A Layered Recurrent Neural Network for Feature Grouping

Heiko Wersing, Jochen J. Steil and Helge Ritter

University of Bielefeld, Faculty of Technology, Neuroinformatics Group
P.O.-Box 10 01 31, D-33501 Bielefeld, Germany

Abstract. We describe a recurrent network, the Competitive Layer Model (CLM) for feature grouping. The model uses a combination of cooperative and competitive interactions to partition a set of input features into salient groups whose number is only restricted by the available layers. We give analytic results on convergence and the attractor states of the model and present simulation results showing grouping by proximity and grouping by symmetry and good continuation.

1 Introduction

In many applications we face the following problem: Given a set of separate features which are presented in parallel, find an appropriate partition into subsets of features underlying a common cause. A typical example is the perceptive grouping problem, where a set of derived image primitives like points and edges have to be grouped according to Gestalt principles like proximity, parallelity etc. A neural implementation that solves this problem has to involve two antagonistic processes: It has to *bind* related features into common groups, and it has to *separate* the groups originating from different sources.

Currently, synchronization models attract much interest, where binding or separating is achieved by temporal synchrony or asynchrony of neuron pools that code for related features [7,6]. However, due to the great dynamical complexity of these models applications are still rather limited to restricted problem domains. A different approach is taken in the brain-state-in-a-box (BSB) model [1], which is a recurrent network based on energy minimization. Upon a given input vector, the BSB network converges to exactly one of possibly many point attractors which forms a single detected group or cluster.

The Competitive Layer Model (CLM) [5] can be viewed as an extension of the BSB model as it allows simultaneous formation of many clusters corresponding to a greater number of salient subgroups in the input. Grouping is achieved by purely spatial interactions, using a set of identical layers of feature-selective neurons. Due to a competitive interaction between layers each feature is unambiguously assigned to one layer, thus separating different groups. The binding of features is caused by a lateral interaction within layers which determines the preferred groupings of compatible features.

appeared in Proc. Int. Conf. on Art. Neur. Netw. ICANN 97, Lausanne, pp.439-444

2 Model Architecture

The CLM consists of a set of L identical layers of feature-selective neurons which are replicas of an input layer (see Fig. 1). The neurons in the input layer are labelled by their position $\mathbf{r}$. Driven by an external input, each input neuron responds with a value $h_\mathbf{r}$ which indicates activity ($h_\mathbf{r} > 0$) in the presence of the corresponding feature or silence ($h_\mathbf{r} \leq 0$). In a simple setting we may think of $h_\mathbf{r}$ as encoding the light intensity at position $\mathbf{r}$ in some "imaginary" retina. (For more abstract features the position vector may be embedded in a higher dimensional spatially organized topographic feature map.)

For each of the input neurons at position $\mathbf{r}$ there is one neuron per replica layer α whose activity is $x_{\mathbf{r}\alpha} \geq 0$ (see Fig. 1). We denote these L activities by the *row* $\mathbf{r}$. The activities are subject to the following constrained gradient dynamics:

$$\dot{x}_{\mathbf{r}\alpha} = -\sigma_{\mathbf{r}\alpha}\frac{\partial E}{\partial x_{\mathbf{r}\alpha}} = \sigma_{\mathbf{r}\alpha} \cdot \Big(J_1 (h_\mathbf{r} - \sum_{\beta=1}^{L} x_{\mathbf{r}\beta}) + \sum_{r'} \mathsf{f}_{\mathbf{r}\mathbf{r}'} x_{\mathbf{r}'\alpha} \Big) \ , \tag{1}$$

where the constraint $x_{\mathbf{r}\alpha} \geq 0$ is dynamically enforced by

$$\sigma_{\mathbf{r}\alpha} = \begin{cases} 0 & \text{if} \quad x_{\mathbf{r}\alpha} = 0, \qquad \frac{\partial E}{\partial x_{\mathbf{r}\alpha}} > 0 \qquad \text{(active constraint)} \\ 1 & \text{else} \qquad\qquad\qquad\qquad\qquad\quad \text{(inactive constraint)} \ . \end{cases} \tag{2}$$

