ART-EMAP: A Neural Network Architecture for Object Recognition by Evidence Accumulation

Gail A. Carpenter, *Member, IEEE,* and William D. Ross

Abstract— **A new neural network architecture is introduced for the recognition of pattern classes after supervised and unsupervised learning. Applications include spatio-temporal image understanding and prediction and three-dimensional (3-D) object recognition from a series of ambiguous two-dimensional views. The architecture, called ART-EMAP, achieves a synthesis of adaptive resonance theory (ART) and spatial and temporal evidence integration for dynamic predictive mapping (EMAP). ART-EMAP extends the capabilities of fuzzy ARTMAP in four incremental stages. Stage 1 introduces distributed pattern representation at a view category field. Stage 2 adds a decision criterion to the mapping between view and object categories, delaying identification of ambiguous objects when faced with a low confidence prediction. Stage 3 augments the system with a field where evidence accumulates in medium-term memory. Stage 4 adds an unsupervised learning process to fine-tune performance after the limited initial period of supervised network training. Each ART-EMAP stage is illustrated with a benchmark simulation example, using both noisy and noise-free data. A concluding set of simulations demonstrate ART-EMAP performance on a difficult 3-D object recognition problem.**

I. Introduction: Object Recognition by Spatial and Temporal Evidence Accumulation

ART-EMAP (adaptive resonance theory with spatial and temporal evidence integration for dynamic predictive mapping) is a neural network architecture that accumulates spatial and temporal evidence to recognize target objects and pattern classes in noisy or ambiguous input environments. During performance, ART-EMAP integrates spatial evidence distributed across coded recognition categories to predict a pattern class. When a decision criterion determines the pattern class choice to be ambiguous, additional input from the same unknown class is sought. Evidence from multiple inputs accumulates until the decision criterion is satisfied, and a high confidence prediction can be made. Accumulated evidence can also be used by the predictive mapping to fine-tune the system during unsupervised rehearsal learning.

In four incremental stages, ART-EMAP both improves performance accuracy and extends the domain of the fuzzy ARTMAP (adaptive resonance theory-supervised predictive mapping) [1] neural network (Fig. 1) to include spatio-temporal recognition and prediction. Applications include a vision system capable of sampling different two-dimensional (2-D) perspectives of three-dimensional (3-D) objects. In this scenario, a sensor generates an organized database of inputs which could be either views of an object from different perspectives or noisy samples of single views. This approach to 3-D object recognition has been successfully used in neural network machine vision applications, particularly the aspect network [2] and [3]. ART-EMAP further develops this strategy.

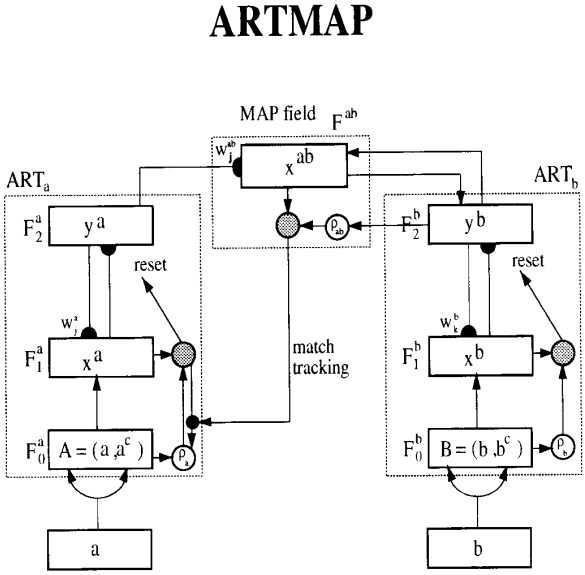


Fig. 1. Fuzzy ARTMAP architecture. The ART_a complement coding preprocessor transforms the vector **a** into the vector $\mathbf{A} = (\mathbf{a}, \mathbf{a}^c)$ at the ART_a field F_0^a. **A** is the input to the ART_a field F_1^a. Similarly, in the supervised mode, the input to the ART_b field F_1^b is the vector $(\mathbf{b}, \mathbf{b}^c)$. When a prediction by ART_a is disconfirmed at ART_b, inhibition of the map field F^{ab} induces the match tracking process. Match tracking raises the ART_a vigilance (ρ_a) to just above the F_1^a-to-F_0^a match ratio $|\mathbf{x}^a|/|\mathbf{A}|$. This triggers an ART_a search, which leads to activation of either an ART_a category that correctly predicts **b** or to a previously uncommitted ART_a category node.

Each stage of the ART-EMAP network is illustrated on an ARPA benchmark simulation problem, circle-in-the-square [4]. This problem requires a system to identify which points of a square lie inside and which lie outside a circle whose area is half that of the square. In the simulations, a single set of training/test exemplars is used to evaluate each of

Manuscript received February 25, 1993; revised March 20, 1994 and August 21, 1994. This work was supported in part by ARPA Grant ONR N00014-92-J-4015, National Science Foundation Grant NSF IRI 94-01659, Office of Naval Research Grant ONR N00014-91-J-4100, and Air Force Office of Scientific Research Grant AFOSR 90-0083.

The authors are with the Center for Adaptive Systems and Department of Cognitive and Neural Systems, Boston University, Boston, MA 02215 USA.

IEEE Log Number 9409152.

the four ART-EMAP stages and to compare performance with that of fuzzy ARTMAP. A training set consists of 100 randomly chosen circle-in-the-square points. Each simulation begins with a supervised learning phase in which these points are presented once, each time in the same order. At each of the four ART-EMAP stages, performance is evaluated on both noise-free and noisy test sets. The noise-free test set consists of a discrete sampling of 11 000 points. The noisy test set is generated by adding random noise to each of these 11 000 inputs. The random noise is Gaussian with standard deviation equal to 0.1 times the length of one side of the square.

To make the discussion of ART-EMAP self-contained, Section II includes a summary of the fuzzy ART and fuzzy ARTMAP neural network architectures. During training, ARTMAP improves its pattern class prediction incrementally through unsupervised clustering of view, or input, categories and supervised learning of a predictive mapping from categories to output classes. On the circle-in-the-square benchmark, fuzzy ARTMAP performs at 93.1% accuracy on the noise-free test set and 86.5% on the noisy test set. ART-EMAP is then introduced as a series of progressive modifications of the fuzzy ARTMAP system. ART-EMAP accomplishes test set pattern class prediction by: Stage 1) the integration of spatially distributed recognition information across view categories, Stage 2) decision criterion control, which delays categorization of ambiguous objects, and Stage 3) temporal integration, or accumulation, of evidence from a sequence of inputs belonging to the same unknown pattern class.

Section III specifies the equations of the spatial evidence integration system (Stage 1) and evaluates this system via parametric simulations. During performance, spatially distributed patterns across several coded recognition categories improve the accuracy of pattern class predictions. An input activates each category in proportion to a measure of match between the input and that category's coded template. This distributed information yields distributed predictive evidence for each pattern class. Choice of the most highly activated pattern class gives a prediction that is more reliable, on average, than the one made by the most active view category alone. ART-EMAP with spatial evidence integration performs at 95.7% accuracy on the noise-free circle-in-the-square test set and at 88.4% accuracy on the noisy test set. Stage 1 alone is thus seen to improve ARTMAP predictive accuracy.

Section IV introduces a strategy for predictive control using a decision criterion (Stage 2). This mechanism allows the system to seek additional views of the same unknown class when the spatial evidence resulting from an input is determined to be ambiguous. The decision criterion provides efficient control of system input and a reliable means of ensuring predictive accuracy in a noisy or unfamiliar environment. In fact, ART-EMAP with a decision criterion performs at 100% accuracy on the noise-free test set and at 93.1% on the noisy test set.

Section V introduces a process that accumulates predictive evidence over a temporal sequence of inputs until the decision criterion is satisfied (Stage 3). In applications, multiple inputs could correspond to noisy images taken from a fixed perspective (samples) or to ambiguous images taken from various perspectives (views). ART-EMAP with evidence accumulation over simulated multiple views performs at 97.6% accuracy on the noisy test set. Evidence accumulation over multiple samples yields 92.2% accuracy.

Section VI describes how unsupervised rehearsal learning (Stage 4) can use Stage 3 predictions to improve future performance. When spatio-temporal evidence integration makes a confident prediction, this knowledge is used by the system to fine-tune its own predictive mapping. Simulations show that unsupervised rehearsal learning yields improved performance on both noise-free and noisy test sets.

Section VII illustrates performance of fuzzy ARTMAP and ART-EMAP, Stage 1 through Stage 4, on a recognition problem which requires a system to identify three similar 3-D objects (pyramid, prism, house) from ambiguous 2-D views. The difficulty of the problem is illustrated by the fact that fuzzy ARTMAP correctly identifies only 64.7% of the objects from a noise-free test set. Stage 1 ART-EMAP raises performance accuracy to 70.6%, while Stage 2 and Stage 3 both boost performance to 98.0%.

Finally, Section VIII compares the structure and dynamics of ART-EMAP with those of the aspect net [2], of Seibert and Waxman, who introduced the use of evidence accumulation processes in ART-based neural network architectures.

II. ARTMAP: A Neural Network Architecture for Self-Organizing Recognition and Prediction

ARTMAP (Fig. 1) is a neural network architecture that performs incremental supervised learning of recognition categories and multidimensional maps in response to input vectors presented in arbitrary order. The first ARTMAP system [5] and [6] was used to classify inputs by the set of features they possess, that is, by a vector of binary values representing the presence or absence of each feature. The more general fuzzy ARTMAP system [1] learns to classify inputs by a fuzzy set of features or a pattern of fuzzy membership values between zero and one that indicate the extent to which each feature is present. This generalization is accomplished by replacing the ART 1 modules [6] and [7] of binary ARTMAP with fuzzy ART modules [8]. Where ART 1 dynamics are described in terms of binary set-theoretic operations, fuzzy ART dynamics are described in terms of analog fuzzy set-theoretic operations [9] and [10] (Fig. 2). Fuzzy ARTMAP accomplishes accuracy, speed, and code compression in both on-line and off-line settings. It has a small number of parameters and requires no problem-specific crafting or choice of initial weights or parameters. In addition, in a growing number of applications, fuzzy ARTMAP has shown better performance than various other neural networks. Examples include automatic analysis of electrocardiogram traces [11], diagnostic monitoring of nuclear plants [12], and prediction of protein secondary structure [13].

