

CBF: A New Framework for Object Categorization in Cortex

Maximilian Riesenhuber and Tomaso Poggio

Center for Biological and Computational Learning and Dept. of Brain and Cognitive
Sciences, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

{max, tp}@ai.mit.edu,

<http://cbcl.mit.edu>

Abstract. Building on our recent hierarchical model of object recognition in cortex, we show how this model can be extended in a straightforward fashion to perform basic-level object categorization. We demonstrate the capability of our scheme, called “Categorical Basis Functions” (CBF) with the example domain of cat/dog categorization, using stimuli generated with a novel 3D morphing system. We also contrast CBF to other schemes for object categorization in cortex, and present preliminary results from a physiology experiment that support CBF.

1 Introduction

Much attention in computational neuroscience has focussed on the neural mechanisms underlying object recognition in cortex. Many studies, experimental [2, 7, 10, 18] as well as theoretical [3, 11, 14], support an image-based model of object recognition, where recognition is based on 2D views of objects instead of the recovery of 3D volumes [1, 8]. On the other hand, class-level object recognition, *i.e.*, *categorization*, a central cognitive task that requires to generalize over different instances of one class while at the same time retaining the ability to discriminate between objects from different classes, has only just recently been presented as a serious challenge for image-based models [19].

In the past few years, computer vision algorithms for object detection and classification in complex images have been developed and tested successfully (*e.g.*, [9]). The approach, exploiting new learning algorithms, cannot be directly translated into a biologically plausible model, however.

In this paper, we describe how our view-based model of object recognition in cortex [11, 13, 14] can serve as a natural substrate to perform object categorization. We contrast it to another model of object categorization, the “Chorus of Prototypes” (COP) [3] and a related scheme [6], and show that our scheme, termed “Categorical Basis Functions” (CBF) offers a more natural framework to represent arbitrary object classes. The present paper develops in more detail some of the ideas presented in a recent technical report [15] (which also applies CBF to model some recent results in Categorical Perception).

2 Other Approaches, and Issues in Categorization

Edelman [3] has recently proposed a framework for object representation and classification, called “Chorus of Prototypes”. In this scheme, stimuli are projected into a representational space spanned by prototype units, each of which is associated with a class label. These prototypes span a so-called “shape space”. Categorization of a novel stimulus proceeds by assigning that stimulus the class label of the most similar prototype (determined using various metrics [3]).

While Chorus presents an interesting scheme to reduce the high dimensionality of pixel space to a veridical low-dimensional representation of the subspace occupied by the stimuli, it has severe limitations as a model of object class representation:

- COP cannot support object class hierarchies. While Edelman [3] shows how similar objects or objects with a common single prototype can be grouped together by Chorus, there is no way to provide a common label for a *group* of prototype units. On the other hand, if several prototypes carry the same class label, the ability to name objects on a subordinate level is lost. To illustrate, if there are several units tuned to dog prototypes such as “German Shepard”, “Doberman”, *etc.*, they would all have to be labelled “dog” to allow successful basic-level categorization, losing the ability to name these stimuli on a subordinate level.

This is not the case for a related scheme presented by Intrator and Edelman [6], where in principle labels on different levels of the hierarchy can be provided as additional *input* dimensions. However, this requires that the complete hierarchy is already known at the time of learning of the first class members, imposing a rigid and immutable class structure on the space of objects.

- Even more problematically, COP does not allow the use of different categorization schemes on the same set of stimuli, as there is only one representational space, in which two objects have a certain, fixed distance to each other. Thus, objects cannot be compared according to different criteria: While an apple can surely be very similar to a chili pepper in terms of color (cf. [15]), they are rather different in terms of sweetness, but in COP their similarity would have exactly one value. Note that in the unadulterated version of COP this would at least be a shape-based similarity. However, if, as in the scheme by Intrator & Edelman [6], category labels are added as input dimensions, it is unclear what the similarity actually refers to, as stimuli would be compared using all categorization schemes simultaneously.

These problems with categorization schemes such as COP and related models that define a global representational space spanned by prototypes associated with a fixed set of class labels has led us to investigate an alternative model for object categorization in cortex, in which input space representation and categorization tasks are decoupled, permitting in a natural way the definition of categorization hierarchies and the concurrent use of alternative categorization schemes on the same objects.

