A Neural Network Model of Invariant Object Identification

Hedwig Wilhelm

1 Motivation and Overview

Invariant object recognition is maybe the most basic and fundamental property of our visual system. It is the basis of many other cognitive tasks, like motor actions and social interactions. Hence, the theoretical understanding and modeling of invariant object recognition is one of the central problems in computational neuroscience. Indeed, object recognition consists of two different tasks: classification and identification. The focus of this thesis is on object identification under the basic geometrical transformations shift, scaling, and rotation. The visual system can perform shift, size, and rotation invariant object identification.

This thesis consists of two parts. In the first part, we present and investigate the VisNet model proposed by Rolls [WR97]. The generalization problems of VisNet trigger our development of a new neural network model for invariant object identification. Starting point for an improved generalization behavior is the search for an operation that extracts images features that are invariant under shifts, rotations, and scalings. Extracting invariant features guarantees that an object seen once in a specific pose can be identified in any pose. We present and investigate our model in the second part of this thesis.

2 Results

2.1 VisNet - The Model of E. Rolls

In the first part of this thesis, we present and investigate the VisNet model proposed by Rolls [WR97]. It is a four layer, feedforward network with lateral inhibition in each layer, a trace learning rule, and an image preprocessing by Difference of Gaussians filters. For the basic geometric transformations Rolls et al. showed that VisNet identify its training exemplars correctly. Additionally, Rolls et al. showed that VisNet generalizes from exemplars trained in all four layers to exemplars trained only in the two lower layers. Since VisNet operates unsupervised, identifying the training exemplars is not trivial for the network. However, these results of Rolls et al. do not answer the question whether VisNet can perform transformation invariant identification.

The experiments, that we conducted, revealed the following properties of VisNet:

1. Under specific circumstances, VisNet shows translation invariant identification. Namely, when trained with one dimensional bars in a sufficient number of different positions VisNet can identify them in any position. The number of training exemplars needed for a perfect translation invariant identification depends on the width of the bars. We derived a formal criterium specifying the minimal number of training exemplars. The necessary number of training exemplars depends marginally on the frequencies used in the Difference of Gaussians filters.

2. The categories build by VisNet out of the training exemplars do not depend on the order of the exemplars during training. Indeed, the external sorting of the training exemplars done to train the support vector machine classifier defines the categories VisNet establishes. This indicates that the support vector machine classifier dominates the association the trace rule is supposed to build between training exemplars. We observe this effect both for simple geometric stimuli and for complex, natural images.

3. The performance of VisNet in a simple translation invariant identification task with one dimensional bars cannot be improved by a regional filter scheme.

4. The performance of VisNet in a simple translation invariant identification task with one dimensional bars depends marginally on the trace rate. The sole exception is $\eta = 1$. In this case the network performance is clearly worse. Especially, the pure Hebbian learning rule, i. e. $\eta = 0$, is as good as the trace rule. Hence, for the translation invariant recognition of one dimensional bars a trace learning rule is not needed. In this case a pure Hebbian rule is sufficient.

5. VisNet is in principle able to perform scale invariant recognition. Namely, when trained with one dimensional bars in a certain width, VisNet can identify one dimensional bars in several, different widths.

In summary, we show that VisNet has good learning abilities. It is very flexible and reliable in building categories. Thereby, a main ingredient is the power of the support vector machine. However, VisNet's generalization behavior displays major problems.

2.2 The Fourier Model

In the second and main part of this thesis, we develop a Fourier filter that processes an image in the three steps:

1. a logarithmic-polar Fourier transform with an adaptive bandpass,

2. a classical Fourier transform,

3. a normalization.

The filter output is identified by a minimal distance classifier with one randomly chosen reference exemplar per object.

We prove that this filter is invariant under shifts and two dimensional rotations within the image plane and approximately invariant under scalings. In addition, we implement the Fourier filter in a neurophysiologically plausible neural network. Due to the limited computational power of neurons, they presumably perform only simple mathematical. Hence, several problems occur when implementing a filter in a neurophysiologically plausible neural network. We present a way to circumvent all problems and develop a neurophysiologically plausible neural network that implements our Fourier filter.

In order to judge the quality of the performance of our network given the results of the minimal distance classifier, we defined two suitable performance measures: the within-category distance and the between-category distance.

Our neural network model shows translation, rotation, and scale invariant identification both for simple geometric objects and for complex, natural images. This result is particulary interesting since the mathematical properties of our model prove exact shift and two dimensional rotation invariance but only an approximate scale invariance. However, experiments demonstrate that the approximation of the scale invariance is sufficiently good to identify even complex objects in different sizes. Experiments reveal that both for simple geometric stimuli and complex, natural images the within-category distance is low. This indicates that the exact choice of the reference exemplar, used in the classifier, has nearly no effect. This is important since the reference exemplar is usually chosen at random. Also, in the human visual system an object identity is not built on the basis of specific object view. Additionally, we show that our model can handle a limited amount of noise in the image. It can identify simple geometric object in the presence of random noise as long as the noise density is not too high. This is surprising. Our model has no built-in mechanism to deal with noise and or to eliminate it from the image. Dealing with noise is a crucial property of the human visual system. The ability to identify images in presence of noise makes our model further realistic as model of the visual system.

References

[WR97] G. Wallis and E. T. Rolls. Invariant Face and Object Recognition in the Visual System. *Progress in Neurobiology*, 51(2):167–194, 1997.