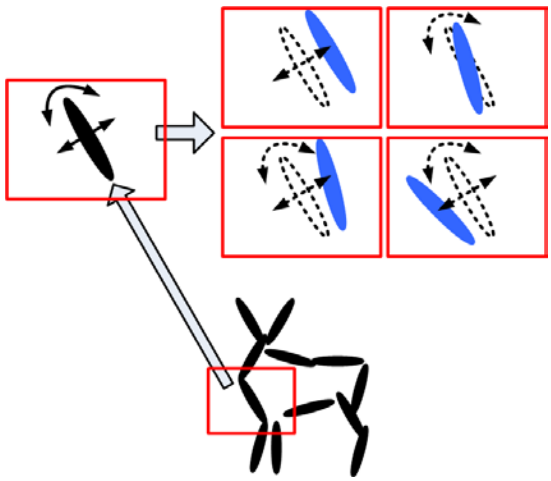
# Findings

## The basic model

Our work for the period July 1, 2010 to December 31, 2010 focuses on further developping the active basis model. The model can be encoded by the following equation:

$$I_m = \sum_{i=1}^{n} c_{m,i} B_{x_i+\Delta x_{m,i},s,\alpha_i+\Delta\alpha_{m,i}} + U_m,$$

where $m = 1, \ldots, M$ indexes the training images, and $I_m$ is the $m$-th training image. Each image $I_m$ is modeled by a linear combination of $n$ basis elements indexed by $i = 1, \ldots, n$. $c_{m,i}$ is the coefficient of the $i$-th basis element $B_{x_i+\Delta x_{m,i},s,\alpha_i+\Delta\alpha_{m,i}}$, where $x_i$ is the nominal location of the $i$-th basis element, $s$ is the scale assumed to be fixed, and $\alpha_i$ is the nominal orientation of the $i$-th basis element. $(\Delta x_{m,i}, \Delta\alpha_{m,i})$ are the hidden variables to accunt for the activity of the $i$-th basis element in location and orientation when representing image $I_m$.



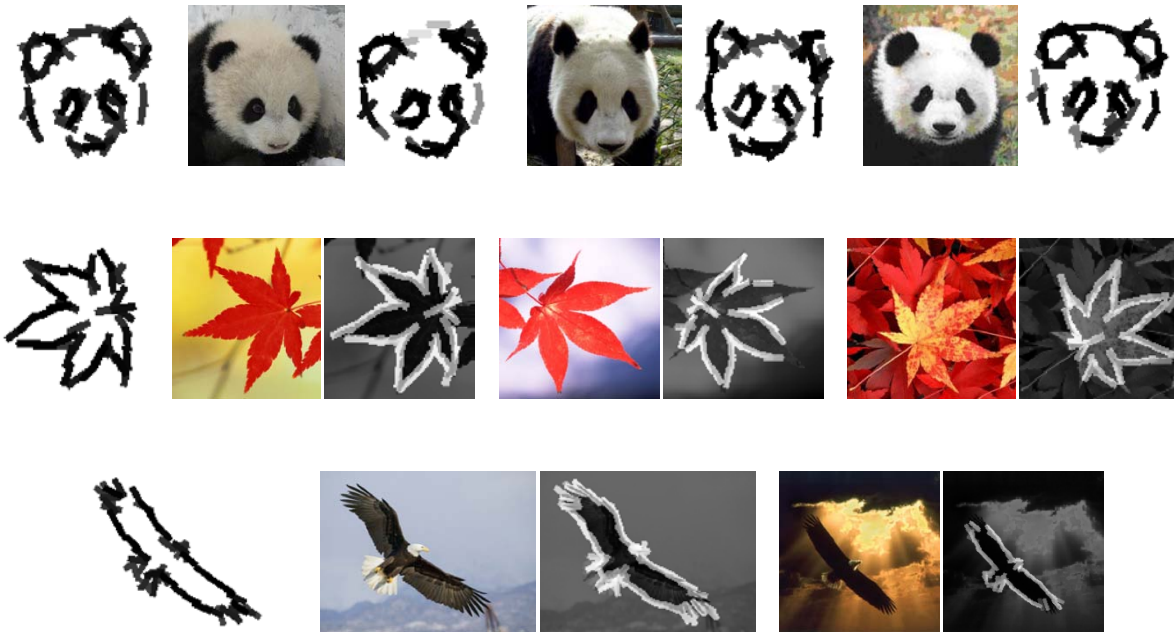
The $n$ basis elements at their nominal locations and orientations form a template. The activities $(\Delta x_{m,i}, \Delta\alpha_{m,i}, i = 1, \ldots, n)$ account for the deformation of the template. The above figure illustrates the basis idea, where each basis element is illustrated by a thin ellipsoid. A black ellipsoid at its norminal location and orientation can be shifted to the blue ellipsoids when coding images.

During the learning process, the basis elements, or more specificially, their nominal locations and orientations, are selected from a dictionary of Gabor wavelet elements at a dense collection of locations and orientations.

