

Learning to Generalize from Single Examples in the Dynamic Link Architecture *

Wolfgang Konen

Christoph von der Malsburg

Institut für Neuroinformatik, Ruhr-Universität Bochum, FRG

Submitted to: Neural Computation

February 25, 1993

Abstract

A large attraction of neural systems lies in their promise of replacing programming by learning. A problem with many current neural models is that with realistically large input patterns learning time explodes. This is a problem inherent in a notion of learning that is based almost entirely on statistical estimation. We propose here a different learning style where significant relations in the input pattern are recognized and expressed by the unsupervised self-organization of dynamic links. The power of this mechanism is due to the very general *a priori* principle of conservation of topological structure. We demonstrate that style with a system that learns to classify mirror symmetric pixel patterns from single examples.

Introduction

Learning is the ability of animals and humans to absorb structure from one scene and apply it to others. The literal storage of whole sensory input fields is of little value since scenes never recur in all detail within our lifetime. Essential for learning is therefore the ability to extract significant patterns from an input field containing mostly patterns with accidental feature constellations, and to apply those significant patterns to the interpretation of later scenes.

How can significant patterns be identified? Theories of learning based on layered neural networks (e.g. back-propagation of errors [Rosenblatt 1962, Rumelhart et al. 1986] or the Boltzmann Machine [Ackley et al. 1985]) are based on the notion that significant patterns are, above all, recurring patterns. Such systems have an input layer, an output layer and hidden units. During a learning phase, many examples are presented to input layer and output layer, and the system is enabled by some plasticity mechanism to pick up and represent patterns that recur with statistical significance in the input training set. This method of identifying significant patterns may be the obvious one — going back to the original definition of significance based on recurrence — but with realistic inputs taken

*Supported by a grant from the Bundesministerium für Forschung und Technologie (413-5839-01 IN 101 B/9), and a research grant from the Human Frontier Science Program.

from natural environments it is far too costly, costly in terms of the number of inputs required to discriminate significant patterns from accidental patterns. The reason for this difficulty lies in the combinatorial explosion in the number of subsets that can be selected from large input fields (there are, for instance, 10^{3432} possible subsets of size 1000 in a set of 10^6). Among those subsets there are only relatively few of significant interest (in vision, for example, the criterion of spatial continuity alone singles out relatively very few subsets). There obviously are potent methods, presumably based on *a priori* knowledge built into the system, to extract significant patterns from a scene.

It is generally recognized that methods based purely on scene statistics must be complemented (if not supplanted) by more powerful ones based on *a priori* structure. One widespread piece of advice is to use input representations that are already adapted to the problem at hand. Down that alley there is, of course, the pitfall of hand-wiring instead of learning the essential structure. The real challenge is to find simple and general architectures that can handle large classes of problems and that can learn with a minimum of scene statistics.

The particular problem we are considering here has originally been proposed by Sejnowski, Kienker and Hinton [Sejnowski et al. 1986]. It consists in learning to classify mirror-symmetrical pixel patterns (see Fig. 1). The authors solved the problem with the help of scene statistics. Their system, consisting of a layer of 12 hidden units and 3 output units corresponding to the 3 symmetry classes, learned as a Boltzmann Machine, which is a variant of supervised learning. With input arrays of 10×10 pixels the system needed about 40,000 training examples in order to reach a success level of 85%.

The system [Sejnowski et al. 1986] demonstrates the strength and the weakness of statistical pattern classification. The strength is full generality with respect to possible patterns. This is best demonstrated with the thought experiment of applying a permutation to the pixels in the input field — the same permutation to all patterns presented. The system would now have the same learning performance, in spite of the complete destruction of topological structure. The weakness is an explosion in the number of examples required when scaling to larger input arrays. This weakness the system shares with a wide class of learning algorithms which all are based on the statistical detection of classes as clusters in input space and their subsequent representation by single prototypes. Prominent examples are the k-nearest neighbor (kNN) algorithm [Fix and Hodges 1951, Cover and Hart 1967], the RCE-algorithm [Reilly et al. 1982], which is a neural version of kNN, and adaptive vector quantization (LVQ, LVQ2) [Kohonen et al. 1988]. None of those algorithms can easily deal with the symmetry classification problem. The reason is that already with modest problem size there are astronomically many patterns in a symmetry class (10^{32} for the 8×8 pixels of 10 features each in Fig. 1) and that these don't at all form clusters in input space and thus cannot be detected in a small training set. It is this which leads to the exploding thirst in learning time and number of prototypes.

Our treatment of the problem is based on the Dynamic Link Architecture (DLA) [von der Malsburg 1981]. The strength of the DLA essential in the present context is its ability to detect pattern correspondences. An application of this ability to the problem of invariant pattern recognition has been reported previously [Bienenstock and von der Malsburg 1987, von der Malsburg 1988, Lades et al. 1993]. Here we demonstrate that with it symmetry classes can be recorded from single examples for later recognition. Our

treatment is based on the *a priori* restriction that significant relations within the input pattern are those which preserve topological structure. It is in this sense less general than the Boltzmann machine, not being able to deal with the permutation symmetries mentioned above. On the other hand its extreme speed of adaptation to new symmetries makes it more potent than the Boltzmann machine. Most of what is achieved in other neural systems with the help of statistical learning is performed here by the self-organization of an explicit representation of the symmetry in the input pattern.

