

Supplementary Materials: Data-Scarce Animal Face Alignment via Bi-directional Cross-Species Knowledge Transfer

This appendix can be divided into the following parts:

- Section 1 emphasizes the significance of animal face alignment.
- Section 2 presents effect of different few-shot animals.
- Section 3 and Section 4 provide an analysis of thresholds T_{neg} and T_{pos} .
- Section 4 presents more visual analysis for Meta-CSKT.

1 SIGNIFICANCE OF ANIMAL FACE ALIGNMENT

The study of animal face alignment is significant because it can help us better understand animals, and promote their well-being by interpreting facial expressions. Understanding the facial expression and behavior of abnormal animals through visual imagery is cheaper and quick alternative to clinical examinations and vital signs monitoring. Animals may express pain through facial behavior, which they might hide from unfamiliar human observers. For example, animals with different anatomical structures, such as horses, rabbits, and sheep, exhibit similar grimace expressions when in pain i.e., tighten eyes, tighten mouth, and flattened cheeks, which can be well captured by facial landmarks. Therefore, we believe that our study has enormous significance for automatic understanding of animal faces.

2 EFFECT OF DIFFERENT FEW-SHOT ANIMALS

We report the performance of Meta-CSKT using different few-shot animals in Figure 1. We test different models on the rest of all training images. We explore the effect of few-shot animals on animal face alignment by varying numbers and species. With a pretrained human face alignment model, species are broadly divided into easy, medium, and hard species classes based on NME errors. Generally, images of the easy species class are similar in shape or appearance to humans, images of the medium/hard species classes are related/unrelated to humans. Figure 1 shows that increasing the number of images generally leads to better performance, except for the easy species class. For easy species class, the improvement is not significant because these images can already be well aligned by using the pretrained human model, and using more of these images cannot facilitate knowledge transfer across species. Moreover, using images from medium species class yields the best performance as they are more likely to transfer knowledge across species in

either easy or hard classes. Thus, Meta-CSKT uses images from the medium species class as few-shot labeled animals.

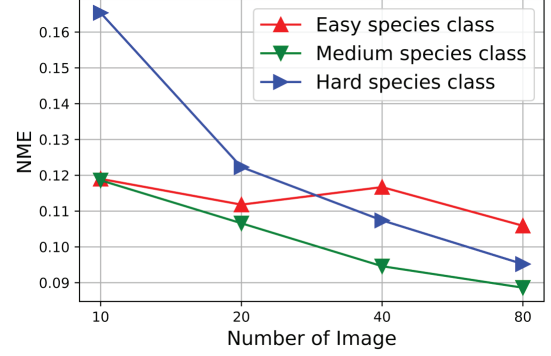


Figure 1: An illustration of the effect of few-shot labeled data in terms of quantity (e.g., 10,20,40,80) and different animal species.

3 ANALYSIS OF THRESHOLD T_{neg}

Figure 2 uses NME error of Meta-CSKT model as a function of the number of excluded negative examples on known and unknown species settings. The thresholds T_{neg} , ranging from 0.5 to 0.1, are marked in the figure. Lower T_{neg} means more hard negative examples are excluded from the large-scale unlabeled data during training. The results show that (1) Meta-CSKT generally obtains better alignment accuracy when excluding large amount of hard negative examples (e.g., 6000, T_{neg} is 0.1) from the large-scale unlabeled data. (2) The performance slightly fluctuates when only a small amount of data is excluded, which is reasonable because the remaining negative examples hinder semi-supervised learning and knowledge transfer between different species in Meta-CSKT.

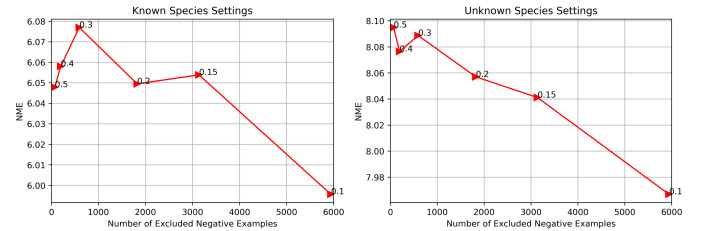


Figure 2: NME error of Meta-CSKT on AnimalWeb with varying numbers of excluded hard negative examples in known/unknown species settings.