# Finding 1: Learning from non-aligned images

We study the problem of learning active basis model from training images where the objects of interest may appear at different locations, scales and orientations. Compared to our past work, we further allow

the uncertainties in orientations. This is important because the objects often appear at different orientations in training images. The following illustrates three experiments of the learning results, where the first picture in each row is the learned template where the selected basis elements are at their nominal locations and orientations. Each selected basis element is illustrated by a bar of the same location, length and orientation. The rest of the pictures in each row display the training images and their corresponding sketches, obtained by deforming the template in the first picture. Due to the space limitation, we only show two to three training images for illustration.



We create a small library of 133 object categories, where each category contains 10 to 50 training images. We are able to learn an active basis template for each object category. The learned templates are all very meaningful. Some of them at listed in the appendix.

# Finding 2: Clustering

It is often the case that there are multiple clusters in the training images. These clusters may correspond to different categories of objects or different poses of the same category. We extend our previous clustering algorithm to allow for uncertainties in locations, scales and orientations of the objects in the training images. This is important because in clustering images, we cannot expect the images are all well-aligned, even if the bounding box of the object is given in each image.

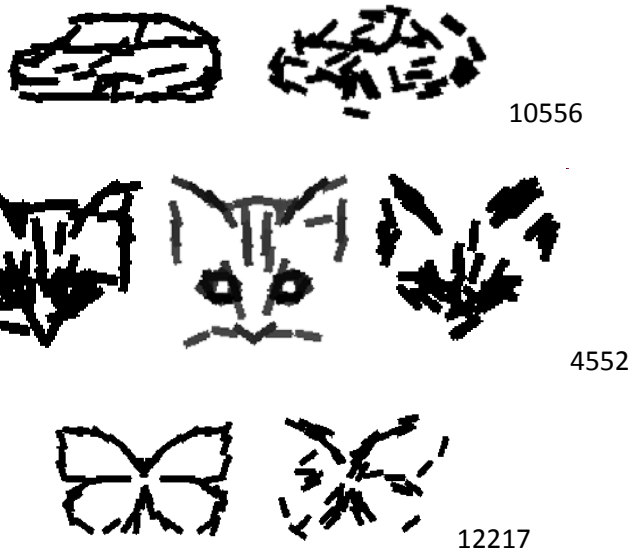
The above figure illustrates an experiment, where our clustering method can learn different animal faces from training images without the knowledge of the types of the objects in the training images.

The above figure illustrates another experiment, where the algorithm is capable of separating two types of faces of cows.

We are currently working on scaling up this experiment, so that we can learn a large dictionary of templates or part-templates from natural scene images or images of different object categories. The ability of learning such “visual words” is an important task of unsupervised learning in vision. Such work may shed light on the neuron cells in V2 area of the visual cortex. The learned dictionary can also be useful for a variety of tasks, including classification.

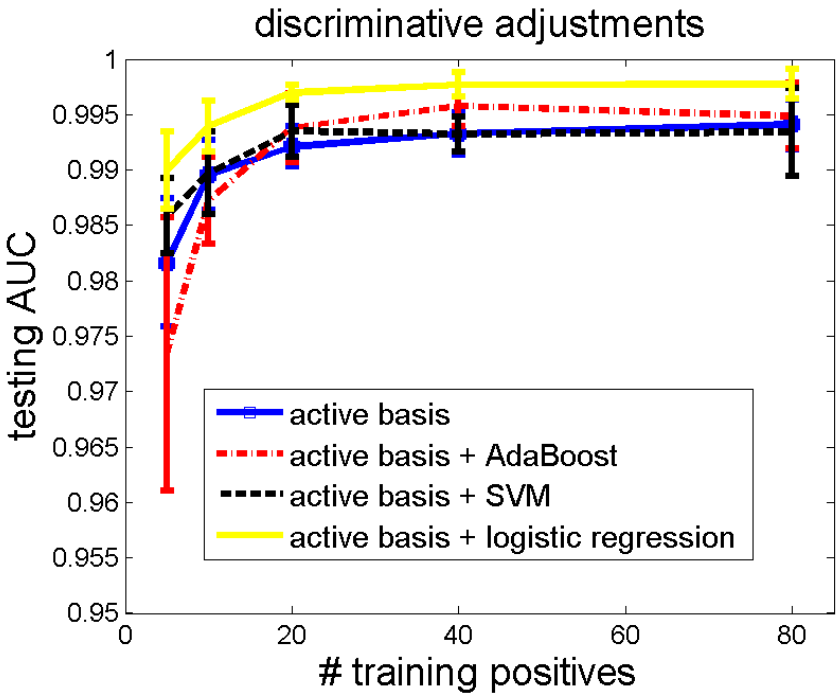
# Finding 3: Classification

We also study the problem of classification, or what is called discriminative learning, and compare it with the generative learning of the active basis model.