J_1 is a positive constant and $\mathsf{f}_{\mathbf{r}\mathbf{r}'} = \mathsf{f}_{\mathbf{r}'\mathbf{r}}$ are the components of a symmetric weight matrix. As a result of the constraints (2), the dynamics (1) defines a nonlinear dynamical system, which consists of a series of continuously connected linear systems. The corresponding energy function E is quadratic and given by

$$E = \frac{J_1}{2} \sum_{r} \Big(\sum_{\alpha=1}^{L} x_{r\alpha} - h_r \Big)^2 - \frac{1}{2} \sum_{\alpha=1}^{L} \sum_{rr'} \mathsf{f}_{\mathbf{r}\mathbf{r}'} x_{\mathbf{r}\alpha} x_{\mathbf{r}'\alpha} \ . \tag{3}$$

The gradient in (1) can be split into $-\frac{\partial E}{\partial x_{\mathbf{r}\alpha}} = J_1 V_\mathbf{r} + F_{\mathbf{r}\alpha}$, where $V_\mathbf{r} = h_\mathbf{r} - \sum_\beta x_{\mathbf{r}\beta}$ and $F_{\mathbf{r}\alpha} = \sum_{\mathbf{r}'} \mathsf{f}_{\mathbf{r}\mathbf{r}'} x_{\mathbf{r}'\alpha}$ are the two basic interactions in the model:

- The *"vertical"* interaction $J_1 V_\mathbf{r}$ causes strict competition between neurons in a row belonging to the same feature. The intensity $h_\mathbf{r}$ gives a constant input to the row. If we consider a single row of the model without lateral connections the competitive dynamics is equivalent to the dynamics of the Winner-Take-All MAXNET [2] with added inhomogeneity.
- The *lateral* interaction $F_{\mathbf{r}\alpha}$ couples activities within each layer by the symmetric weight matrix $\mathsf{f}_{\mathbf{r}\mathbf{r}'}$. The interactions $\mathsf{f}_{\mathbf{r}\mathbf{r}'}$ determine which pattern configurations, if elicited as activity pattern within a single layer, will be mutually supporting among their constituent parts (for $\mathsf{f}_{\mathbf{r}\mathbf{r}'} > 0$) or instead suffer mutual inhibition (for $\mathsf{f}_{\mathbf{r}\mathbf{r}'} < 0$).

As will be shown below, the interplay between the vertical and the lateral interactions leads to a dynamic partitioning of an input pattern (offered via the

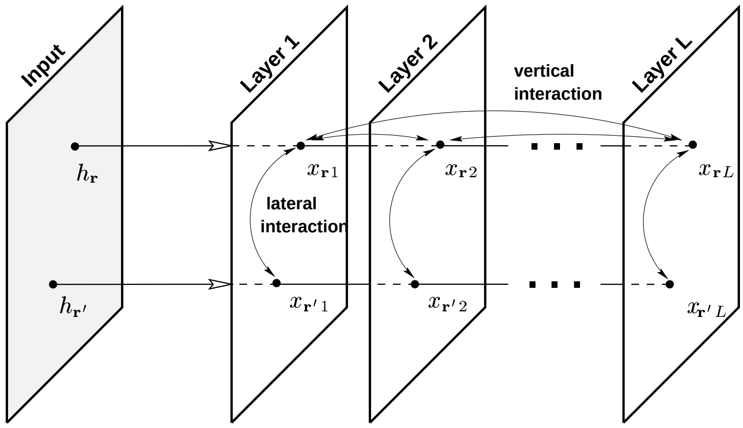


Fig. 1. The Competitive Layer Model architecture.

$h_{\mathbf{r}}$-inputs of the leftward input layer) such that each layer responds only to a pattern fragment, with the union of all fragments giving the entire pattern. In this way the full pattern becomes grouped into smaller constituents which are determined by the lateral interaction $\mathsf{f}_{\mathbf{rr}'}$.

