–Supplementary Material– Chimpanzee Faces in the Wild: Log-Euclidean CNNs for Predicting Identities and Attributes of Primates

Alexander Freytag[1,2], Erik Rodner[1,2], Marcel Simon[1], Alexander Loos[3], Hjalmar S. Kühl[4,5], and Joachim Denzler[1,2,5]

[1]Computer Vision Group, Friedrich Schiller University Jena, Germany
[2]Michael Stifel Center Jena, Germany
[3]Fraunhofer Institute for Digital Media Technology, Germany
[4]Max Planck Institute for Evolutionary Anthropology, Germany
[5]German Centre for Integrative Biodiversity Research (iDiv), Germany

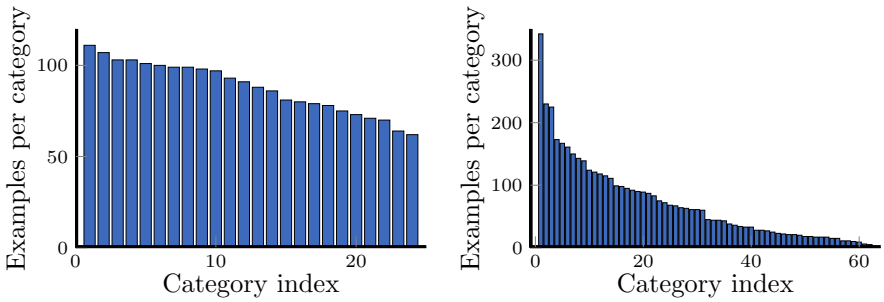
Abstract. In the following, we present our chimpanzee face dataset in detail and give additional results for age and age group prediction. Although the content of this document is not essential for the main paper, it contains additional details to support our contributions.

S1 Statistics of Our Chimpanzee Datasets

For a detailed understanding of experimental results on both datasets, it is essential to analyze their statistics in advance to check for potential dataset bias. In the following, we provide several views on the provided data of both datasets. As in the experimental setup of the main paper, we exclude categories with less than 5 examples for evaluation. These categories are either artifacts of incorrect annotations in meta-data (*e.g.,* chimpanzees identified as "Allex" instead of "Alex") or contain rare individuals for which classification estimates are hardly reliable. In consequence, we obtain all 24 categories for C-Zoo and 62 of 78 categories for C-Tai.

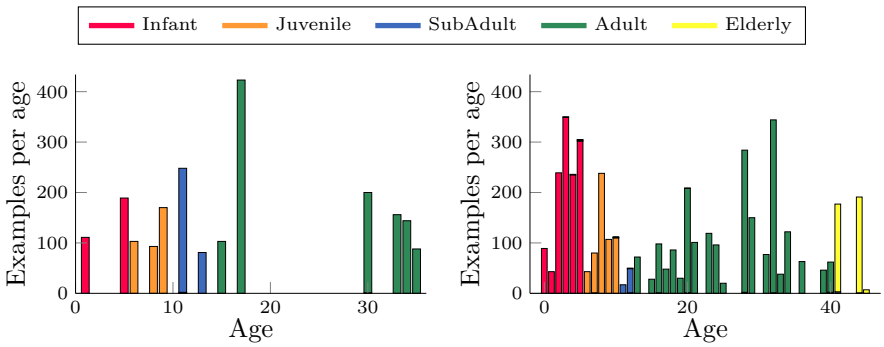
Statistics of Individuals The distribution of annotated faces per individual is shown in Fig. S1. As can be seen, the C-Zoo dataset offers a moderately balanced setting. In contrast, the C-Tai dataset is strongly heavy-tailed. Thus, models and evaluation metrics need to consider this aspect, *e.g.,* by reporting averaged class-wise recognition rates instead of overall recognition rates.