First we compare the templates learned by the active basis model and the adaboost method. The active basis model seeks to explain the positive training images. The learning does not require negative images. The adaboost method seeks to separate the positive training images from the negative training images. In each row of the above figure, the picture on the left shows the template learned by the active basis model. The picture on the right shows the template learned by adaboost. In the second row, the picture in the middle is the template learned by the active basis by allowing the local shifts in locations, orientations and scales of the objects in training images. The number on each row is the number of negative training images used to train the adaboost classifier. It appears that the active basis model learns cleaner and more meaningful templates than the adaboost method. It is also easier to incorporate hidden variables such as unknown locations and orientations into the active basis model as illustrated by the cat example in the

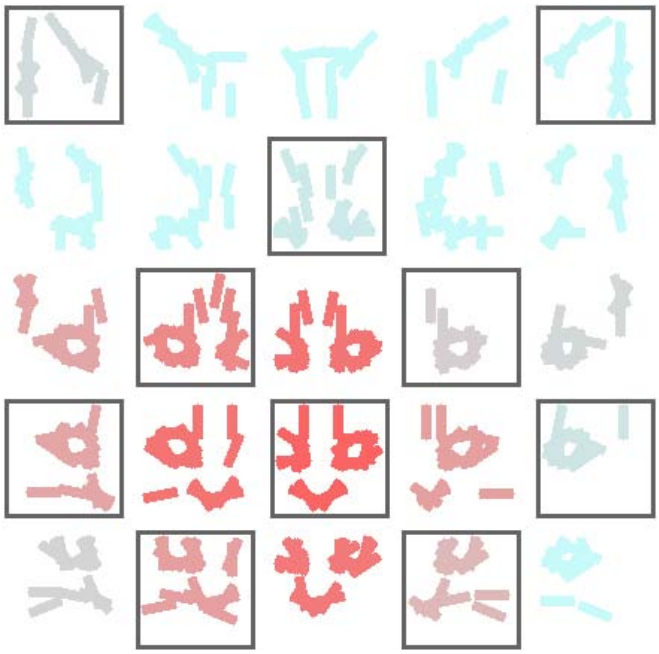
second row. We believe that the active basis model may be more useful for unsupervised learning, such as the clustering problem mentioned in Finding 2.

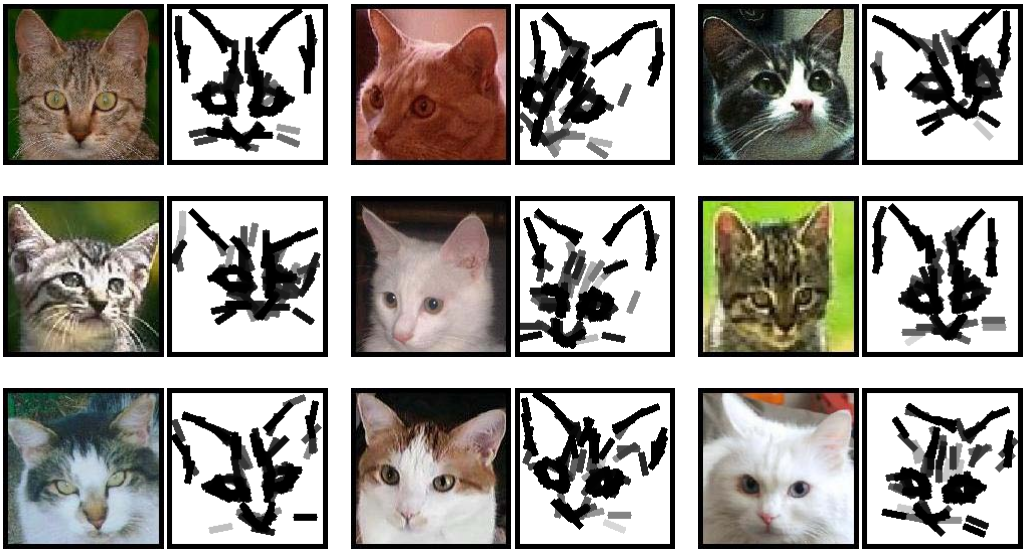



Second, we study the issue of adjusting the active basis model by logistic regression. In learning the active basis model, the basis elements are selected by approximately maximizing the likelihood of the positive training images. The learning does not require negative examples. But if our goal is classification, we can include the negative examples, and adjust the parameters of the learned active basis model by logistic regression. Given the hidden variables, the probability distribution in the active basis model is in the form of an exponential family model relative to a reference distribution. If we take the reference distribution to be the distribution of all negative examples, then this leads naturally to a logistic regression if we include negative examples. However, as is commonly the case, fitting logistic regression without regularization often leads to over-fitting. So we include the L-2 penalty term to regularize the logistic regression. The adjusted model leads to better classification performance on testing images, as shown in the above figure. The L-2 regularized logistic regression also outperforms adjustment by SVM.