3 Categorical Basis Functions (CBF)

Figure 1 shows a sketch of our model of object categorization. The model is an extension of our model of object recognition in cortex (shown in the lower part of Fig. 1), that explained how view-tuned units (VTUs) can arise in a processing hierarchy from simple cell-like inputs. As discussed in [13, 14], it accounts well for the complex visual task of invariant object recognition in clutter and is consistent with several recent physiological experiments in inferotemporal cortex. In the model, feature specificity and invariance are gradually built up through different mechanisms. Key to achieve invariance and robustness to clutter is a MAX-like response function of some model neurons which selects the maximum activity over all the afferents, while feature specificity is increased by a template match operation. By virtue of combining these two operations, an image is represented through an (overcomplete) set of features which themselves carry no absolute position information but code the object through a combination of local feature arrangements. At the top level of the model of object recognition, view-tuned units (VTUs) respond to views of complex objects with invariance to scale and position changes (to perform view-invariant recognition, VTUs tuned to different views of the same object can be combined, as demonstrated in [11]).

VTU receptive fields can be learned in an unsupervised way to adequately cover the stimulus space, *e.g.*, through clustering [15], and can in the simplest version even just consist of all the input exemplars. These view-tuned, or *stimulus space-covering units* (SSCUs) (see caption to Fig. 1) serve as input to *categorical basis functions* (CBF) (either directly, as illustrated in Fig. 1, or indirectly by feeding into view-invariant units [11] which would then feed into the CBF — in the latter case the view-invariant units would be the SSCUs) which are trained in a supervised way to participate in categorization tasks on the stimuli represented in the SSCU layer. Note that there are no class labels associated with SSCUs. CBF units can receive input not only from SSCUs but also (or even exclusively) from other CBFs (as indicated in the figure), which allows to exploit prior category information during training (*cf.* below).

3.1 An Example: Cat/Dog Categorization

In this section we show how CBF can be used in a simple categorization task, namely discriminating between cats and dogs. To this end, we presented the system with 144 randomly selected morphed animal stimuli, as used in a very recent physiological experiment [4] (see Fig. 2).

The 144 stimuli were used to define the receptive fields (*i.e.*, the preferred stimuli) of 144 model SSCUs by appropriately setting the weights of each VTU to the C2 layer (results were similar if a k-means procedure was used to cluster the input space into 30 SSCUs [15]). The activity pattern over the 144 SSCUs was used as input to train an RBF categorization unit to respond 1 to cat stimuli and -1 to dog stimuli. After training, the generalization performance of the unit was tested by evaluating its performance on the same testing stimuli as used in a