Each ARTMAP system includes a pair of ART modules (ART_a and ART_b) that creates stable recognition categories in response to arbitrary sequences of input patterns (Fig. 1). During supervised learning, ART_a receives a stream of input patterns $\{\mathbf{a}^{(l)}\}$ and ART_b receives a stream $\{\mathbf{b}^{(l)}\}$ of input

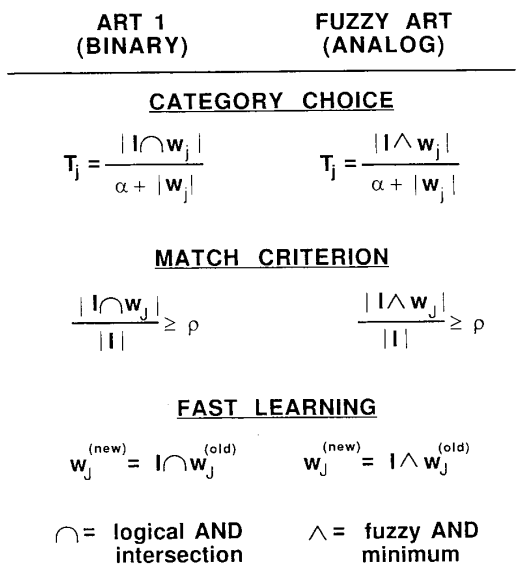


Fig. 2. Comparison of ART 1 and fuzzy ART logic.

patterns, where $\mathbf{b}^{(l)}$ is the correct prediction given $\mathbf{a}^{(l)}$. These modules are linked by an associative learning network and an internal controller that ensures autonomous system operation in real time. The controller is designed to create the minimal number of ART_a recognition categories, or "hidden units," needed to meet accuracy criteria. It does this by realizing a minimax learning rule that enables ARTMAP to conjointly minimize predictive error and maximize code compression. Predictive success is automatically linked to category size on a trial-by-trial basis using only local operations through increasing the matching criterion, or vigilance parameter (ρ_a), of ART_a by the minimal amount needed to correct a predictive error at ART_b. Vigilance ρ_a calibrates the minimum confidence that ART_a must have in a recognition category, or hypothesis, activated by an input $\mathbf{a}^{(l)}$ for ART_a to accept that category, rather than search for a better one (and perhaps establish a new category). Lower values of ρ_a enable larger categories to form, leading to a higher degree of code compression. A predictive failure at ART_b increases ρ_a by the minimum amount needed to trigger alternative hypothesis testing at ART_a, via a mechanism called match tracking [5]. Match tracking sacrifices the minimum amount of compression necessary to correct a predictive error. The combination of match tracking and fast learning allows an ARTMAP system to learn a correct prediction for a rare event embedded in a cloud of featurally frequent events that make a different prediction.

A. Fuzzy ART Summary

The fuzzy ARTMAP system incorporates two fuzzy ART modules, ART_a and ART_b. Each fuzzy ART subsystem includes a field, F_0, of nodes that represent a current input vector; a field, F_1, that receives both bottom-up input from F_0 and top-down input from a field; F_2, that represents the active code, or category (Fig. 1). The F_0 activity vector is denoted $\mathbf{I} = (I_1, \cdots, I_M)$, with each component I_i in the interval [0, 1] ($i = 1, \cdots, M$). The F_1 activity vector is denoted $\mathbf{x} = (x_1, \cdots, x_M)$, and the F_2 activity vector is denoted $\mathbf{y} = (y_1, \cdots, y_N)$. The number of nodes in each field is arbitrary.

Weight Vector: Associated with each F_2 category node j ($j = 1, \cdots, N$) is a vector $\mathbf{w}_j \equiv (w_{j1}, \cdots, w_{jM})$ of adaptive weights, or LTM (long-term memory) traces. Initially, when each category is said to be uncommitted

$$w_{j1}(0) = \cdots = w_{jM}(0) = 1. \tag{1}$$

After a category is selected for coding it becomes committed. Each LTM trace w_{ji} is monotonically nonincreasing through time and hence converges to a limit. The fuzzy ART weight vector $\mathbf{w}_j$ formally represents both the bottom-up and top-down weight vectors of ART 1.

Parameters: Fuzzy ART dynamics are determined by a choice parameter $\alpha > 0$ a learning rate parameter $\beta \in [0, 1]$ and a vigilance parameter $\rho \in [0, 1]$.

Category Choice: For each input $\mathbf{I}$ and F_2 node j, the choice function T_j is defined by

$$T_j(\mathbf{I}) = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|} \tag{2}$$

where the fuzzy AND, or intersection, operator ($\wedge$) is defined by

$$(\mathbf{p} \wedge \mathbf{q})_i \equiv \min(p_i, q_i) \tag{3}$$

and where the norm $|\cdot|$ is defined by

$$|\mathbf{p}| = \sum_{i=1}^{M} |p_i| \tag{4}$$

for any M-dimensional vectors $\mathbf{p}$ and $\mathbf{q}$. For notational simplicity, $T_j(\mathbf{I})$ is written as T_j when the input $\mathbf{I}$ is fixed.

The system is said to make a category choice when at most one F_2 node can become active at a given time. The category choice is indexed by J, where

$$T_J = \max\{T_j : \ j = 1 \cdots N\}. \tag{5}$$

If more than one T_j is maximal, the category j with the smallest index is chosen. In particular, nodes become committed in order $j = 1, 2, 3, \cdots$. When the Jth category is chosen, $y_J = 1$ and $y_j = 0$ for $j \neq J$. In a choice system, the F_1 activity vector $\mathbf{x}$ is characterized by the equation

$$\mathbf{x} = \begin{cases} \mathbf{I} & \text{if } F_2 \text{ is inactive} \\ \mathbf{I} \wedge \mathbf{w}_J & \text{if the } J\text{th } F_2 \text{ node is active.} \end{cases} \tag{6}$$

Resonance or Reset: Resonance occurs if the match function $|\mathbf{I} \wedge \mathbf{w}_J|/|\mathbf{I}|$ of the chosen category J meets the vigilance criterion

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho. \tag{7}$$

That is, when the Jth category is chosen, resonance occurs if

$$|\mathbf{x}| = |\mathbf{I} \wedge \mathbf{w}_J| \geq \rho |\mathbf{I}|. \tag{8}$$

Learning then ensues, as defined below. Mismatch reset occurs if

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} < \rho. \tag{9}$$

That is, when

$$|\mathbf{x}| = |\mathbf{I} \wedge \mathbf{w}_J| < \rho|\mathbf{I}| \tag{10}$$

the value of the choice function T_J is set to zero for the duration of the input presentation to prevent the persistent selection of the same category during search. A new index J is then chosen by (5). The search process continues until the chosen J satisfied (7).

Learning: Once search ends, the weight vector $\mathbf{w}_J$ is updated according to the equation

$$\mathbf{w}_J^{(\text{new})} = \beta\left(\mathbf{I} \wedge \mathbf{w}_J^{(\text{old})}\right) + (1-\beta)\mathbf{w}_J^{(\text{old})}. \tag{11}$$

Fast learning corresponds to setting $\beta = 1.0$.

B. ARTMAP Summary

In the ARTMAP system, ART_a and ART_b are linked via an inter-ART module, F^{ab}, called a map field, as follows.

ART_a and ART_b: Inputs to ART_a and ART_b are in the complement code form: for ART_a, input $\mathbf{I} = \mathbf{A} = (\mathbf{a}, \mathbf{a}^c)$; and for ART_b, input $\mathbf{I} = \mathbf{B} = (\mathbf{b}, \mathbf{b}^c)$ (Fig. 1). Variables in ART_a or ART_b are designated by superscripts a or b. For ART_a, $\mathbf{x}^a \equiv (x_1^a, \cdots, x_{2M_a}^a)$ denotes the F_1^a output vector; $\mathbf{y}^a \equiv (y_1^a, \cdots, y_{N_a}^a)$ denotes the F_2^a output vector; and $\mathbf{w}_j^a \equiv (w_{j,1}^a \cdots w_{j,2M_a}^a)$ denotes the jth ART_a weight vector. For ART_b, $\mathbf{x}^b \equiv (x_1^b, \cdots, x_{2M_b}^b)$ denotes the F_1^b output vector; $\mathbf{y}^b \equiv (y_1^b, \cdots, y_{N_b}^b)$ denotes the F_2^b output vector; and $\mathbf{w}_k^b \equiv (w_{k,1}^b, \cdots, w_{k,2M_b}^b)$ denotes the kth ART_b weight vector. For the map field, $\mathbf{x}^{ab} \equiv (x_1^{ab}, \cdots, x_{N_b}^{ab})$ denotes the F^{ab} output vector and $\mathbf{w}_j^{ab} \equiv (w_{j1}^{ab}, \cdots, w_{jN_b}^{ab})$ denotes the weight vector from the jth F_2^a node to F^{ab}. Components of vectors $\mathbf{x}^a$, $\mathbf{y}^a$, and $\mathbf{x}^{ab}$ are reset to zero between input presentations. Initially, each weight is set equal to one. Note, that $|\mathbf{A}| = M_a$ and $|\mathbf{B}| = M_b$ for all input vectors $\mathbf{a}$ and $\mathbf{b}$.

Map Field Activation: Map field F^{ab} is activated when one of the ART_a or ART_b categories becomes active. When the Jth F_2^a node is chosen, $F_a^2 \rightarrow F^{ab}$ input is proportional to the weight vector $\mathbf{w}_J^{ab}$. When the Kth F_2^b node is chosen, the F^{ab} node K is activated by one-to-one pathways between F_2^b and F^{ab}. If both ART_a and ART_b are active, as in supervised learning, then F^{ab} activity reflects the degree to which a correct prediction has been made. With fast learning, F^{ab} remains active only if ART_a predicts the same category as ART_b, via the weight vector $\mathbf{w}_J^{ab}$, or if the chosen ART_a category J has not yet learned an ART_b prediction. In summary, the F^{ab} output vector $\mathbf{x}^{ab}$ obeys

$$\mathbf{x}^{ab} = \begin{cases} \mathbf{y}^b \wedge \mathbf{w}_J^{ab} & \text{if the } J\text{th } F_2^a \text{ node is active and } F_2^b \text{ is active} \\ \mathbf{w}_J^{ab} & \text{if the } J\text{th } F_2^a \text{ node is active and } F_2^b \text{ is inactive} \\ \mathbf{y}^b & \text{if } F_2^a \text{ is inactive and } F_2^b \text{ is active} \\ \mathbf{0} & \text{if } F_2^a \text{ is inactive and } F_2^b \text{ is inactive.} \end{cases} \tag{12}$$

If the prediction $\mathbf{w}_J^{ab}$ is disconfirmed by $\mathbf{y}^b$, this mismatch event triggers an ART_a search for a new category, as follows.

