BARC: Learning to Regress 3D Dog Shape from Images by Exploiting Breed Information – Supplementary Material

Nadine Rüegg[1,2], Silvia Zuffi[3], Konrad Schindler[1] and Michael J. Black[2]

[1]ETH Zürich, Switzerland
[2]Max Planck Institute for Intelligent Systems, Tübingen, Germany
[3]IMATI-CNR, Milan, Italy

1. AMT Perceptual Studies

3D shape evaluation based on 2D reprojection errors can be misleading. Figure 1 shows an example where the IoU score is high, but the estimated 3D shape of the dog not accurate. In order to better evaluate predicted shapes in 3D, we propose an evaluation based on breed prototype consistency as well as perceptual studies. While results of all evaluation methods are shown in the main paper, we elaborate here more on our procedure to perform perceptual studies. Controlled perceptual tasks are designed to evaluate our method relative to (1) the SOTA or (2) to an ablated model. Workers on Amazon Mechanical Turk (AMT) judge which of two rendered 3D body shapes better fits a query dog image. Figure 2 shows the framework that we provide to the AMT workers. We show each worker an image that contains a dog, our predicted 3D model in T-pose and the model in T-pose from SOTA or ablated method. We do not present the predicted 3D posed models in order to focus workers on shape. The left-right ordering of the rendered meshes is random. We let each worker first process 8 samples to get used to the task and then use the next 30 hits. The task is split in 4 batches with 30 samples each. We have 10 workers for each batch. This gives us a total of 1200 hits. In order to verify the workers understand the task and perform it diligently, we include two catch trials in each batch. These are extreme cases where one 3D shape is so far off that only one answer is plausible. For all quantitative results reported, votes from workers who failed one or both catch trials are ignored.

Figure 1. *Misleading reprojection errors* Both IoU and PCK are sometimes misleading, as they can be high for poor 3D estimates.

Figure 2. *AMT Framework.* The picture shows an example screenshot from the perceptual studies that we ran on Amazon Mechanical Turk.

2. 3D CG Models

We propose to use 3D CG models to help training our network, in case such models are available. See Tab. 1 for a list of 3D CG models and corresponding breeds which BARC uses in its 3D model loss.

Breed	Stanford Extra Name
American Staffordshire Terrier	n02093428-American_ Staffordshire_terrier
Boxer	n02108089-boxer
German Shepherd	n02106662-German_shepherd
Doberman	n02107142-Doberman
Staffordshire Bullterrier	n02093256-Staffordshire _bullterrier
French Bulldog	n02108915-French_bulldog
Bull Mastiff	n02108422-bull_mastiff
Great Dane	n02109047-Great_Dane
Italian Greyhound	n02091032-Italian_greyhound
Rottweiler	n02106550-Rottweiler
Siberian Husky	n02110185-Siberian_husky

Table 1. *3D CG models.* Models used for our 3D model loss L^B_{3D}

3. Keypoint Weights

Table 2 shows for each keypoint the weight that was used as part of the weighted keypoint loss.

keypoint	w	keypoint	w
left front leg, paw	3	right rear leg, top	2
left front leg, middle	2	tail start	3
left front leg, top	2	tail end	3
left rear leg, paw	3	base left ear	2
left rear leg, middle	2	base right ear	2
left rear leg, top	2	nose	3
right front leg, paw	3	chin	1
right front leg, middle	2	left ear tip	2
right front leg, top	2	right ear tip	2
right rear leg, paw	3	left eye	1
right rear leg, middle	2	right eye	1

Table 2. *Keypoint weights.* Weights that are used within the weighted keypoint loss.

4. Failure Case Analysis

We divide the failure cases in two main groups: shape and pose failures.

Pose Failure Cases: At development time we have trained our network with various pose priors, such as a mixture of gaussians prior as in [1, 3], a variational auto-encoder as in [2] and our final normalizing flow pose prior. One failure mode that goes through all priors is the erroneous prediction of dogs not facing the camera. The Stanford Extra training set is unbalanced in the sense that it shows many dogs from a front- or side-view. Furthermore, most of the dogs do not bend the front legs as they are either sitting, laying or standing, this leads to challenges when predicting poses for dogs with heavily bent wrists. As training with different pose priors lead to similar error cases, we believe that those challenges are not structural problems of the pose prior, but rather of the image dataset. Nevertheless, it might be worth examining different training schedules such that rare poses obtain higher weights or are repeated more often. One more thing worth mentioning is, that often perceived 3D quality from front view is considerably higher than from side-views. A strong 3D regularization is inevitable. Predictions for laying and sitting dogs could be improved by training a pose prior on a more suitable 3D pose dataset. Furthermore, BARC has troubles predicting poses for dogs that are only partly visible.