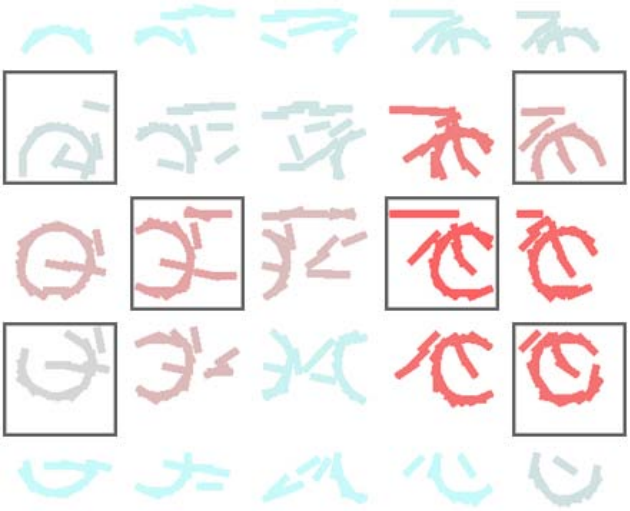
# Finding 4: Hierarchical active basis model

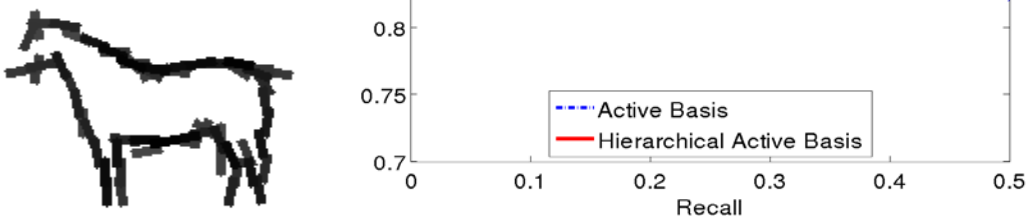
The active basis model is a composition of Gabor basis elements that are allowed to shift their locations and orientations. We can use active basis models as part-templates, and further compose them into a whole template. This leads to a recursive hierarchical model, where the whole template consists of part-templates that can shift and rotate relative to each other, while each part-template consists of basis elements that can also shift and rotate relative to each other. The model consists of two layers of activities. One is the activities of the part-templates. The other is the activities of the basis elements of each part-template on top of the overall activity of this part-template. Such a hierarchical active basis model has more flexibility for modeling large deformations and articulations in objects.





The above two figures illustrate an example of learning a hierarchical template of cat heads. The learning algorithm learns part-templates, such as ears, eyes, nose etc. Then the learning algorithm selects the part-templates and composes them into a whole-template. The learned hierarchical template can account for large deformations in the training images caused by changes in views and poses. The following two figures show the learning result for tandem-bikes.



