Tiger Re-ID in the Wild *DLCV Final Project 2020*

Javier Sanguino[1], Carlos Marzal[1], Julia Maricalva[1], Céline Nauer [2]

[1]Universidad Politecnica de Madrid; [2]ETH Zürich

1 Re-ID

Monitoring the geospatial distribution of endangered species is a crucial task of wildlife conservation. Traditionally, tagging methods are used to identify individuals. However, this approach is not scalable to large populations. This projects aim is to explore deep learning strategies to re-identify individuals amongst a population of Amur Tigers. Our work is based on the Aligned Re-ID [1] network, where we added an additional vertical local feature extractor. In addition, we propose a semi-supervised approach taking into account unlabeled data.

2 Data

The basis of this work are images of the LILA BC Amur Tiger Re-Identification dataset. The dataset contains around 3,393 images of different sizes of 92 Amur Tigers in zoos in China. The images are cropped to bounding boxes, thus all of different size. The images are labelled with the respective ID of the tiger.

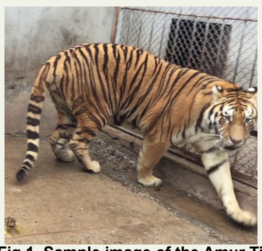

Fig.1. Sample image of the Amur Tiger Re-Identification Dataset

Pre-Processing

The following steps are taken to preprocess the images:

1. Resize images to 224 (shortest length)
2. Randomly crop the images to a 224x224 square (Centre crop for test images).
3. Random whitening
4. Randomly add white squares onto images

After augmentation, the images of the same tiger are grouped to batches of three or four, in order to get an appropriate input for a network with a triplet loss.

Generating labels for unlabeled data

Around ⅓ of the data set is unlabeled. In order to make use of this data, we train a simple classifier on the labeled data and predict the missing labels.

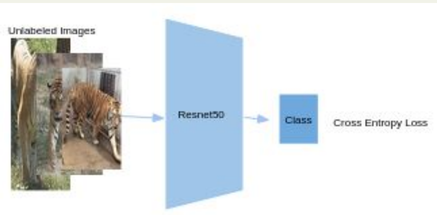

Fig.2. Classification model to predict labels for unlabeled data.

The threshold is set such that only the top 20% of the matching images get used in the labeling for the new data.

3 Model

Architecture

The architecture of the baseline model is inspired by the architecture of the Aligned Re-Identification Network[1]. The feature extractor used is a DenseNet121, which reduces the images feature tensors. The vertical, horizontal and global features are further extracted. Via hard sample mining, the most similar and most dissimilar image within a group are determined and from there, the triplet loss is calculated for the global, vertical and horizontal features.

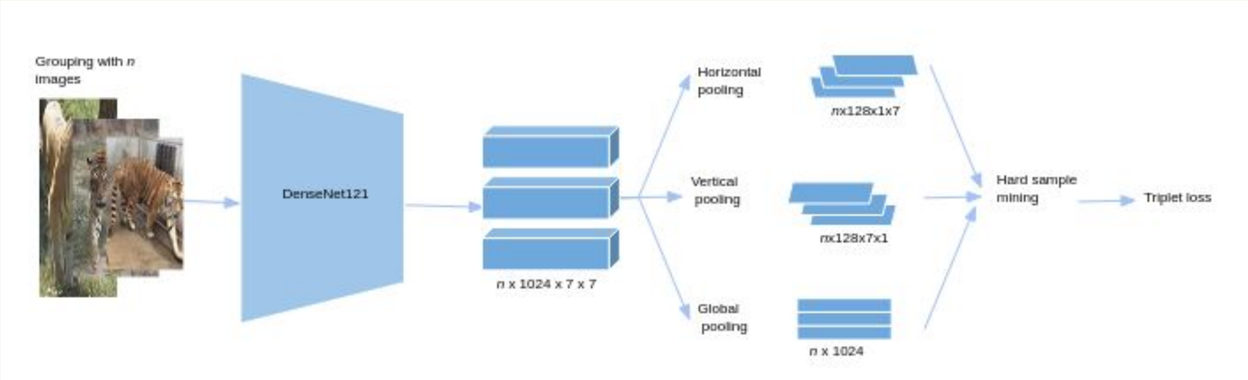

Fig.3. An overview over the model architecture. The grouping used is 3 or 4.

Loss Function

The Triplet Loss is simultaneously measuring the intra-class and the inter-class distances within a batch. Thereby, we try to minimize the former and maximize the latter as follows.

$$L(A,P,N) = max(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0)$$

4 Results

The following parameters are set to train an optimal model. The weights for the contribution of the global loss is 2, to the local loss 1, the classification loss 1.5 and the vertical loss 0.2. The learning rate is 0.00025. We achieve the following results on Tiger Re-Identification:

ACC MSE	ACC Cosine Distance	rank1	top5
0.711	0.702	0.711	0.845

Visualisation

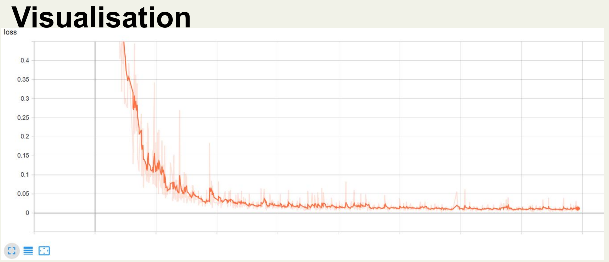

Fig.4. Losses over iterations for the best trained model.

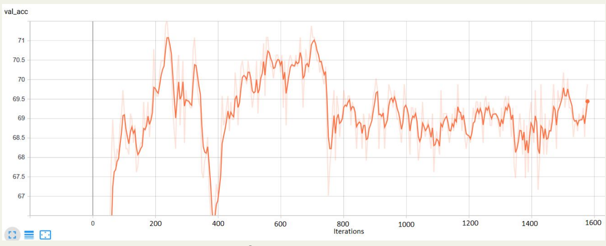

Fig.5. Accuracy over iterations for the best trained model.

5 Evaluation

The essential components of the proposed model are the different losses. In order to investigate the importance of each feature extraction, we create simplified models, where we only consider the contribution of only one feature to the loss function at a time. The following table gives an insight into the performance based on only the local, global or classification loss.

Loss	Weights			
Global	1	0	0	0
Local Horizontal	0	1	0	0
Local Vertical	0	0	1	0
Classification	0	0	0	1
MSE	**0.703**	**0.562**	**0.585**	**0.642**

6 Conclusion

We propose an extended model of the Aligned ReID, taking into account global, local horizontal & vertical as well as classification losses. We achieve an accuracy of 71.1% on Tiger Re-Identification. A closer analysis of the individual components suggests that global pooling is the most powerful feature extractor for this task.

7 References

1. Alexander Hermans et al., *In Defense of the Triplet Loss for Person Re-Identification*, arXiv:1703.07737v4, 2017
2. Xuan Zhang et al., *Aligned Re-Id: Surpassing Human-Level Performance in Person Re-Identification*, arXiv:1711.08184, 2017