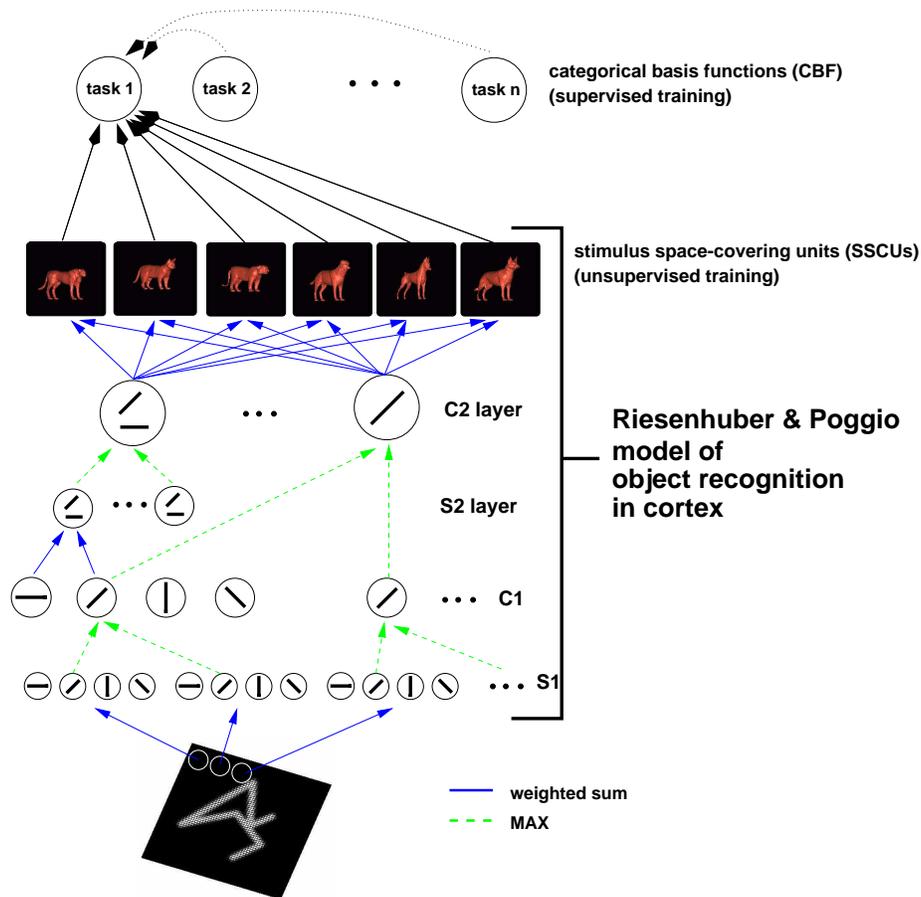


Fig. 1. Our model of object categorization. The model builds on our model of object recognition in cortex [14], which extends up to the layer of view-tuned units (VTUs). These VTUs can then serve as input to categorization units, the so-called categorical basis functions (CBF), which are trained in a supervised fashion on the relevant categorization tasks. In the proof-of-concept version of the model described in this paper the VTUs (as stimulus space-covering units, SSCUs) feed directly into the CBF, but input to CBFs could also come from view-invariant units, that in turn receive input from the VTUs [11], in which case the view-invariant units would be the SSCUs. A CBF can receive inputs input not just from SSCUs but also from other CBFs (illustrated by the dotted connections), permitting the use of prior category knowledge in new categorization tasks, *e.g.*, when learning additional levels of a class hierarchy.

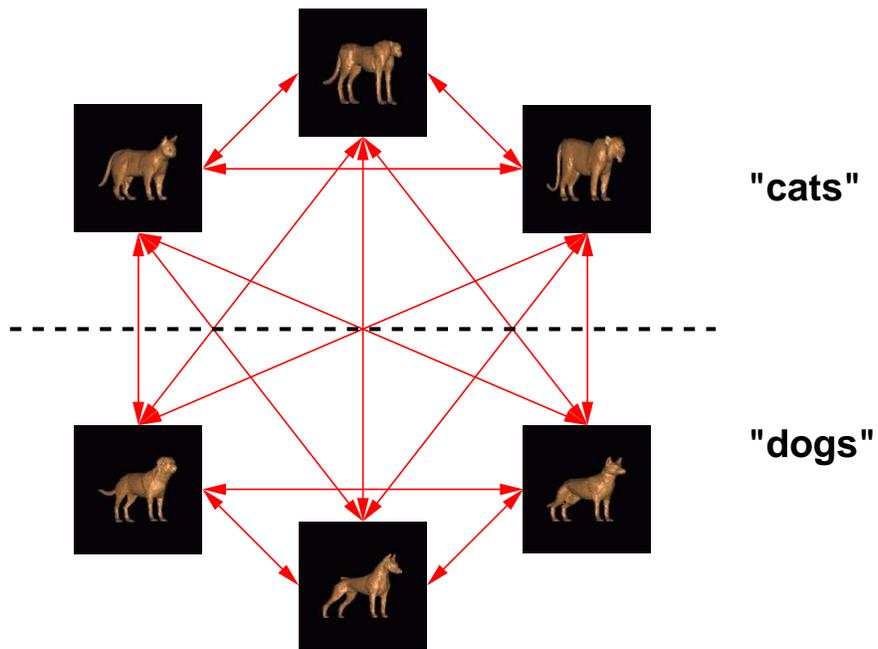


Fig. 2. Illustration of the cat/dog morph space [15]. The stimulus space is spanned by six prototypes, three "cats" and three "dogs". Using a 3D morphing system developed in our lab [17], we can generate objects that are arbitrary combinations of the prototype objects, for instance by moving along prototype-connecting lines in morph space, as shown in the figure and used in the test set. Stimuli were equalized for size and color, as shown (cf. [4]).

recent physiology experiment [4]: Test stimuli were created by subdividing the prototype-connecting lines in morph space into 10 intervals and generating the corresponding morphs (with the exceptions of morphs on the midpoint of each line, which would lie right on the class boundary in the case of lines connecting prototypes from different categories), yielding a total of 126 stimuli.

The response of the categorization unit to stimuli on the boundary-crossing morph lines is shown in Fig. 3. Performance (counting a categorization as correct if the sign of the categorization unit agreed with the stimulus' class label) on the training set was 100% correct, performance on the test set was 97%, comparable to monkey performance, which was over 90% [4]. Note that the categorization errors lie right at the class boundary, *i.e.*, occur for the stimuli whose categorization would be expected to be most difficult.

