

A 5-Chunk Developmental Brain-Mind Network Model for Multiple Events in Complex Backgrounds

Juyang Weng *Fellow, IEEE*

Abstract—There has been no prior general purpose brain-mind model for multiple events in complex backgrounds. I first discuss that although the age of brain-mind seems to have arrived, the current infrastructure does not well fit the need of research and peer-review for the challenging and important subject of brain-mind. Then, I present a general purpose model of the brain-mind, called the Epigenetic Developer (ED) network model. The model proposes five necessary “chunks” for the brain picture: development, architecture, area, space and time. The *development* chunk means that any practical brain, natural or artificial, needs to autonomously develop through interactions with the natural environments without any previously given set of tasks. The *architecture* chunk handles (1) multiple objects in complex backgrounds; (2) reasoning under abstract contexts; (3) multiple sensory modalities and multiple motor modalities and their integration. The *area* chunk addresses the issue of feature development and area representation, without rigidly specifying what each neuron does. The *space* chunk deals with spatial attention to individual objects in complex backgrounds, to satisfy the invariance and specificity criteria for type, location, and other concepts. The *time* chunk indicates that the brain uses its intrinsic spatial mechanisms to deal with time, without dedicated temporal components. The model copes with temporal contexts of events, to satisfy invariance and specificity criteria for time warping, time duration, temporal attention, and long temporal length. The theory and mechanisms are presented and some related experimental results are summarized.

I. INTRODUCTION

TWO types of models have been used to model cognitive architectures, symbolic and connectionist.

For the symbolic type, we assume that there is a one-to-one correspondence between every selected symbol and their meaning in a specified domain — each symbol has only one or a fixed set of meaning and each meaning has only one or a fixed set of symbol [20], [24], [26]. Because of the symbol use, the modelers use “skull-open” approaches — the holistically-aware central controller is the outside human designer. He defines each internal entity (e.g., a module) using a symbolic meaning, designs or trains each entity separately, and then manually links the modules.

For the connectionist type, we draw inspiration from brain-like emergent internal representation (e.g., [19], [9], [12], [36]). By emergent representation, we mean that the internal representation is not directly specified by symbolic meanings, but rather, firing patterns. Because the representation is indirectly related to symbolic meanings (e.g., the direction

of edge features), the network learning mechanisms are not specific to particular meanings either. This important property is useful for understanding development, as the set of learning mechanisms must work through different ages, through which the agent (natural or artificial) learns an open number of skills and performs an open number of tasks. Often, we use some symbolic meanings as examples (e.g., the direction of edge features for V1) to assist understanding of the functions of emergent representations, but such explanations are not necessarily precise.

As pointed out [20], connectionist approaches are bottom-up (e.g., from pixels) and symbolic approaches are top-down (from abstract concepts). Between the concrete (e.g., an edge or an edge grouping) and the abstract (e.g., goal), “much in-between” is missing.

In theory, the symbolic approaches may also link an abstract symbol with concrete sensory signals (e.g., design a “V1” module as one that generates symbolic edges). However, the resulting systems are brittle in dealing with real environments during its real applications. The holistically-aware human controller has left the system and it is difficult to guarantee that the human controller’s domain restrictions are all met by real applications. This is because a symbolic system is not able to generate internal representations to go beyond the representations that have been handcrafted. This is also true for hybrid approaches that use a mixture of both types of approaches [5], [26].

On the other hand, purely connectionist models have shown some power in processing sensory information, such as classification. However, connectionist models have shown their limitations in reasoning.

Both types have not been effective in reasoning with the concrete and the contrast, e.g., given an abstract goal (e.g., object type), find the object in concrete images that contain complex backgrounds.

On the surface, this situation is due to the current lack of engineering grade models of brain-mind. However, this view is superficial. The true reason for this systemic situation is deeper.

Supported by this author’s experience in the past 10 years, this situation is further due to the need of improvement of the infrastructure for cross-disciplinary education beyond one’s home discipline. Thomas Kuhn’s convincing account needs to be better taken account by the improvement of the infrastructure.

The remainder material of this paper is organized as follows. The following section discusses some major concepts of the

Juyang Weng is with Michigan State University, East Lansing, MI, USA (email weng@cse.msu.edu).

The author would like to thank Z. Ji, M. Luciw, M. Solgi, Q. Zhang, M. Chi, X. Xue for the discussions and the experimental data cited. Detail of the experimental work has appeared elsewhere.

brain architecture. The section III explains the top-scale 5 “chunks” of the brain-mind puzzle. To facilitate understanding of these 5 chunks, the section IV presents a basic model that contains all these chunks. Section V presents the principled architecture of two integrated visual pathways, ventral and dorsal. Section VI outlines a cortical area, which consists of layers. Section VII describes the model for a general-purpose layer. Section VIII deals with general reasoning — reasoning with the abstract and the concrete. The experimental results are discussed in Section IX. Section X gives concluding remarks.

II. IMPROVING INFRASTRUCTURE

For many years, although there have been debates between the symbolic approaches and the connectionist approaches, the debates have been inclusive [20]. The “skull open” nature of the former and the “skull close” nature of the latter have been hardly raised. The lack of analyzable mechanisms of the latter was sometimes called “black box”.

As far as we know, Cresceptron 1993 [32], [33] was the first developmental model for visual learning from complex natural backgrounds. By developmental, we mean that the internal representation is emergent from interactions with environment, without allowing a human to manually instantiate a task-specific representation. Throughout nearly 20 years of cross-disciplinary research along this line, we have faced various doubts from domain experts. Examples include: (1) Artificial intelligence does not need to follow the brain’s way. (2) Modeling the human mind does not need to follow the brain’s way. (3) Your commitment to understanding the brain is laudable but naive; (4) Regardless a clear advance of knowledge and its importance, I need to see X. (5) There is a need to improve the infrastructure as documented by Thomas Kuhn [16].

