Visual Re-identification of Wildlife using Mega Descriptor

Arun Yadav
arun20033@iiitd.ac.in

Mohammad Shariq
shariq20220@iiitd.ac.in

Kunal Maurya
kunal20215@iiitd.ac.in

Indraprastha Institute of Information Technology, Delhi

1. Abstract

This paper presents WildlifeDatasets, a pioneering open-source toolkit designed to enhance wildlife research through the re-identification of individual animals using camera trap images and other visual data sources. The toolkit's primary purpose is to support developing and validating machine learning models that recognize animals based on distinctive features like patterns, scars, and colorations. By integrating these capabilities into wildlife research methodologies, WildlifeDatasets aims to streamline the process of tracking animal movements, monitoring population dynamics, and studying behavioral patterns, thus facilitating a more efficient and accurate approach to ecological monitoring and biodiversity studies.

WildlifeDatasets includes a comprehensive suite of tools and extensive datasets curated from diverse geographical locations featuring a variety of species. These resources enhance the accuracy and applicability of the machine learning models across different ecological zones. As an open-source platform, it encourages collaboration among scientists, tech developers, and conservationists, fostering a community dedicated to advancing wildlife conservation through technology. By automating the identification process and reducing human error, WildlifeDatasets significantly improves the efficiency of data analysis, allowing researchers to quickly assess wildlife populations and health, thereby supporting informed conservation strategies and promoting a broader sharing of knowledge and innovative practices in the field.

2. Introduction

Animal re-identification is critical in various wildlife research areas, including population monitoring, animal movements, behavioral studies, and wildlife management. Although the exact definitions and methods for animal re-identification can differ across studies, the fundamental aim is the same: to reliably and efficiently identify individual animals within a species by their unique traits, such as markings, patterns, or distinctive physical features.

Automating the process of identifying and tracking individual animals allows for the gathering of accurate and extensive data on aspects like population dynamics, migration patterns, habitat use, and behaviors. This data is crucial for monitoring animal movements, estimating population sizes, and observing demographic changes, providing deeper insights into species dynamics, identifying threats to biodiversity, and shaping conservation strategies based on solid evidence.

The growing volume of data collected and the need for manual processing, which is often time-consuming, have underscored the necessity for automated methods. These methods significantly reduce the need for labor-intensive manual oversight in animal identification. Consequently, numerous automated re-identification datasets and methods have been developed for various animal groups, including primates, carnivores, reptiles, whales, and other mammals.

However, the field faces challenges such as a lack of standardization in algorithmic procedures, evaluation metrics, and dataset usage. These issues affect the comparability and reproducibility of results, which in turn impedes progress in the field. It is vital to categorize and reassess general re-identification approaches, align them with real-world applications, and suggest suitable algorithmic configurations for specific scenarios. By critically evaluating the methodologies used in various studies, we aim to identify trends and offer insights into the most effective techniques for different situations. This analysis will help researchers and practitioners choose the best algorithms for their specific re-identification needs, thereby pushing forward the field of animal re-identification and its contributions to wildlife conservation and research.

To tackle these challenges, we have developed an open-source toolkit called WildlifeDatasets, designed for ecologists and researchers in computer vision and machine learning. In this paper, we not only provide a concise overview of our toolkit's main features but also (i) compile a comprehensive list of all publicly available wildlife re-identification datasets, (ii) conduct the most extensive experimental comparison of these datasets and re-identification methods, (iii) introduce a foundational model, the MegaDescriptor, which utilizes various Swin architectures and is trained on a newly assembled dataset, and (iv) offer a selection of pre-trained models available on the HuggingFace hub.

3. Problem Statement

A significant issue in **wildlife management and conservation** is the ability to distinguish individual animals in their natural habitats using unique characteristics such as spots or stripes. This differentiation is crucial as it aids in monitoring animal populations, which in turn helps to ascertain their specific locations and enhances their protection. However, the challenge lies in identifying these animals accurately and efficiently without confusion and without consuming excessive time. Developing an effective method to achieve this is essential for the effective conservation and study of wildlife.

4. Approaches

Advances have largely influenced the development of automated animal re-identification methods in ecological research in machine learning. Here's a technical summary of the three primary approaches to wildlife re-identification as discussed.

4.1. Tools And Methods

- **Local-Feature-Based Methods:**

  Key Techniques: These methods utilize algorithms like SIFT, SURF, or ORB to detect unique keypoints in images and extract local descriptors.

  Matching Process: The extracted descriptors are used to match images against a database containing known identities, identifying the individual animal with the highest number of descriptor matches.