Statistics of Age We further analyze the available faces within each age group and visualize the resulting statistics in Fig. S2. First of all, we notice tiny inconsistencies of age group labels, especially at the border between age groups. In addition, we observe that the Sub-Adult group is rarely recorded in the dataset C-Tai. Furthermore, the recordings in wild life (dataset C-Tai) contain substantially more infants. Again, imbalance of distributions should be reflected in the chosen evaluation metric.



(a) C-Zoo (b) C-Tai

Fig. S1: Distributions of annotated **faces per individual** for C-Zoo (*left*) and C-Tai (*right*).



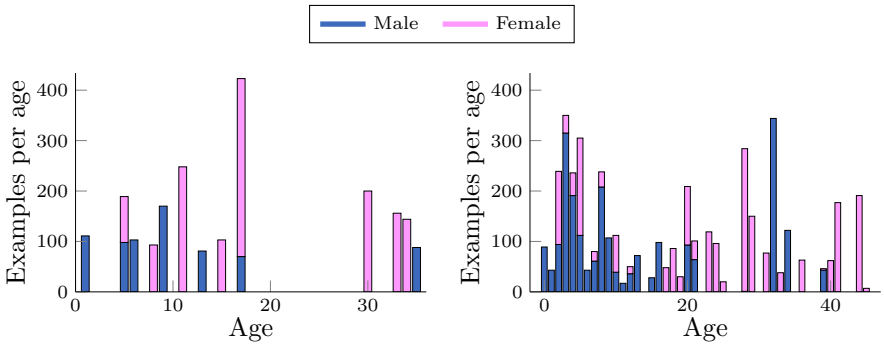
(a) C-Zoo (b) C-Tai

Fig. S2: Distributions of annotated **faces within age groups** for C-Zoo (*left*) and C-Tai (*right*).

Statistics of Gender We finally analyze the distribution among male and female individuals for both datasets. The obtained statistics are shown in Fig. S3. It can be clearly seen that recordings of several ages are only covered by single genders. However, both genders are widely spread over almost all ages.

S2 Age and Age Group Prediction

In addition to identification and gender classification shown in the main paper, our approach also allows for age and age group prediction. In the following, we present results for both tasks which are not contained in the main paper due to the lack of space.



(a) C-Zoo (b) C-Tai

Fig. S3: Distributions of annotated **faces among male and female** for C-Zoo (*left*) and C-Tai (*right*).

S2.1 Evaluation of Chimpanzee Age Estimation

Setup – Data For each dataset, we divide the entire range of ages into five equally sized age intervals. From each interval, we randomly sample 100 individuals for training. All remaining data serve for hold-out testing. As in the remaining evaluations, we repeat the process of random sampling five times to obtain reliable results.

Setup – Baselines As for the task of gender estimation (Sect. 6.2 in the main paper), we are not aware of a published baseline for predicting ages of chimpanzees. Thus, we compare our results against two straight-forward solutions for the task. The simplest baseline is to always return the average age of all chimpanzee recordings within training data ("baseline naive"). In addition, we use our identification approach (Sect. 6.1) to predict the ID of an individual and return its mean age within training data ("Identification + attribute query"). Accuracy of results is measured as L_2 error between predicted age and ground truth data.

Setup – Approaches To directly estimate the age of chimpanzees reliably, we apply Gaussian process (GP) regression models. Hence, we treat the ordinal attribute as continuously valued which is consistent with the continuous process of aging. For a detailed introduction to GPs, we refer to Rasmussen and Williams [3]. An RBF-kernel serves as covariance function with shared variance across all dimensions. We did not apply automatic relevance determination techniques due to the insufficient amount of data given the high dimensionality of representations. Hyperparameters of covariance function and regularizer are found by exhaustive five-fold cross validation (parameter ranges $1 \dots 10$ and $2^{-7} \dots 2^{2}$). We use only a single fold for training and the remaining four folds for validation which resulted in better parameter estimates. We also experimented with marginal likelihood optimization (which is theoretically appealing) but only found inferior behavior. Training of GP models is done using the GPML toolbox

Table S1: **Age estimation** results on C-Zoo and C-Tai. Results are averaged over five random splits. We report L_2-errors between predicted and ground truth age. Lower is better.