Symmetry Detection by Dynamic Link Matching — Qualitative Model

Dynamic link matching is capable of finding and representing topological, feature-preserving mappings between parts of the input plane. Such mappings are systems of pair-wise links that are neighborhood-preserving and that connect pairs of points with the same local properties in the input pattern. In this section we describe the network and its function qualitatively and establish its relationships to other, previously published models [von der Malsburg 1988, Bienenstock and von der Malsburg 1987, Buhmann et al. 1989, Lades et al. 1993, von der Malsburg and Buhmann 1992], and to the circuitry of cortical structures. In the next section we will describe an explicit quantitative, though somewhat simplified, model.

The network resembles primary visual cortex in representing visual images in a columnar fashion: Each resolution unit (“pixel”) of the sensory surface is subserved by a collection (“column”) of neurons, each neuron reacting to a different local feature. (In our concrete model, local features will simply be gray values. In a more realistic version, features would refer to texture, color and the like.) There are intra-columnar connections, whose function will be explained below, and inter-columnar connections. The latter are what we will refer to as “dynamic links”, are of rather large range in visual space and are restricted to pairs of neurons with the same feature type. (In our explicit model the connections will run between cells responding to the same gray value in the image.) When a pattern is presented as visual input, those neurons in a column are selected which code for a feature that is present in the corresponding pixel. We refer to the selected cells as “pre-activated” neurons. The set of all pre-activated neurons represents the input image.

During the presentation of an image, the pre-activated cells are actually not allowed to fire all at the same time. Rather, activity in the network takes the form of a sequence of “blob activations.” During a blob activation, only those pre-activated neurons are permitted to fire which lie in a group of neighboring columns. A blob activation corresponds to the “flash of the searchlight of focal attention” discussed, for instance, by Crick [Crick 1984]. In the absence of any other control of attention, blob activations are created spontaneously in random positions in a rapid sequence of “cycles.”

When a blob is active, its active cells send out signals that excite pre-activated neurons of the same feature type in other locations. Thus, within the total network those pre-activated neurons are excited whose type is represented in the active blob. Most of these cells form a diffuse spray over the image domain. If there is a symmetry in the image, however, there will be a location where all the feature types in the active blob are assembled locally again. With appropriate dynamics, those neurons are activated as

well, forming a “shadow blob.” The network thus has discovered the significant relationship between two symmetrical regions in the image, and with the help of rapid synaptic plasticity in the inter-columnar connections (“dynamic links”) it is possible to record it, simply strengthening the synaptic connections between all pairs of neurons lying one in each blob. During a sequence of many blob pairs, a full consistent system of point-to-point connections will get established, forming a topological mapping between the symmetric parts of the image.

This sequence of events constitutes the dynamic link mapping mechanism. It is very robust. Occasional erroneous blob pairs are of little consequence whereas all correct blob pairs form a cooperative system of mutual reinforcement. Once the covering of the image with blobs is fairly complete the plexus of reinforced connections stabilizes signal correlations between symmetric points and, as our simulations show, false blob pairs do no longer occur. For each new image (or for each new fixation of an image, for that matter), a new mapping of dynamic links has to be built up.

A slow, and simpler, version of the dynamic link mapping mechanism was first described in [Willshaw and von der Malsburg 1976] to account for the ontogenetic establishment of retinotopic mappings from retina to tectum. A dynamic link mapping system using feature labels has later been proposed as a solution to the problem of invariant object recognition [von der Malsburg 1988, Bienenstock and von der Malsburg 1987, Buhmann et al. 1989, Lades et al. 1993]. As a mapping system, the present model goes beyond previous work in needing dramatically fewer activation cycles. The columnar connectivity pattern described here was introduced as part of a proposed solution to the figure ground segmentation problem [Schneider 1986, von der Malsburg and Buhmann 1992].

In the explicit model described below some network details are just necessary to realize the qualitative behavior described above. Others, however, we introduced to simplify the dynamics of our system. Prominent among these is the introduction of an “activator cell” (or *X*-cell) and a “collector cell” (or *Y*-cell) for each column, see Fig. 2A. The activator cells spontaneously create the active blob and activate all sensorily pre-activated neurons in their column. The collector cells sum up all activity that arrives in the pre-activated neurons of their column and that comes from the active blob, and they interact to form the shadow blob. Also, active collector cells gate the pre-activated neurons in their columns into full activity. The presence of activator cells and collector cells makes sure that all pre-activated neurons in a column make their firing decision together. Global inhibition between all activator cells and between all collector cells ensures that there is exactly one active blob and exactly one shadow blob at any one time. An activator cell is kept by a compensating inhibitory connection from exciting the collector cell of its own column via its feature cells.

In our explicitly simulated network described below we make the simplifying assumption that during the presentation of an image exactly one of the feature cells in a column is active (corresponding to one of a number of possible gray values). As a consequence, at most one inter-columnar connection is active between two columns at any one time (exactly when the two columns are pre-activated with the same gray value). This justifies our introduction of “compound connections” from the activator cells to the collector cells, treating all columnar quality cells and their connections implicitly, see Fig. 2B.

Symmetry Detection by Dynamic Link Matching — Explicit Model

After these preparatory heuristic discussions we are ready to introduce the explicit dynamic link mapping network that we have simulated. It has the following parts, cf. Fig. 2B. Our image domain is formed by a grid of 8×8 pixels. Positions in the image domain are designated by letters a, b, \dots . An input image is described by a distribution of features F_a over the pixel positions a , where $F_a \in \{1, \dots, F_{max}\}$, see Fig. 1. The image domain is covered by two layers of cells, the X -layer and the Y -layer. The connection from cell a in the X -layer to cell b in the Y -layer is controlled by the dynamic link variable J_{ba} , which is subject to the dynamic described below. The constraint of feature specificity is formulated with the help of the similarity constraint matrix

$$T_{ba} = \begin{cases} 1 & \text{if } F_a = F_b \text{ and } b \neq a \\ 0 & \text{else.} \end{cases} \quad (1)$$

The total connection from cell a in the X -layer to cell y in the Y -layer is described by the “effective coupling” $J_{ba}T_{ba}$.