The parametric model for $\mathsf{f}_{\mathbf{rr}'}$ has to be chosen such that desired grouping states correspond to energy minima of (3), where the quality of a grouping is measured by the second term, summing up all pairwise contributions weighted by $\mathsf{f}_{\mathbf{rr}'}$. In this paper we suggest a heuristic approach and consider two interactions of the form $\mathsf{f}_{\mathbf{rr}'} = M_{\mathbf{rr}'} - k$ where $M_{\mathbf{rr}'}$ is a positive measure of the compatibility between two features $\mathbf{r}, \mathbf{r}'$ and $k > 0$ is an inhibitory threshold which controls the required degree of overall conformity within a group. Larger values for k result in a more fragmented grouping, while for $k = 0$ the global minimum of (3) is attained by forming just one large group.

3 Convergence and Stability Analysis

The following theorem guarantees the global convergence of the CLM:

Theorem 1. *The CLM system is convergent if* $J_1 > \sum_{r' \wedge \, \mathsf{f}_{\mathbf{rr}'} > 0} \mathsf{f}_{\mathbf{rr}'}$ *for all* $\mathbf{r}$.

Proof. (Outline) The behaviour of E at infinity is dominated by the second-order terms which can be shown to be bounded from below by $E' = \frac{1}{2} \sum_{\mathbf{r}\alpha} x_{\mathbf{r}\alpha}^2 \big(J_1 - \sum_{\mathbf{r}' \wedge \, \mathsf{f}_{\mathbf{rr}'} > 0} \mathsf{f}_{\mathbf{rr}'}\big)$. For sufficiently large J_1, E is then bounded from below and $E \to \infty$ for $\|\mathbf{x}\| \to \infty$. Since E is also strictly decreasing on nonstationary trajectories E is a global Lyapunov function and according to the Invariant Set Theorem [3] the system converges to the set of equilibria. $\square$

This condition implies that the coupling to the input, controlled by J_1, must be sufficiently strong compared to the self-excitation within each layer.

The essential property of the attractors of the CLM is the unambiguous assignment of each input feature to just one layer. We prove this property by showing that it is a necessary condition on local minima of E. Since the dynamics (1) is a gradient descent system, attractors are situated at local energy minima.

Theorem 2. *If the lateral interaction is positive self-coupling,* $\mathfrak{f}_{\mathbf{rr}} > 0$ *for all* $\mathbf{r}$*, then an attractor has in each row* $\mathbf{r}$ *either*
i) at most one positive activity $x_{\mathbf{r}\alpha(\mathbf{r})} = h_{\mathbf{r}} + F_{\mathbf{r}\alpha(\mathbf{r})}/J_1,\ x_{\mathbf{r}\beta\neq\alpha(\mathbf{r})} = 0,$ *where* $\alpha(\mathbf{r})$ *is the index of the maximally supporting layer with* $F_{\mathbf{r}\alpha(\mathbf{r})} > F_{\mathbf{r}\beta\neq\alpha(\mathbf{r})}$ *or*
ii) for all activities in a row $x_{\mathbf{r}\alpha} = 0,\ F_{\mathbf{r}\alpha} \leq 0$*.*

Proof. (Outline) Suppose an equilibrium has two positive activities $x_{\mathbf{r}\alpha,\beta} > 0$ in a row $\mathbf{r}$. Then the constraint is inactive, $\sigma_{\mathbf{r}\alpha,\beta} = 1$, hence $\frac{\partial E}{\partial x_{\mathbf{r}\alpha,\beta}} = 0$. Now consider a small perturbation of the form $\epsilon_{\mathbf{r}\alpha} = \eta,\ \epsilon_{\mathbf{r}\beta} = -\eta,\ \epsilon_{\mathbf{r}'\alpha'} = 0$ for all other $\mathbf{r}', \alpha'$. Expanding E about the equilibrium we obtain for ΔE under this perturbation with all linear components vanishing at the equilibrium

$$\Delta E = \frac{1}{2} \sum_{\mathbf{r}\alpha\mathbf{r}'\alpha'} (J_1 \delta_{\mathbf{rr}'} - \delta_{\alpha\alpha'} \mathfrak{f}_{\mathbf{rr}'}) \epsilon_{\mathbf{r}\alpha} \epsilon_{\mathbf{r}'\alpha'} = -\eta^2 \mathfrak{f}_{\mathbf{rr}} < 0 \ . \tag{4}$$