Match Tracking: At the start of each input presentation the ART_a vigilance parameter ρ_a equals a baseline vigilance, $\bar{\rho}_a$. The map field vigilance parameter is ρ_{ab}. Match tracking is triggered by a mismatch at the map field F^{ab}, that is, if

$$|\mathbf{x}^{ab}| < \rho_{ab}|\mathbf{y}^b| = \rho_{ab} \tag{13}$$

as in (10). Match tracking increases ρ_a until it is slightly larger than the ART_a match value, $|\mathbf{A} \wedge \mathbf{w}_J^a||\mathbf{A}|^{-1}$, where $\mathbf{A}$ is the input to F_1^a and J is the index of the active F_2^a node. After match tracking, therefore

$$|\mathbf{x}^a| = |\mathbf{A} \wedge \mathbf{w}_J^a| < \rho_a|\mathbf{A}| = \rho_a M_a. \tag{14}$$

When this occurs, ART_a search leads either to ARTMAP resonance, where a newly chosen F_2^a node J satisfies both the ART_a matching criterion

$$|\mathbf{x}^a| = |\mathbf{A} \wedge \mathbf{w}_J^a| \geq \rho_a|\mathbf{A}| \tag{15}$$

and the map field matching criterion

$$|\mathbf{x}^{ab}| = |\mathbf{y}^b \wedge \mathbf{w}_J^{ab}| \geq \rho_{ab}|\mathbf{y}^b| = \rho_{ab} \tag{16}$$

or, if no such F_2^a node exists, to the shutdown of F_2^a for the remainder of the input presentation. Since $w_{ij}^a(0) = w_{jk}^{ab}(0) = 1$ and $0 \leq \rho_a, \rho_{ab} \leq 1$, ARTMAP resonance always occurs if J is an uncommitted node.

Map Field Learning: A learning rule similar to (11) determines how the map field weights w_{jk}^{ab} change through time, as follows. Weights w_{jk}^{ab} in $F_2^a \rightarrow F^{ab}$ paths initially satisfy

$$w_{jk}^{ab}(0) = 1. \tag{17}$$

During resonance with the ART_a category J active, $\mathbf{w}_J^{ab}$ approaches the map field vector $\mathbf{x}^{ab}$. With fast learning, once J learns to predict an ART_b category K, that association is permanent; i.e., $w_{JK}^{ab} = 1$ and $w_{Jk}^{ab} = 0 (k \neq K)$ for all time.

C. ART-EMAP

ART-EMAP augments ARTMAP to improve recognition of a target object in noisy or ambiguous situations by integrating information across space and time. During performance, F_2^a activity that is spatially distributed across several coded ART_a categories helps disambiguate a noisy view or input. ART-EMAP can also perform temporal integration of information across multiple views until a target object is recognized with confidence. Multiple inputs might correspond to noisy images from a fixed perspective or to ambiguous images from various perspectives. Expansion of the map field into a multi-field EMAP module (Fig. 3) enables the accumulation of evidence from distributed patterns of activation across a series of input patterns. This capability allows ART-EMAP to make an accurate class prediction from a sequence of individually ambiguous input patterns, as follows.

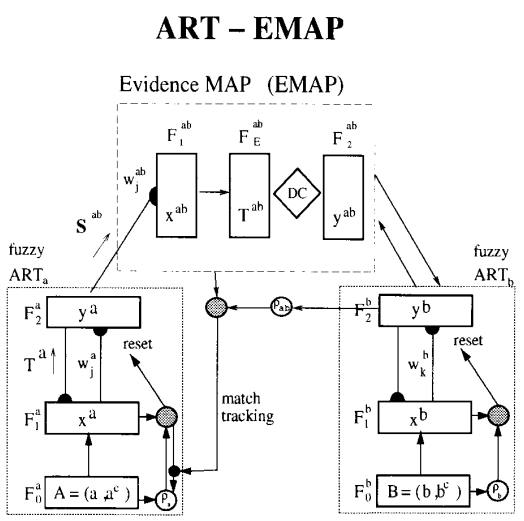


Fig. 3. ART-EMAP architecture. The ARTMAP map field F^{ab} is replaced with a multi-field EMAP module. During testing, a distributed F_2^a output pattern $\mathbf{y}^a$, resulting from partial contrast enhancement of $F_1^a \rightarrow F_2^a$ input $\mathbf{T}^a$, is filtered through EMAP weights w_{jk}^{ab} to determine the F_1^{ab} activity $\mathbf{x}^{ab}$. If a predictive decision criterion is not met, additional input can be sought, until the decision criterion is met at F_2^{ab}.

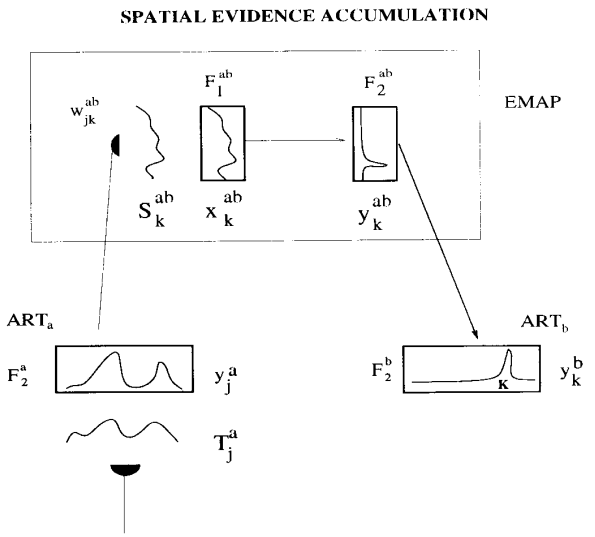


Fig. 4. Partial contrast enhancement of $F_1^a \rightarrow F_2^a$ input $\mathbf{T}^a$ results in a distributed F_2^a output pattern $\mathbf{y}^a$. Filtered through EMAP weights w_{jk}^{ab}, $\mathbf{y}^a$ determines a distributed F_1^{ab} pattern $\mathbf{x}^{ab} = \mathbf{S}^{ab}$. Choice at the EMAP field F_2^{ab} predicts a pattern class K at ART$_b$.

III. Stage 1: Spatial Evidence Accumulation

ART-EMAP employs a spatial evidence accumulation process that integrates a distributed pattern of activity across coded category nodes to help disambiguate a noisy or novel input. In contrast, previous ART [6]–[8] and ARTMAP [1], [5] simulations chose only the most highly activated category node at the field F_2^a (Fig. 1) as the basis for recognition and prediction.

In the fast-learn fuzzy ARTMAP system, the input from F_1^a to the jth F_2^a node is given by

$$T_j^a = \frac{|\mathbf{A} \wedge \mathbf{w}_j^a|}{\alpha + |\mathbf{w}_j^a|} \tag{18}$$

as in (2). Fuzzy ARTMAP uses a binary choice rule (5), so only the F_2^a category J that receives maximal $F_1^a \rightarrow F_2^a$ input T_j^a delivers output to F^{ab}

$$y_J^a = \begin{cases} 1 & \text{if } T_J^a > T_j^a \quad \text{for all } j \neq J \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

ART-EMAP also uses binary choice rule (19) during the initial period of supervised training. During performance, however, F_2^a output $\mathbf{y}^a$ is determined by less extreme contrast enhancement of the $F_1^a \rightarrow F_2^a$ input pattern $\mathbf{T}^a$ (Fig. 4). Limited contrast enhancement extracts more information from the relative activations of F_2^a categories than does the all-or-none choice rule (19). To make consistent comparisons across ART-EMAP Stages 1–4, the algorithm that determines F_2^a activation $\mathbf{y}^a$ was held fixed across all simulations. The following discussion shows how this algorithm was selected. Parameter selection is also shown to be robust under wide perturbations. This procedure typifies all parameter selection methods: systems are specified completely, so that simulations may be replicated, but parameter selection is also determined to be robust throughout.

A. Contrast Enhancement Algorithm

Two algorithms that approximate partial contrast enhancement for spatial evidence accumulation at F_2^a are a power rule and a threshold rule as outlined below.

Power Rule: Raising the input T_j^a of the jth F_2^a category to a power $p > 1$ is a simple way to implement contrast enhancement. Equation (20) defines a normalized power rule

$$y_j^a = \frac{(T_j^a)^p}{\sum_{n=1}^{N_a} (T_n^a)^p}. \tag{20}$$

Normalization constrains the F_2^a output values to a manageable range without altering relative values or subsequent predictions. Power rule (20) approximates the dynamics of a shunting competitive short-term-memory (STM) network that contrast-enhances its input pattern [14]. The power rule is equivalent to choice rule (19) when p is large. For smaller p, distributed activity pattern (20) uses information from the relative F_2^a category activations to improve test set predictive performance at ART$_b$.

Threshold Rule: An alternative contrast enhancement algorithm introduces a threshold T

$$y_j^a = \frac{[T_j^a - T]^+}{\sum_{n=1}^{N_a} [T_n^a - T]^+} \tag{21}$$

where $[x]^+ \equiv \max\{x, 0\}$.

After contrast enhancement by either (20) or (21), the F_2^a output $\mathbf{y}^a$ is filtered through the weights w_{jk}^{ab} to activate the

EMAP field F_1^{ab}. The input S_k^{ab} from F_2^a to the kth F_1^{ab} node obeys the equation

$$S_k^{ab} = \sum_{j=1}^{N_a} w_{jk}^{ab} y_j^a. \tag{22}$$

Since distributed F_2^a activity generally determines distributed EMAP field F_1^{ab} input, some means of choosing a winning prediction is required. The simplest method is to choose the EMAP category K that receives maximal input from F_2^a. This can be implemented by letting $x_k^{ab} = S_k^{ab}$ and defining F_2^{ab} activity by

$$y_K^{ab} = \begin{cases} 1 & \text{if } x_K^{ab} > x_k^{ab} \quad \text{for all } k \neq K \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

Other methods for predicting an ART_b category will be discussed below.