Academically, this situation is mainly caused by a lack of appropriate curriculum infrastructure for academic degrees and a lack of lifelong educational programs for domain experts. Many *symbolic* AI and Cognitive Science researchers do not have sufficient knowledge in electrical engineering, neuroscience and biology to understand signal-based processing and learning in the brain. Many *neural network* engineering researchers and Cognitive Science researchers do not have adequate knowledge in computer science (e.g., formal reasoning and complexity), neuroscience and biology to understand signal-based abstract reasoning and learning-to-reason in the brain. Many *computational neuroscience* modelers do not have a solid electrical engineering and computer science background to understand how the signal-based brain learns to deal with large-size, high-dimensional practical problems. Many *neuroscience and psychological* researchers do not have a sufficient background for, and therefore interest in, quantitative and computational aspects of their subjects of study.

The required knowledge to solve the grand puzzle of brain-mind includes biology, neuroscience, psychology, computer science, electrical engineering, robotics and mathematics. None can be left out. The traditional curriculum designs in developed countries were never meant for such a wide scope of knowledge. Consequently, reviewers do not have

sufficient width and depth in their knowledge to evaluate papers and proposals for brain-mind issues as confidently as those for relatively more conventional well studied issues. Serious improvements and investments for the infrastructure for converging research on intelligence are urgently needed. Such infrastructure is necessary for the healthy development of science and technology in the modern time.

III. FIVE CHUNKS OF THE BRAIN-MIND PUZZLE

It seems that the grand picture of the brain-mind includes five chunks at its top scale: development, architecture, area, space, and time.

The “development” chunk has task-nonspecificity, emergent representation, and skull-closedness as necessary conditions for the brain.

The “architecture” chunk of the puzzle outlines how basic units are put together.

The “area” chunk serves as the basic unit, the general-purpose unit for the architecture, like the bricks for a building.

The “space” chunk describes how the brain deals with spatial information, for elements in the external environment (outside the brain) and elements in the internal environment.

The “time” chunk schematically clarifies how the brain deals with time beyond the short time span of 50ms characterized by the intrinsic biological properties of a single neuron (Bi & Poo 2001 [2] and Dan & Poo 2006 [6]).

There are many other details, which seem to be under these five chunks. For example, neuromodulation includes reinforcement learning and emotion. The spiking model of the neuron is useful to deal with finer temporal resolution. However, all such more detailed mechanisms seem to be embedded into the above 5 basic chunks.

All further phenotypes, such as behaviors due to the external environment outside the brain, including learning to use the limbs, skills for social interactions, emergence of language and culture, should all autonomously emerge from the living and developing experience of such developmental agents in a human society.

IV. A BASIC BRAIN-MIND MODEL: ED NETWORK

First, I introduce a basic network, called Epigenetic Developmental (ED) network. “Epi” means post. “Genetic” means conception — the origination of a life. It is useful as a starting basic model to explain the above five chunks. An ED network is not task-specific [37] (e.g., not concept-specific) at the birth time. Its concepts are learned dynamically and incrementally from its motor end, through interactions with the environments.

Exemplified in Fig. 1(b), an ED network lives and learns autonomously in the open-ended, dynamic, real physical world. Its motor vector \mathbf{z} , is not only an output vector to drive effectors (muscles) to act on the real world R but also an input vector from the motor end because the environment can set \mathbf{z} , e.g., by a teacher in R . The ED network was taught to attend to a spatiotemporal object in a complex background