Approach	C-Zoo	C-Tai
Baseline: naive		
3-a) average age	8.73 ± 0.02	12.78 ± 0.03
Identification + attribute query		
3-b) using 1-k)	$\mathbf{2.55 \pm 0.21}$	8.30 ± 0.46
CNN codes + GP		
3-c) VGGFaces `pool5`	5.83 ± 0.06	8.41 ± 0.12
3-d) VGGFaces `fc7`	5.93 ± 0.13	8.35 ± 0.13
3-e) BVLC AlexNet `pool5`	4.51 ± 0.06	6.79 ± 0.08
3-f) BVLC AlexNet `fc7`	4.75 ± 0.17	$\mathbf{6.61 \pm 0.05}$
CNN codes + Pooling + GP		
3-g) BVLC AlexNet `pool5` + bilinear	5.13 ± 0.10	7.06 ± 0.10
3-h) BVLC AlexNet `pool5` + bilinear + norm	5.03 ± 0.10	6.90 ± 0.09
3-i) BVLC AlexNet `pool5` + bilinear + norm + logm	6.74 ± 0.05	9.78 ± 0.09
Cross-Dataset		
3-j) using 3-f)	6.99 ± 0.03	9.88 ± 0.01

of Rasmussen and Nickisch [2]. We clipped predicted ages at zero from below to prevent negative age estimates. As representations of face regions, we apply CNN codes of the VGGFaces network and the Caffe BVLC reference net. The only difference to the setup in Sect. 6.1 and Sect. 6.2 is that we do not L_2-normalize CNN activations which otherwise lowered regression accuracy considerably. Furthermore, we evaluate the effect of bilinear pooling and our LOGM-operation.

Setup – Generalization As for gender estimation in the main paper, we are finally interested in the generalization abilities of learned models across datasets. Therefore, we use all data of one dataset for model training and evaluate it on the five splits of the remaining dataset. Parameters are estimated as described before.

Results All results are shown in Table S1. On C-Zoo, the identification baseline remarkably outperforms all remaining approaches. Hence, we conclude that regression seems to be substantially more difficult than identification of individuals if training data of each individual is sufficiently representative. Similar to previous experiments, we again observe that the faces network is not able to compete with the object network. For the C-Tai dataset, we observe a slightly different trend. Since identification is more difficult, the baseline is clearly outperformed by direct age regression.

In total, findings for age regression seem to be inverse to the classification tasks: CNN codes obtained from the last layer seem to work well but normalization reduces accuracy. Similarly, bilinear pooling only results in mediocre accuracy and our LOGM operation reduces accuracy further. Hence, we conclude that regression and identification are substantially different. In consequence, further research needs to be done to understand involved effects.

Table S2: **Age group classification** results on C-Zoo and C-Tai. Results are averaged over five random splits. We report averaged class-wise recognition rates (ARR in %). Higher is better.

Approach	C-Zoo	C-Tai
Baseline: naive		
4-a) majority age group	25.00 ± 0.00	20.00 ± 0.00
Identification + attribute query		
4-b) using 1-k)	$\mathbf{94.73 \pm 1.20}$	77.88 ± 4.09
CNN codes + SVM		
4-c) VGGFaces `pool5`	86.10 ± 1.84	84.02 ± 1.25
4-d) VGGFaces `fc7`	78.78 ± 2.33	76.43 ± 1.41
4-e) BVLC AlexNet `pool5`	89.95 ± 1.90	85.33 ± 0.97
4-f) BVLC AlexNet `fc7`	86.67 ± 2.73	83.79 ± 1.48
CNN codes + Pooling + SVM		
4-g) BVLC AlexNet `pool5` + bilinear	92.68 ± 1.84	87.58 ± 1.06
4-h) BVLC AlexNet `pool5` + bilinear + norm	93.15 ± 1.35	$\mathbf{88.85 \pm 1.49}$
4-i) BVLC AlexNet `pool5` + bilinear + norm + logm	91.08 ± 1.29	85.33 ± 1.51
Cross-Dataset		
4-j) using 4-i)	55.96 ± 1.26	49.39 ± 0.37

S2.2 Evaluation of Chimpanzee Age Group Classification

In contrast to directly predicting the age of a chimpanzee, age group classification only distinguishes between a few categories. Details about the distribution of these classes across both datasets were given in Section S1.