The activities of cells are designated x_a or y_a . Both layers have homogeneous internal connections of the form

$$K_{aa'} = G_{aa'} - \beta. \quad (2)$$

Here, $G_{aa'}$ is a short-range excitatory connection kernel, and β is the strength of a long-range (here: global) inhibitory connection. For both X and Y we assume wrap-around boundary conditions.

The dynamic of the X -layer is governed by the differential equations

$$\dot{x}_a = -\alpha x_a + \sum_{a'} K_{aa'} S(x_{a'}) + \rho. \quad (3)$$

Here, $S(x)$ is a sigmoidal non-linearity which saturates at $S(x) = 0$ for low x and at $S(x) = 1$ for high x , whereas ρ is a constant excitation.

The dynamic of the Y -layer is governed by the differential equations

$$\dot{y}_b = -\alpha y_b + \sum_{b'} K_{bb'} S(y_{b'}) + e \sum_a J_{ba} T_{ba} S(x_a). \quad (4)$$

With given effective connections and small noisy initial values (as a model for spontaneous activity) for the x_a , the activator and collector cell activities evolve on a fast time scale towards an equilibrium distribution in the form of local blobs of active cells ($S \approx 1$), with the rest of the cells in the layer X or Y inactive, ($S \approx 0$). The size of the blob is controlled by the parameters α , β and σ , whereas their position is determined by the noise input in the case of X and by the profile of the activation in the case of Y .

Once the activity in X and Y has settled, the dynamic link variables J_{ba} are modified by three consecutive substitutions:

$$\begin{aligned} J_{ba} &\rightarrow J_{ba} + \epsilon J_{ba} T_{ba} S(y_b) S(x_a) \\ &\rightarrow J_{ba} / \sum_{a'} J_{ba'}, \\ &\rightarrow J_{ba} / \sum_{b'} J_{b'a}. \end{aligned} \quad (5)$$

The first step encapsulates the general idea of Hebbian plasticity, though regulated here by the constant ϵ for the rapid time scale of a single image presentation. After the second and third steps the new connections conform to divergent and convergent sum rules.

When an image is presented, the full sequence of events is the following. First, the connections J_{ba} are initialized with a constant value conforming to the sum rules. Then a number of activity-and-modification cycles are carried through. For each of these, the X -activities are initialized with noise distribution, the Y -activities are reset to 0 and the dynamics of X and Y are run according to (3) and (4) to reach stationary values. Then the dynamic links are updated according to (5). After typically 50-80 such cycles the dynamic links J relax into a stable configuration displaying the underlying symmetry of the actual input image. For a typical result see Fig. 4. The network is now ready for permanently recording the symmetry type if it is new, or for recognizing it according to a previously recorded type.

If a link J_{ba} is active, the activity dynamics of (3)–(4) produces correlated activity in the connected cells: In the stationary state towards the end of each cycle, cells a and b are always active or inactive together. In comparison to the dynamic links, activity correlations have the distinction of graceful degradation: Even if a single link is corrupted, the correlation between the corresponding x and y cells is high if there are strong links in the neighborhood (remember that an activity blob always covers a neighborhood along with a given cell).

Recording and Recognizing a Symmetry

The main task necessary for solving the symmetry recognition problem is solved for our model by the unsupervised process of dynamic link mapping described in the last section. For a given symmetric pattern it constructs a temporary representation in the form of a set of active links. This set is the same for all input patterns belonging to the same symmetry class. In order to record a symmetry type it is now simply necessary to create hidden units as permanent representatives for some of the active links (or rather the correlations created by them) and to connect them to appropriate output units. Once a symmetry type has been represented by such a network, its second occurrence can be detected and the system is ready to recognize all patterns of this symmetry type as such.

Our recognition network structure is similar to the one used in [Sejnowski et al. 1986] and is shown in the upper panel of Fig. 3. It consists of three output units C_k , $k = 1, 2, 3$ (sufficient for three symmetry types) and, connected to each output unit, 6 hidden units. Each hidden unit i has a randomly chosen fixed reference cell $a(i)$ in X and plastic synapses w_{ib} from all cells b in Y ¹. The output h_i of a hidden unit is driven by a coincidence of activity $x_{a(i)}$ of its reference cell in X and activity within its receptive field w_{ib} in Y :

$$h_i = S(x_{a(i)}) \sum_b w_{ib} S(y_b). \quad (6)$$

¹In principle, the number of hidden units per output cell could be one. Recognition is more reliable and faster, however, if the density of reference cells $a(i)$ is large enough so that most of the active blobs in X hit at least one of them.

In *recording mode*, hidden units modify their Y -connections w_{ib} at the end of each activity cycle according to the Hebbian plasticity rule:

$$\Delta w_{ib} = \eta S(y_b) \quad \text{if } h_i > \theta \quad \text{and} \quad C_k > 0. \quad (7)$$

Synaptic plasticity is supervised in the sense that only those hidden units modify their connections whose output unit C_k is currently activated by a teacher (the role of the teacher simply being to fixate attention on one group of hidden units during the presentation of one pattern). In this way, a hidden unit whose X connection is hit by a blob learns to associate with it the corresponding blob in the Y plane. The whole process is completed for a symmetry type during one presentation (or in some cases two presentations, see below).