We further show the family distribution of the selected hard negative examples in Figure 3. The results show that most species of hard negatives are located in the tail of the histogram, which is

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2023 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

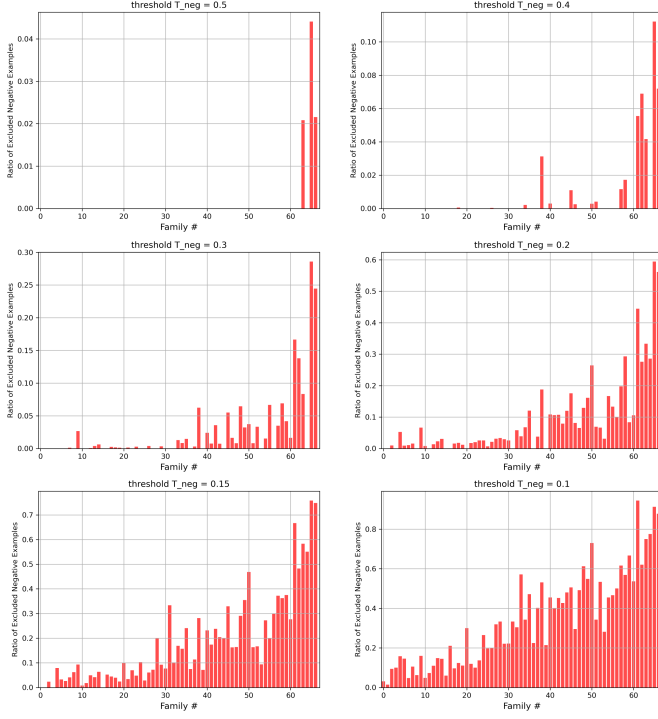


Figure 3: The animal family distribution of excluded hard negative examples by different threshold T_{neg} .

consistent with Fig.1 of the main paper. As T_{neg} decreases, some animal species in the middle of the histogram are also selected. The reason is that some animals have large intra- and inter-species variations and cannot satisfy the flipping constraints. These hard negatives are defined in our positive example mining module.

4 ANALYSIS OF THRESHOLD T_{pos}

Figure 4 uses Meta-CSKT model NME error as a function of the number of included positive examples on known and unknown species settings. The thresholds T_{pos} , ranging from 0.01 to 0.07, are marked in the figure. As results show, (1) including more positive examples as labeled data leads to better accuracy in known and unknown species settings. (2) The accuracy drops when only a small number of positive examples (e.g., 944, T_{pos} is 0.02) are included. The decline is because most positive examples are human-like species that can be perfectly aligned with the human model and Meta-CSKT cannot benefit a lot from bi-directional cross-species knowledge transfer.

We further show the family distribution of the selected positive examples in Figure 5. The results show that most species are located at the head of the histogram, which is in consistent with Fig.1 of the main paper. Specifically, when threshold is very small, only animal species similar to humans can produce low NME and be selected. As T_{pos} increases, more unlabeled data are mined as Meta-CSKT positive examples. Specifically, some animal species in the middle of the histogram are selected. They can facilitate bi-directional cross-species transfer as these animal species are generally positively

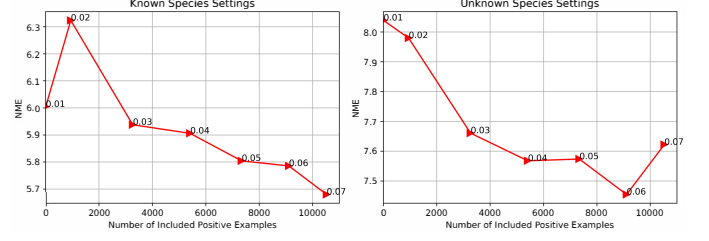


Figure 4: NME error of Meta-CSKT on AnimalWeb with varying numbers of included positive examples in known/unknown species settings.

correlated with many other animals and knowledge is shared among them.