Shape Failure Cases: Our breed losses help to regularize dog shape. BARC can predict more reasonable shapes, especially for dogs that are not fully visible from the side. Never the less, we do sometimes observe shortened limbs when they are difficult to predict due to poses such as a dog laying and facing the camera. As discussed in the main paper, working with a single shape for each dog breed is not an option, as there is not negligible intra-class variability. Another challenge is dog hair. First, shape variability can become enormous, consider for example differently sheared poodles. Secondly, long hair does swing and the shape that we want to predict for a dog with fluffy hair is not clearly defined. In such cases, representing a dog with a mesh is not ideal.

Some Visual Examples of Failure Cases: We show four failure cases in Figure 3: (1) a dog which is not fully visible, our prediction shows a shrunken body. (2) most training images show dogs that face the camera. When the dog is turned away, pose prediction fails. (3) a Japanese Spaniel with lots of hair. Shape prediction for such breeds is difficult. (4) A dog that is hard to recognize and where, in part, the difficult pose is compensated by a wrong shape – instead of bending the back, the dog is given a stouter body.

Figure 3. *Failure Cases.*Pose and shape failure cases.

5. Qualitative Results

In this section we present additional qualitative results. Figure 4 shows results for ablated versions of BARC. To the left we render results from our method without any of the breed related losses, in the middle results with the breed similarity loss only and to the right with the breed similarity loss as well as the 3D CG model loss. For each of the three versions we show front as well as a side view. Finally, we test BARC on images of previously unseen breeds. All of those images are downloaded from the American Kennel Club web page. Figure 5 illustrates an overlay of our prediction on the input image as well as front and side view for each of the seven dogs. We observe that BARC can generalize well to new breeds. Furthermore it generalizes to puppies, as illustrated in Figure 6. For the figures in the paper, we select results to illustrate variety. Last, in Figure 7 we present results on completely randomly sampled Stanford Extra test set images. For each input image we show the overlap of our prediction with this image, a 3D visualization of our prediction and a 3D visualization of the previous state-of-the-art method WLDO.

References

[1] Benjamin Biggs, Ollie Boyne, James Charles, Andrew Fitzgibbon, and Roberto Cipolla. Who left the dogs out: 3D animal reconstruction with expectation maximization in the loop. In *ECCV*, pages 195–211, 2020.

[2] Silvia Zuffi, Angjoo Kanazawa, Tanya Berger-Wolf, and Michael J Black. Three-D safari: Learning to estimate zebra pose, shape, and texture from images ”in the wild”. In *ICCV*, pages 5359–5368, 2019.

[3] Silvia Zuffi, Angjoo Kanazawa, and Michael J Black. Lions and tigers and bears: Capturing non-rigid, 3d, articulated shape from images. In *CVPR*, pages 3955–3963, 2018.

Figure 4. *Ablation Study.* Qualitative comparison of from left to right (1) our method trained without any breed losses (2) our method trained with similarity breed loss only (3) BARC (our method). We show for various input images front views as well as side views.

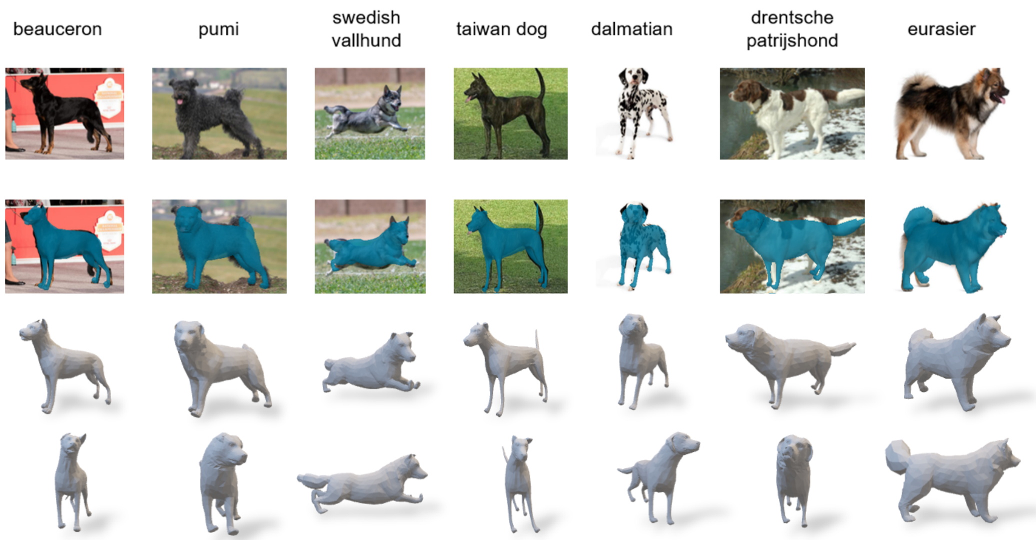

Figure 5. *Results for Unseen Breeds.* Qualitative results of BARC (our method) on images of previously unseen breeds. All test images are downloaded from the American Kennel Club web page. We show for various input images an overlay, front view as well as side view of our predicted dog.