In *recognition mode*, the output units perform a leaky integration of the sum of the activities (6) of their group of hidden units. After a number of cycles, the output unit with maximal signal is taken to indicate the class into which the input pattern falls.

Simulation Results

Simulations of the model were carried out for input patterns of size 8×8 . The parameters for the blob formation process in Eqs. (3) and (4) were adjusted to let the equilibrium blobs cover between 25% and 40% of the layer area; for example, with $\{\alpha, \beta, e, \eta, \epsilon, \rho, \theta\} = \{.3, .85, 1.8, .02, .8, .6, .125\}$ blobs cover 25% of their layer. As convolution kernel $G_{aa'}$ we used a Gaussian of width 4 and strength 2.1, restricted, however, to a window of 5×5 pixels. For almost all input patterns, self-organization of the correct mapping J from X to Y was observed. Fig. 4 shows a typical example in some stages of the organization process. The degree of organization can be measured quantitatively by the correlation between corresponding cells, which is shown in Fig. 5 for a specific input example. During the first 40–50 activation cycles the correlation builds up and reaches almost the theoretical optimum 1. Thus, during all further cycles symmetrically corresponding points in X and Y are marked by strong correlations, whereas pairs of units in far-from-symmetrical positions would have correlation -1.

After learning the specific symmetries from either one or two training examples, the network can generalize almost perfectly to new input patterns of the same symmetry class. Fig. 6A shows the classification performance on 200 new examples. There is a clear trade-off between the reliability of recognition and the required time (in terms of activation cycles). In principle, one example can supply sufficient information for this performance. However, with our parameter settings two examples gave slightly more reliable results, see Fig. 6A.

Our network achieves a recognition reliability of 98%. Its level of reliability is only weakly affected by perturbations to the feature similarity matrix T up to $t = 40\%$ (Fig. 6B). This is due to the robustness of the dynamic link mapping mechanism, see Fig. 5, which creates near-to-perfect signal correlations between symmetric points. Since the hidden units are trained by these correlations, the presence of perturbations in the matrix T even during learning does not affect the performance of the system. We have verified numerically that after training the hidden units with $t = 40\%$ the performance is

virtually the same as in Fig. 6B, e. g. 93% reliability if the recognition is forced after 100 cycles.

Discussion

We have presented here a network which is able to discover a system of relations in individual input patterns and to immediately generalize to further examples of the same type. The network is based on dynamic link mapping. The self-organization of dynamic links in our model is extremely fast, requiring much less than 100 update cycles. This is due to the use of local feature information in conjunction with a topology constraint. For simplicity, we have used gray-values of single pixels as our visual features. In applications to large pixel arrays this would be impractical. The number of dynamic links in the matrix J would have to grow with the fourth power of the linear extent of the input plane. However, if one replaced the gray-level sensitivity of our feature cells by extended receptive fields (e.g., of the Laplace type with a hierarchy of spatial scales, in analogy to the feature jets of [Buhmann et al. 1991]) one could cover the input plane with a fairly low-density set of sampling points and correspondingly operate with manageably small X and Y planes.

The main goal of our paper is to make a point about the learning issue, symmetry detection merely playing the role of an example. It may be interesting, though, to briefly discuss symmetry detection in humans and animals. Infants can detect symmetry at the age of four months [Bornstein et al. 1981]. Pigeons learn to discriminate symmetry in very few trials [Delius and Nowak 1982], although one may suspect that they already come equipped with the ability of detecting symmetry and only have to be conditioned for the appropriate response. Our system may shed new light on the old discussion of nature vs. nurture with respect to the symmetry detection issue: Our demonstration that learning time could be extremely short makes it impossible to decide the issue by demonstrating the capability in naive or new-born subjects.

At first sight it is tempting to take our system directly as a model for symmetry detection in primary visual cortex, identifying all of our cell types (X and Y cells, feature cells and hidden units) with neurons found in cortical hypercolumns in area V1. This view would run, however, into a number of difficulties. One of them is the need, in our model, for long range connections (inter-columnar links and the w_{ib} connections from Y cells to hidden units). With respect to area V1 this requirement creates a conflict, especially in view of the fact that humans are better at detecting symmetry around the vertical midline than around the horizontal [Barlow and Reeves 1979], and callosal connections are absent within V1 except for a narrow seam along the vertical meridian. (This difficulty is mitigated, though, by the fact that symmetry detection in humans relies mainly on a small strip along the symmetry axis of the input pattern, at least in random dot patterns [Julesz 1975].) The problem can be largely avoided by placing our model in a later area in which larger visual angles are spanned by horizontal fibers. Another hint to this effect is the observation that symmetry detection in humans may be based not on distributions of gray levels directly but rather on object shapes reconstructed from shading [Ramachandran 1988]. Another difficulty for a direct biological application of our model (which it shares with the one of [Sejnowski et al. 1986]) is its lack of invariance with

respect to displacement of the symmetry axis, as for instance caused by eye movements during inspection of a pattern. All of these difficulties point to a slightly more complicated model, which would, however, obscure our main point.

Our system is not limited to mirror symmetry. It could equally record and recognize other pattern relations such as simple duplication with or without rotation (or, in a system of only slightly more general form, expansion). Humans, on the other hand, perform much worse on these [Corballis and Roldan 1974]. The reason for such differences may be a rather superficial one. Even if the ontogeny of symmetry detection is of the nature we are putting forward here, the system will after some experience be dominated by the hidden units it has acquired. Once these have sufficient density, the dynamic link mechanism is no longer necessary for the recognition of familiar pattern relations (the same way the correct hidden units in [Sejnowski et al. 1986] are activated directly by the known symmetries). The relative performance on different problem types is then dominated by experience rather than by the nature of the ontogenetic mechanism. This would explain our bias towards vertical mirror symmetry. The heavy reliance of humans on a strip around the symmetry axis mentioned above may point to a mechanism relying on memorized symmetric shapes, such as butterfly patterns and the like, formed on the basis of a general learning mechanism but soon supplanting it by being more rapidly detected.

