Biologically Feasible Generative Networks and Evolutionary Learning

Serge Dolgikh (sdolgikh@nau.edu.ua)
National Aviation University https://orcid.org/0000-0001-5929-8954

Research Article

Keywords: Machine learning, unsupervised learning, representation learning, concept learning, clustering

Posted Date: May 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1622673/v1

Abstract

Representations play an essential role in the learning of artificial and biological systems due to their capacity to identify characteristic patterns in the sensory environment. In this work we examined latent representations of several sets of images, such as basic geometric shapes and handwritten digits, produced by generative models in the process of unsupervised generative learning. A biologically feasible neural network architecture based on bi-directional synaptic connection equivalent in training and processing to a symmetrical autoencoder was proposed and defined. It was demonstrated that conceptual representations with good decoupling of concept regions can be produced with generative models of minimal complexity; and that incremental evolution of architecture can result in improved ability to learn data of increasing conceptual complexity, including realistic images such as handwritten digits. The results presented in this work demonstrate the potential of conceptual latent representations produced in unsupervised generative learning as a natural platform for conceptual modeling of the sensory environment and possibly, other intelligent behaviors.

1 Introduction

Representation learning with the objective to identify the informative patterns in the observable data has a well-established record in the field of Machine Learning. Identification of stable patterns of attribute combinations, or concepts [1,2], in the observed environment can provide significant advantages to the learner due to significant reduction in the cost of processing of sensory inputs, memory, as well as the ability to associate beneficial behaviors to a general class of observations, rather than specific observed instance. For these reasons, intelligent systems capable of successful identification of concepts in the observable data find applications in an increasing number of domains. In the commonly used approach, concepts can be identified in the process of representation learning [3] that produces a transformed representation of observable data in a latent space created by a learning model in the process of training under certain constraints.

1.1 Related Work

Hierarchical representations of observable data were obtained in a completely unsupervised training process with Restricted Boltzmann Machines (RBM) and Deep Belief Networks (DBN) [4,5] offering a noticeable improvement in the quality of subsequent supervised learning [6]. Different types, architectures and flavors of generative models were investigated since including autoencoder neural networks, Generative Adversarial Networks (GAN) [7–10] and other types and flavors of generative models to name only a few in a rapidly expanding field, resulting in improved accuracy and versatility of the models with a wide range of applications. In the theory of learning systems, the relations between learning and statistical thermodynamics were studied in [11,12] leading to understanding of a deep connection between learning processes in artificial generative models and principles of information theory and statistical thermodynamics.

In experimental studies with generative neural network models, a number of interesting results was reported including the “cat experiment” that demonstrated spontaneous emergence of concept sensitivity on a single neuron level in unsupervised deep learning with images [13]. Disentangled representations were produced and discussed with a deep variational autoencoder architecture and several sets of artificial image data [14] pointing at the possibility of a general nature of structured latent representations emergent in the process of unsupervised generative learning. Geometric and topological structure of conceptual representations of images of basic geometric shapes clearly correlated with characteristic patterns in the training data was studied in [15].

In a growing number of results, concept-associated structure has been observed with real-world data of different types and origin, including general and medical imaging [16,17], linguistic applications [18], Internet and network security [19,20] and other applications.

These results demonstrated that self-learning with generative models can produce structured representations of the sensory data associated with characteristic patterns, or “concepts”, in an entirely unsupervised process, based on the ability of the learning models to compress the observations and restore them into and from an informative low-dimensional representation.

Interestingly, alongside these results, very recent advances in the research of biologic sensory networks [21,22] demonstrated commonality of low-dimensional neural representations in processing sensory information by animals, including humans. These results suggest intriguing parallels in learning processes between artificial and biologic systems.

To summarize this brief overview of a wide and rapidly expanding field of general concept learning, while significant progress has been made in interpretation and learning of concepts in different environments, with growing number of models, methods and applications with increasing efficiency, there remain some essential limitations in our understanding of how general, abstract concepts emerge and are learned. The first one relates to the generality of results obtained with models of increasing specialization and complexity, where it can be questioned whether the observed effects related to specific selection of model architecture or properties, and to what extent they can be expected to apply to different models and data in a more general way.

The second set of questions relates to the origin of concepts as general classes of characteristic patterns in the sensory data. To observe, measure and learn higher-level concepts, models in these studies needed confidently identified instances of the concepts. This knowledge was not a part of the learning model and had to be provided externally. Then an investigation into the origin of concepts becomes less clear, as to identify concepts one would need their known instances and the result may depend on a specific choice of concepts selected for experiments.

