounds and viewpoints of each animal image can vary substantially, adding to the real-world challenge. The data was split into a training set (80%) and a test set (20%) based on individual identities to ensure zero-overlap of images.

4.2 Metrics

The baseline models (ResNet-50 and EfficientNet-B3) were trained for 10 epochs using the Ohio Super Computer (OSC). For the evaluation method, we utilized an open-set recognition approach, utilizing 2,000 images of known individuals and 500 images of individuals not present in the training set. We report the rank accuracy of the top 1 and top 20 (R-1 and R-20). Rank-1 accuracy serves as our primary metric, indicating the percentage of queries where the correct individual was the top prediction.

4.3 Results and Analysis

When training the standard out of the box ResNet-50 model from torch, the authors trained one model to perform species classification and another to perform individual animal identification. The species classification achieved near perfect accuracy, while the individual animal classification performed poorly. This verified that more advanced approaches must be applied to successfully perform animal re-ID, the results of which are discussed below. First, the results for closed-set identification are presented and discussed followed by insights generated for the open set identification task.

Our experiments revealed a significant performance disparity between the two baseline architectures. Figure 1 shows the training loss and the R-1 accuracy for ResNet-50 and EfficientNet-B3 over 10 epochs. The ResNet-50 baseline achieved an R-1 accuracy of approximately 56% and an R-20 accuracy of 76%. In contrast, the EfficientNet-B3 architecture significantly outperformed ResNet-50, achieving an R-1 accuracy of 72% and an R-20 accuracy of 86%.

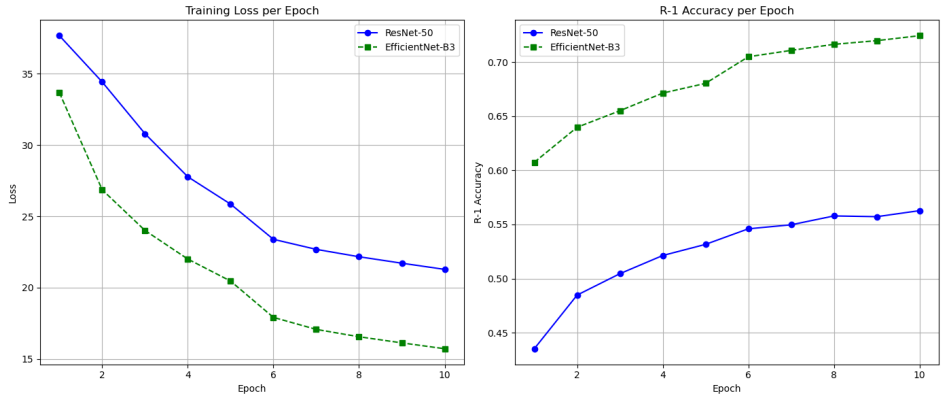

Figure 1: Shows the training and loss of both baseline models ResNet50 and Efficient-B3 over 10 epochs fine-tuned on the Wildlife-10k dataset.

After adding LighGlue to the model, the accuracy of the ResNet-50 with ArcFace model was improved, but the accuracy of the EfficientNet-B3 with ArcFace model was degraded. Adding LightGlue to the ResNet-50 with ArcFace architecture increased the R-1 accuracy to 64% (from 56%) and the R-25 accuracy to 79% (previous R-20 accuracy 76%). Adding LightGlue to the EfficientNet-B3 with ArcFace architecture decreased R-1 accuracy to 71% (previously 72% accuracy) and the R-25 accuracy to 85% (previous R-20 accuracy 86%). Please note that due to an error, top-20 accuracies were computed for models without LightGlue, but top-25 accuracies were computer for models with LightGlue, inhibiting direct comparison. The fact that LightGlue increased accuracy for the ResNet-50 model but decreased accuracy for the EfficientNet model may suggest that LightGlue can only do “so much" to improve a model but will not always increase performance of well-suited neural networks.