The structure of our model fits very well the general character of cortical columnar organization (as also employed in [von der Malsburg and Buhmann 1992]). Of central importance to our system is the encoding of significant relations with the help of temporal signal correlations. Candidate correlations of an appropriate nature have been observed in visual cortex [Gray et al. 1989, Eckhorn et al. 1988]. The model may thus give support to the idea that correlations play an important functional role in cortical information processing.

The central point that we would like to make here refers to the general learning issue. The origin of knowledge in our mind has puzzled philosophers for centuries. Extreme empiricism is not tenable. Its most concrete formulation, non-parametric estimation, shows that it requires astronomical learning times. The opposite extreme, assuming all knowledge to be present in the brain at birth, is equally untenable, not doing justice to the flexibility of our mind, and just putting the burden of statistical estimation on evolution. The only possible way out of this dilemma is the existence of general principles simple enough to be discovered by evolution and powerful enough to make learning a very fast process. This can only be imagined in a universe with profound regularities. The one we are exploiting here is the wide-spread existence of similarities between simultaneously visible patterns. This regularity is captured in the rather simple structure of our network, enabling it to generalize from single examples of symmetrical patterns, in striking contrast to the system of [Sejnowski et al. 1986] which is based on statistical estimation. With small modifications, dynamic link mapping can be used for the purpose of object recognition invariant with respect to translation, rotation and distortion, making the step from the correspondence of simultaneous patterns to those of consecutive patterns. Again, those transformations could be learned from few examples.

The very simple *a priori* principles incorporated in the learning system that we have presented are feature correspondence, topology, and rapid synaptic plasticity. We feel

that it is structural principles of this general style that make natural brains so extremely efficient in extracting significant structure from complex scenes. Although statistical estimation certainly plays a role for animal learning, it can evidently not be its only basis — natural scenes are too complex, and it is impossible to keep track of the whole combinatorics of subpatterns. Potent mechanisms are required to identify significant patterns already within single scenes. Ours may be a candidate.

References

- [Ackley et al. 1985] D.H. Ackley, G.E. Hinton, and T.S. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9:147–149, 1985.
- [Barlow and Reeves 1979] H.B. Barlow and B.C. Reeves. The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, 19:783–793, 1979.
- [Bienenstock and von der Malsburg 1987] E. Bienenstock and C. von der Malsburg. A neural network for invariant pattern recognition. *Europhysics Letters*, 4:121–126, 1987.
- [Bornstein et al. 1981] M.H. Bornstein, K. Ferdinandsen, and C.G. Gross. Perception of symmetry in infancy. *Developmental Psychology*, 17:82–86, 1981.
- [Buhmann et al. 1989] Joachim Buhmann, Jörg Lange, and C. von der Malsburg. Distortion invariant object recognition by matching hierarchically labeled graphs. In *IJCNN International Conference on Neural Networks, Washington*, pages I 155–159. IEEE, 1989.
- [Buhmann et al. 1991] J. Buhmann, J. Lange, C. von der Malsburg, J.C. Vorbrüggen, and R.P. Würtz. Object recognition in the dynamic link architecture – parallel implementation on a transputer network. In B. Kosko, editor, *Neural Networks: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, New York, 1991.
- [Coolen and Kuijk 1989] A.C.C. Coolen and F.W. Kuijk. A learning mechanism for invariant pattern recognition in neural networks. *Neural Networks*, 2:495, 1989.
- [Corballis and Roldan 1974] M.C. Corballis and C.E. Roldan. On the perception of symmetrical and repeated patterns. *Perception Psychophysics*, 16:136–142, 1974.
- [Cover and Hart 1967] T.M. Cover and P.E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, IT-13:21–27, 1967.
- [Crick 1984] F. Crick. Function of the thalamic reticular complex: the searchlight hypothesis. *Proc. of the Nat. Acad. of Sciences*, 81:4586 – 4590, 1984.
- [Delius and Nowak 1982] J.D. Delius and B. Nowak. Visual symmetry recognition by pigeons. *Psychological Research*, 44:199–212, 1982.

- [Eckhorn et al. 1988] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, and H. Reitboeck. Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60:121, 1988.
- [Fix and Hodges 1951] E. Fix and J.L. Hodges. Discriminatory analysis, non-parametric discrimination. Technical report, USAF School of aviation medicine, Project 21-49-004, Rept. 4, 1951.
- [Gray et al. 1989] C. M. Gray, P. König, A. K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337, 1989.
- [Julesz 1975] B. Julesz. Experiments in the visual perception of texture. *Scientific American*, 4, 1975.
- [Kohonen et al. 1988] T. Kohonen, G. Barna, and R. Chrisely. Statistical pattern recognition with neural networks: benchmarking studies. In *Proceedings of the IEEE ICNN, San Diego*, 1988.
- [Lades et al. 1993] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *to appear in: IEEE Transaction on Computers*, 1993.
- [Ramachandran 1988] V. Ramachandran. Perceiving Shape from shading. *Scientific American*, 10:76–83, 1988.
- [Reilly et al. 1982] D.L. Reilly, L.N. Cooper, and C. Elbaum. A neural model for category learning. *Biological Cybernetics*, 45:35–41, 1982.
- [Rosenblatt 1962] F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington, D.C., 1962.
- [Rumelhart et al. 1986] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning representations by backpropagating errors. *Nature*, 323:533–536, 1986.
- [Schneider 1986] W. Schneider. Anwendung der Korrelationstheorie der Hirnfunktion auf das akustische Figur-Hintergrund-Problem (Cocktailparty-Effekt). Ph.d. thesis, Universität Göttingen, 3400 Göttingen, F.R.G., 1986.
- [Sejnowski et al. 1986] T.J. Sejnowski, P.K. Kienker, and G.E. Hinton. Learning symmetry groups with hidden units: Beyond the perceptron. *Physica*, 22D:260–275, 1986.
- [von der Malsburg 1981] C. von der Malsburg. The correlation theory of brain function. Internal report, 81-2, Max-Planck-Institut für Biophysikalische Chemie, Postfach 2841, 3400 Göttingen, FRG, 1981.
- [von der Malsburg 1988] C. von der Malsburg. Pattern recognition by labeled graph matching. *Neural Networks*, 1:141–148, 1988.