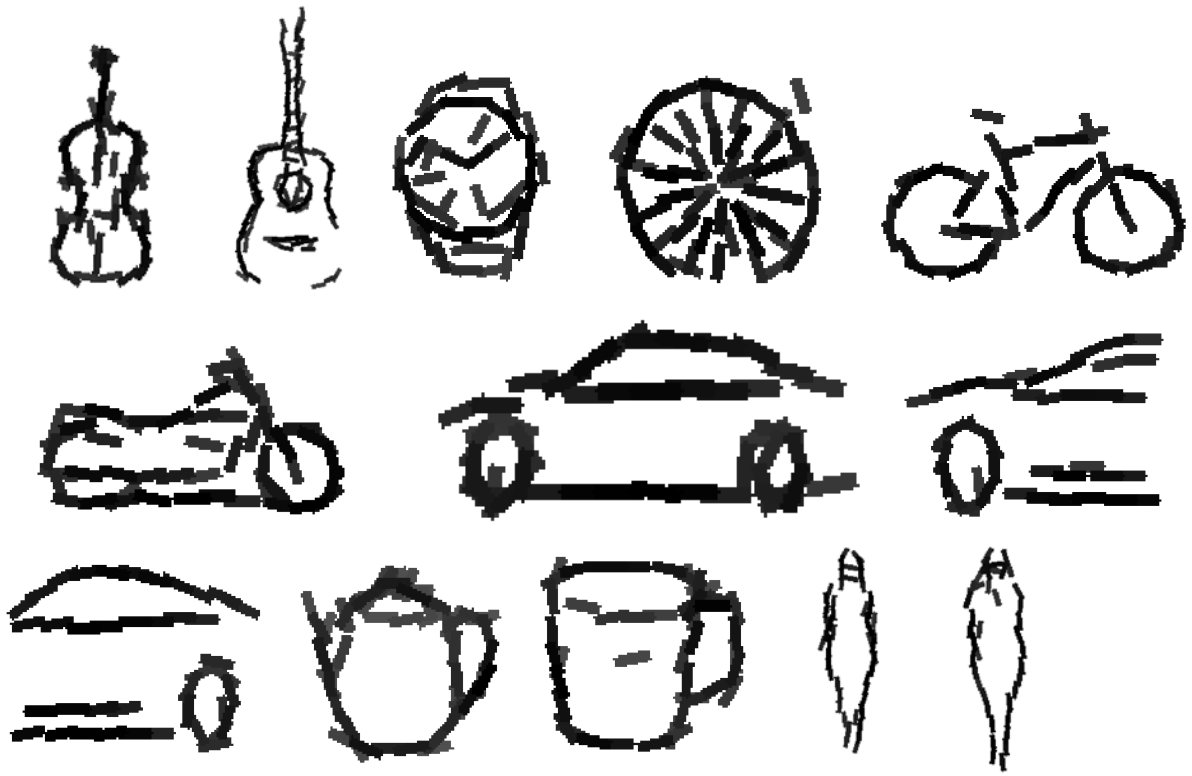



We also conduct an experiment on classification performance of the hierarchical active basis model and the flat active basis model without parts. In the above figure, the image on the left is the whole-template

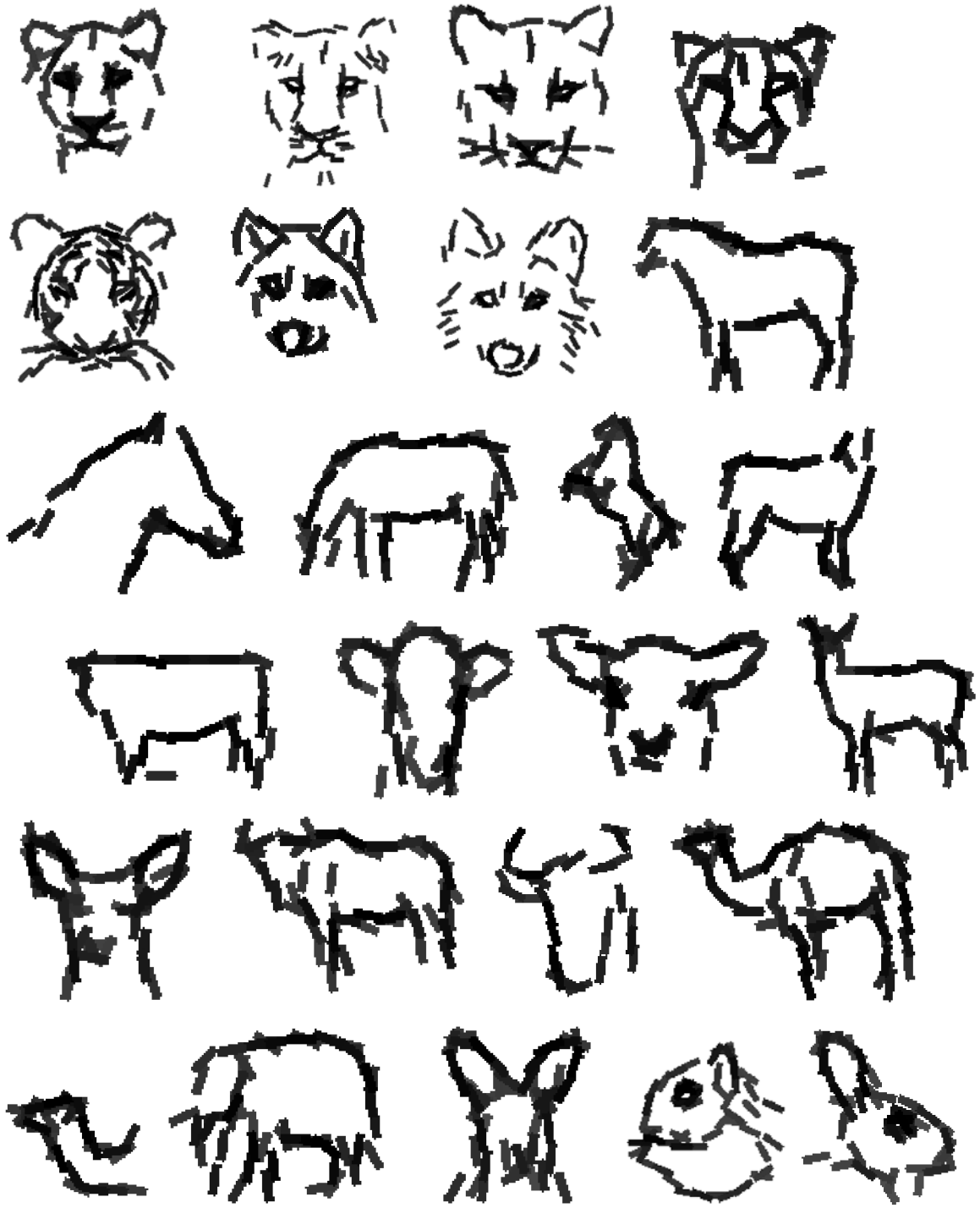
learned by hierarchical model from the Weizmann horse dataset. The plot on the right shows the precision-recall results of the two models. The hierarchical model outperforms the flat model.