Another essential question is how generative models capable of successfully learning complex real-world data could have emerged in natural systems? It entails that architectures of the learning models need to

be feasible for biological systems, including an essential ability to evolve incrementally with an improvement of the learning success with sensory data of increasing complexity.

In this work we attempted to approach these questions first by using generative models of limited complexity, well within the range of simplest biologic systems and without advanced or specific architectural features. By using models of limited complexity and plain, “generic” architecture the problem of generality can be addressed, and it can be expected that the observed effects could be reproduced with models and data of similar complexity. In addition, should this stage of investigation produce positive results with realistic though possibly, simple data, such “minimal” models can serve as a reference, or starting point for incremental adaptation of architectures for learning with data of higher complexity.

To address the origin problem, methods of analysis of latent representations were developed that are entirely unsupervised, do not depend on or require information on the content of the learning data. These methods can be used to study and describe the “native” information structure in the latent representations of generative models that does not depend on pre-defined external concepts. This approach allows to advance toward an understanding of concept origin as a characteristic structure in the informative representations of generative models that does not depend on external definition of concepts.

The rest of the paper is organized as follows: section 2 provides a description of the biologically feasible generative architecture of the studied models, data and the methods used in this work; in section 3 the results on the structure and topology of unsupervised latent representations obtained with models and data of increasing complexity are reported and discussed. Section 4 offers a summary of the results with a discussion, including the origins of abstraction in generative learning and connections to fields and areas in machine and general learning.

2 Bi-directional Generative Neural Network Architecture

To overcome challenges of some conventional generative models outlined earlier, a novel architecture that avoids duplication of resources, such as neurons and synapses in generative neural network models, such as autoencoders [7,23], is proposed. The advantages of the proposed solution are that it is compatible with the requirements of biological feasibility, generality and evolvability while, as discussed further in this paper, being capable of successful learning concepts in the data of minimal, but realistic conceptual content.

2.1 Bi-Directional Generative Architecture

The proposed network architecture is based on bi-directional synapses and a two-phase training cycle.

A bi-directional synapse differs from a neural synapse in the conventional feed-forward neural network architecture by its ability to operate in both directions, having two sets of trainable parameters, “forward” and “back”: $W^f = \{w_{ab}{}^f, b^f\}$, $W^b = \{w_{ab}{}^b, b^b\}$.

Use of bi-directional synapses allows to avoid the redundancy problems of conventional generative models such as autoencoders, by reducing the resources, both neurons and synapses effectively by close to a half, while fully retaining universal approximation capacity of these models.

2.2 Training

In the training phase the model effectively “unfolds” as with subsequent training equivalent to conventional generative architectures such as autoencoder. In the first, forward pass through the network the forward parameters are used first in the direction $W_{i,f}$ to $W_{l,f}$, then backward ones are used in the reverse direction, i.e. from the latent segment to the input, $W_{l,b}$ to $W_{i,b}$ resulting in generated output:

$$y = T_b \times T_f(x)$$

1

where y is the output generated by the model from observable input x, T_f and T_b, tensors obtained with forward and backward sets of model parameters as described in Section 2.1, respectively.

In the second, learning phase, the parameters are adjusted based on the deviation of the output from the input, defined by the cost function, in the opposite direction, i.e. $W_{i,b}$ and to $W_{i,f}$. Any type of Bayesian method can be used, including gradient descent, or its biologically feasible implementations.

By definition, training of a bi-directional model of this type would be equivalent to that of a dual conventional symmetrical model of an approximately double size (excluding the latent component) with parameters $\{ W_{i,f} .. W_{l,f}, W_{l,b} .. W_{i,b} \}$ in the sense that under the same conditions of training it would produce the same configuration of training parameters, and consequently, distributions of data in the latent representation component of the model. In the simplest case, as in the examples discussed further, it can be a single encoding layer producing latent representation described by the coordinates of neuron activations; in other cases, it can be a set of layers with a more complex structure of the latent representation.

In the training phase a model can be trained in an unsupervised generative process with a subset of sensory data that can be sampled randomly or via an enhanced process of selection of training samples. In the operation phase of a trained model, the encoding and generative parts of model described by tensors $T^{(f)}$ and $T^{(b)}$ respectively, are effectively “disconnected” producing the encoding transformation E from the observable space to latent representation, and the generative one, G, operating in the opposite direction:

$$r = E(X) = T^{(f)}(X); y = G(l) = T^{(g)}(l)$$

2

where *X*, an observable sample; *r*, the latent image of X; and *y*, the observable interpretation of a latent position *l* generated by the model.