[von der Malsburg and Buhmann 1992] C. von der Malsburg and J. Buhmann. Sensory segmentation with coupled neural oscillators. *Biological Cybernetics*, 67:233–242, 1992.

[Willshaw and von der Malsburg 1976] D.J. Willshaw and C. von der Malsburg. How patterned neural connections can be set up by self-organization. *Proc. R. Soc. London*, B194:431–445, 1976.

Figures

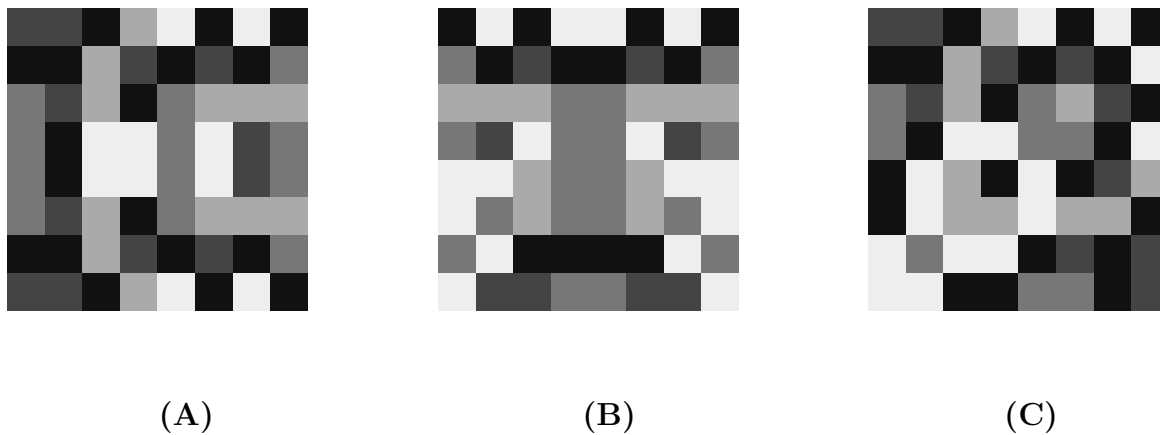


Figure 1: **Symmetrical Pixel Patterns.** Input patterns are arrays of $N \times N$ pixels, here $N = 8$. Pixel a has gray level feature value $F_a \in \{1, \dots, F_{max}\}$. In most of our simulations, $F_{max} = 10$. In each input image, pixel values are random, but equal for points symmetrical with respect to one of three axes: (A) horizontal, (B) vertical, (C) diagonal. The system has to solve the task of assigning input patterns to classes according to these symmetries, and to learn this performance from examples.

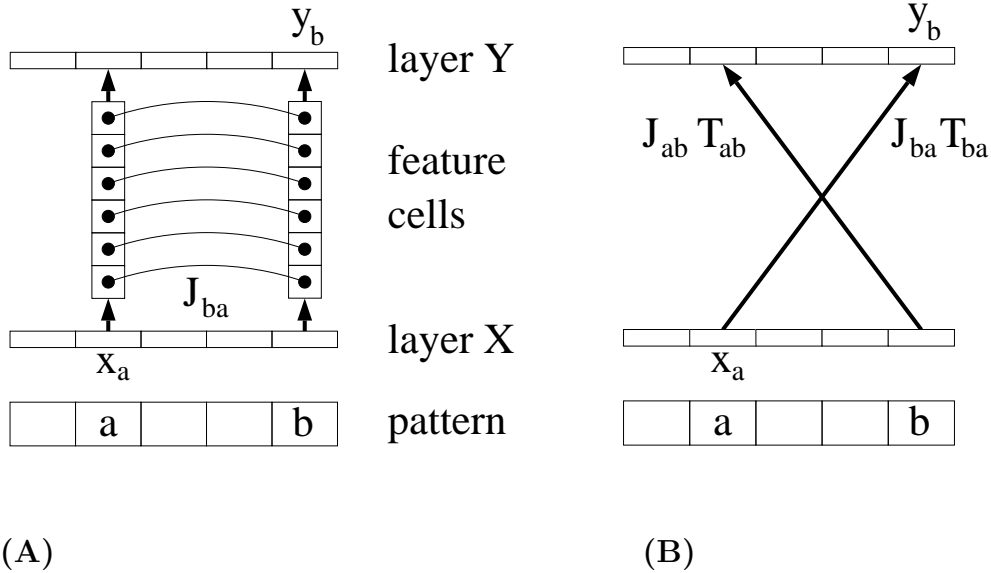


Figure 2: **Architecture of the dynamic link network.** (A) The complete architecture. The columns in two positions, a and b are shown. Feature cells are pre-activated by the pattern presented. Columns are connected with each other by feature-preserving links. These links are rapidly modifiable (“dynamic links”). Both the activator cells (layer X) and the collector cells (layer Y) have short-range excitatory and long-range inhibitory connections (not shown) and each have the tendency to form a local blob. Coupling from an X -cell a to a Y -cell b is via the pre-activated cells in column a , inter-columnar links, and the pre-activated cells in column b . (B) In our case, where only one feature is active per column, a functionally equivalent description uses the effective connections $J_{ba}T_{ba}$, where T_{ba} encodes the feature similarity between image positions a and b , *cf.* (Eq. (1)), and J_{ba} is the rapidly modifiable strength of the dynamic link.