Because this perturbation decreases the energy this equilibrium must be instable and there can be at most one positive activity in a row for a stable equilibrium. Now consider case i): Let $\alpha(\mathbf{r})$ denote the layer index of the positive activity in row $\mathbf{r}$. The constraint is then inactive for $x_{\mathbf{r}\alpha(\mathbf{r})}$ and active for $x_{\mathbf{r}\beta\neq\alpha(\mathbf{r})}$. Hence

$$0 = \frac{\partial E}{\partial x_{\mathbf{r}\alpha(\mathbf{r})}} = J_1 V_{\mathbf{r}} + F_{\mathbf{r}\alpha(\mathbf{r})}, \qquad 0 > -\frac{\partial E}{\partial x_{\mathbf{r}\beta\neq\alpha(\mathbf{r})}} = J_1 V_{\mathbf{r}} + F_{\mathbf{r}\beta\neq\alpha(\mathbf{r})} \ , \tag{5}$$

which proves case i). Case ii) gives just the remaining possible stable equilibria which are not covered by i). □

This ensures that each feature is either mapped into the layer with maximal lateral feedback or completely removed from the grouping result. For large J_1 the superposition of activity patterns in all layers reproduces the input pattern. If J_1 is just above the threshold given in Theorem 1 the resulting activity pattern will show a strong deviation from the input intensity due to the lateral feedback $F_{\mathbf{r}\alpha}$. This allows a controllable auto-associativity of the model. An example would be the "filling in" of a feature that was not detected in the input, $h_{\mathbf{r}} = 0$, but receives a strong lateral feedback $F_{\mathbf{r}\alpha}$ in one layer.

4 Simulation Results for Perceptive Grouping Tasks

In the following, we illustrate the analytical results with some computer simulations for the task of grouping visual features. In the simplest possible case, we consider simple point features (characterized only by their location $\mathbf{r}$) and use an "On-Center-Off-Surround" interaction $\mathfrak{f}_{\mathbf{rr}'} = \exp(-|\mathbf{r} - \mathbf{r}'|^2/(2R)) - k$ with a range R and inhibitory strength k. A CLM with this interaction splits input patterns into compact clusters of a typical radial size R. Fig.2a shows a simple image of a scene used as input for the CLM. In Fig.2b we see the grouping result achieved by a CLM of 10 layers, operating on the input image after sub-sampling to 50x40 pixels (each thresholded pixel intensity serving as a point feature in this simple demonstration).

Usually, one desires a more sophisticated grouping that requires to take additional Gestalt principles into account. Further important Gestalt laws are the "law of good continuation", which requires to consider in addition to position also local edge orientation as a visual feature, and laws that consider symmetry in various ways. Such additional Gestalt laws can be incorporated into the CLM by choosing a more sophisticated interaction function $\mathsf{f}_{\mathbf{rr}'}$. We can use, e.g., a suggestion made by Reisfeld [4] in a different context for an interaction between edge elements to detect local symmetries. Edges are parametrized by a position vector $\mathbf{r}$ and a unit orientation vector $\hat{\mathbf{n}}$. This leads to an interaction given by

$$\mathsf{f}_{\mathbf{rr}'\hat{\mathbf{n}}\hat{\mathbf{n}}'} = \exp\big(-(|\hat{\mathbf{n}}\hat{\mathbf{d}}| - |\hat{\mathbf{n}}'\hat{\mathbf{d}}|)^2\sigma\big)\exp\big(-\mathbf{d}^2/R\big) \quad - \quad k \ , \tag{6}$$