An interesting observation was that sometimes LightGlue would focus only on the background and not on the features of the animal itself. In Figure 2, the model shows how it occasionally will focus on features of the background instead of the animal. This indicates that LightGlue may perform better against similar or simple backgrounds, or even if image segmentation can be performed as a pre-processing step to force the model only to look at the animal. In Figure 3, LightGlue successfully matches features from the Whale Shark, showing how the same features in one view can be used to identify the same individual in a different view, which is a strength of the model.

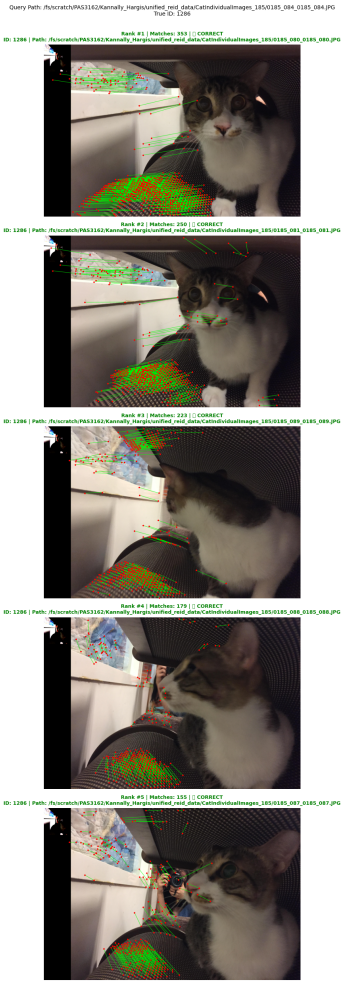

Figure 2: LightGlue focused too much on the background, indicating that this model may perform better with segmented images or against a neutral background.

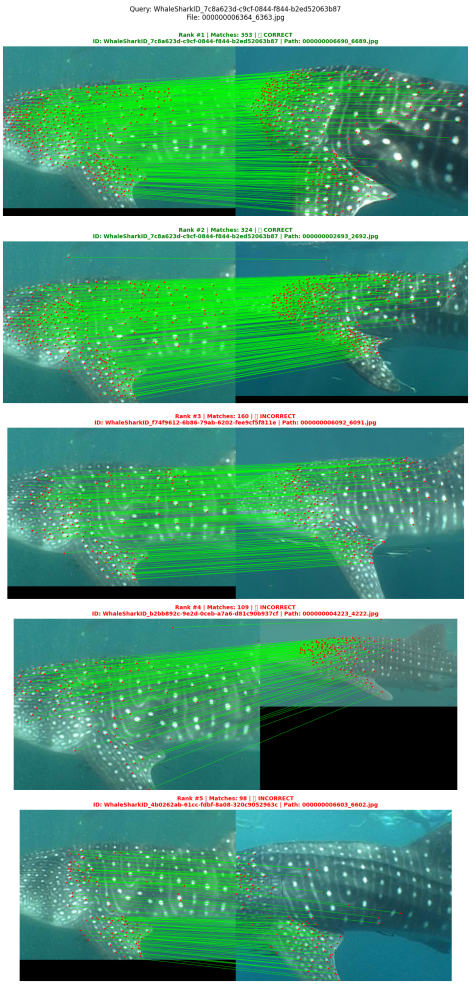

Figure 3: LightGlue successfully identified the individual Whale Shark by matching features from two different viewpoints.

Analysis of the LightGlue performance suggests that while it can help re-rank when the baseline model accuracy is lower, it struggles with the complex backgrounds inherent in wildlife photography. We observed instances where feature matching focused on background elements rather than the individual animal.