We are also investigating the completely unsupervised learning of the part-templates from natural scene images or images of different object categories by our clustering method.

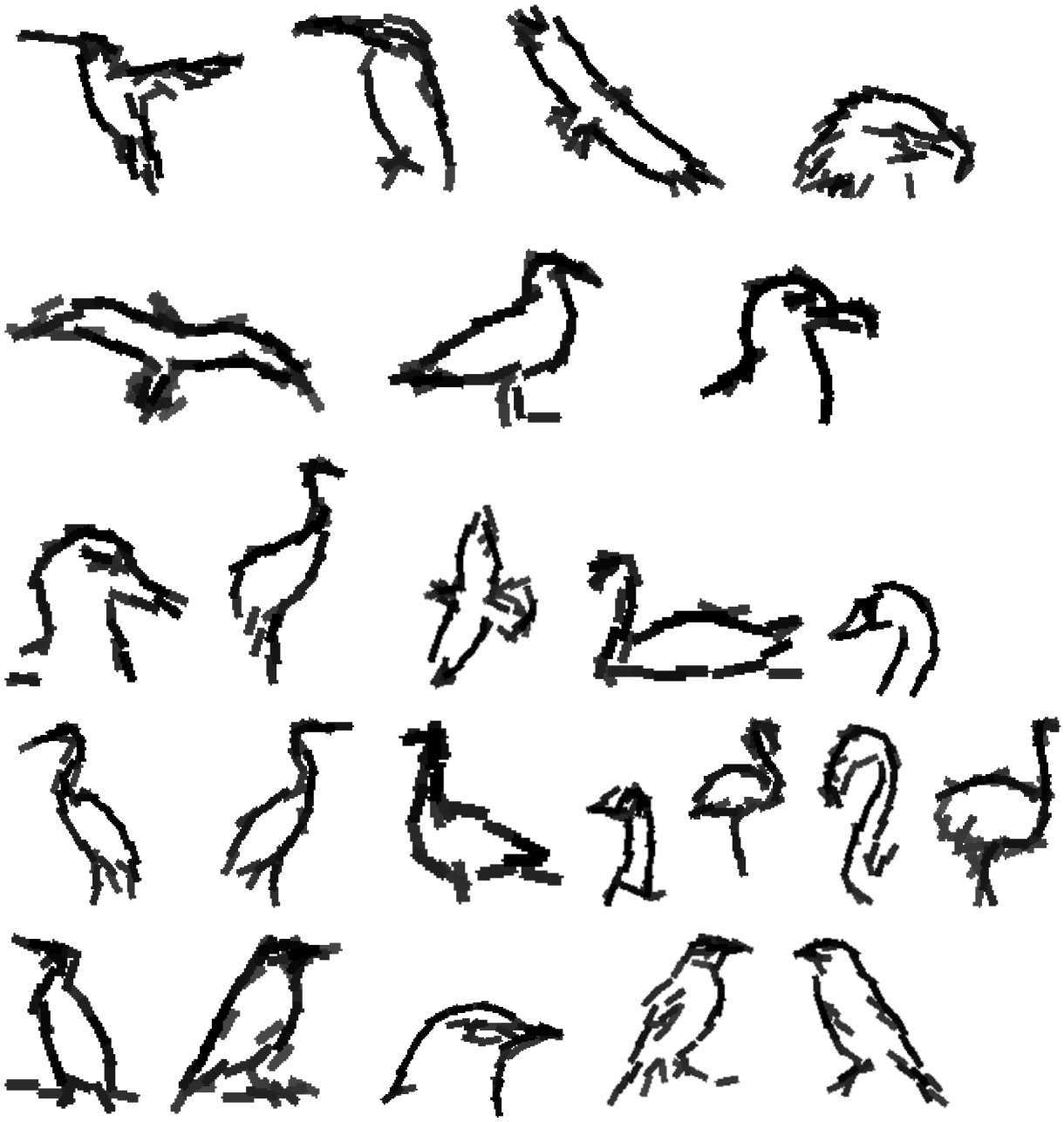
## Appendix: **Active basis templates learned from non-aligned training images, each basis element is illustrated by a bar of the same location, length and orientation.**

