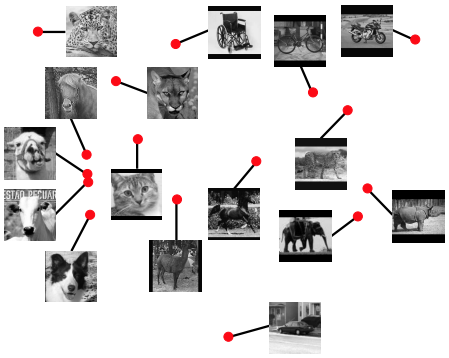

Figure 3. MDS embedding of pairwise distances between the learned part distributions for 16 object categories.

of Sec. 3.2 were not needed. For our Matlab implementation, each sampling iteration requires roughly 0.1 seconds per training image. The learning procedure showed little sensitivity to the part distribution hyperparameters, which were set to provide a weak ($\nu_p = 6$) bias towards moderate covariances and sparse ($\beta = 0.1$) appearance densities. The object–specific part distribution hyperparameter, α, was set via cross–validation as discussed below.

Following 200 iterations of the Gibbs sampler, we used the final assignments $\mathbf{z}$ to estimate each part's posterior distribution over feature appearance and position (Sec. 3.3). In Fig. 2, we visualize these distributions for seven parts. Only two parts seem specialized to a single category: a spotted texture part used by the "leopard face" category, and another part devoted to the extremely well aligned "side car" category. The next three parts model features from animal mouths, animal legs, and vehicles, respectively. We also show two of several parts which seem to model background clutter around image boundaries, and are widely shared between categories.

To further investigate these shared parts, we used the symmetrized KL divergence [15] to compute a distance between all pairs of object–specific part distributions:

$$D\left(\theta_k, \theta_\ell\right) = \sum_{j=1}^{P} \theta_k(j) \log \frac{\theta_k(j)}{\theta_\ell(j)} + \theta_\ell(j) \log \frac{\theta_\ell(j)}{\theta_k(j)} \quad (15)$$

Fig. 3 shows the two–dimensional embedding of these distances produced by multidimensional scaling (MDS) [16]. Except for cars, these part distances seem to closely match our own intuitive notions of category similarity.

4.2. Detection and Recognition

To evaluate our model, we consider two sets of experiments. In the detection task, we use 100 training images to learn an 8-part background appearance model, and then use probabilities computed as in Sec. 3.3 to classify test images as object or background. To facilitate comparisons, we also consider a recognition task in which test images are classified as either their true category, or one of the 15 other

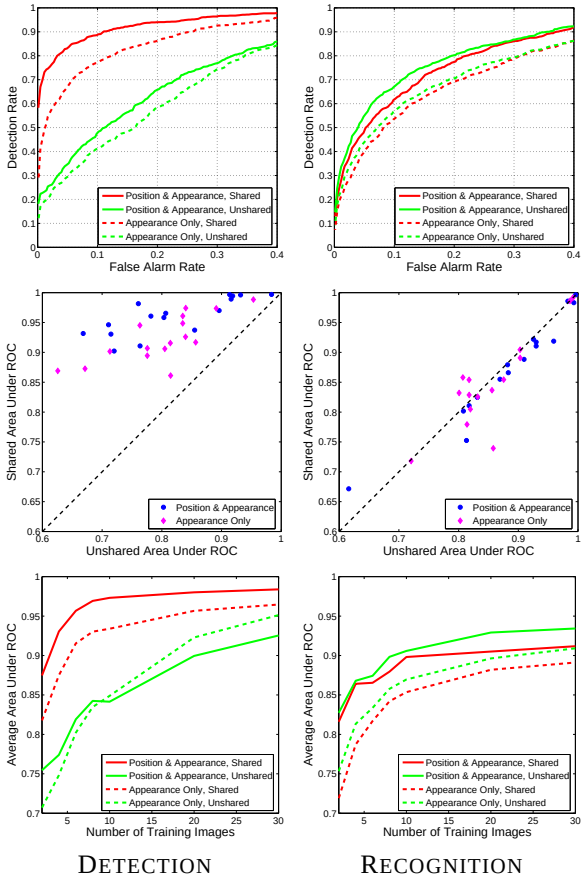

Figure 4. Performance for the tasks of detection (left) and recognition (right) of 16 object categories. TOP: **Average of ROC curves across all categories (6 training images).** MIDDLE: **Scatter plot of areas under ROC curves for the shared and unshared models of each category (6 training images).** BOTTOM: **Area under average ROC curves for different numbers of training images per category.**

categories. For both tasks, we compare our *shared* model of all object categories to a set of 16 *unshared* models trained on individual categories. We also consider versions of both models which neglect the spatial location of features, as in recent "bag of keypoints" approaches [3, 17]. Performance curves average over three randomly chosen training sets of the given size, and use all other images for testing.

As shown in Fig. 4, we find that shared parts lead to consistent, significant improvements in detection performance. These improvements are greatest when fewer than 10 training examples per category are available. For the recognition task, the shared and unshared models perform similarly, with the shared model becoming slightly less effective when many training examples are available. Confusion matrices (not shown) confirm that this slight performance degradation is produced by pairs of categories with very similar part distributions (see Fig. 3). For both tasks, feature po-