  Advantages: They are easy to implement ("plug-and-play") and do not require model fine-tuning, making them suitable for zero-shot settings similar to large foundation models like CLIP or DINOv2.

  Limitations: These methods may not scale efficiently to larger datasets and their performance may not be optimal due to simpler computational techniques.

  Typical Applications: Widely used by ecologists, especially those without deep technical backgrounds, often facilitated by intuitive graphical user interfaces in software products like WildID, HotSpotter, and I3S.

- **Deep Feature-Based Approaches:**

  Key Techniques: These involve training deep neural networks to learn vector representations (deep embeddings) of images, typically resulting in 1024 or 2048-dimensional vectors.

  Matching Process: Similar to local-feature methods, deep embeddings are matched against an identity database.

  Advantages: Potentially higher accuracy and adaptability to different scenarios, mirroring techniques used in human or vehicle re-identification.

  Limitations: Requires fine-tuning on species-specific data, making the model's performance highly dependent on the particular species trained on.

  General Applications: Often repurposed from models initially designed for broad computer vision tasks.

- **Species-Specific Methods:**

  Key Techniques: These methods are customized for individual species or groups of closely related species, focusing on unique visual features like patterns or markings.

  Matching Process: Techniques might involve comparing specific image regions or calculating similarities through measures such as Chamfer distance for polar bear whiskers, or correlation between cheetah spots.

  Advantages: High specificity and potentially accurate identification within the target species.

  Limitations: Their applicability is restricted to the species or groups they are designed for, and they often require substantial manual preprocessing, which limits scalability and flexibility.

  Typical Applications: Best suited for studies where precise identification of individuals within a species is critical and where visually distinctive features are prevalent.

These approaches demonstrate the range of methods available in wildlife re-identification, each with specific strengths and limitations based on the desired application and the ecological and technical resources available.

4.2. Accessible feature extraction and matching

The WildlifeDatasets toolkit offers an extensive range of tools for wildlife re-identification using various feature extraction and matching algorithms, which are easily accessible and ready for immediate use. It includes:

- **Local Descriptors:**

  These are included as a baseline due to their extensive use and effectiveness in animal re-identification. The toolkit features implementations of SIFT descriptors and a matching algorithm that leverages local descriptors using the latest insights. GPU-based FAISS for nearest neighbor search eliminates the need for approximate neighbors, addressing concerns about time complexity.

- **Pre-trained Deep Descriptors:**

  The toolkit also allows the use of any pre-trained model from the HuggingFace hub for feature extraction across re-identification datasets. Integrated with the Timm library, it supports modern CNN and transformer-based architectures such as ConvNeXt, ResNext, ViT, and Swin for both feature extraction and model fine-tuning.

- **MegaDescriptor:**

  This is a novel foundation model designed specifically for individual animal re-identification across various species. The toolkit provides various versions of MegaDescriptor (Small, Medium, Large), which have shown superior performance over other models like CLIP and DINOv2.

- **Matching:**

  A high-level API for matching query and reference sets is also provided. It simplifies the process of computing pairwise similarity between images, returning the most visually similar identity and the corresponding image once initialized with the identity database.

4.3. MegaDescriptor Methodology

The methodology section for MegaDescriptor focuses on adapting wildlife re-identification to a real-life scenario by comparing a set of known identity images (reference set) with newly acquired images (query set) to identify individual animals. The methodology draws from existing literature and evaluates two main approaches: local feature descriptors and metric learning, using the WildlifeDataset toolkit's datasets.

- **Local Feature Approaches :**

  Local descriptors, such as SIFT and Superpoint, are used due to their proven effectiveness in wildlife re-identification. The process involves, Extracting keypoints and descriptors from all images in both the reference and query sets.Calculating the distance between descriptors of all possible image pairs.Using a ratio test with a threshold to discard likely false matches, with the threshold optimized based on performance.Determining the identity from the reference set that has the highest number of correspondences with the query image.

- **Metric Learning Approaches :**

  Two metric learning methods, ArcFace and Triplet Loss, are chosen based on their success in other re-identification domains. These methods aim to map images into a deep embedding space where the distance reflects visual similarity:

  Triplet Loss: Trains using triplets (anchor, positive, negative) to minimize the distance between similar images (anchor and positive) and maximize the distance between dissimilar images (anchor and negative), with different strategies for selecting triplets.

  ArcFace: Enhances softmax loss by adding an angular margin to improve the discriminative power of the embeddings, with embeddings normalized and scaled to fit on a hypersphere.

- **Matching and Training Strategies :**

  A simplified matching strategy is used, relying on the closest match within the reference set determined by a 1-nearest-neighbor classifier using cosine similarity.