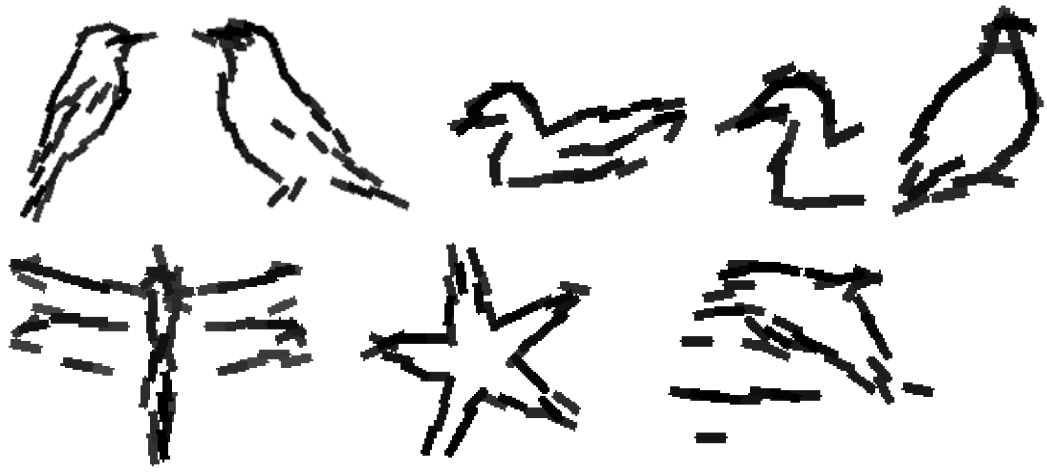


## Man-made objects



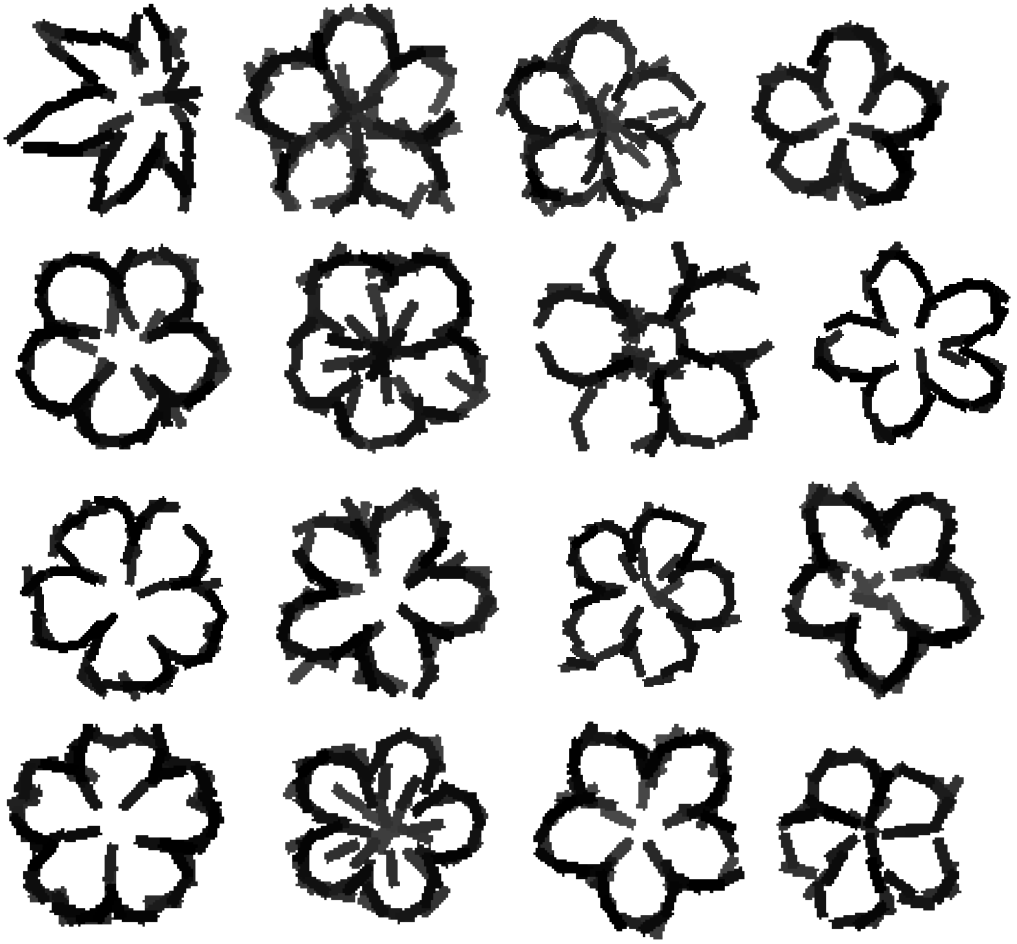
## Animal faces and