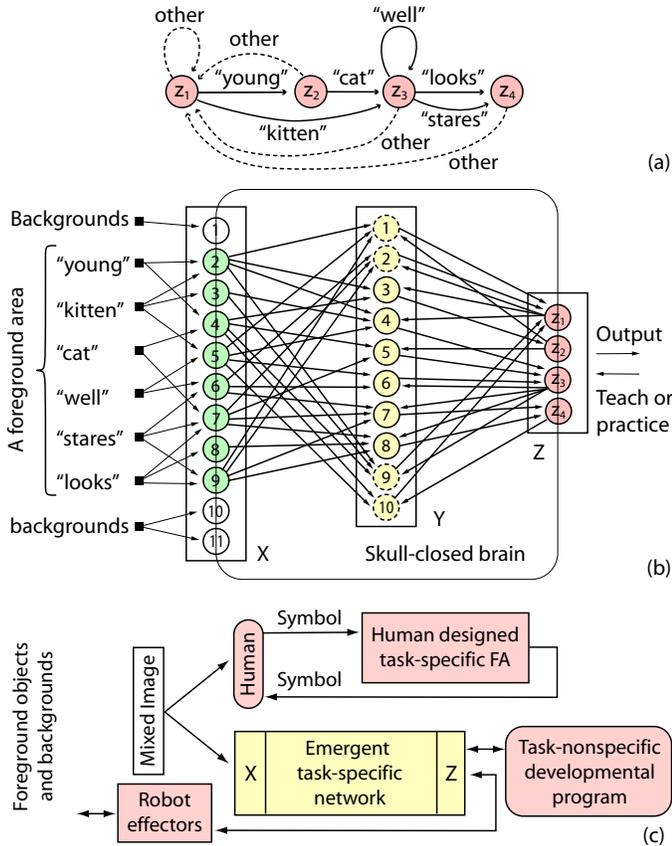


Fig. 1. Relate a “skull-open” Finite Automaton (FA) with a “skull-closed” ED network. (a) An FA, handcrafted, static, and reason from symbol to symbol. (b) A corresponding ED network that lives and learns autonomously in the real world. Each Y neuron represents a cluster in the $X \times Z$ space. It was taught to produce the same equivalent actions as (a). The top-matched neuron in Y almost perfectly matches its input $\mathbf{p} = (\mathbf{x}, \mathbf{z})$ with its weight $(\mathbf{v}_x, \mathbf{v}_z)$ which requires both parts to match well. This each top-matched Y neuron represents the transition from the corresponding state represented by \mathbf{z} and the input represented by \mathbf{x} . Such nearest-neighbor match is realized by the Lobe Component Analysis [35]. Each object in the real-world environment generates a sample foreground patch. Any human teacher cannot intervene directly inside its closed “skull” other than indirectly through its sensory port X and motor port Z . A circle represents a neuron. “Green” means a foreground image patch among possibly many such patches in a practical system, although only one foreground patch is illustrated here. In the brain, all possible patches are present at different resolutions according to the experience. The ED network autonomously self-organizes its internal (inside the “skull”) representation through interactions with the real-world environment. When the motor end is not supervised (imposed or “clamped”), the network autonomously practice without changing its operational mode. The pattern in Z not only affects the external environment, but also serves as a concept (location, type, goal, etc.) that operates on the internal environment — “skull-closed brain” as top-down attention, which affects which neuron in Y wins and further affects which object in X is attended or boosted among typically large complex backgrounds. For simplicity, top-down connections from Y to X are omitted. But for input attention, they should be present and learned. (c) A comparison of the developmental and operational mode of the FA and the ED network. Pink areas are human designed or human taught. Yellow areas are autonomously developed.

according to the motor \mathbf{z} which also represents any concept (e.g., location, type, goal).

Given any FA, there is an ED network that, from the same sequence of inputs, generates exactly the same output sequence as the FA for infinitely long time.

As illustrated in Fig. 1, between the FA and the real physical world is a human operator, who attends to an object from the real world and interprets it as a symbolic label which is fed into the FA. The FA outputs a symbolic output. The human interprets the meaning of the output label from the FA. This process continues. In contrast, a general-purpose ED network deals with directly the real physical real world, including human teachers. Its autonomous development is regulated by a human designed or evolved, task-nonspecific developmental program. A human in the environment can teach the ED network to produce the exact equivalent action sequence that the human-FA combination produces, while the ED network gradually matures though its autonomous “living” in the real world. The learning mode used in our experiment was type 2 — communicative motor-supervised learning. Additional components of ED are needed for it to learn in 3 additional modes of developmental learning [34], type 3 — reinforcement motor-supervised, type 1 — reinforcement motor-autonomous, and type 0 — communicative motor-autonomous.

The following sections provide more realistic examples of such a basic model.

V. BRAIN ARCHITECTURES

A. Architecture considerations

Symbolic architectures use human handcrafted instantiation of internal representation, as illustrated in Fig. 2(a). Artificial neural networks have been largely used as classifiers or regressors, as indicated in Fig. 2(b). In artificial intelligence, an agent is modeled as something that senses the external environment and acts on the external environment.

How the brain deals with internal attention is a fascinating subject. Treisman 1980 [27] proposed the existence of a master feature map inside the brain. This idea of internal master map has been used by others [1], [23], [28].

The major new concepts of our brain architecture include: (1) There exists no master map in the brain. (2) A receptive field is in general dynamic, not appearance kept or even topology kept. (3) Effective fields are also dynamic. (4) Motor areas are hubs for abstraction and top-down attention through action learning. (5) Abstraction, reasoning, generalization, planning, decision making, multimodal integration, self-awareness, and consciousness are all different aspects of the emergent responses of the same tightly integrated brain network. (6) The brain is a giant skull-closed statistical machine that efficiently records statistical relationships throughout lifetime.

B. Visuomotor example

Since the work of Ungerleider & Mishkin 1982 [29], [21], a widely accepted description of visual cortical areas is illustrated in Fig. 3 [10], [23]. A ventral or “what” stream that runs from V1, to V2, V4, and IT areas TEO and TE computes

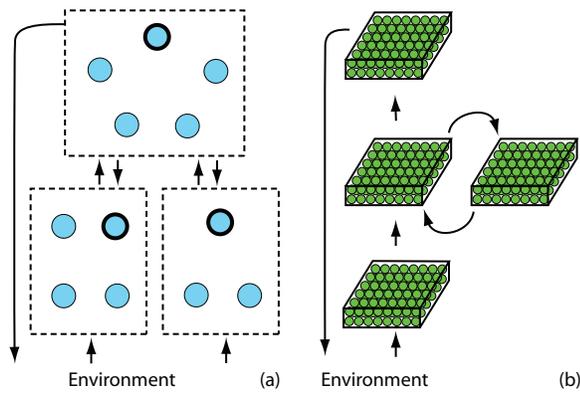


Fig. 2. Three types of agent architectures: (a) symbolic, (b) connectionist and (c) the SASE brain model. (a) Symbolic “open-skull” architecture where each rectangular block corresponds to a module with a handcrafted (defined) meaning, e.g., a word at the lower level and a sentence at the higher level. A heavy node is the end node of the module. Thus, this type cannot autonomously learn. (b) Connectionist network with fully emergent representation. This type can autonomously learn, like many variety of neural networks. This type has not found ways to deal with (1) complex backgrounds, (2) deliberative reasoning, (3) any desired temporal contexts. (c) An example of the SASE model for the brain. Only major connections are shown. For every unit, its sensory area is also an output port for its top-down attention (self-effecting), and its motor area is also an input port for its top-down sensing (self-aware). The brain is skull-closed since no direct internal manipulation by the teacher is permitted after the birth. The internal self-organization and representation emergence is fully autonomous through the agent’s life. A lower brain is developed earlier, so that the higher brain as basic areas can innervate into lower ones later.

properties of object identity such as shape and color. A dorsal or “where and how” stream that runs from V1, to V2, V3, MT and the medial superior temporal areas MST, and on to the posterior parietal cortex (PP) computes properties of the location of the stimulus on the retina or with respect to the animal’s head.