2.3 Learning and Generative Ability

The statement of equivalence of training of a bi-directional generative model and the dual symmetric autoencoder-type model will be used extensively throughout this work given the significant number of results obtained with such models [13–20,24].

With regards to training and learning success, as has been discussed in a number of earlier results [7,20], a success of generative learning can be verified with unsupervised methods that do not require prior knowledge about conceptual content in the sensory data, such as:

- monitoring the values of training parameters such as cost function, during the training process;
- correlation measures of the input and generated output samples during and resulting from training;
- verification of generative ability of trained models via generating a subset of samples of the types represented in the training set.

2.4 Experimental Datasets

In evolvable learning approach, the objective differs from conventional methods as what is sought is not s specific and highly specialized, perhaps very complex architectural solutions as a number of recent neural architectures [25,26] that proved to be successful in analyzing real-world data of high complexity; but rather establishing possible pathways in which models of plain architecture and limited complexity would be able to evolve in success of analysis and representation of data of increasing complexity.

To follow this objective, we start with generic models or limited complexity, and data of limited, though still realistic in some simple environments, conceptual content. Two essential objectives of this study were to demonstrate that such standard “vanilla” models are capable of successful self-learning with simple conceptual data; and then, demonstrate their ability to evolve toward successful learning of more complex data in an incremental way that does not require massive addition of resources or architectural modifications.

In following this program, several datasets of images of basic geometric shapes such as: circles, triangles and backgrounds with differences in content and complexity, measured by variety of content were created. While images represented simple shapes, the intent was for the data to have certain realistic context for simple learning systems, for example, different types of shapes can be associated with sources of food versus predators and general background in some simple natural environments.

The first dataset of grayscale images, Shapes-1 (G1), consisted of 600 images of circles, triangles and grayscale backgrounds with two representative samples per shape with variation in the size and contrast of fore / background.

The second dataset, Shapes-2 (G2), contained 1,000 grayscale images of circles, triangles and backgrounds with variation of the size in the range 0.3–1.0 of the image size (i.e., 0.3 × 64 pixels), with variation of contrast of fore- vs. background for each size.

The third dataset, Shapes-C (C) contained 1,200 color images of circles and other shapes as described in Table 2. of two colors, red and blue, of different size and contrast to the background; triangles of two colors with the same characteristics; horizontal stripes of types: wide red and narrow blue; and empty grayscale backgrounds.

In artificially generated datasets the images were centered, symmetrical and had no rotation based on the argument [27] that a separate orientation function could be effective in producing sufficient quality of observations without significant cost associated with neural networks of higher depth and complexity.

Finally, as an example of real-world image data the MNIST dataset of handwritten digits [28] was used.

The range of datasets used in the study allowed us to evaluate learning ability of the models in the learning success, generative quality and the ability to generalize. One can introduce a measure of conceptual complexity of the observable data as the number of characteristic patterns that at least in some cases can be identified without prior knowledge about the content and semantics of the data. The difference in conceptual complexity of the image data in the datasets allowed us to make a number of observations on learning success of generative models with respect to data of increasing complexity.

2.5 Generic Generative Architecture

An essential expectation for a biologically-feasible architecture is evolvability, that is, a possibility of an incremental change or sequence of changes from simpler to more complex architecture associated with improved learning capacity.

A generative neural network architecture can be described by three essential components, performing the functions of physical adaptation or rendering; deep processing; and latent representation, or R-D-L.

In the rendering stage the observed sensory data is transformed to an invariant numerical representation. This stage is specific to the data and sensory mechanisms, for example, light sensitive elements for visual data, auditory or olfactory neurons and similar for other senses, producing output in the format of a numeric vector that is fed into the next stage of processing.

The deep stage of a standard architecture consists of a number of deep layers with possible additional features such as sparsity constraint, residual layers and so on. This stage represents the brawn of the model, allowing to produce features at different scales with effective description of the observed data.

In the final, representation stage, the output of the deep stage is compressed to produce effective low-dimensional representation of the data, i.e. a small set of effective features that can be related to activations of neurons in the representation block of the model. The representation stage can be a single layer in simplest models, or a more complex combination of layers.

In following the objectives of the study, generic models of minimal complexity were used in this work. For visual data, rendering phase (R) consisted of a number of convolutional – pooling layers to capture features of higher scale. The deep stage consisted of a single interconnected layer of a constant size *N* (D_N, e.g. D_{30}). Finally, the latent phase was represented by a single layer of a constant size *M* (L_M, e.g. L_{10}) with the latent coordinates in the representation space defined by activations of the neurons in the latent layer, { l_1, .. l_M}.