bodies

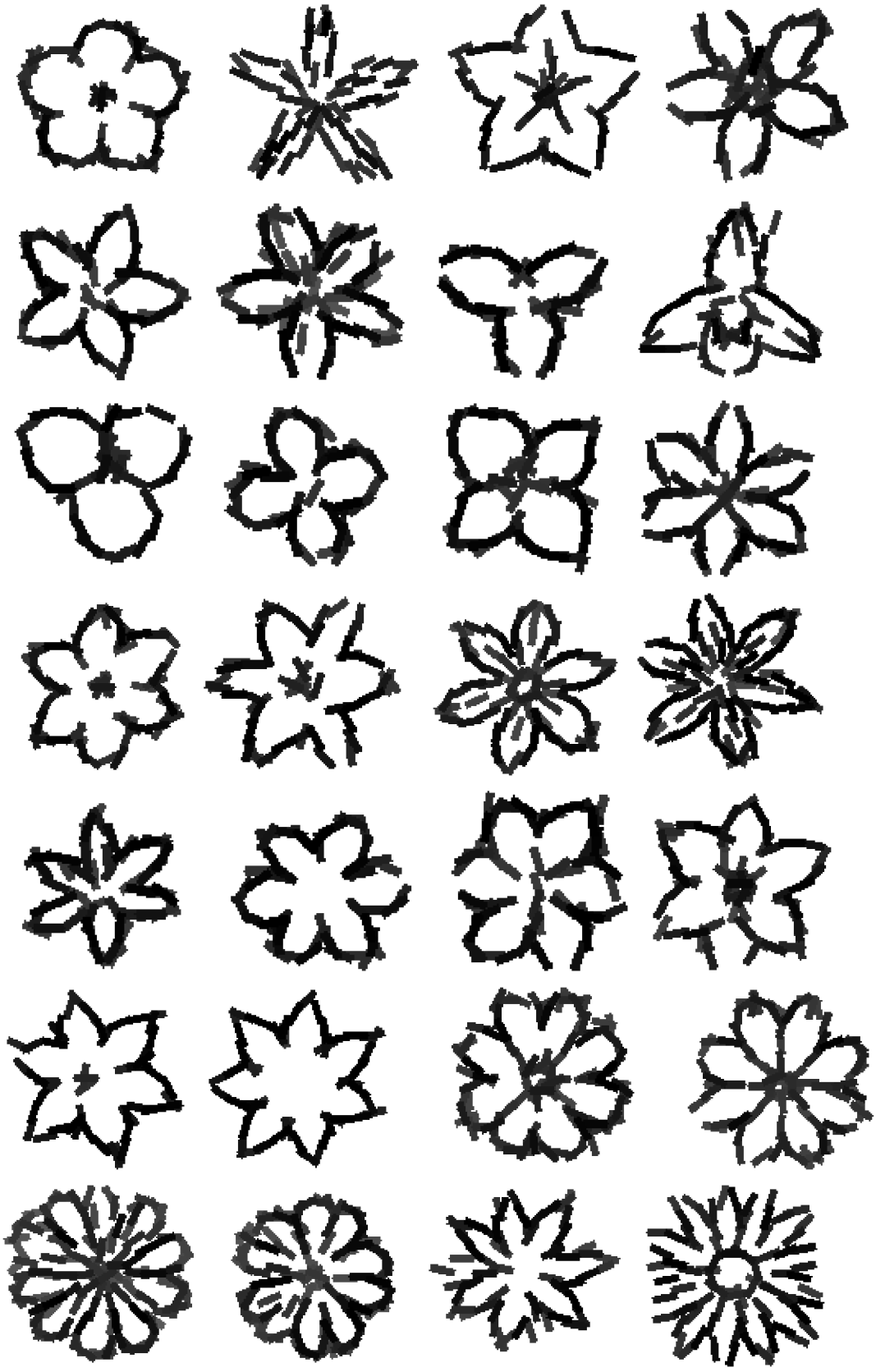


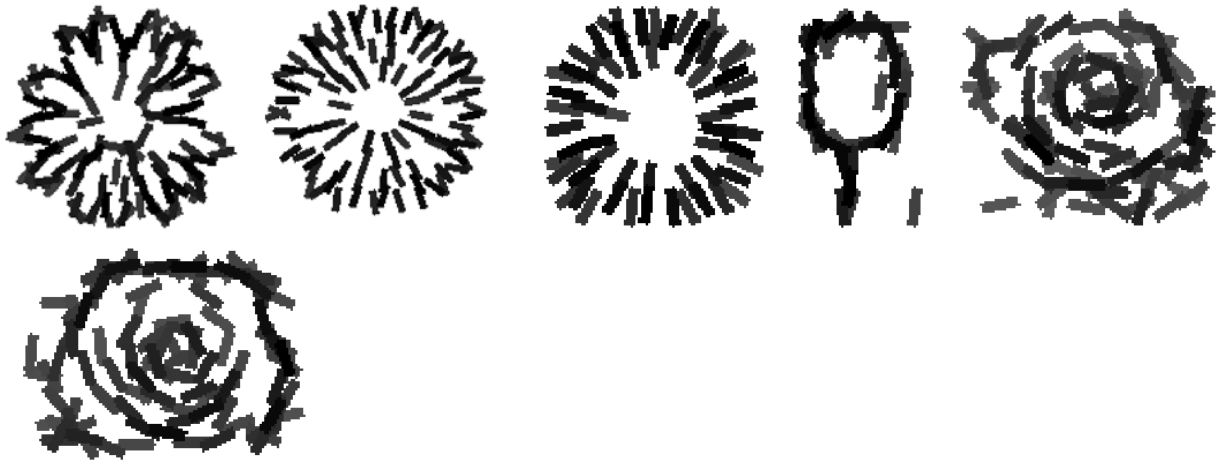


**Birds etc.**

---







# Flowers