.	Shape	Appearance		Classification Accuracy (%)				
		layout type	using ground truth	family	breed (S. 4.2)		both (S. 4.3)	
				(S. 4.1)	cat	dog	hierarchical	flat
1	✓	–	–	94.21	NA	NA	NA	NA
2	–	Image	–	82.56	52.01	40.59	NA	39.64
3	–	Image+Head	–	85.06	60.37	52.10	NA	51.23
4	–	Image+Head+Body	–	87.78	64.27	54.31	NA	54.05
5	–	Image+Head+Body	✓	88.68	66.12	57.29	NA	56.60
6	✓	Image	–	94.88	50.27	42.94	42.29	43.30
7	✓	Image+Head	–	95.07	59.11	54.56	52.78	54.03
8	✓	Image+Head+Body	–	94.89	63.48	55.68	55.26	56.68
9	✓	Image+Head+Body	✓	95.37	66.07	59.18	57.77	59.21

Table 4. **Comparison between different models.** The table compares different models on the three tasks of discriminating the family, the breed given the family, and the breed and family of the pets in the Oxford-IIIT Pet dataset (Sect. 2). Different combinations of the shape features (deformable part model of the pet faces) and of the various appearance features are tested (Sect. 3.2, Fig. 4).

for the cat class. The same is done to obtain a score for the dog class. Then a linear SVM is learned to discriminate between cats and dogs based on these two scores. The classification accuracy of this model on the Oxford-IIIT Pet test data is 94.21%.

Appearance only. Spatial histograms of visual words are used in a non-linear SVM to discriminate between cats and dogs, as detailed in Sect. 3.2. The accuracy depends on the type of spatial histograms considered, which in turn depends on the layout of the spatial bins. On the Oxford-IIIT Pet test data, the *image layout* obtains an accuracy of 82.56%; adding head information using *image+head layout* yields an accuracy of 85.06%. Using *image+head+body layout* improves accuracy by a further 2.7% to 87.78%. An improvement of 1% was observed when the ground-truth segmentations were used in place of the segmentations estimated by grab-cut (Sect. 3.2). This progression indicates that the more accurate the localization of the pet body, the better is the classification accuracy.

Shape and appearance. The appearance and shape information are combined by summing the exp-χ^2 kernel for the appearance part (Sect. 3.2) with a linear kernel on the cat scores and a linear kernel on the dog scores. The combination boosts the performance by an additional 7% over that of using appearance alone, yielding approximately 95.37% accuracy (Table 4, rows 5 and 9), with all the variants of the appearance model performing similarly.

The ASIRRA data. The ASIRRA data does not specify a training set, so we used models trained on the Oxford-IIIT Pet data and the ASIRRA data was used only for testing. The accuracy of the shape model on the ASIRRA data is 92.9%, which corresponds to a 42% probability of breaking the test in a single try. For comparison, the best accuracy reported in the literature on the ASIRRA data is 82.7% [25], which corresponds to just a 9.2% chance of breaking the test. Due to lack of sufficient training data to train appearance models for ASIRRA data, we did not evaluate these models on ASIRRA dataset.

Method	Mean Class. Accuracy
Golle *et al.* [25]	82.7%
This paper (Shape only)	92.9%

Table 5. **Performance on ASIRRA Data.** Table shows performance achieved on task of pet family classification posed by the ASIRRA challenge. Best results obtained by Golle [25] were obtained using 10000 images from the data. 8000 for training and 2000 for testing. Our test results are shown on 24990 images in the ASIRRA dataset.

4.2. Breed discrimination

This section evaluates the models on the task of discriminating the different breeds of cats and dogs given their family. This is done by learning a multi-class SVM by using the 1-Vs-rest decomposition [37] (this means learning 12 binary classifiers for cats and 25 for dogs). The relative performance of the different models is similar to that observed for pet family classification in Sect. 4.1. The best breed classification accuracies for cats and dogs are 63.48% and 55.68% respectively, which improve to 66.07% and 59.18% when the ground truth segmentations are used.

4.3. Family and breed discrimination

This section investigates classifying both the family and the breed. Two approaches are explored: *hierarchical classification*, in which the family is decided first as in Sect. 4.1, and then the breed is decided as in Sect. 4.2, and *flat classi-*