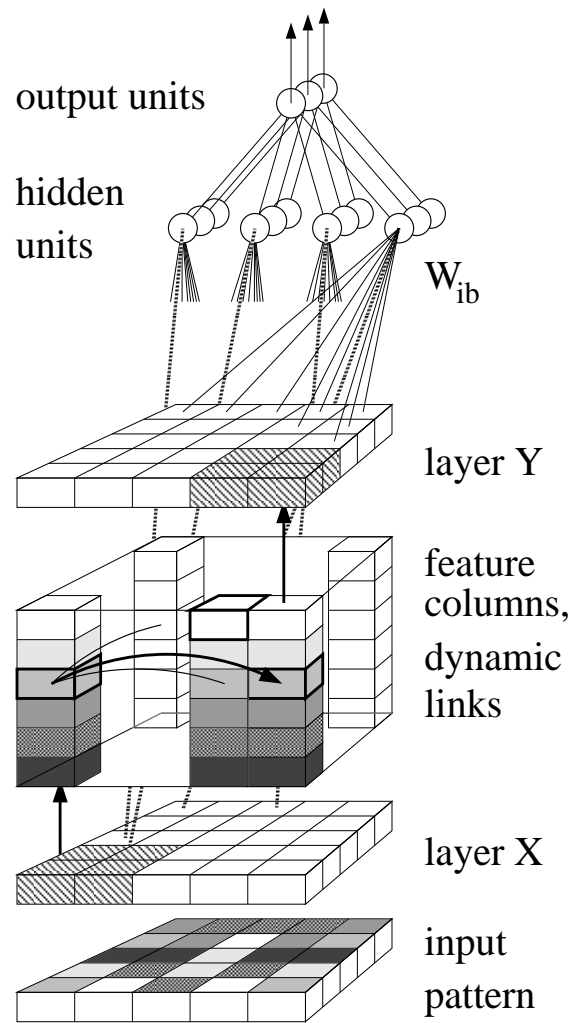


Figure 3: **The complete system.** An input pattern (lowest layer) is represented by sets of pre-activated neurons in the feature columns (marked here by heavy outline, on the basis of gray levels). Columns are connected by feature-preserving dynamic links (inter-columnar arrows). The dynamic link mechanism creates simultaneous blobs in layers X (“active blob”) and Y (“shadow blob”) in symmetrically corresponding positions (hatched areas). The symmetry type is then recorded (when it is new) or recognized (when already known) in the classification network (upper part). There are 6 hidden units per output unit (only 4 of which are shown). Each hidden unit has one fixed connection to its output unit, one connection from a randomly chosen X -cell, and has plastic connections w_{ib} from all Y -cells. These latter connections record a symmetry type permanently, by being concentrated into the Y -region lying symmetrically to the location of the X -input of the hidden unit.