ART-EMAP, using either power rule (20) or threshold rule (21) for F_2^a output, was evaluated on the circle-in-the-square benchmark (Section I). To select a single ART-EMAP system for further development, parametric comparisons were made between a fuzzy ARTMAP system with a choice rule at F_2^a and ART-EMAP systems with power or threshold rules at F_2^a. Each system was trained on the 100-sample circle-in-the-square training set during a single supervised learning epoch, and then tested on 900 other randomly chosen points. All three systems are functionally equivalent in the supervised learning phase. After a single epoch, each performed at 98% accuracy on the 100-sample training set. Fuzzy ARTMAP, with choice at F_2^a, had a 92.9% correct prediction rate on the test set. ART-EMAP with power rule (20) had a significantly improved 94%–96% correct prediction rate for p between five and 35 [Fig. 5(a)]. A power rule with p set equal to 10 at F_2^a was thus selected and fixed for each subsequent circle-in-the-square simulation.

Threshold rule (21) did not show comparable performance improvements [Fig. 5(b)]. It may still be useful for certain problems, as it is computationally simpler than the power rule, but is not used here.

B. Stage 1 Simulations

Fig. 6 shows circle-in-the-square simulation response plots for fuzzy ARTMAP, with the choice rule at F_2^a and for Stage 1 ART-EMAP, with spatial evidence accumulation at F_2^a. On both noise-free and noisy test sets, ART-EMAP significantly outperforms fuzzy ARTMAP. Each system underwent supervised learning on the standard 100-point training set, each input being presented once, always in the same order (Section I). Each response plot shows a system's predictions on the standard 11 000-point test set. A response plot shows the circle and all point predicted by the given system to be "outside the circle."

Response plots in Fig. 6(a)–(b) show fuzzy ARTMAP and Stage 1 ART-EMAP decision boundaries for a noise-free test set, in which each test set point was accurately labeled as being inside or outside the circle. Response plots in Fig. 6(d)–(e) show performance of the same systems with test set inputs degraded by Gaussian noise. Noisy test set inputs were generated

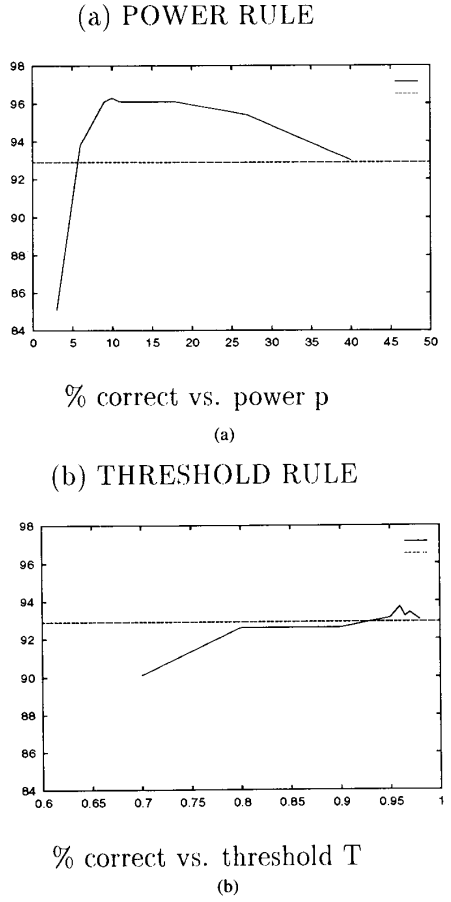


Fig. 5. Evaluation of F_2^a contrast enhancing rules power rule (a) and threshold rule (b) on the circle-in-the-square problem with 100 training points and 900 test points. Fuzzy ARTMAP choice rule performance of 92.9% is plotted for comparison. In all subsequent simulations, ART-EMAP uses a power rule (20) with $p = 10$ at F_2^a.

by adding random noise to each ART_a input point before the system made its in/out prediction. If a point a was inside the circle, the actual input to the ART_a system would be another point selected at random according to a Gaussian distribution centered at a, with standard deviation 0.1 (Fig. 7). Thus, many points near the boundary of the circle were actually mislabeled during testing. As expected, errors are most frequent near the boundary of the circle. Spatially distributed F_2^a activation is seen to correct some of these errors, allowing the system to accumulate evidence across several categories when the decision is ambiguous.

IV. Stage 2: EMAP Predictive Decision Criterion

An alternative to Stage 1 predictive choice rule (23) uses a decision criterion (DC) at the field F_2^{ab}. The decision criterion may permit ART_b choice only when the most active EMAP category K becomes a minimum proportion more active than the next most active EMAP category. Thus

$$y_K^{ab} = \begin{cases} 1 & \text{if } x_K^{ab} > (\text{DC}) x_k^{ab} \quad \text{for all } k \neq K \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

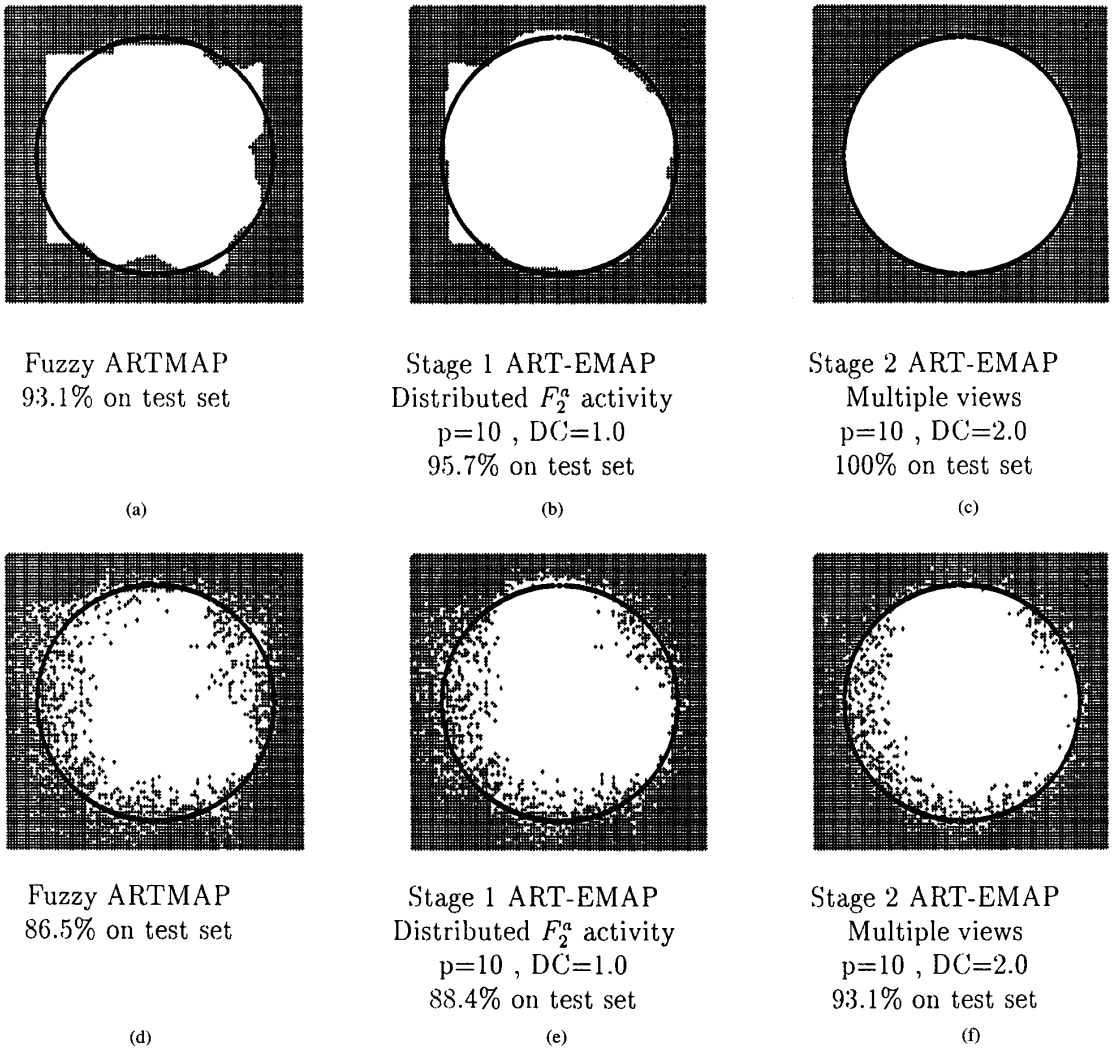


Fig. 6. Response plot decision boundaries and performance accuracy for 100/11 000 training/test exemplars. Plotted points are those predicted to be outside the circle. Plots (a), (b), and (c) show decision boundaries in a noise-free test environment. Plots (d), (e), and (f) show performance with Gaussian noise (SD = 0.1) added to each test set input vector **a**. Plots (a) and (d) show fuzzy ARTMAP performance, using the choice rule. Plots (b) and (e) show Stage 1 ART-EMAP performance, using the power rule with $p = 10$. Plots (c) and (f) show Stage 2 ART-EMAP performance using the power rule with $p = 10$, an EMAP decision criterion DC = 2.0, and multiple views.

where DC ≥ 1 (Fig. 8). When DC = 1, Stage 2 decision criterion rule (24) reduces to Stage 1 F_2^{ab} choice rule (23). When DC > 1, the decision criterion prevents prediction in cases in which multiple EMAP categories are about equally activated at F_1^{ab}, representing ambiguous predictive evidence. As the DC increases, both accuracy and the number of required input samples per decision tend to increase. For computational convenience, activity at F_1^{ab} can also be contrast enhanced by a normalized power rule

$$x_k^{ab} = \frac{(S_k^{ab})^q}{\sum_{n=1}^{N_b}(S_n^{ab})^q}. \tag{25}$$

Setting $q = 3$ in (25) makes performance less sensitive to the DC value than in the case $q = 1$ (no contrast enhancement at F_1^{ab}). The value of q does not change system function.

When the decision criterion fails and (24) implies that $y_k^{ab} = 0$ for all k, additional input is sought to resolve the perceived ambiguity. In an application, additional input might be obtained from multiple views or from multiple samples of a single view, as illustrated in the following simulations.

A. Stage 2 Simulations

Multiple Views: If inputs are noise free, only additional object views (or their analog) can provide more information when the decision criterion fails [Fig. 9(a)]. In the case of a stationary object, this could correspond to moving a sensor to obtain a new perspective. In the case of a moving object, a subsequent reading from a fixed sensor could be used. To emulate a multiple views strategy in circle-in-the-square

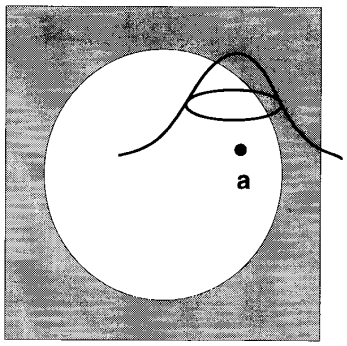


Fig. 7. Gaussian noise contamination of the circle-in-the-square noisy test set inputs. ART-EMAP spatial and temporal evidence accumulation helps compensate for input noise.