The conceptual complexity of the data represented in datasets varied from C = 3, datasets Shapes-1,2 to well over 10 and possibly, up to 100 of the MNIST dataset of handwritten numbers.

The architecture and parameters of models used in the study are shown in Table 3.

Table 1
Minimal generative architecture

Architecture	Data	Rendering	Depth	Latent
Minimal	Grayscale shapes, Shapes-1,2	Convolution, 2–3 stages	1 layer (D_{50-100})	1 flat layer, L_{3-10}
Incremental 1	Color shapes, Shapes-C	-	-	1 sparse layer, L_{5-10S}, $l_1 = 10^{-5}$
Incremental 2	Handwritten digits, MNIST	-	-	1 sparse layer, L_{10-24S}, $l_1 = 10^{-4..-5}$

Generative neural models used in the study were implemented in Keras / Tensorflow [29], several common data analysis and machine learning packages and libraries were used.

3 Results

3.1 Concept Learning with Minimal Generative Models

Minimal generative models with single deep and latent layers (D = 1 (50 .. 100), L = 1 (3 .. 10)) demonstrated successful generative learning with the data of minimal conceptual complexity in grayscale shapes datasets Shapes-1,2.

High level of learning success, in the range of 80–90% was observed in unsupervised training with geometrical shapes datasets [15].

Clear concept-correlated structure with good separation of concept regions was observed in the latent representations of successful models. A generative scan taken from a grid of latent positions g_t,

propagated with generative transformation *G(I)* (2) shows concept regions clearly associated with the shapes in the training dataset in the form of slanted columns (Fig. 3).

Structured character of latent representations created by generative models in these experiments can be used as a basis for a number of methods, including entirely unsupervised, to identify latent regions associated with characteristic patterns in the sensory data.

While models and data used in these experiments were of minimal complexity, they demonstrated that minimal generative models with the number of active neurons in the range of a hundred are already capable of learning to differentiate simple but meaningful and realistic in some simple environments, data. Following the arguments in Section 2.2, these results are fully compatible with the bi-directional generative architecture.

3.2 Incremental Learning: Sparse Representations

The results of experiments with the datasets of increasing conceptual complexity offer insights on incremental evolution of generative architecture for improved ability to learn concepts in the sensory data. It was found [30] that minimal generative models with a single low-dimensional latent layer discussed in the previous section are not able to maintain learning success with the data of higher conceptual complexity, such as colored shapes in the Shapes-C dataset. In training with colored images, models that were successful in differentiating and interpreting black and white geometric shape images had significantly lower learning success (less than 20%). Moreover, even successful learners were not able to comply with essential constraints in realistic biological learning, such as a limit on high activations. A significant increase in the dimensionality of the latent layer, from L = 3 to L = 10 and above restored learning success, however representations produced by such models were of higher dimensionality with no clear disentanglement of latent concept regions, complicating learning of concepts.

A solution was found in imposing a sparsity constraint on the latent activations, i.e. $L_{10} \lozenge L_{5..10S}(I_1 = 10^{-5}..10^{-6})$, I_1 being Level 1 activation penalty imposed in training. Sparse models of this type achieved several objectives simultaneously: a) restored training success with color data to the level of up to 90%; b) produced low-dimensional representations with clear separation of concept regions; and not in the least, c) due to low dimensionality of representations, did not require additional utilization of resources that is, neurons and synapses in interpretation of inputs compared to minimal models.

Latent representations of successful models retained a highly structured character closely correlated with types of shapes in the training dataset as illustrated in Fig. 4. An analysis of the representations produced by successful models indicated that they used “stacking” strategy, effectively distributing concepts between low-dimensional “slices” produced by activations of participating neurons.

Summarizing the results in this section, a conceptual learning limit was found for minimal models discussed in the preceding section and overcome with an incremental evolution of generative architecture

$L_3 \lozenge L_{5..10S}$, resulting in restoration of learning success, structured low-dimensional representations and minimal increase in operational resources. The evolved “Incremental 1” generative models were capable of successfully learning data of significantly higher conceptual complexity: from $C = 3$ (Shapes-1,2 data) to $C = 7–10$ (Shapes-C data).

3.3 Incremental Learning: Handwritten Digits

Sparse models introduced and discussed in Section 3.2 can be seen as an incremental variation of minimal architecture that has demonstrated a significant improvement in the ability of the models to learn and interpret data of higher conceptual complexity. In this section sparse models of this type were applied to learn MNIST dataset of handwritten digits.