The ventral and dorsal pathways themselves are not sufficient for understanding the developmental causality of the pathways. I propose that we need to look further beyond

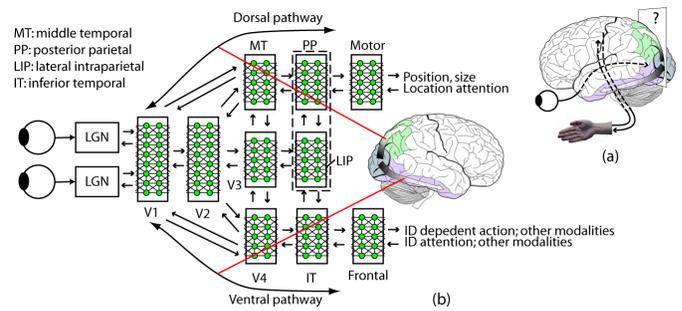


Fig. 3. (a) How does the brain generate internal representation? The only external sources are sensors and effectors. The imaginary page slices the brain to “peek” into its internal representation. (b) The dorsal “where and how” pathway and the ventral “what” pathways. The nature of the processing along each pathway is shaped by not only sensory inputs but also the motor outputs.

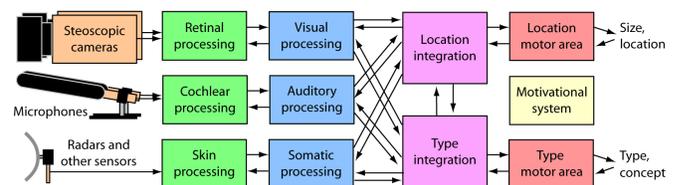


Fig. 4. The system diagram: multi-sensory and multi-effector integration through learning.

the two pathways, as the development of the functions of the “where” and “what” pathways is largely due to: (1) Signals flow from motors. (2) Top-down connections. Put in a short way, *motor is often abstract*. Any meaning that can be communicated between humans is motorized: spoken, written, hand-signed, etc. Of course, “motor is abstract” does not mean that every stage of every motor action sequence is abstract. However, the sequences of motor actions provide statistically crucial information for the development of internal abstractive representation, e.g., in the premotor area.

In this model, much of the frontal cortex is for multimodal integration, not for abstraction. For example, the visual “car” and auditory “cat” do not form sufficient abstraction till they reach a motor area that produces motor signals to say “cat” or another motor area that produces motor signals to write “cat”.

A system level multi-sensory multi-motor architecture is illustrated in Fig. 4 with details illustrated for the visuomotor model in Fig. 5.

VI. CORTEX SCALE

It is known that cortical regions are typically inter-connected in both directions [10], [4]. However, computational models that incorporate both directions have resisted full analysis [22], [13], [7], [11], [3], [15]. Informed by the neuro-anatomic work of [10], [4], our computational model illustrated in Fig. 6, provides further details about the major connections within a cortical area and across cortical areas.