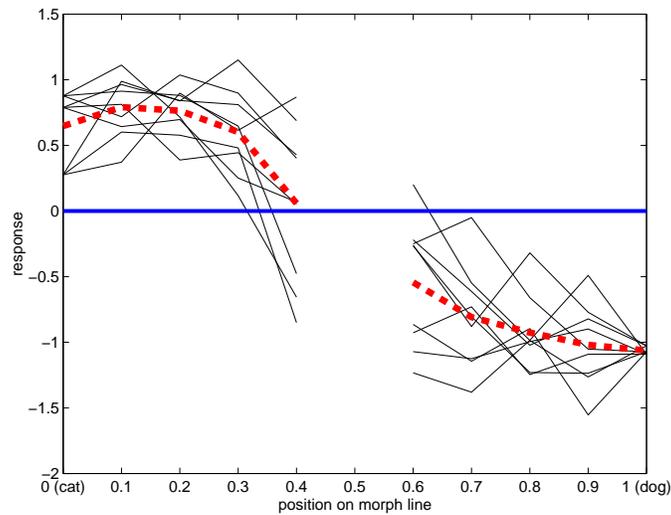


Fig. 3. Response of the categorization unit (based on 144 SSCU, 256 afferents to each SSCU, $\sigma_{SSCU} = 0.7$) along the nine class boundary-crossing morph lines. All stimuli in the left half of the plot are “cat” stimuli, all on the right-hand side are “dogs” (the class boundary is at 0.5). The network was trained to output 1 for a cat and -1 for a dog stimulus. The thick dashed line shows the average over all morph lines. The solid horizontal line shows the class boundary in response space. From [15].

3.2 Use of Multiple Categorization Schemes in Parallel

As pointed out earlier, the ability to allow more than one class label for a given object, *i.e.*, the ability to categorize stimuli according to different criteria, is a crucial requirement of any model of object categorization. Here we show how an alternative categorization scheme over the same cat/dog stimuli (*i.e.*, using

the same SSCUs) used in the previous section, can easily be implemented using CBF.

To this end, we regrouped the stimuli into three classes, each based on one cat and one dog prototype. We then trained three category units, one for each class, on the new categorization task, using 180 stimuli, as used in an ongoing physiology experiment (D. Freedman, M. Riesenhuber, T. Poggio, E. Miller, in progress). Each unit was trained to respond at a level of 1 to stimuli belonging to “its” class and -1 to all other stimuli. Each unit received input from the same 144 SSCU used before.

Performance on the training set was 100% correct (defining a correct labeling as one where the label of the most strongly activated CBF corresponded to the correct class). On the testing set, performance was 74%. The lower performance as compared to the cat/dog task reflects the increased difficulty of the three-way classification task. This was also borne out in an ongoing experiment, where a monkey has been trained on the same task using the same stimuli — psychophysics are currently ongoing that will enable us to compare monkey performance to the simulation results. In the monkey training it turned out to be necessary to emphasize the (complex) class boundaries, which presumably also gives the boundaries greater weight in the SSCU representation, forming a better basis to serve as inputs to the CBFs (cf. [15]).

3.3 Learning Hierarchies

The cat/dog categorization task presented in the previous section demonstrated how CBF can be used to perform basic level categorization, the level at which stimuli are categorized first and fastest [16]. However, stimuli such as cats and dogs can be categorized on several levels. On a superordinate level, cats and dogs can be classified as mammals, while the dog class, for instance, can be divided into dobermans, German shepherds and other breeds on a subordinate level.

Naturally, a model of object categorization should be able not just to represent class hierarchies but also to exploit category information from previously learned levels in the learning of new subdivisions or general categories. CBF provides a suitable framework for these tasks: Moving from, *e.g.*, the basic to a superordinate level, a “cat” and a “dog” unit, resp., can be used as inputs to a “mammal” unit (or a “pet” unit), greatly simplifying the overall learning task. Conversely, moving from the basic to the subordinate level (reflecting the order in which the levels are generically learned [16]), a “cat” unit can provide input to a “tiger” unit, limiting the learning task to a subregion of stimulus space.

4 Predictions and Confirmations of CBF

A simple prediction of CBF would be that when recording from a brain area involved in object categorization with cells tuned to the training stimuli we would expect to find object-tuned cells that respect the class boundary (the

CBF) as well as cells that do not (the SSCUs). In COP, or in the Intrator & Edelman scheme, however, the receptive fields of all object-tuned cells would be expected to obey the class boundary.