Setup – Data For each dataset, we randomly split data of each age group into 90% for training and 10% hold-out for testing. As previously, we repeat the process of random sampling five times to obtain reliable results. Model accuracy is reported as averaged age-group-wise recognition rates to account for imbalanced age groups.

Setup – Baselines, Approaches, and Generalization We apply the same setup as for gender classification in Sect. 6.2. Hence, we obtain two baselines by predicting either the most common age group ("baseline naive") or using our identification approach ("Identification + attribute query"). We further train linear SVMs on CNN codes and evaluate the effect of bilinear pooling and our LOGM-operation. Finally, we evaluate the generalization ability across both datasets.

Results All results are shown in Table S2. As for the regression of age the identification baseline leads to the best accuracy on C-Zoo. However, results with bilinear pooling are only little behind. Hence, we conclude that the complexity of age group classification is in between identification and age regression. For C-Tai, where identification of individuals is more difficult, direct prediction of age groups pays off and leads to significantly increased accuracy. Again, we observe that VGGFaces is clearly inferior to the BVLC AlexNet. Furthermore, bilinear pooling gives additional improvements. Regarding our LOGM-operation, we do not observe noticeable benefits from the embedding in a sound (Euclidean)

vector space. Finally, we observe that the generalization across datasets is only partially possible. We attribute this to the different number of categories and strong changes of data distributions.

S3 Normalization of Bilinear CNN Activations

We already noted in the main that a normalization of bilinear activations can be beneficial when parsing it to the LOGM transformation to ensure numerical stability. In the following, we outline the applied normalization technique which has been inspired by [1]. Instructions are given as Matlab code.

1. **Obtain feature map of CNN layer**:
```
features = net.blobs( s_layer ).get_data();
```
2. **Vectorize spatial response map**:
```
i_channelCount = size ( features, 3);
features  = reshape ( features, [],i_channelCount);
```
   ⇒ reshape responses into a (width · height) × i_channelCount matrix
3. **Response normalization at each location**:
```
features = bsxfun(@times, features, 1./sqrt(sum(features,2).^2));
```
   ⇒ increase comparability of activations at different locations
   ⇒ thereby improve the condition of the bilinear matrix
4. **Second order matrix with sum pooling**:
```
features = features'*features;
```
   ⇒ sum pooling realized as inner product
5. **Apply logm transformation**:
```
features = logm( features + f_sigma*eye( size(features) ) );
```
6. **Take lower triangle**:
```
features     = features ( logical(tril(ones(size(features)))));
```
   ⇒ remove redundant information from symmetric matrix
7. **Signed square root**:
```
features     = sign(features).*sqrt(abs(features));
```
   ⇒ similar to "power normalization" and "RootSIFT"
8. **L_2-normalization**:
```
features     = features / sqrt(sum(features.^2));
```
   ⇒ pre-conditioning for SVM solver

References

1. Lin, T.Y., RoyChowdhury, A., Maji, S.: Bilinear cnn models for fine-grained visual recognition. In: IEEE International Conference on Computer Vision (ICCV). pp. 1449–1457 (2015)
2. Rasmussen, C.E., Nickisch, H.: Gaussian processes for machine learning (gpml) toolbox. Journal of Machine Learning Research (JMLR) 11, 3011–3015 (2010)
3. Rasmussen, C., Williams, C.: Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning, The MIT Press, Cambridge, MA, USA (01 2006)