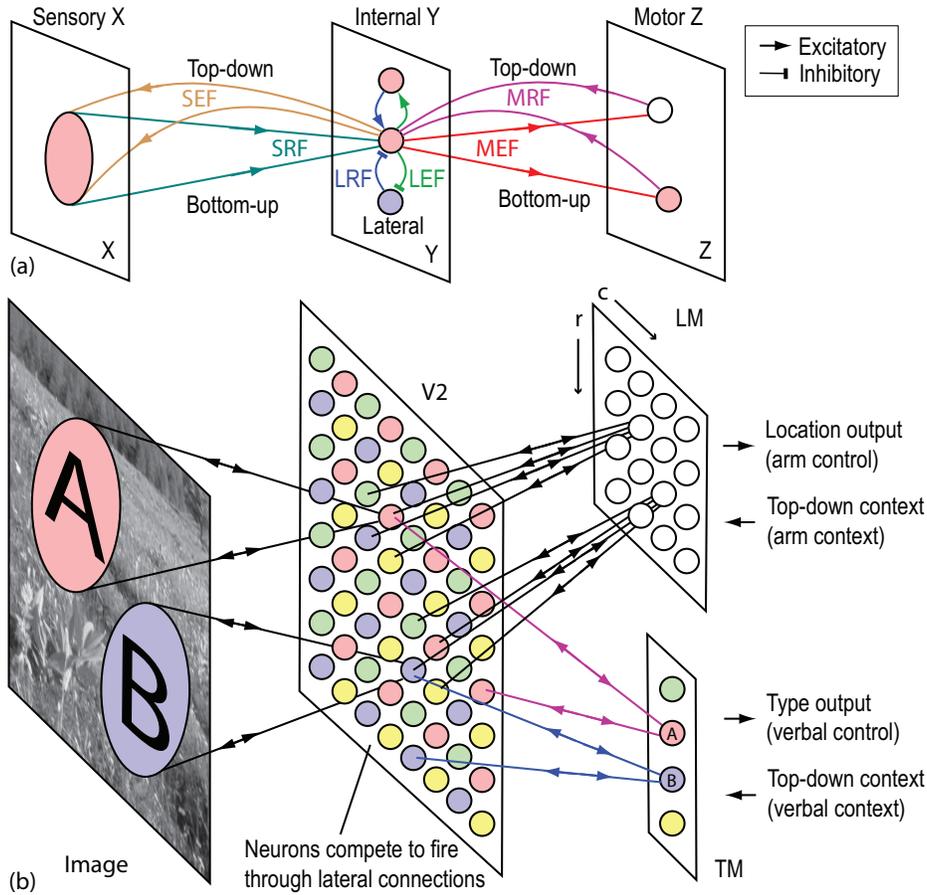


Fig. 5. The basic unit of the hextuple representations and an simple Where-Where Network (WWN) example. (a) A basic unit of the hextuple representation that bridges any two arbitrary brain areas. SRF, MRF, LRF, SEF, MEF, and LEF, thus are highly recurrent. (b) A simple WWN with four areas (image, V2, LM and TM) and its hextuple network representation. Each wire connects if the pre-synaptic and post-synaptic neurons have co-fired. The weight is the frequency of pre-synaptic co-firing when the post-synaptic neuron fires. Within each cortical area, each neuron connects with highly correlated neurons using excitatory connections (e.g., NMDA-ergic) but connect with highly anti-correlated neurons using inhibitory connections (e.g., GABA-ergic). This forces neurons in the same area to detect different features in SRF and MRF. These developmental mechanisms result in the shown connections. Every V2 neuron is *location-specific* and *type-specific*, corresponding to an object type (marked by its color) and to a location block (2×2 size each). Each LM neuron is location-specific and type-invariant (more invariance, e.g., lighting-direction invariance, in more mature networks). Each TM neuron is type-specific and location-invariant (more invariance in more mature networks). Each motor neuron pulls all applicable cases from V2. It also top-down boosts all applicable cases in V2 as top-down context. A two-way arrow means two one-way connections. All the connections within the same area are omitted for clarity. Since V2 is the first area from the image here, V2 does not need explicit SEF connections but all LM and TM neurons have global SEFs.

A. Prescreening

Prescreening for both bottom-up and top-down signal sources is necessary so as to disregard irrelevant responses that are distractors before the bottom-up and top-down integration.

The cerebral cortex contains six layers: layer L1 is the superficial layer and layer L6 is the deep layer. Weng et al. [36] reasoned that L4 and L2/3 each has a prescreening layer as shown in Fig. 6 with L5 assisting the one in L2/3 (red) and L6 assisting L4 (green), in the sense of enabling long range lateral inhibition. Such long range inhibitions encourage different neurons to detect different features. However, there needs to be an integration layer (yellow) in L2/3, integrating the prescreened bottom-up and top-down inputs. The model illustrated in Fig. 6 was informed by the work of Felleman &

Van Essen [10], Callaway and coworkers [4], and others (e.g., [8]).

B. Three source of information for each cortical area

The integration layer in L2/3 based on its current own content $L(t_{n-1})$ takes three signal sources: prescreened bottom-up input $x(t_{n-1})$ as lower features, lateral input $y(t_{n-1})$ from its own layer as its last response, and top-down input $z(t_{n-1})$ from prescreened top-down input, all at time t_{n-1} . Through the feature development function modeled as the Lobe Component Analysis (LCA) [35], the integration layer generates its next response $y(t_n)$ at time t_n as the attention-

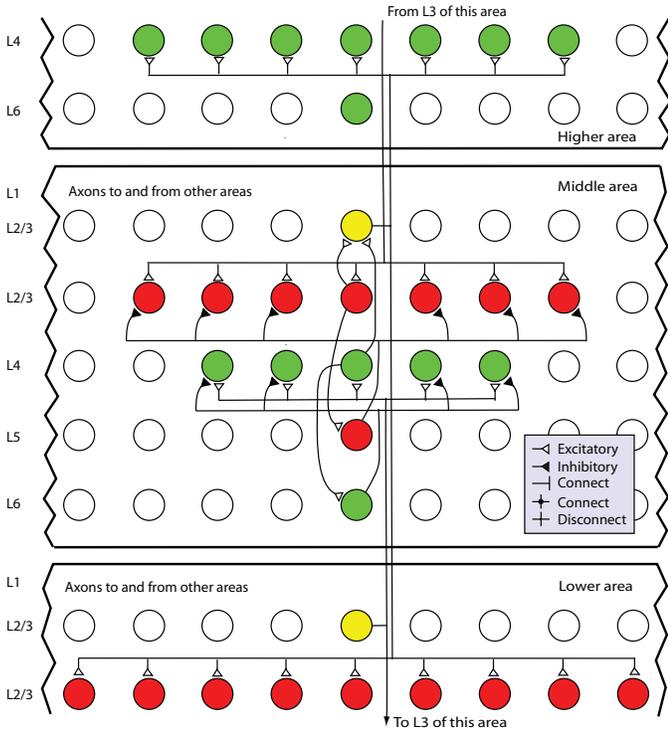


Fig. 6. Cortex scale: A schematic illustration of the major laminar cortical circuits that enable the pre-screening for the bottom-up input (green) and two-down input (red) before two-way integration (yellow). L5 assists L2/3 and L6 assists L4. Other minor connections not shown here are due to less significant statistical co-firing.

selected response and to update its level to $L(t_n)$:

$$(\mathbf{y}(t_n), L(t_n)) = f(\mathbf{x}(t_{n-1}), \mathbf{y}(t_{n-1}), \mathbf{z}(t_{n-1}) \mid L(t_{n-1})) \quad (1)$$

where f denotes the function of LCA. We call this process *attentive context folding*, folding the spatiotemporal information from the three sources into one response vector and the updated cortical layer. LCA models the feature layer and its assistive layer in the laminar cortex. For the bottom-up prescreening in L4 and L6 combination, $\mathbf{z}(t_{n-1})$ is not needed in the above expression. For the top-down pre-screening in L2/3, $\mathbf{x}(t_{n-1})$ is not needed.

C. Temporal cognition and temporal action

Fig. 7 describes a generic unit for spatiotemporal processing. As illustrated in Fig. 1, such a processing unit can develop to connect any two cortical areas that needs input-to-output attended regression and top-down attention.

