CV Project

Abhinav Ujjawal (2021120)
Abhishek Sushil (2021441)
Aditya Jain (2021511)

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Problem statement

This project aims to explore and evaluate the performance of two MegaDescriptors (Swin image feature models), on wildlife datasets.

Specifically, we will replicate the results of recent papers in this domain by conducting inference using pre-trained models on one of the datasets used in the papers.

Additionally, we will design and execute an analysis experiment to assess the performance differences and types of errors between the two approaches using an alternative wildlife dataset.

Dataset

1. Training
   a. Dataset Name: wildlife-datasets. It is a collection of 33 publicly available dataset each containing one or more animals.
   b. Publication Year: 2024

2. Testing
   a. Dataset Name: Animal Image Dataset. It has 90 different types of animals and each animal has 60 images.
   b. Publication Year: 2023

Approaches investigated (Papers)

The main Wildlife MegaDescriptor paper uses 2 approaches :

- **Local Features Approaches:**
  SIFT and Superpoint descriptors, in wildlife re-identification. The process involves the following steps:

  1. **Keypoint Extraction and Descriptor Calculation**: Keypoints and their corresponding descriptors are extracted from all images in both the reference and query sets.
  2. **Descriptor Distance Computation**: The distances between descriptors of all possible pairs of reference and query images are computed.
  3. **Ratio Test**: A ratio test with a threshold is employed to eliminate potentially false matches. The optimal threshold values are determined by evaluating the matching performance on the reference set.
  4. **Identity Determination**: The identity is determined based on the absolute number of correspondences. The identity with the highest number of correspondences from the reference set is predicted.

Approaches investigated (Papers)

- **Metric Learning Approaches**:
  1. Further two methods, **ArcFace and Triplet loss**, are selected for ablation studies. (Explained later)
     These methods are inspired by (then) recent advancements in human and vehicle re-identification. ArcFace and Triplet loss both aim to learn a representation function that maps objects into a deep embedding space, where the distance reflects visual similarity.
  2. **Matching Strategy** : To determine the identity of query (i.e., test) images, relying solely on the closest match within the reference set. Essentially, creating a 1NN algorithm.
  3. **Training** : Models were optimized using the SGD optimizer with momentum (0.9) for 100 epochs using the cosine annealing learning rate schedule and mini-batch of 128.

Evaluation metrics

It majorly takes uses two losses

- Triplet loss: It compares triplet of images, an anchor image(x_a), a positive image(x□) and a negative image (x□). x_a shares the same label as x□ but different from x□. It tries to learn an embedding where distance between x_a and x□ is smaller than x_a and x□ by at least a margin m. So the model learns embeddings where similar images have low distance while dissimilar images have large distance.
- ArcFace: It adds a margin to the standard softmax loss to improve the discriminative power of the learned embeddings. The embeddings are normalized and scaled to place them at a hypersphere of radius s which is selected as hyperparameter.

Inference time hardware requirements

- CPU
- GPU
- RAM
- Disk space

References

- Link to paper
- Training dataset
- Testing dataset