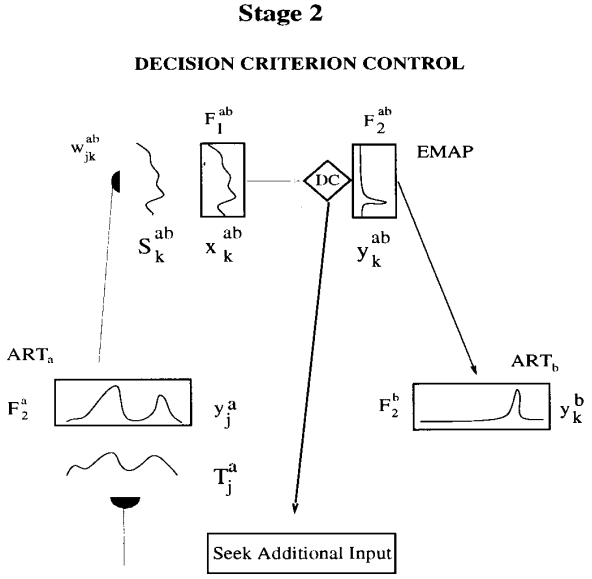


Fig. 8. Stage 2 ART-EMAP: If predictive evidence is ambiguous, according to a decision criterion, then additional ART$_a$ input is sought. If the decision criterion is satisfied, choice at the EMAP field F_2^{ab} predicts an ART$_b$ pattern class.

simulations, a new ART$_a$ input would be presented when decision criterion (24) failed during testing. This new input would predict the same ART$_b$ class as the previous input, but the identity of that class would be unknown to the system. An in/out class prediction would be made only when one input was found that met the decision criterion. For DC = 2.0, this strategy dramatically improved performance, compared to Stage 1, where DC = 1.0. In fact, Fig. 6(c) demonstrates perfect test set prediction on the noise-free test set, and Fig. 6(f) shows significantly improved performance on the noisy test set.

Multiple Samples: If a test set input is contaminated by noise, additional samples of the same view, or view analogue, can provide new information [Fig. 9(b)]. This might correspond to seeing a single object view under variable conditions, such as occlusion by fog or by other objects. For circle-in-the-square Stage 2 simulations with DC = 2.0, multiple samples showed improved performance over Stage 1 single samples (DC = 1.0), but the improvement is not as great as in the multiple views case. On the other hand, multiple samples may be available at times when multiple views are not.

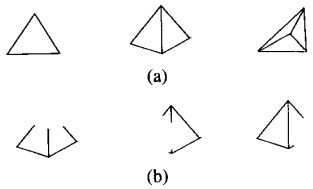


Fig. 9. Types of additional input, when the EMAP decision criterion fails. (a) Multiple views represent different sensor perspective on the same object. (b) Multiple samples represent different sensor inputs measured from a single perspective.

Both the multiple views and the multiple samples strategies for disambiguating information can be simulated in a variety of database examples. During testing, multiple views require presentation of data known to belong to the same as yet unidentified class. Multiple samples can be simulated by adding noise to the database, producing multiple versions of each input. Stage 2 requires a certain amount of *a priori* knowledge of the structure of database test set input patterns. In pattern recognition applications this requirement does not unusually constrain the method: multiple samples can be obtained in most sensor domains, and multiple views can be obtained whenever multiple perspectives or multiple sensors are present. The only constraint is that the still-unknown identity of the target itself remain constant across multiple views or samples.

V. Stage 3: Temporal Evidence Accumulation

The predictive decision criterion strategy (Stage 2 ART-EMAP) searches multiple views or samples until one input satisfies the decision criterion. A single noisy input vector **a**, however, might produce map field activity that satisfies a given decision criterion but still makes an incorrect prediction. This strategy makes no use of the partial evidence provided by the series of views that failed to meet the decision criterion. Further performance improvement in a noisy input environment is achieved through the application of a decision criterion to time-integrated predictions that are generated by multiple inputs. Stage 3 ART-EMAP accumulates evidence at a map evidence accumulation field F_E^{ab} (Fig. 10). The time scale of this medium-term memory (MTM) process is longer than that of the short-term memory (STM) field activations resulting from the presence of a single view, but shorter than that of the LTM stored in adaptive weights.

Additive Evidence Integration: A simple way to implement evidence accumulation at the EMAP module is to sum a sequence of F_1^{ab} map activations at the evidence accumulation field F_E^{ab}

$$\left(T_k^{ab}\right)^{(\text{new})} = \left(T_k^{ab}\right)^{(\text{old})} + x_k^{ab}. \tag{26}$$

At F_E^{ab}, evidence accumulating MTM (T_k^{ab}) starts at zero and is reset to zero when the DC is met. Activities y_k^{ab} at field

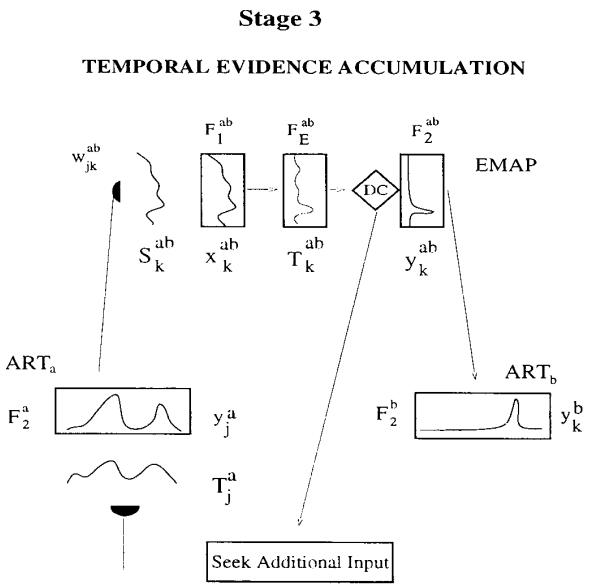


Fig. 10. Stage 3 ART-EMAP evidence accumulation. F_1^{ab} patterns $\mathbf{x}^{ab}$ are integrated at field F_E^{ab}. If the accumulated predictive evidence $\mathbf{T}^{ab}$ is ambiguous according to a decision criterion, additional input is sought. If the decision criterion is satisfied, choice at F_2^{ab} predicts an ART$_b$ pattern class K.

F_2^{ab} obey

$$y_K^{ab} = \begin{cases} 1 & \text{if } T_K^{ab} > (\mathrm{DC}) T_k^{ab} \quad \text{for all } k \neq K \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

A decision will eventually be made if the DC starts large and gradually decreases toward one. As in Stage 2, larger DC values tend to covary with both greater accuracy and longer input sequences.

A. Stage 3 Simulations

In simulations, the DC decreased exponentially from 6.0 to 1.0

$$\mathrm{DC}(l) = 5.0(1.0 - r)^{l-1} + 1.0 \tag{28}$$

where $\mathbf{a}(l)$ is the lth input in a same-class sequence ($l = 1, 2, \cdots$). The decay rate (r) was set equal to 0.2. Additive integration is equivalent to applying the decision criterion to a running average of map field activations $\mathbf{x}^{ab}$ rather than to $\mathbf{x}^{ab}$ itself, as in Stage 2 (24). This averaging process tends to factor out random noise over multiple inputs.

In the circle-in-the-square problem, an average of 3.1 noisy samples of a given input a were needed to exceed the decision criterion, using Stage 3 ART-EMAP with DC function (28) and additive integration rule (26). Fig. 11 shows response plots for the circle-in-the-square problem with multiple view [Fig. 11(b)] and multiple sample [Fig. 11(c)] strategies for noisy test set inputs. Comparing Fig. 11(b) (Stage 3) with Fig. 6(f) (Stage 2), which uses a fixed DC =2.0, shows that evidence accumulation can significantly reduce test set errors in a noisy environment; in this case the error rate is reduced from 6.9% to 2.4%.

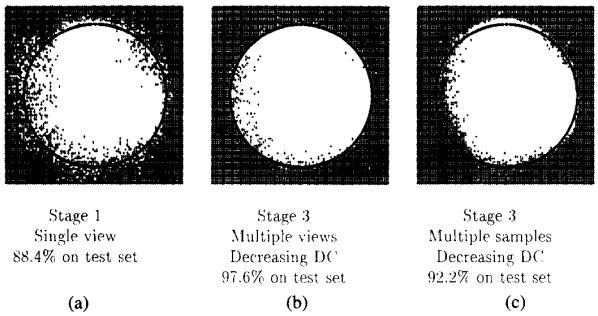


Fig. 11. ART-EMAP response plots for 11 000 noisy test set points. In each simulation, F_2^a activity is distributed, with $p = 10$. (a) Stage 1: Single view with DC = 1.0. (Fig. 6(e) shows the same simulation.) (b) Stage 2: Multiple views. (c) Stage 3: Multiple samples. In (b) and (c), the DC decreases from 6.0 toward 1.0 by (28), as evidence accumulates.

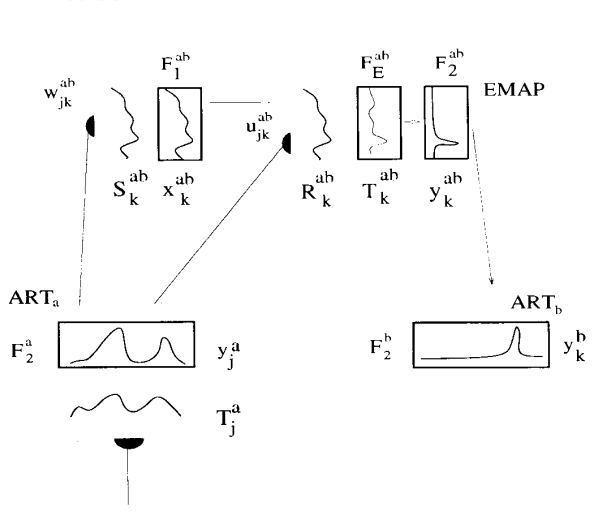


Fig. 12. Stage 4 ART-EMAP, with unsupervised rehearsal learning. During rehearsal, after ART-EMAP determines object identity from a sequence of input patterns, weights u_{jk}^{ab} are trained to improve future performance on individual input patterns of the sequence.