Fig. 7 shows the temporal mechanism called Motor Assisted Recursive Abstraction (MARA). LCA in area 2 enables the area to develop (optimal) feature detectors for the input space (\mathbf{x}, \mathbf{z}) . The response vector from area 2 indicates the goodness of match by each feature detector. The competition using LCA allows only few neurons to fire, as “experts” for the current input context (\mathbf{x}, \mathbf{z}) . What is critically important is that area 3 is supervised with desired (I care) output $\phi(\mathbf{p})$ for $\mathbf{p} = (\mathbf{x}, \mathbf{z})$, as the equivalent class (or action) of the spatially

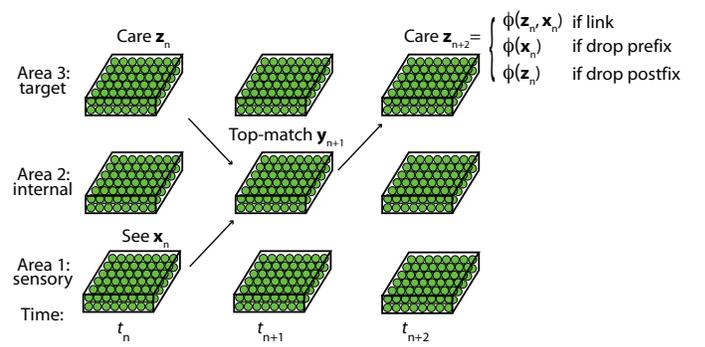


Fig. 7. The Motor Assisted Recursive Abstraction (MARA) temporal scheme. Area 2 learns features for (\mathbf{x}, \mathbf{z}) . Area 3 learns the desired output. Top-down projections from area 3 to area 2 shape the representation in area 2. The output \mathbf{z} can attend to subsets of temporal contexts.

and temporally attend part of \mathbf{p} , just like the state in an FA. Suppose that the previous temporal context is represented recursively by \mathbf{z}_n and the current context is $\mathbf{p}_n = (\mathbf{x}_n, \mathbf{z}_n)$. Fig. 7 illustrates three cases, link, drop prefix \mathbf{z}_n , and drop postfix \mathbf{x}_n , depending on which subparts of \mathbf{p}_n that \mathbf{z}_{n+2} depends on.

It can be proved [31] that the MARA scheme can learn to attend to complex temporal subsets of a temporal sequence $\mathbf{p}_{n-m}, \mathbf{p}_{n-m+1}, \dots, \mathbf{p}_n$ for any bounded $m > 0$. This proof is consistent to the fact that any FA can be a special case of an ED network.

VII. LAYER SCALE: THE DUALY OPTIMAL LCA

As shown in Fig. 7, given parallel input space consisting of the bottom-up space X and the top-down input space Z , represented as $X \times Z$, the major developmental goal of each cortical level (L4 or L2/3 in Fig. 6) is to have different neurons in the level to detect different features, but nearby neurons should detect similar features.

The cortex inspired Candid Incremental Covariance-free (CCI) LCA [35] has the desired dual optimality: spatial and spatiotemporal, as illustrated in Fig. 8.

The spatial optimality requires that the spatial resource distribution in the cortical level is optimal in minimizing the representational error. For this optimality, the cortical-level developmental program modeled by CCI LCA computes the best feature vectors $V = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c)$ so that the expected square approximation error $\|\hat{\mathbf{p}}(V) - \mathbf{p}\|^2$ is statistically minimized:

$$V^* = (\mathbf{v}_1^*, \mathbf{v}_2^*, \dots, \mathbf{v}_c^*) = \arg \min_V E \|\hat{\mathbf{p}}(V) - \mathbf{p}\|^2. \quad (2)$$

where E denotes statistical expectation. The minimum error means the optimal allocation of limited neuronal resource: frequent experience is assigned with more neurons (e.g., human face recognition) but rare experience is assigned with fewer neurons (e.g., flower recognition for a nonexpert). This optimization problem must be computed incrementally, because the brain receives sensorimotor experience incrementally.