  The models are trained on 29 datasets from the WildlifeDataset toolkit, split into 80/20 for reference and query sets, using a closed set setting. Training is done using SGD with momentum and a cosine annealing learning rate schedule over 100 epochs.

- **HyperParameter Tuning :**

  Extensive hyperparameter searches are conducted to optimize performance across all datasets for both metric learning methods and different backbone architectures, comparing traditional CNNs (EfficientNet-B3) and transformer-based models (Swin-B). With configurations such as :- Backbone: Swin B, Learning rate: 0.001, ArcFace margin: 0.5, ArcFace scale: 64, Triplet mining: semihard, Triplet margin: 0.2 .

4.4. Abalation Studies

This section of the document outlines a series of ablation studies conducted to empirically validate the design decisions made during the development of the MegaDescriptor, a foundational model for animal re-identification. These studies focused on selecting the model's best methods, architectures, and hyperparameters.

Loss and Backbone Components: The studies compared the performance of two metric learning loss functions—ArcFace and Triplet loss—using two types of backbone architectures: a transformer-based Swin-B and a CNN-based EfficientNet-B3. The findings indicate that combining Swin-B with ArcFace generally yields competitive or superior performance compared to the other configurations. Detailed performance data, presented through box plots, show that ArcFace consistently outperforms Triplet loss, regardless of the backbone used.

Hyperparameter Tuning: Extensive grid searches were conducted to identify optimal hyperparameters for minimizing the performance variability inherent to metric learning methods. For the Swin-B backbone using ArcFace, the optimal settings were identified (learning rate of 0.001, margin m of 0.5, and scale s of 64), achieving a median performance of 87.3 percent. Notably, some configurations underperformed dramatically, likely due to issues in training convergence. In comparison, settings using the Triplet loss demonstrated higher variability in performance, suggesting that while Triplet loss can achieve competitive results, its performance is more sensitive to hyperparameter settings.

Overall, these ablation studies validate the design choices made for the MegaDescriptor, demonstrating its effectiveness and robustness in zero-shot wildlife re-identification scenarios compared to other methods like SIFT.

5. Results

Below are the results of our experiment in the form of evaluation metrics:

Table 1. Evaluation Performance Metrics with seen dataset (Values in %)

Dataset	Accuracy	Precision	Recall	F1score
MacaqueFaces	100.00	1.00	1.00	1.00
Nyala	11.00	0.54	0.11	0.11
Lion	8.00	0.51	0.08	0.07
Stripe Spotter	98.00	0.98	0.98	0.97
Ipanda50	100.00	1.00	1.00	1.00
CZoo	100.00	1.00	1.00	1.00

Table 2. Evaluation Performance Metrics with unseen dataset (Values in %)

Dataset	Accuracy	Precision	Recall	F1score
Cow	99.00	0.99	0.99	0.98
DogNet	56.00	0.98	0.56	0.57
MPDD	3.00	0.99	0.03	0.03
PolarBear	100.00	1.00	1.00	1.00

In our study, the accuracy/model performance for the seen data is similar to what is reported in the paper. For unseen data, the accuracy for the Cow and PolarBear datasets is high because the model was previously trained on similar datasets (Cows2021 and Open Cows2020). In contrast, the performance on DogFaceNet and MPDD is lower.

5.1. Performance Evaluation

Insights from our ablation studies led to the creation of MegaDescriptors – the Swin-transformer-based models optimized with ArcFace loss and optimal hyperparameters using all publicly available animal re-id datasets. The proposed MegaDescriptor with Swin-L/p4-w12-384 backbone

performs consistently on 5 datasets and outperforms all methods in all datasets.

Table 3. Results of Hyperparameter Tuning after Ablation of animal re-id methods. We compare the local feature (SIFT) method with two metric learning approaches (Triplet and ArcFace). Metric learning approaches outperformed the local-feature methods on most datasets. ArcFace provides more consistent performance. For metric learning, we list the median from the previous ablation.

Dataset	Th	SIFT	ArcFaceLoss	TripletLoss
MacaqueFaces	0.8	66	100	99.98
IPanda50	0.8	25	82	85
LionData	0.6	21	10	6
NayalData	0.7	5	19	21
StripeSpotter	0.7	95	55	79

Table 4. Hyperparameter For New Dataset

Dataset	Th	SIFT	ArcFaceLoss	TripletLoss
DogFaceNet	0.7	11	57	67

6. Individual Analysis

6.1. Kunal Maurya Analysis:-

The paper tell us about several key insights into the performance of the model and methodologies.