VI. Stage 4: Unsupervised Rehearsal Learning

Temporal evidence accumulation allows the Stage 3 ART-EMAP system to recognize objects from a series of ambiguous views. The system learns nothing from the final outcome of this decision process, however. If, for example, an input sequence $\mathbf{a}^{(1)}, \cdots, \mathbf{a}^{(L)}$ predicts an ART$_b$ category K, by (26)–(27), the entire sequence would need to be presented again before the same prediction would be made.

Unsupervised rehearsal learning (Stage 4) fine-tunes performance by feeding back to the system knowledge of the final prediction. Specifically, after input $\mathbf{a}^{(L)}$ allows ART-EMAP to choose the ART$_b$ category K, the sequence $\mathbf{a}^{(1)}, \cdots, \mathbf{a}^{(L)}$ is re-presented, or rehearsed. Weights in an adaptive filter u_{jk}^{ab} from F_2^a to F_E^{ab} (Fig. 12) are then adjusted, shifting category decision boundaries so that each input $\mathbf{a}^{(l)}$ in the sequence

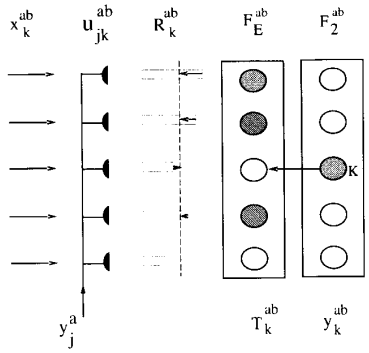


Fig. 13. Stage 4: Unsupervised rehearsal learning during presentation of input $\mathbf{a}^{(l)}$, part of a sequence that led to prediction of ART_b category K. Selected category K, clamped on at F_2^{ab}, acts as a self-generated training input to F_E^{ab}. Weights u_{jK}^{ab} increase, and weights u_{jk}^{ab} to one or more F_E^{ab} nodes with maximal input R_k^{ab} decrease. Learning ceases when $\mathbf{a}^{(l)}$ would predict K on its own, with $\mathbf{DC} = 1.0$. When weights stop changing, $R_K^{ab} = \max\{R_k^{ab}\}$ for as long as $\mathbf{a}^{(l)}$ is active.

becomes more likely, on its own, to predict category K. That is, at F_E^{ab}

$$\left(T_k^{ab}\right)^{(\text{new})} = \left(T_k^{ab}\right)^{(\text{old})} + R_k^{ab} \tag{29}$$

where

$$R_k^{ab} = x_k^{ab} + \sum_{j=1}^{N_a} y_j^a u_{jk}^{ab}. \tag{30}$$

Weights $\mathbf{w}_j^a$, $\mathbf{w}_k^b$, and $\mathbf{w}_j^{ab}$, trained during the supervised learning interval, remain constant during unsupervised rehearsal learning.

Weights u_{jk}^{ab} in the $F_2^a \rightarrow F_E^{ab}$ adaptive filter are updated as follows. After decision criterion (27) is met and the ART_b category K is predicted, the sequence $\mathbf{a}^{(1)}, \cdots, \mathbf{a}^{(L)}$ is re-presented to ART_a. This time, however, activity F_2^{ab} is clamped at prediction K

$$y_k^{ab} = \begin{cases} 1 & \text{if } k = K \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

(Fig. 13.) During the rehearsal of each input pattern $\mathbf{a}^{(1)}, \cdots, \mathbf{a}^{(L)}$, activity $\mathbf{T}^{ab}$ at the field F_E^{ab} obeys a binary equation

$$T_k^{ab} = \begin{cases} 1 & \text{if } R_k^{ab} = \max\{R_{k'}^{ab}\} \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

During rehearsal learning, some weights u_{jk}^{ab} can decrease while others increase. Equation (32) can then allow more than one F_E^{ab} node to become active, as follows.

Weights u_{jk}^{ab} start at zero and adapt according to the equation

$$\frac{d}{dt} u_{jk}^{ab} = \left(y_k^{ab} - T_k^{ab}\right) y_j^a. \tag{33}$$

By (31)

$$y_k^{ab} - T_k^{ab} = \begin{cases} 1 & \text{if } k = K \text{ and } T_k^{ab} = 0 \\ -1 & \text{if } k \neq K \text{ and } T_k^{ab} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

By (27) and (30)–(34), each weight u_{jK}^{ab} to the correct F_E^{ab} category node K increases at a rate proportional to F_2^a activity y_j^a just until the active input $\mathbf{a}^{(l)}$ would have predicted K on its own, with $\text{DC} = 1.0$. Thereafter, $T_K^{ab} = 1$ and weights u_{jK}^{ab} remain constant until the next input is presented. At the same time, each weight u_{jk}^{ab} to an incorrectly predicted F_E^{ab} node, where $T_k^{ab} = 1$ and $k \neq K$, decreases (Fig. 13). Rehearsal learning ceases when total input R_k^{ab} to each such node becomes slightly less than R_K^{ab}. Note that, as weights u_{jk}^{ab} decrease, an increasing number of F_E^{ab} nodes may come to receive maximal total input R_k^{ab}, by (32)–(34). Weights to these nodes then also begin to decrease, until K becomes the system prediction for input $\mathbf{a}^{(l)}$.

By (33), the rate of adaptation for each weight u_{jk}^{ab} is proportional to y_j^a. Thus, more active portions of the distributed F_2^a pattern $\mathbf{y}^a$ are assigned more credit as predictors of the boundary shift. Parametric simulations show that this credit-assignment strategy eliminates most unstable learning and wide oscillations in $\mathbf{u}_j^{ab}$. Learning at $\mathbf{u}_j^{ab}$ shifts the EMAP decision boundary just enough to include each input pattern in the predicted class K. Basic prediction rules, embodied in the ART_a, ART_b, and $F_2^a \rightarrow F_1^{ab}$ weights, are held fixed following the supervised learning interval. Unsupervised rehearsal learning allows the system to learn from its own experience "in the field," adapting to local statistics or correcting errors due to gaps in the training set. This type of adaptation can improve performance marginally, while basic system predictions remain stable. The following Stage 4 simulations show that this boundary shift improves test performance on single input patterns when $\text{DC} = 1.0$ (Stage 1).

A. Stage 4 Simulations

Fig. 14 shows Stage 4 ART-EMAP test set response patterns for both noise-free and noisy test sets. Each simulation has an initial supervised training on 100 inputs, and each uses distributed F_2^a activity during performance, with $p = 10$. As in Fig. 6, no evidence is accumulated during performance, except during unsupervised rehearsal learning. Test set inputs are noise-free in Fig. 14(a)–(c) and noisy in Fig. 14(d)–(f). In Fig. 14(b) and (e), rehearsal learning adapted weights u_{jk}^{ab} after temporal evidence accumulation resulted in identification of 50 randomly chosen test sequences. On average 3.1 inputs were necessary to exceed decision criteria (27)–(28). Note the small shift in the decision boundary from Fig. 14(a)–(b). Similarly, Fig. 14(c) and Fig. 14(f) show that rehearsal learning on 900 randomly chosen test sequences can further improve performance on the remaining test set.

VII. 3-D Object Recognition Simulations

Circle-in-the-square simulations illustrate incremental performance improvements, from fuzzy ARTMAP to ART-EMAP Stages 1–4, on a benchmark problem. ART-EMAP performance in a more realistic setting will now be demonstrated. Simulations of a 3-D object recognition problem take as inputs 2-D image views of a geometric object, observed from various angles. During performance a system, presented with a previously unseen view, is required to identify an object as

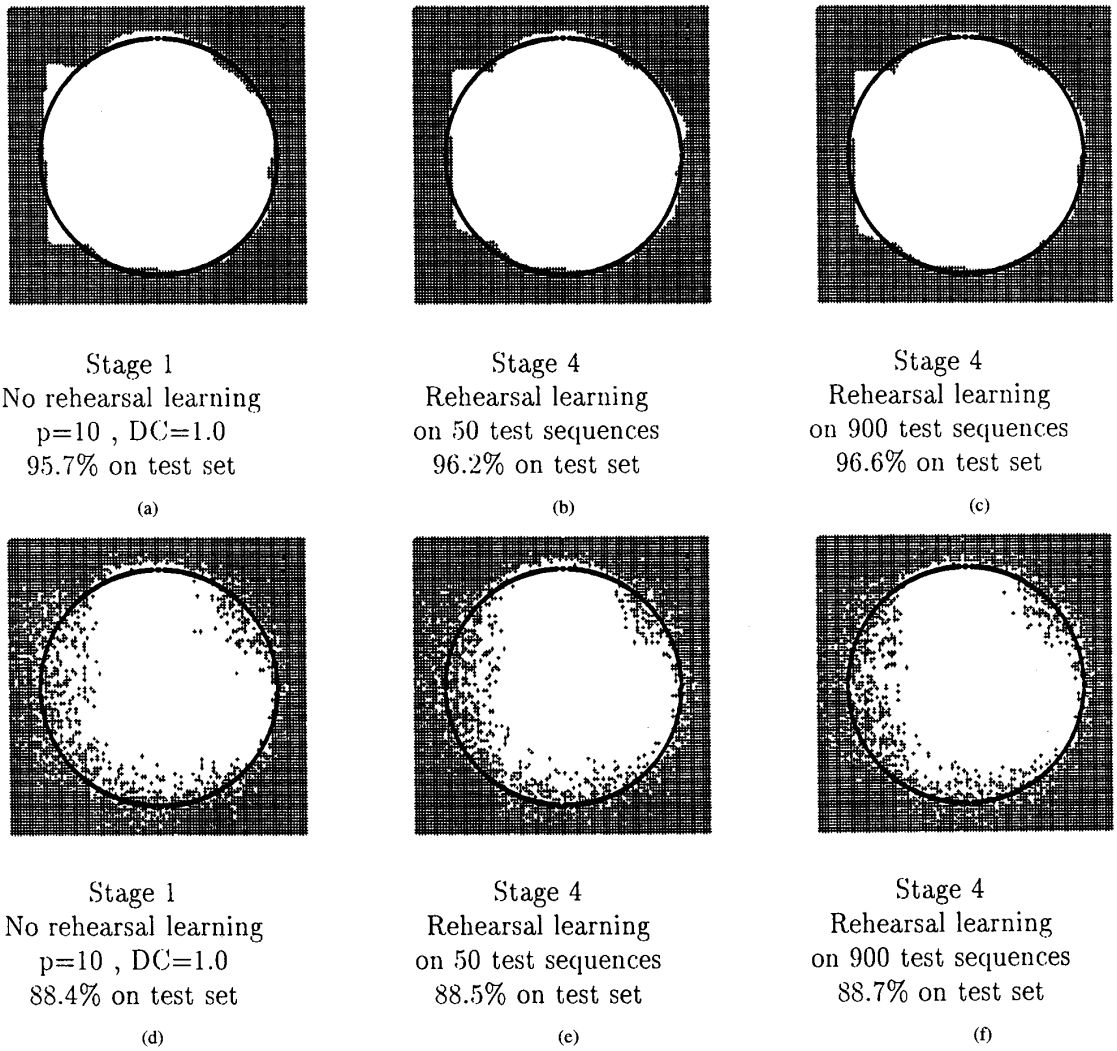


Fig. 14. Each plot shows ART-EMAP performance from 100 training and 11 000 test exemplars. During testing, F_2^a activity is distributed, with $p = 10$ and DC = 1.0. Plots (a), (b), and (c) show decision boundaries for the noise-free test set. Plots (d), (e),and (f) show performance for the noisy test set. (a) No unsupervised rehearsal learning. (Fig. 6(b) shows the same simulation.) (b) Unsupervised rehearsal learning on 50 randomly chosen test set sequences. (c) Unsupervised rehearsal learning on 900 randomly chosen test set sequences. (d) No unsupervised rehearsal learning. [Fig. 6(e) shows the same simulation.] (e) Unsupervised rehearsal learning on 50 randomly chosen test set sequences. (f) Unsupervised rehearsal learning on 900 randomly chosen test set sequences.