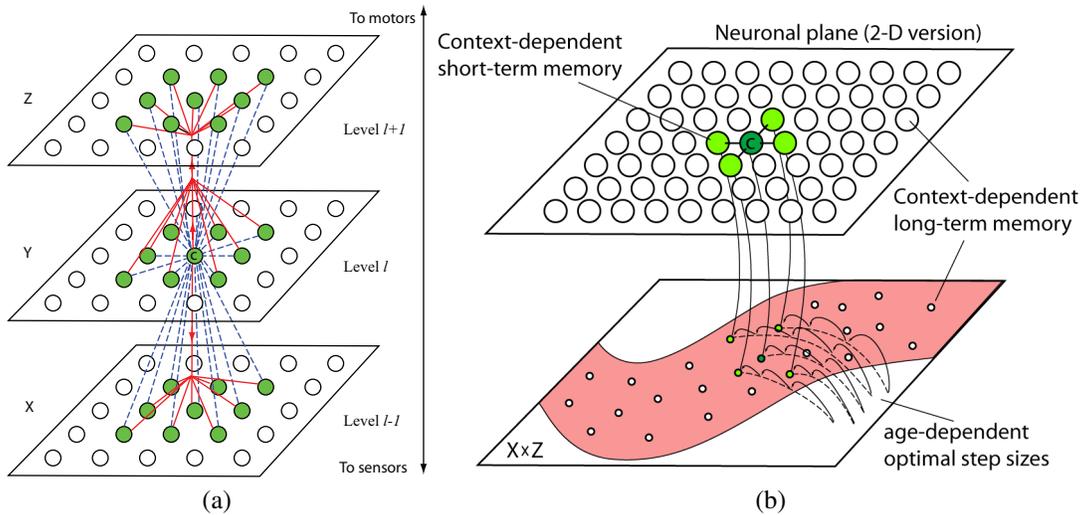


Fig. 8. (a) The default connection pattern of every neuron in cortical layer L2/3 and L4 (no Z input for L4). The connections are local but two-way. Blue: neuronal input; red: axonal output. (b) For each neuron in a layer, near neurons (e.g., green for the center neuron) are connected to the neuron by excitatory connections (for layer smoothness) and far neurons (white ones) are connected to the center neuron by inhibitory connections (competition resulting in detection of different features by different neurons). The upper layer indicates the positions for the neurons in the same layer: firing neurons are (context-dependent) working memory and those do not fire are (context dependent) long-term memory. The lower layer indicates the very high dimensional input space of the cortical layer ($X \times Z$) but illustrated in 2-D. The purple area indicates the manifold of the input distribution. The connection curve from the upper neuron and lower small circle indicates the correspondence between the upper neuron and the feature that it detects. The neuronal weight vectors must quickly move to this manifold as the inputs are received and further the density of the neurons in the purple area should reflect the density of the input distribution. The challenge of fast adaptation at various maturation stages of development: The updating trajectory of every neuron is a highly nonlinear trajectory. The statistical efficiency theory for neuronal weight update (amnesic average) results in the nearly minimum error in each age-dependent update, meaning not only the direction of each update is nearly optimal, but also every step length.

The spatiotemporal optimality gives optimal step sizes of learning. Each neuron takes response weighted input $\mathbf{u}(t) = r(t)\mathbf{x}(t)$ at time t (i.e., Hebbian increment). From the mathematical theory of statistical efficiency, CCI LCA determines the optimal feature vectors $V^*(t) = (\mathbf{v}_1^*(t), \mathbf{v}_2^*(t), \dots, \mathbf{v}_c^*(t))$ for every time instant t starting from the conception time $t = 0$, so that the distance from $V^*(t)$ to its target V^* is minimized:

$$V^*(t) = \arg \min_{V(t)} E \|V(t) - V^*\|^2. \quad (3)$$

CCI LCA aims at this deeper optimality — the smallest average error from the starting time (birth of the network) up to the current time t , among all the possible estimators, under some regularity conditions. A closed form solution was found that automatically gives the optimal retention rate and the optimal learning rate (i.e., step size) at each synaptic update [35]

In summary, the spatial optimality leads to Hebbian incremental direction: response weighted pre-synaptic activity ($y\mathbf{p}$). The deeper spatiotemporal optimality leads to the best step sizes, automatically determined by the update age of each neuron.

VIII. REASONING FOR CONCRETE AND ABSTRACT

Interestingly, Fig. 7 reveals a process of autonomous sensory motor process of ED network through time:

$$\begin{aligned} \dots &\Rightarrow \text{Abstract context } \mathbf{z}(t_n) \\ &\Rightarrow \text{Internal updates with } \mathbf{y} \text{ for } (\mathbf{x}(t_n), \mathbf{z}(t_n)) \\ &\Rightarrow \text{Abstract context } \mathbf{z}(t_{n+2}) \text{ updated } \Rightarrow \dots \end{aligned}$$

Our above discussion indicates that the action \mathbf{z} can shape which neurons in area 2 fires: Consider that only one neuron in area 2 is allowed to fire. Then, the firing neuron must match both components in $\mathbf{p} = (\mathbf{x}, \mathbf{z})$ well. In other words, the top-down signal \mathbf{z} is used as a top-down attention selector, one that boosts neurons in area 2 that have matched \mathbf{z} parts. As motor output \mathbf{z} is the abstract outcome of previous sensorimotor context, the top-down input from \mathbf{z} can be interpreted as the goal, or any other abstract context needed at this time for future processing. As \mathbf{z} can be abstract and \mathbf{x} is typically concrete, the above type of reasoning is for the abstract \mathbf{z} and the concrete \mathbf{x} both, not simply abstract-to-abstract in symbol based mathematical logic.

IX. EXPERIMENTS

Due to space limit, some experimental results published elsewhere are summarized here. A visual WVN-2 [14] reached 92.5% in object recognition rate and 1.5 pixels in average position error with 75% of the area in each image filled with unknown natural backgrounds. The WVN-3 [17]

has shown a capability to deal with multiple learned objects in complex backgrounds. The user can specify either goal (location or type) and WWN-3 reports the reasoning results for other concepts (from location goal to type, or from type goal to location). A temporal version for visual recognition [18] has reached an almost perfect recognition rate for centered objects viewed from any of the 360° object views. The stereo version of WWN [25] has shown that pre-screening is truly necessary for the temporal mechanisms to improve the result. A text processing version [38] has been tested for part-of-speech tagging problem (assigning the words in a sentence to the corresponding part of speech, about 99% correct); and chunking (grouping sequences of words together and classify them by syntactic labels, about 96% success rate) using text corpus from the Wall Street Journal.

X. CONCLUSIONS

This seems the first general-purpose brain-mind model for multiple events in complex backgrounds. The detailed algorithm is available in the cited publications. This brain-mind picture has 5 chunks, development, architecture, area, space and time. Every feature in the brain is a mix of sensory and motor information, not just sensory or just motoric. The presented 5-chunk model has been supported by engineering-grade experimental results. It seems that the age of brain-mind has arrived, although there are other brain-mind problems.