Preliminary results from Freedman *et al.* [4] support a CBF-like representation: They recorded from 136 cells in the prefrontal cortex of a monkey trained on the cat/dog categorization task. 55 of the 136 cells showed modulation of their firing rate with stimulus identity ($p < 0.01$), but only 39 of these 55 stimulus-selective cells turned out to be category-selective as well.

One objection to this line of thought could be that cells recorded from that are object-tuned but do not respect the class boundary might be cells upstream of the Chorus units. However, this would imply a scheme where stimulus spaced-out units feed into category units, which would be identical to CBF.

Regarding the representation of hierarchies, Gauthier *et al.* [5] have recently presented interesting results from an fMRI study in which human subjects were required to categorize objects on basic and subordinate levels. In that study, it was found that subordinate level classification activated additional brain regions compared to basic level classification. This result is compatible with the prediction of CBF alluded to above that categorization units performing subordinate level categorization can profit from prior category knowledge by receiving input from basic level CBFs. Note that these results are not compatible with the scheme proposed in [6] (nor COP), where a stimulus activates that same set of units, regardless of the categorization task.

5 Discussion

The simulations above have demonstrated that CBF units can generalize within their class and also permit fine distinction among similar objects. It will be interesting to see how this performance extends to even more naturalistic object classes, such as the photographic images of actual cats and dogs as used in studies of object categorization in infants [12]. Moreover, infants seem to be able to build basic level categories even without an explicit teaching signal [12], possibly exploiting a natural clustering of the two classes in feature space. We are currently exploring how category representations can be built in such a paradigm.

Acknowledgments

Fruitful discussions with Michael Tarr and email exchanges with Shimon Edelman are gratefully acknowledged. Thanks to Christian Shelton for MATLAB code for k-means and RBF training and for the development of the correspondence program [17].

Bibliography

- [1] Biederman, I. (1987). Recognition-by-components : A theory of human image understanding. *Psych. Rev.* **94**, 115–147.
- [2] Bülthoff, H. and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Nat. Acad. Sci. USA* **89**, 60–64.
- [3] Edelman, S. (1999). *Representation and Recognition in Vision*. MIT Press, Cambridge, MA.
- [4] Freedman, D., Riesenhuber, M., Shelton, C., Poggio, T., and Miller, E. (1999). Categorical representation of visual stimuli in the monkey prefrontal (PF) cortex. In *Soc. Neurosci. Abs.*, volume 29, 884.
- [5] Gauthier, I., Anderson, A., Tarr, M., Skudlarski, P., and Gore, J. (1997). Levels of categorization in visual recognition studied with functional mri. *Curr. Biol.* **7**, 645–651.
- [6] Intrator, N. and Edelman, S. (1997). Learning low-dimensional representations via the usage of multiple-class labels. *Network* **8**, 259–281.
- [7] Logothetis, N., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552–563.
- [8] Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. Freeman, San Francisco, CA.
- [9] Papageorgiou, C., Oren, M., and Poggio, T. (1998). A general framework for object detection. In *Proceedings of the International Conference on Computer Vision, Bombay, India*, 555–562. IEEE, Los Alamitos, CA.
- [10] Perrett, D., Oram, M., Harries, M., Bevan, R., Hietanen, J., Benson, P., and Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Exp. Brain Res.* **86**, 159–173.
- [11] Poggio, T. and Edelman, S. (1990). A network that learns to recognize 3D objects. *Nature* **343**, 263–266.
- [12] Quinn, P., Eimas, P., and Rosenkrantz, S. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception* **22**, 463–475.
- [13] Riesenhuber, M. and Poggio, T. (1999). Are cortical models really bound by the “Binding Problem”? *Neuron* **24**, 87–93.
- [14] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2**, 1019–1025.
- [15] Riesenhuber, M. and Poggio, T. (1999). A note on object class representation and categorical perception. Technical Report AI Memo 1679, CBCL Paper 183, MIT AI Lab and CBCL, Cambridge, MA.
- [16] Rosch, E. (1973). Natural categories. *Cognit. Psych.* **4**, 328–350.
- [17] Shelton, C. (1996). Three-dimensional correspondence. Master’s thesis, MIT, (1996).

- [18] Tarr, M. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonom. Bull. & Rev.* **2**, 55–82.
- [19] Tarr, M. and Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition* **67**, 73–110.