a pyramid, a prism, or a house (Fig. 15). The problem is made difficult by the similarity of several views across objects and by the fact that several test set views do not resemble any training set views of the same object.

Database Inputs: The simulation database was constructed using Mathematica to generate shaded 2-D projections of 3-D objects illuminated by an achromatic point light source. For each of the three objects, 24 training set views were obtained from perspectives spaced 30–60 degrees apart around a viewing hemisphere [Fig. 15(a)]. For each object, 17 test set views, spaced at 45-degree intervals, were obtained from perspectives between those of the training set [Fig. 15(b)]. Each 2-D view was then preprocessed, using Gabor filters [15], [16] to recover boundaries, competitive interactions to sharpen boundary locations and orientations [17], and coarse coding, yielding a 100-component input vector **a**. The preprocessing algorithm is a typical feature extractor, chosen to illustrate comparative performance of different recognition systems and was not selected to optimize performance of any one of these systems.

As in the circle-in-the-square simulations, fuzzy ARTMAP and ART-EMAP Stages 1–4 were evaluated using both a noise-free test set and a noisy test set. The noisy test set was constructed by adding Gaussian noise (SD = 0.2) to each input component. Each system was initially trained under one standard supervised learning regime, with the training set presented once. Since the training set views were deliberately selected to be sparse and nonredundant, a situation of minimal code compression was simulated during training. This was achieved by assigning a high value to the baseline vigilance

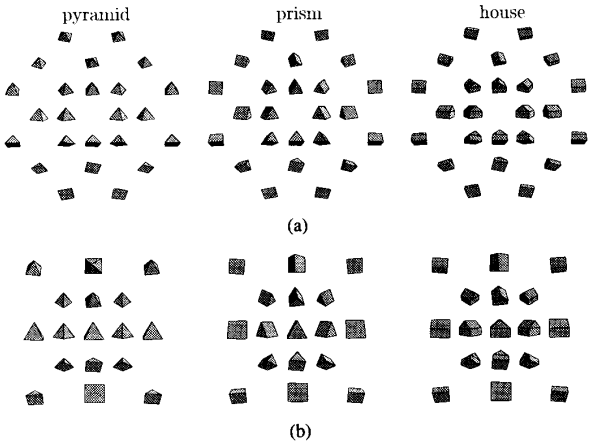


Fig. 15. 3-D object database images. (a) The training set consisted of 24 views spaced 30–60 degrees apart within the front viewing hemisphere of each object. The topmost training images for each ordered set are views taken from above the object, the bottommost from beneath the object, etc. (b) The test set was constructed of 17 views from within the front hemisphere, between the training set views, spaced 45 degrees apart.

($\bar{\rho}_a = 0.9$), which established 58 ART_a recognition categories for the 72 training set pairs.

FuzzyARTMAP: Performance measures of fuzzy ARTMAP and ART-EMAP on the 3-D object recognition database the summarized in Fig. 16, for noise-free test set inputs [plots (a)–(c)] and for noisy set inputs [plots (d)–(f)]. The prediction of each test set view is represented graphically, on shaded viewing hemispheres. Each hemisphere shows 17 faces, which correspond to the 17 test set viewing angles [Fig. 15(b)]. For each simulation, three hemispheres show object class predictions made by the system in response to the corresponding input, with shading of a face indicating a prediction of pyramid (black), prism (gray), or house (white).

Fuzzy ARTMAP made only 64.7% correct object class predictions on the noise-free test set and 60.8% correct predictions on the noisy test set. This poor performance indicates the difficult nature of the problem, when prediction must be made on the basis of a single view. Note, for example, that many of the test set inputs from the lower left part of the pyramid view hemisphere were incorrectly identified as prism views [Fig. 16(a)–16(d)]. The reason for these errors can be inferred by observing the close similarity between the corresponding pyramid and prism 2-D views in the test set [Fig. 15(b)].

Stage 1: Like fuzzy ARTMAP, Stage 1 ART-EMAP, with its spatially distributed activity pattern at F_2^a (Fig. 4), is required to make a prediction from each single test set view. Nevertheless, predictive accuracy is significantly improved, from 64.7% to 70.6% on the noise-free test set [Fig. 16(b)] and from 60.8% to 64.7% on the noisy test set [Fig. 16(e)]. As in circle-in-the-square simulations, power rule algorithm (20) was used to approximate contrast enhancement at F_2^a. With the input dimension $M_a = 2$ in the circle-in-the-square, a power rule with $p = 10$ gave adequate contrast enhancement. For the 3-D object recognition simulations, where $M_a = 100$, a higher power ($p = 24$) approximated contrast enhancement at F_2^a.

Stage 2: Stage 1 spatial evidence accumulation improves performance by causing a novel view to activate categories of two or more nearby training set views, which then strongly predict the correct object. Many single view errors, caused by similar views across different objects, however, remain. Most of these errors are corrected at Stage 2 or Stage 3, when multiple views of the unknown object are made available. With a high fixed decision criterion (DC = 5.0) and an average of 4.8 test set views, Stage 2 ART-EMAP achieves 98.0% accuracy on the noise-free test set. Even on the noisy test set, object identification remains at 90.2% accurate, with an average of 6.8 test set views. Table I shows how both performance and the average number of views decrease continuously as the fixed decision criterion decreases from five to one.

TABLE I
STAGE 2 ART-EMAP 3-D OBJECT RECOGNITION SIMULATION RESULTS AS THE FIXED DECISION CRITERIA (DC) DECREASE FROM FIVE TO ONE. WHEN DC = 1, STAGE 2 REDUCES TO STAGE 1

	Noise-free test set		Noisy test set	
DC	Percent	Average # views	Percent	Average # views
5	98.0	4.8	90.2	6.8
4	92.2	4.3	96.1	6.0
3	92.2	4.2	94.1	4.1
2	88.2	2.7	80.4	3.2
1	70.6	1.0	64.7	1.0

Stage 3: For a two-class prediction problem such as circle-in-the-square, evidence accumulation improves performance primarily by averaging across noisy inputs. Stage 3 ART-EMAP becomes increasingly useful as the number of predicted classes increases, since evidence accumulation can also help solve the difficult problem of disambiguating nearly identical views of different objects. With three or more object classes, when equal predictive evidence exists for both the correct object and an incorrect one, the identity of the erroneous class tends to vary. As the sequence of views grows, erroneous evidence is quickly overwhelmed by evidence for the correct object. In the Stage 3 ART-EMAP three-object simulations, with decreasing DC function (28), an average of 9.2 views were needed to reach 98.0% correct performance on the noise-free test set [Fig. 16(c)]. On the noisy test set, an average of 11.3 views allowed the system to reach 92.2% correct performance [Fig. 16(f)].

ART-EMAP performance is consistently robust across wide variations in parameters as shown, for example, in Fig. 5(a) and in Table I. This robustness is now also illustrated for Stage 3 noise-free simulations. Fig. 17 shows increasing performance accuracy as the DC curve shifts up, which causes the average number of required views to increase as well. In the limiting case, where DC =1.0, Stage 3 reduces to single-view Stage 1, where performance is just 70.6% [Fig. 16(b)]. The trade-off between the number of required views and criterion predictive accuracy can be adjusted for different applications.

Stage 4: Unsupervised rehearsal learning improves single view test set performance only marginally on the 3-D object simulations. Stage 4 rehearsal learning was conducted on the 51 noise-free test set views. Temporal evidence accumulation drew from an enlarged test set that included 72 additional views. Accessing exemplars from this larger test set allows stable fine-tuning by decreasing the percentage of ambiguous test views. After this fine-tuning, performance on individual