The final step was to modify the architectures with LightGlue to perform open set recognition of individual animals, which requires enabling the model to distinguish between known individuals and new individuals. The idea here was to calculate a normalized feature matching scale, which computed the score of features in the query image and the number of matched features to the candidate images enabling the two sets to be distinguished, but this method proved unsuccessful. Furthermore, in the context of open-set recognition, we analyzed the LightGlue score distribution and found that there was no clear threshold to distinguish between known and new individuals, as the score densities for known identities and singletons overlapped significantly. Figure 4 shows these distributions and how finding an appropriate threshold involves making a trade-off that will be difficult to navigate for real-world data where the model needs to recognize if the individual in question is a new animal, or simply an existing individual being seen from a new viewpoint or against a different background.

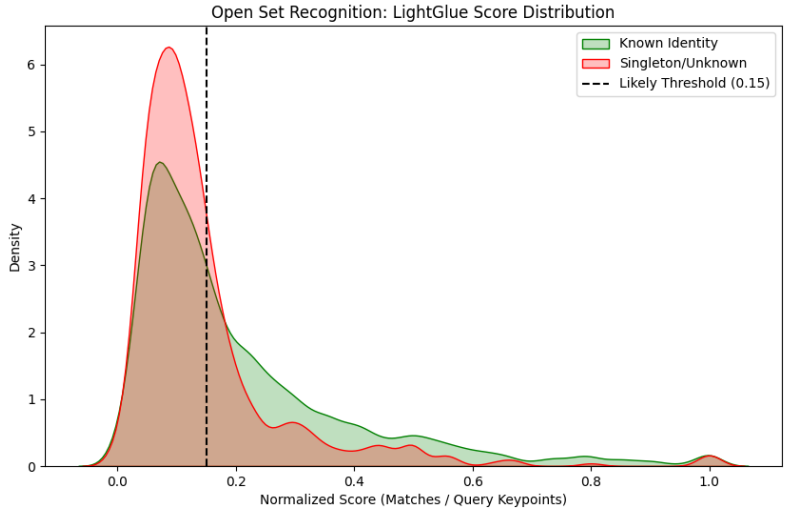

Figure 4: Distributions for where to delineate a new individual not in the training set, as opposed to a familiar individual from a significantly new viewpoint. A good threshold is difficult to find.

5 Discussion and Conclusion

5.1 Insights

This study demonstrates that a two-stage pipeline using ArcFace and LightGlue can enhance re-identification performance for weaker baseline models, but may not be sufficient for identifying new individuals or improving already strong baselines. While ResNet-50 benefited from geometric re-ranking, the more optimized EfficientNet-B3 performed better as a standalone global embedding model.

5.2 Key Challenges

The primary challenges identified included the time-consuming nature of training these models and the difficulty in generating good candidate images for LightGlue without excessive computational overhead. Future work will focus on data pre-processing, specifically image segmentation to remove backgrounds so the model learns only the features of the individual. Additionally, we aim to explore data augmentation to generate alternative views and investigate different feature matching algorithms and loss functions beyond ArcFace.

5.3 Workload and Distribution

Both authors contributed equally to the project. Sal Hargis contributed to the baseline approaches, identifying models for advanced approches such as efficient net and LightGlue, developing code for training and testing models, running training and testing scripts, working on the final deliverable (e.g., presentation and final report). Connor Kannally worked on setting up the OSC environment, preparing data to be input into the pipeline, baseline approaches, training and testing baseline and advanced approaches, and worked on the final deliverables as well. Most tasks were not completed in isolation but through in-person meetings between the two authors to discuss approaches, challenges, next steps, and work on code together. At the end of the day, both authors contributed equally.

Both authors would like to thank Wei-Lun (Harry) Chao for a very well organized and structured class. Although the content was challenging for non-CSE majors, we learned a great amount and really appreciate that non-majors can take this computer vision course. Cheers and have a wonderful holiday break!

References

[1] AnimalCLEF25 @ CVPR-FGVC & LifeCLEF.

[2] Lukáš Adam, Vojtěch Čermák, Kostas Papafitsoros, and Lukas Picek. WildlifeReID-10k: Wildlife re-identification dataset with 10k individual animals, April 2025. arXiv:2406.09211 [cs].

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[4] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[5] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019.