Figure 6. *Puppies.* Qualitative results on puppies from the Stanford Extra test set.

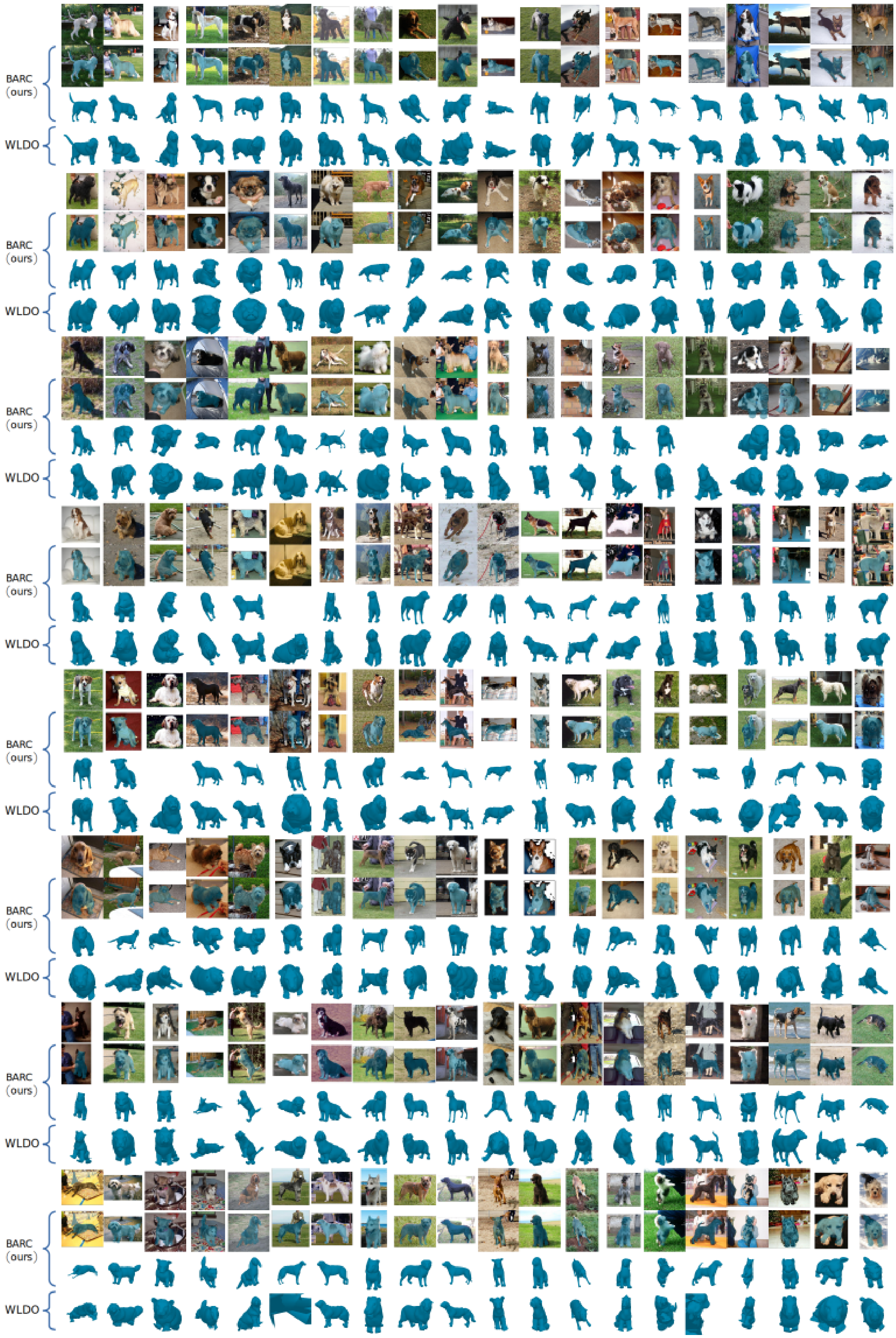

Figure 7. *Randomly sampled results.* We show qualitative results on the Stanford Extra test set: for each sample an input image, the overlay of our prediction (BARC) with that image, our prediction and previous state-of-the-art (WLDO).