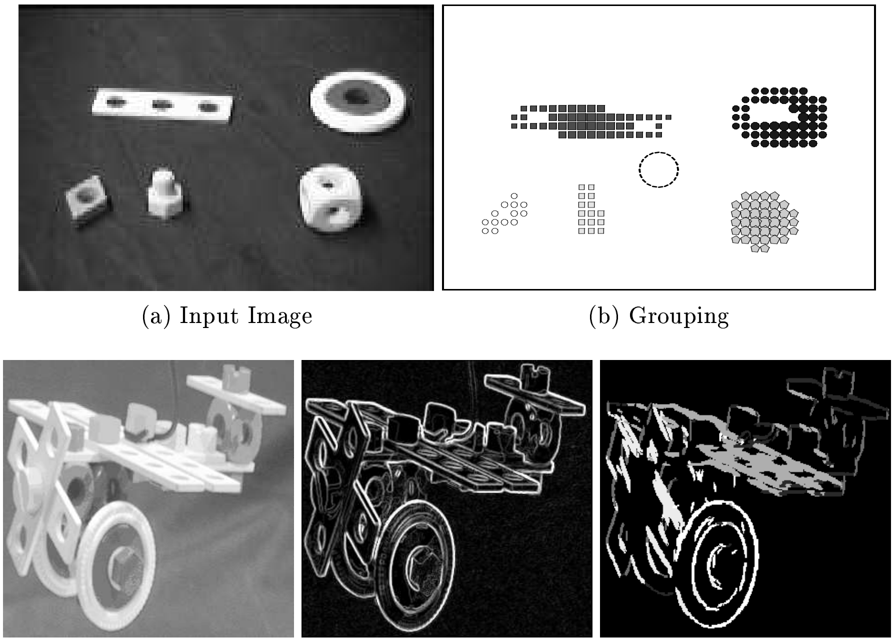
where $\mathbf{d} = \mathbf{r} - \mathbf{r}'$, $\hat{\mathbf{d}} = \mathbf{d}/|\mathbf{d}|$ is the spatial (normalized) difference vector, σ controls the angular selectivity, R gives the spatial range and k controls the degree of conformity within groups. Fig.2 shows that a CLM with this interaction is capable of successfully grouping edge elements (obtained from a real world image) according to the principles of good continuation and symmetry. The spatial range R was chosen to be of the same magnitude as the desired group size. We observed that for noisy input data like the output of a gradient edge detection preprocessor the selectivity σ must be low to allow the detection of salient groups.

5 Conclusion

The CLM is able to break a set of input features into a number of groups by two different kinds of interactions: A competitive, topographically structured interaction between layers, and a lateral interaction within layers. The lateral interaction defines an energy functional, where grouping states correspond to minima of this energy. With a stability analysis we show that the attractors of the model form a disjoint decomposition of the input feature pattern for arbitrary symmetric and positive self-coupling lateral interactions. A necessary condition on attractors implies that each feature has to be grouped into the layer with maximal lateral feedback. We apply the model to perceptive grouping where we report results for point clustering and symmetry and good continuation grouping of edge elements.

References

1. S. A. Ritz J. A. Anderson, J. W. Silverstein and R. S. Jones. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 84:413–451, 1977.



(a) Input Image (b) Grouping

(c) Input Image (d) Edge Map (e) Grouping

Fig. 2. Application to perceptive grouping. a) Input image. b) Grouping result for an "On Center Off Surround" interaction with 10 layers on a collection of point features derived from a) where layers are represented by different symbols. The range R is displayed by the radius of the dashed circle and the inhibition was given by $k = 0.1$. c) Input image and d) edge map obtained from a gradient operator. e) Grouping result of a lateral interaction for symmetry and good continuation with 5 layers on a set of 1200 derived (sub-sampled) edge features, displayed as (overlapping) ellipses of corresponding orientation and with size representing activity. Interaction parameters were $\sigma = 100, R = 0.1, k = 0.2$ and the displayed square was normalized to unit length. Wheel, wings and parallel parts of the model plane's propeller form separate groups.

2. R. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Mag.*, pages 4–22, 1987.
3. D. G. Luenberger. *Introduction to dynamical systems: Theory, models, and applications.* New York: Wiley, 1979.
4. Yeshurun H. W. Reisfeld, D. Context-free attentional operators: The generalized symmetry transform. *Int. Journal of Computer Vision*, 14:119–130, 1995.
5. H. Ritter. A spatial approach to feature linking. In *Int. Neur. Netw. Conf. Paris*, 1990.
6. R. Ritz, W. Gerstner, U. Fuentes, and J.L. van Hemmen. A biological motivated and analytically soluble model of collective oscillations in the cortex. II. Application to binding and pattern segmentation. *Biol. Cybern.*, 71:349–358, 1994.
7. C. v.d. Malsburg. The correlation theory of brain function. Technical Report 81-2, MPI Göttingen, 1981.