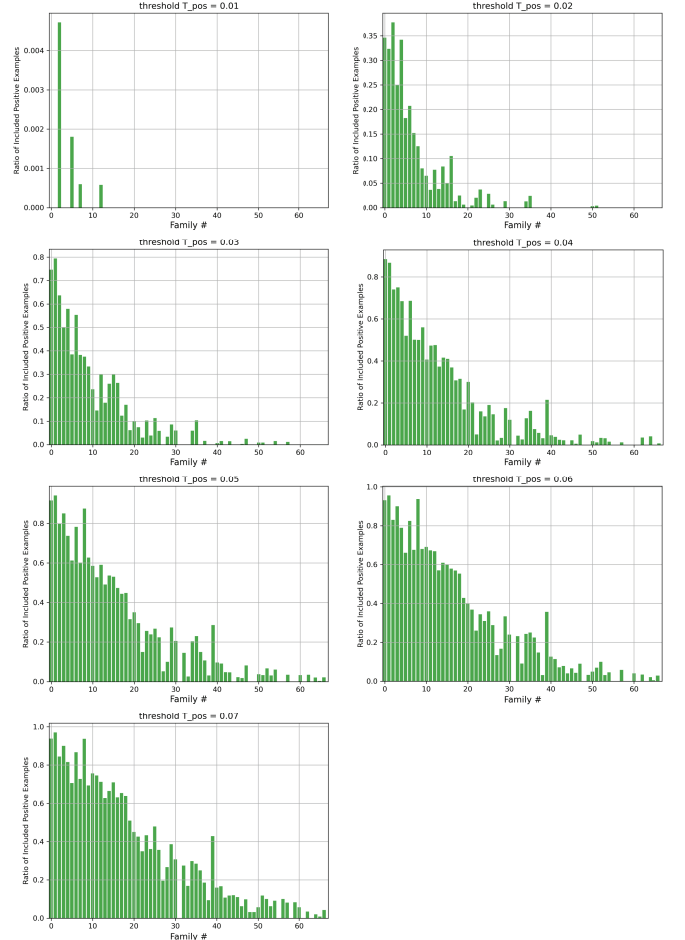


Figure 5: The animal family distribution of included positive examples by different threshold T_{pos} .

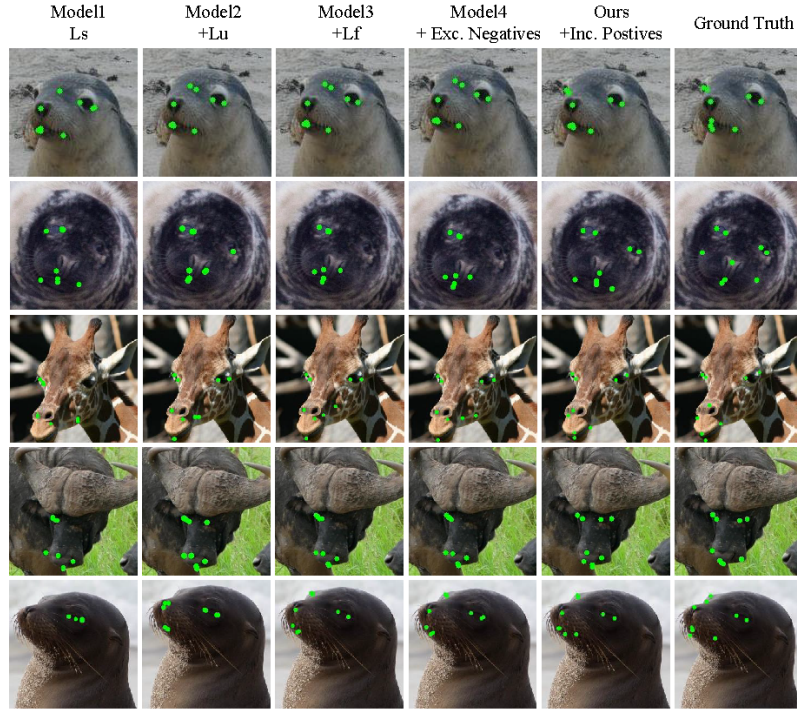


Figure 6: Qualitative examples comparing our Meta-CSKT and our baselines on Animalweb of unknown specie setting. Compared to the left column, the Baseline model in the right column adds a new loss or mining operation by using ‘+’.

5 MORE VISUAL ANALYSIS

In Figure 6, we compare qualitative examples of predicted landmarks generated by different baselines on AnimalWeb. Similar to the results shown in the main paper, Meta-CSKT is more successful in predicting accurate landmarks in important facial regions such as the mouth, nose, and eyes. The first three columns explore the

effect of Meta-CSKT using different loss functions which verifies the bi-directional cross-species knowledge transfer works. From four to five columns explore the effect of using proposed positive example mining and indicate the consistent performance improvement of using the proposed mining method. Despite animal faces exhibiting large variations due to extremely high variability in facial appearance, our method produces reasonably accurate landmarks.