REFERENCES

- [1] C. H. Anderson and D. C. Van Essen. Shifter circuits: A computational strategy for dynamic aspects of visual processing. *Proc. Natl. Acad. Sci. USA*, 84:6297–6301, Sept. 1987.
- [2] G. Bi and M. Poo. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, 24:139–166, 2001.
- [3] T. J. Buschman and E. K. Miller. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315:1860–1862, 2007.
- [4] E. M. Callaway. Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, 21:47–74, 1998.
- [5] B. Cho, P. S. Rosenbloom, and C. P. Dolan. Neuro-soar: A neural-network architecture for goal-oriented behavior. In P. S. Rosenbloom, J. E. Laird, and A. Newell, editors, *The Soar Papers*, pages 1199–1203. MIT Press, Cambridge, Massachusetts, 1993.
- [6] Y. Dan and M. Poo. Spike timing-dependent plasticity: From synapses to perception. *Physiological Review*, 86:1033–1048, 2006.
- [7] G. Deco and E. T. Rolls. A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 40:2845–2859, 2004.
- [8] R. J. Douglas and K. A. C. Martin. Neural circuits of the neocortex. *Annu. Rev. Neurosci.*, 27:419–451, 2004.
- [9] J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. *Rethinking Innateness: A connectionist perspective on development*. MIT Press, Cambridge, Massachusetts, 1997.
- [10] D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47, 1991.
- [11] M. D. Fox, M. Corbetta, A. Z. Snyder, J. L. Vincent, and M. E. Raichle. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. National Academy of Sciences U S A*, 103(26):10046–10051, 2006.
- [12] G. E. Hinton. Learning multiple layers of representation. *Trends in Cognitive Science*, 11(10):428–434, 2007.
- [13] L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2:194–203, 2001.
- [14] Z. Ji and J. Weng. WWN-2: A biologically inspired neural network for concurrent visual attention and recognition. In *Proc. IEEE International Joint Conference on Neural Networks*, Barcelona, Spain, July 18-23 2010.
- [15] Z. Ji, J. Weng, and D. Prokhorov. Where-what network 1: “Where” and “What” assist each other through top-down connections. In *Proc. IEEE International Conference on Development and Learning*, pages 61–66, Monterey, CA, Aug. 9-12 2008.
- [16] T. S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL, second edition, 1970.
- [17] M. Luciw and J. Weng. Where What Network 3: Developmental top-down attention with multiple meaningful foregrounds. In *Proc. IEEE International Joint Conference on Neural Networks*, Barcelona, Spain, July 18-23 2010.
- [18] M. Luciw, J. Weng, and S. Zeng. Motor initiated expectation through top-down connections as abstract context in a physical world. In *IEEE International Conference on Development and Learning*, pages 1–6, Monterey, CA, Aug. 9-12 2008.
- [19] J. L. McClelland, D. E. Rumelhart, and The PDP Research Group, editors. *Parallel Distributed Processing*, volume 2. MIT Press, Cambridge, Massachusetts, 1986.
- [20] M. Minsky. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Magazine*, 12(2):34–51, 1991.
- [21] M. Mishkin, L. G. Unterleider, and K. A. Macko. Object vision and space vision: Two cortical pathways. *Trends in Neuroscience*, 6:414–417, 1983.
- [22] J. Moran and R. Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715):782–784, 1985.
- [23] B. A. Olshausen, C. H. Anderson, and D. C. Van Essen. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11):4700–4719, 1993.
- [24] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Upper Saddle River, New Jersey, 2nd edition, 2003.
- [25] M. Solgi and J. Weng. Developmental stereo: Emergence of disparity preference in models of visual cortex. *IEEE Trans. Autonomous Mental Development*, 1(4):238–252, 2009.
- [26] R. Sun and X. Zhang. Accounting for a variety of reasoning data within a cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, 18:169–191, 2006.
- [27] A. M. Treisman. A feature-integration theory of attention. *Cognitive Science*, 12(1):97–136, 1980.
- [28] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nufflo. Modeling visual attention via selective tuning. *Artificial Intelligence*, 78:507–545, 1995.
- [29] L. G. Ungerleider and M. Mishkin. Two cortical visual systems. In D. J. Ingel, editor, *Analysis of visual behavior*, pages 549–586. MIT Press, Cambridge, MA, 1982.
- [30] J. Weng. On developmental mental architectures. *Neurocomputing*, 70(13-15):2303–2323, 2007.
- [31] J. Weng. The SWWN connectionist models for the cortical architecture, spatiotemporal representations and abstraction. In *Proc. Workshop on Bio-Inspired Self-Organizing Robotic Systems, IEEE International Conference on Robotics and Automation*, pages 1–6, Anchorage, Alaska, May 3-8 2010.
- [32] J. Weng, N. Ahuja, and T. S. Huang. Learning recognition and segmentation of 3-D objects from 2-D images. In *Proc. IEEE 4th Int'l Conf. Computer Vision*, pages 121–128, May 1993.
- [33] J. Weng, N. Ahuja, and T. S. Huang. Learning recognition and segmentation using the Cresceptron. *International Journal of Computer Vision*, 25(2):109–143, Nov. 1997.
- [34] J. Weng and W. Hwang. From neural networks to the brain: Autonomous mental development. *IEEE Computational Intelligence Magazine*, 1(3):15–31, 2006.
- [35] J. Weng and M. Luciw. Dually optimal neuronal layers: Lobe component analysis. *IEEE Trans. Autonomous Mental Development*, 1(1):68–85, 2009.
- [36] J. Weng, T. Luwang, H. Lu, and X. Xue. Multilayer in-place learning networks for modeling functional layers in the laminar cortex. *Neural Networks*, 21:150–159, 2008.
- [37] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291(5504):599–600, 2001.
- [38] J. Weng, Q. Zhang, M. Chi, and X. Xue. Complex text processing by the temporal context machines. In *Proc. IEEE 8th International Conference on Development and Learning*, Shanghai, China, June 4-7 2009.