In the paper first its give a full toolkit for animal re-identification is introduces, backbone architectures, focusing on the metric learning approaches, hyperparameter tuning, and evaluation metrics. different backbone architectures with ArcFace and triplet loss function to compare the effectiveness and emphasizes the importance of selecting the right combination for get optimal performance. The hyperparameter tuning process give us the significance for the tuning model to get good result for different datasets.

The second part of paper compares the megaDescriptor model with other techniques like SIFT, Superpoint, ImageNet, CLIP, and DINOv2.In the paper its shows that metric learning techniques outperform the local-feature based on different dataset with ArcFace and Swin-B backbone.

While analysis of the unseen data I see that model achieved good accuracy for the Cow dataset, most like model is train for the similar kind of dataset like Cows2021, Opencows2020, FriesianCattle2015 and FriesianCattle2017. However, For the DogFaceNet and MPDD datasets the model accuracy is low due to model is not train for that dataset.

Overall, the studies give us the valuable insights of the field animal Re-identification and continuous refinement of algorithms to address the challenges in wildlife monitoring and conservation efforts.

6.2. Arun's Analysis

The paper provides a comprehensive overview of a full toolkit designed for animal re-identification, introducing various backbone architectures and focusing on metric learning approaches, which include techniques like ArcFace and triplet loss functions. It highlights the importance of carefully selecting the right combination of methods and performing extensive hyperparameter tuning to optimize performance across different datasets. In comparing the megaDescriptor model with other techniques such as SIFT, Superpoint, ImageNet, CLIP, and DINOv2, the paper demonstrates that metric learning approaches generally outperform local-feature based methods when tested across various datasets, especially with the use of ArcFace and the Swin-B backbone.

The analysis of unseen data reveals significant insights into the model's performance: it achieved good accuracy on the Cow dataset, likely because it was trained on similar datasets like Cows2021, OpenCows2020, FriesianCattle2015, and FriesianCattle2017. However, the model's performance on the DogFaceNet and MPDD datasets was lower, indicating that it had not been trained on data similar to these datasets. Overall, this study offers valuable insights into the field of animal re-identification and underscores the ongoing need to refine algorithms to meet the challenges of wildlife monitoring and conservation efforts effectively. The paper underlines how these continuous improvements in machine learning models are crucial for adapting to diverse environmental conditions and ensuring the reliability of these tools in real-world conservation scenarios.

6.3. Shariq's Analysis

The paper explains why the MegaDiscriptor Methodology outperforms other methodologies like SIFT, SuperPoint, etc. via performing ablation studies and hyperparameter tuning. MegaDiscriptor is a transformer-based model based on Swin-B architecture.

The studies compared the performance of two metric learning loss functions—ArcFace and Triplet loss—using transformer-based Swin-B as a backbone.

The findings indicate that combining Swin-B with ArcFace generally yields competitive or superior performance compared to the other configurations. Detailed performance data show that ArcFace consistently outperforms Triplet loss, regardless of the backbone used.

Hyperparameter Tuning: Extensive grid searches were conducted to identify optimal hyperparameters for minimizing the performance variability inherent to metric learning methods. For the Swin-B backbone using ArcFace, the optimal settings were identified (learning rate of 0.001, margin of 0.5, and scale s of 64), achieving a median performance of greater than 80 percent.

Notably, some configurations unperformed dramatically, likely due to issues in training convergence due to less epoch training. In comparison, settings using the Triplet loss demonstrated higher variability in performance, suggesting that while Triplet loss can achieve competitive results, its performance is more sensitive to hyperparameter settings.

Overall, these ablation studies validate the design choices made for the MegaDescriptor, demonstrating its effectiveness and robustness in zero-shot wildlife re- identification scenarios compared to other methods, like SIFT

7. Refernces and Resources

1. https://openaccess.thecvf.com/content/WACV2024/html/Cermak_WildlifeDatasets_An_Open-Source_Toolkit_for_Animal_Re-Identification_WACV_2024_paper.html
2. https://openaccess.thecvf.com/content/WACV2024/supplemental/Cermak_WildlifeDatasets_An_Open-Source_WACV_2024_supplemental.pdf
3. https://huggingface.co/BVRA/MegaDescriptor-T-224/tree/main
4. https://paperswithcode.com/paper/wildlifedatasets-an-open-source-toolkit-for/review/
5. https://github.com/WildlifeDatasets/wildlife-tools/tree/main
6. https://github.com/WildlifeDatasets/wildlife-datasets/tree/main
7. https://wildlifedatasets.github.io/wildlife-datasets/datasets/
8. https://openaccess.thecvf.com/content_WACVW_2020/papers/w2/Haurum_Re-Identification_of_Zebrafish_using_Metric_Learning_WACVW_2020_paper.pdf