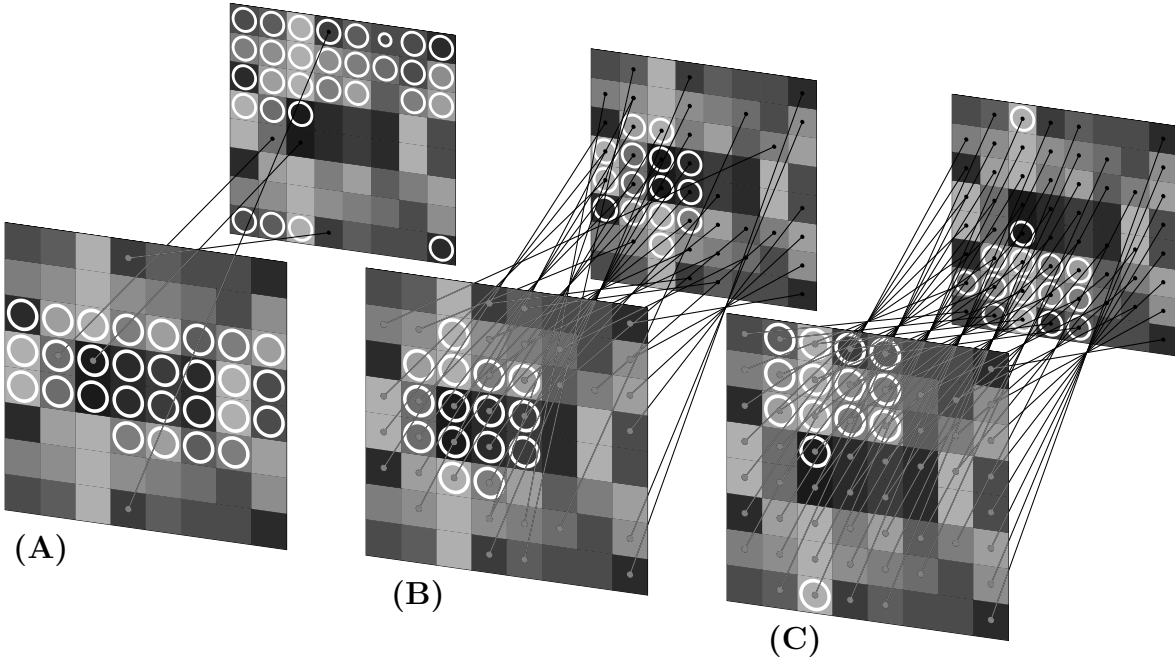


Figure 4: **Dynamic link mapping.** The network, with layers X (in front) and Y (in the rear) in different activation states, after 15 (A), 50 (B) and 80 (C) activity cycles, all generated for a fixed input pattern of symmetry class **A** (cf. Fig. 1). The dynamic link mapping process is based on a sequence of blob activations (white circles). Dynamic links $J_{ba} \in [0, 1]$ grow between temporally correlated cells. Only links with $J_{ba} \geq 0.4$ are shown in the figure.

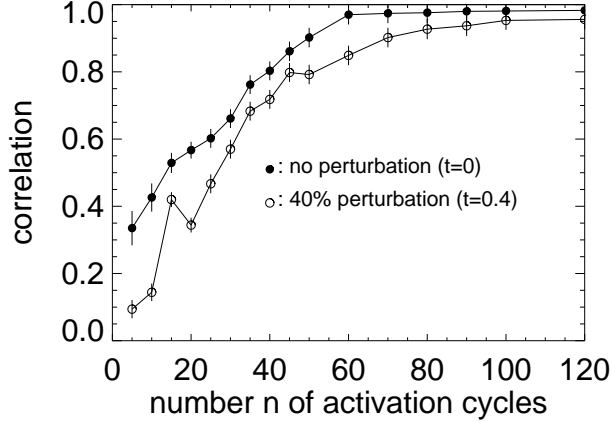
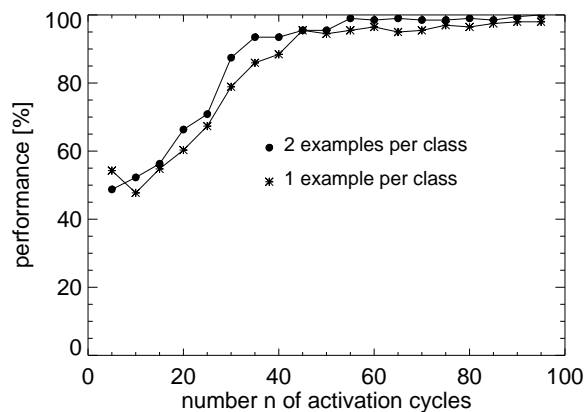


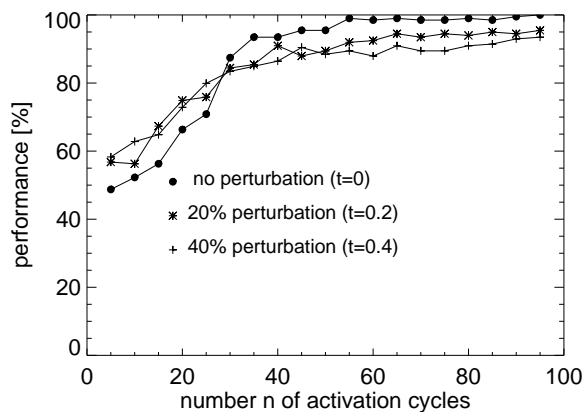
Figure 5: **Mean correlation between pairs of corresponding cells in layer X and layer Y for a given state of the dynamic links J after n activation cycles (blob pairs).** Correlation is computed as

$$C(x_a, y_{s(a)}) = \frac{\langle x_a y_{s(a)} \rangle - \langle x_a \rangle \langle y_{s(a)} \rangle}{\Delta x_a \Delta y_{s(a)}}$$

with $\Delta x = \sqrt{\langle (x - \langle x \rangle)^2 \rangle}$, and $s(a)$ denoting the the cell that lies symmetrically to a . To measure the correlation after n activation cycles, the link state $\{J_{ba}\}$ is frozen after n cycles (by setting $\epsilon = 0$), while the blob activation cycles continue. x_a and $y_{s(a)}$ are the equilibrium activities of cells a and $s(a)$, respectively, and $\langle \cdot \rangle$ denotes averaging over many cycles. Possible correlation values range from -1 for perfect anticorrelation to 1 for perfect correlation. What is displayed is the mean of $C(x_a, y_{s(a)})$ with respect to all possible positions a , and error bars denote the statistical errors when averaging over 900 cycles. Filled circles: Perfect feature similarity matrix $T_{ba} \in \{0, 1\}$. Open circles: All matches $T_{ba} = 1$ are replaced by random values $T_{ba} \in [1 - t, 1]$, all non-matches $T_{ba} = 0$ by a random $T_{ba} \in [0, t]$, to mimic the effects of noisy feature information. The correlations are robust against this perturbation.



(A)



(B)

Figure 6: **Symmetry Recognition Performance.** A total of 200 input patterns are classified according to one of three possible symmetries, cf. Fig. 1. The symmetry types have been recorded previously. The percentage of correct decisions is displayed as a function of the number n of activation cycles until the decision is forced. **(A)** Unperturbed features, $T_{ba} \in \{0, 1\}$, training with $k = 1$ or $k = 2$ examples per class, $120/k$ learning steps according to Eq. (7) for each example. **(B)** Influence on performance of perturbations in the feature similarity matrix T during recognition: The network can tolerate perturbations of $t = 20\%$ or even $t = 40\%$, where t is defined as in Fig. 5.