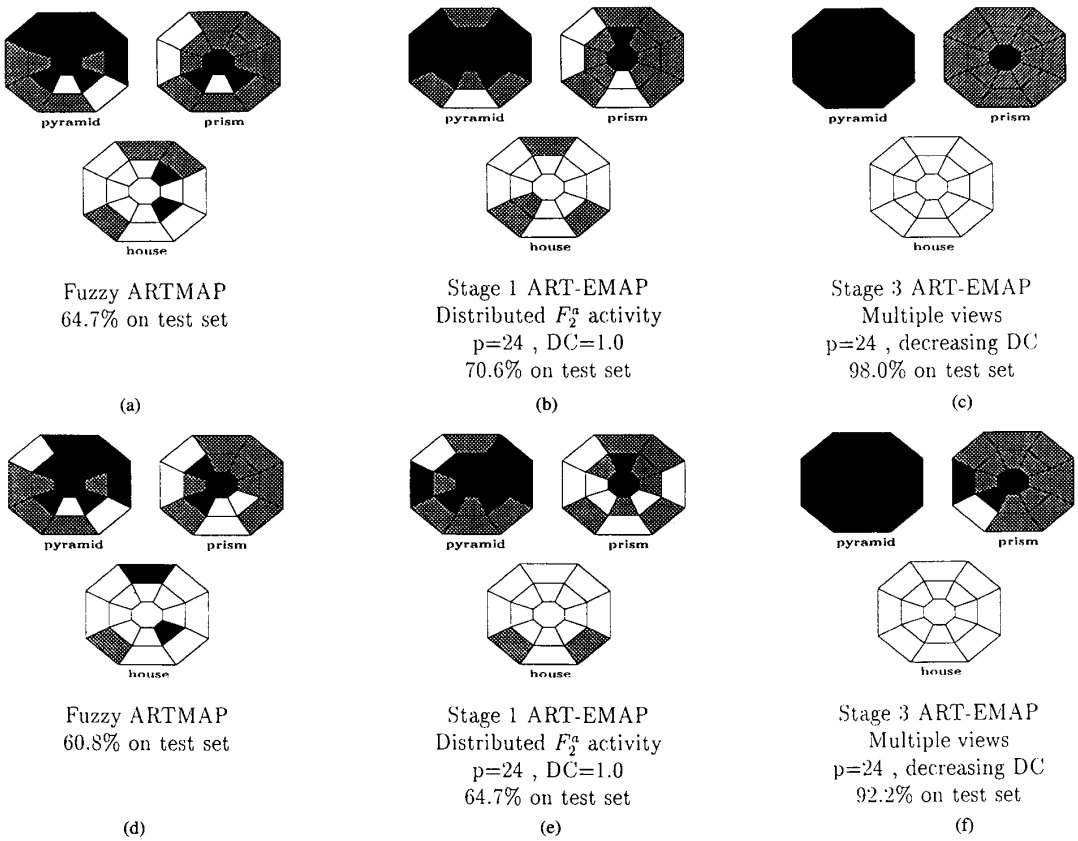


Fig. 16. 3-D object simulations: Response viewing hemispheres for each object show object predictions from each test set view. Each window in the hemisphere corresponds to one of the 17 test views [Fig. 15(b)]. Plots (a), (b), and (c) show noise-free test set results, and plots (d), (e), and (f) show noisy test set results. Plots (a) and (d) show fuzzy ARTMAP performance, using the F_2^a choice rule. Plots (b) and (e) show Stage 1 ART-EMAP performance using the power rule (20) with $p = 24$. Plots (c) and (f) show Stage 3 ART-EMAP performance using power rule (20) with $p = 24$, temporal evidence accumulation with the decreasing decision criterion (28) and multiple views.

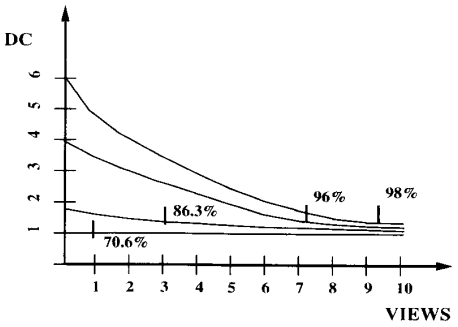


Fig. 17. Stage 3 ART-EMAP—Parametric comparison of system performance for four DC functions. On noise-free 3-D image simulations, test set performance is seen to degrade as the average number of views needed to exceed the DC decreases.

TABLE II
COMPARISON OF THE ASPECT NETWORK [2] AND ART-EMAP

	Aspect Network	ART-EMAP
Preprocessing	Boundary detection Coarse coding	Boundary detection Coarse coding
View categorization	ART 2	Fuzzy ART
Connectivity/ Complexity	second-order $N^2 \times M$	first-order $N \times M$
Learning	view transition to object	view to object
View field output (performance)	winner-take-all (competitive choice)	distributed (partial contrast enhancement)
Spatial evidence accumulation	none	across view categories
Predictive control	confidence threshold	decision criterion
Temporal evidence accumulation	across transitions	across views
Problem domain	structured input sequences	single inputs or input sequences

views from the original 51 test set inputs was 73%, compared to 70.6% at Stage 1 [Fig. 16(b)].

VIII. THE ASPECT NETWORK

Seibert and Waxman [2], [18], and [19] were the first to use temporal evidence accumulation in an ART-based network for 3-D object recognition. Aspect network learning strengthens a second-order weight when a transition between two view categories occurs (Fig. 18). As a 3-D object is examined, observed view transitions provide accumulating evidence leading to object identification. ART-EMAP uses different evidence accumulation processes (Table II). In particular, ART-EMAP spatial evidence accumulation allows the network to retain the first-order connectivity of fuzzy ARTMAP, rather than the

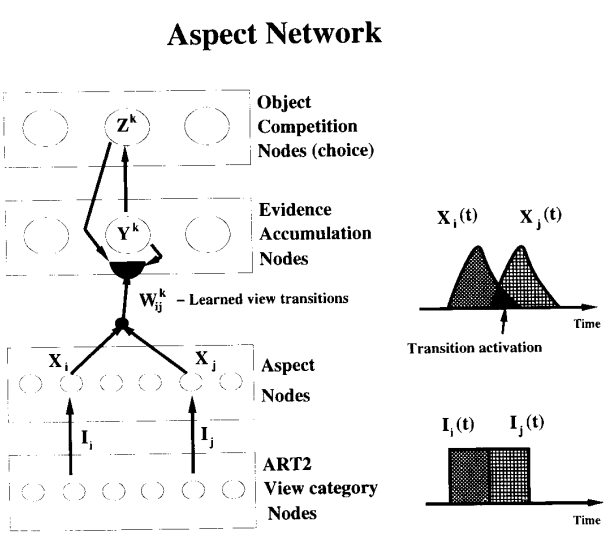


Fig. 18. The Seibert–Waxman aspect network [2]. ART 2 recognition (choice) of a view category i followed by category j causes transition activation of aspect nodes i and j. Evidence accumulates until the object network chooses a category k, which allows the second-order weights W_{ij}^k to be strengthened.

second-order learned view transition structure of the aspect network.

IX. Conclusion

Spatial and temporal evidence accumulation by ART-EMAP has been shown to improve fuzzy ARTMAP performance on both the circle-in-the-square benchmark and on a 3-D object recognition problem. Unsupervised rehearsal learning illustrates how self-training can fine-tune system performance. ART-EMAP is a general purpose algorithm for pattern class prediction based on the temporal integration of predictive evidence resulting from distributed recognition across a small set of trained categories. The system promises to be of use in a variety of applications, including recognition of 3-D objects from ambiguous 2-D views and spatio-temporal image analysis and prediction.

Acknowledgment

The authors wish to thank R. Locke for her valuable assistance in the preparation of the manuscript.

References

[1] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Networks*, vol. 3, pp. 698–713, 1992.

[2] M. Seibert and A. M. Waxman, "Learning aspect graph representations from view sequences," *Neural Inform. Process. Syst.*, in *Proc. NIPS Conf.*, 1990, pp. 258–265.

[3] A. A. Baloch and A. M. Waxman, "Visual learning, adaptive expectations, and behavioral conditioning of the mobile robot MAVIN," *Neural Networks*, vol. 4, pp. 271–302, 1991.

[4] G. Wilensky, "Analysis of neural network issues: Scaling, enhanced nodal processing, comparision with standard classification," *DARPA Neural Network Program Rev.*, Oct. 29–30, 1990.

[5] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Networks*, vol. 4, pp. 565–588, 1991.

[6] G. A. Carpenter and S. Grossberg, Eds., *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA: MIT Press, 1991.

[7] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Comput. Vision, Graphics, Image Process.*, vol. 37, pp. 54–115, 1987.

[8] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, pp. 759–771, 1991.

[9] B. Kosko, "Fuzzy entropy and conditioning," *Inform. Sci.*, vol. 40, pp. 165–174, 1986.

[10] L. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338–353, 1965.

[11] F. M. Ham and S. W. Han, "Quantitative study of the QRS complex using fuzzy ARTMAP and MIT/BIH arrythmia database," in *Proc. World Congr. Neural Networks*, vol. I, pp. 207–211, 1993.

[12] S. Keyvan, A. Durg, and L. C. Rabelo, "Application of artificial neural networks for development of diagnostic monitoring system in nuclear plants," *Trans. Amer. Nuclear Soc.*, vol. 1, pp. 515–522, 1993.

[13] B. V. Mehta, L. Vij, and L. C. Rabelo, "Prediction of secondary structures of proteins using fuzzy ARTMAP," in *Proc. World Congr. Neural Networks*, vol. I, 1993, pp. 228–232.

[14] S. Grossberg, "Contour enhancement, short term memory, and constancies in reverberating neural networks," *Studies Appl. Math.*, vol. LII, pp. 213–257, 1973.

[15] D. Gabor, "A theory of communication," *J. Inst. Elec. Eng.*, vol. 93, pp. 429–457, 1946.

[16] J. G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 1169–1179, 1988.

[17] S. Grossberg and E. Mingolla, "Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading," *Psych. Rev.*, vol. 92, pp. 173–211, 1985.

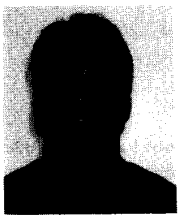
[18] M. Seibert and A. M. Waxman, "Adaptive 3-D object recognition from multiple views," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 107–124, 1992.

[19] A. M. Waxman, M. Seibert, A. M. Bernardon, and D. A. Fay, "Neural systems for automatic target learning and recognition," *Lincoln Lab. J.*, vol. 6, pp. 77–116, 1993.

Gail A. Carpenter (M'95) received the B.A. degree in mathematics from the University of Colorado, Boulder, in 1970 and the M.A. and Ph.D. degrees in mathematics from the University of Wisconsin, Madison, in 1972 and 1974, respectively.

She is Professor of Cognitive and Neural Systems (CNS) and Mathematics at Boston University and the CNS Director of Graduate Studies. Her research interests include the development, analysis, and application of neural models of nerve impulse generation (Hodgkin–Huxley equations), biological rhythms, vision, and distributed coding; adaptive resonance theory (ART) architectures for self-organizing category learning and pattern recognition; and systems that incorporate ART modules into neural network architectures for incremental supervised learning (ARTMAP and fuzzy ARTMAP).

Dr. Carpenter was the 1989 Vice President of the International Neural Network Society (INNS); organization chair of the 1988 INNS annual meeting; and a member of the editorial boards of *Brain Research*, IEEE Transactions on Neural Networks, *Neural Computation*, and *Neural Networks*.



William D. Ross received the B.S. degree in electrical engineering from Cornell University in 1988 and the M.A. and Ph.D. degrees from Boston University in 1993 and 1994, respectively.

He is currently a Postdoctoral Research Associate in the Boston University Department of Cognitive and Neural Systems and has worked at the Raytheon Corporation. His research interests include development and application of models of attentive visual search and object recognition.