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

Nadine Rüegg [1, 2, 4] Silvia Zuffi [3] Konrad Schindler [2] Michael J. Black [1]

(1) Max Planck Institute for Intelligent Systems – Dept. Perceiving Systems
(2) ETH Zürich - Photogrammetry and Remote Sensing Lab
(3) Consiglio Nazionale delle Ricerche – IMATI
(4) Max Planck ETH Center for Learning Systems

Motivation

Goal

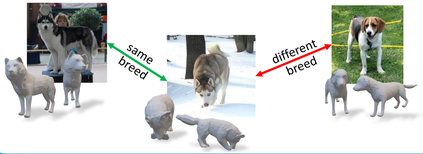
Reconstruct detailed, realistic 3D dogs, represented as 3D meshes, directly from monocular images.

Problem

Shape variations for dogs are high and no ground truth 3D shapes are available.

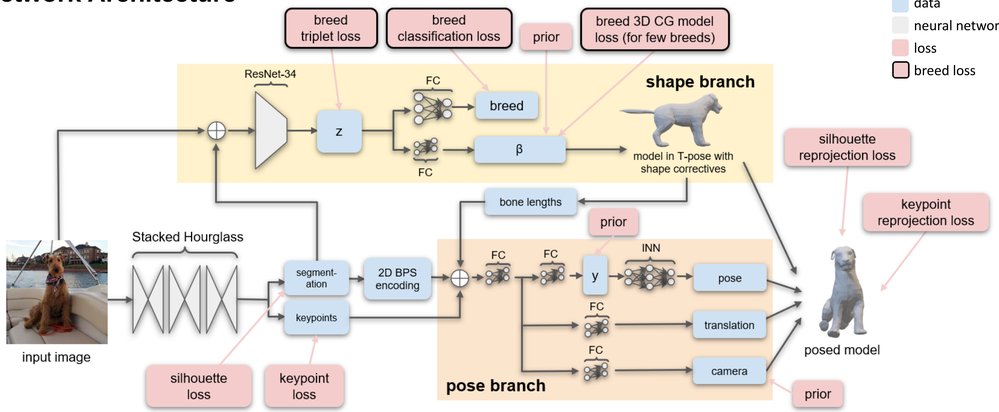
Key Idea

Use side information: dogs of the same breed have more similar shapes than dogs belonging to different breeds.



Method

Network Architecture



Breed Losses

- *Breed similarity loss* L^B_{sim}: consists of a triplet loss $L^B_{triplet}$ and a classification loss L^B_{cs}.

$$L^B_{triplet} = \sum_{i=1}^{N_{triplets}} \max(d(z_{a,i}, z_{p,i}) - d(z_{n,i}, z_{a,i}) + m,\ 0)$$

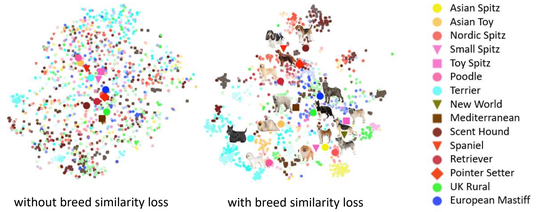
$$L^B_{cs} = -\sum_{c=1}^{N_{classes}} y_{o,c} \log(p_{o,c})$$

- *Breed 3D model loss*: auxiliary loss that penalizes differences from a 3D model of the same breed, when available.

$$L^B_{3D} = \sum_j^{N_{\beta pca}} (\beta^{pred}_{pca,j} - \beta^{breed}_{pca,j})^2 + \sum_k^{N_\kappa} (\kappa^{pred}_k - \kappa^{breed}_k)^2$$

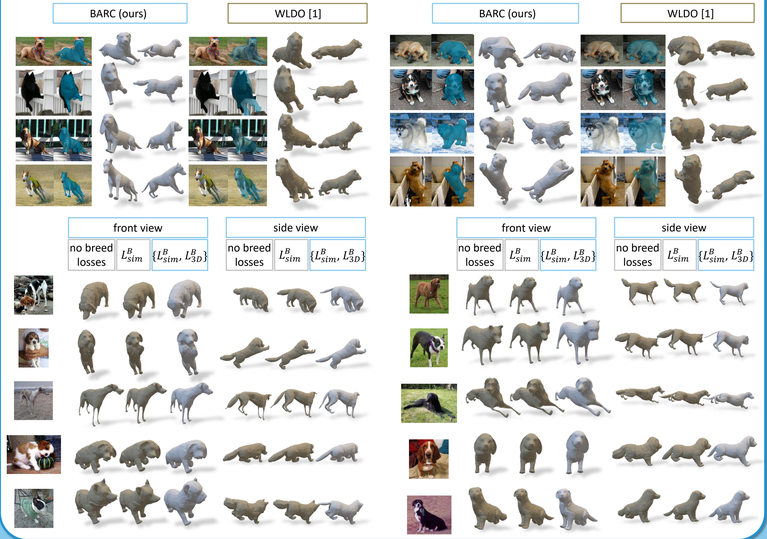
Results

T-SNE Plot of the Latent Shape Space



without breed similarity loss with breed similarity loss

Qualitative Results



Quantitative Results

All results are calculated on the Stanford Extra dataset [1].

Reprojection Errors

Method	IoU	PCK @ 0.15				
		Avg	Legs	Tail	Ears	Face
3D-M [2]	69.9	69.7	68.3	**68.0**	57.8	**93.7**
CGAS [3]	63.5	28.6	30.7	34.5	25.9	24.1
WLDO [1]	74.2	78.8	76.4	63.9	78.1	92.1
BARC	**75.1**	**82.8**	**82.3**	63.3	**83.3**	91.3

Breed Prototype Evaluation

Method	WLDO	BARC		
		no breed losses	L^B_{sim}	$\{L^B_{sim}, L^B_{3D}\}$
Error [m]	0.1155	0.0858	0.0776	**0.0695**

Perceptual Shape Evaluation

Experiment Settings	AMT Results	
	Votes	Percentage
L^B_{sim} vs. no breed losses	638 : 382	**62.55%** : 37.45%
$\{L^B_{sim}, L^B_{3D}\}$ vs. L^B_{sim}	670 : 440	**60.36%** : 39.64%
$\{L^B_{sim}, L^B_{3D}\}$ vs. WLDO	998 : 82	**92.41%** : 7.59%

All results are updated w.r.t. the CVPR paper, we refer to the arXiv version.

References: [1] B. Biggs et al., Who left the dogs out: 3D animal reconstruction with expectation maximization in the loop. In ECCV, 2020. [2] S. Zuffi et al., 3D menagerie: Modeling the 3D shape and pose of animals. In CVPR, 2017. [3] B. Biggs et al., Creatures great and SMAL: Recovering the shape and motion of animals from video. In ACCV, 2018.

Model available at: https://barc.is.tue.mpg.de/