This dataset contains real-world image data, rather than artificially created visual data in the earlier experiments, with significantly higher conceptual complexity. One can recall that “a digit” is an abstract notion that can be represented in the real data with at least several if not many types of interpretations (for example, the digit “1” can be long or short, upright or with different inclination and so on). For this reason, the conceptual complexity of MNIST data can be estimated as several distinct characteristic patterns per digit, i.e., in the order of several dozen and up to 100. A more precise estimation would be an interesting topic for another research but was not attempted in this study.

Initial results showed interesting parallels with those seen in Section 3.2: while the models of type $L_{10S}(l_1 = 10^{-5})$ were generally successful in learning and interpreting visual data (Fig. 5), the resulting representations were of higher dimensionality, up to six active neurons.

An attempt to reduce latent dimensionality of activations in learning with MNIST data by imposing a stronger activation penalty resulted in loss of learning success. The success was restored by another incremental change of generative architecture, by expanding the latent layer with simultaneous strengthening of the activation penalty: $L_{5..10S}(10^{-5}) \lozenge L_{20..24S}(10^{-4})$.

This modification restored learning success (Table 2) and low-dimensionality in the representations of MNIST images. Representations demonstrated clear structure correlated with images of digits in the input data with a clear separation of concept regions (Fig. 6).

Table 2
Learning success, sparse generative architecture “Incremental-2”, MNIST dataset

Architecture	Digits: 0,1,6,9	3,5,8	4,7	2	Overall
$L_{10S,}\ l_1 = 10^{-5}$	≥ 95%	≥ 90%	≥ 87%	78%	93.8%
$L_{24S,}\ l_1 = 10^{-4}$	≥ 95%	≥ 90%	≥ 85%	73%	90.4%

The results in this section demonstrated that the next incremental modification of generative architecture, L_{5-10S} (Incremental-1) to L_{20-24S} (Incremental-2) produced models of successful learning and with informative structured representations with a real data of significantly higher conceptual complexity.

3.4 Unsupervised Concept Learning and Concept Origin

Highly structured character of representations of sensory data produced by minimal generative models in the preceding sections points to a possible direction in resolution of the concept origin paradox.

A possibility to learn characteristic concept regions in the latent representations of generative models with unsupervised methods was demonstrated in [15,20] with unsupervised methods such as density clustering [31], nearest neighbor novelty detection [32] and others. For example, an identified latent structure R_i such as a density cluster, can define a relation of containment of a sensory input X in the concept region as $E(X) \in R_i$.

A "native" or natural concept N_k can be interpreted then as the latent region associated via encoding and generative transformation (2) with a certain characteristic pattern in the sensory data, such as in this work, a type of geometrical shape. A symbolic token or identifier can be associated with identified concept regions by individual learners allowing to distinguish between them, for example, indexing clusters identified by a clustering method. Then, definition of a native concept is equivalent to identification of its characteristic region R_i that can be described by a concept classifier in the observable data space: $K_i(X) : E(X) \in R_i$.

A natural concept N_i thus can be defined by a 3-tuple R_i, K_i, q_i (region, classifier, internal unique index or symbol).

$$N_i : (R_i, K_i, q_i)$$

3

The results obtained in Sections 3.1–3.3 with the geometric shape and handwritten digits data illustrate well-defined structure of latent concept regions produced in the process of unsupervised generative learning with minimal models.

4 Evolutionary Learning

In the study, incremental evolution of generative architecture as described in Sections 3.1–3.4 was observed (4).

$$L_{3F} \rightarrow L_{5-10S}\left(10^{-5}\right) \rightarrow L_{20-24S}\left(10^{-4}\right)$$

4

The incremental evolution produced noticeable improvement in learning ability of generative data with the data of increasing conceptual complexity from 3, grayscale shapes data to up to 100 characteristic patterns of handwritten digits in the MNIST data, as summarized in Table 3.

Table 3
Incremental evolution of generative architecture

Architecture	Data	Complexity	Latent layer	Effective dimensions	Conceptual structure
Minimal	Artificial, Shapes-1,2	3	Flat L_3	3	Yes
Incremental 1	Artificial, Shapes-C	7–10	Sparse L_{5-10}, $l_1 = 10^{-5}$	3	Yes
Incremental 2	Real, MNIST	up to 100	Sparse L_{20-24}, $l_1 = 10^{-4}$	3–4	Yes [(1)]
[(1)] $L_{24}(l_1 = 10^{-4})$ architecture					

These results point to a general approach of evolutionary generative learning that can be outlined as follows.

A model satisfying the definition of “standard” generative architecture in Section can be described by a set of parameters of the second order, $F = \{f_1 .. f_N\}$ (parameters of the first order conventionally describe individual models such as, in the case of neural network-based models, weights and biases). The parameter space defined by F on the other hand, describes the architecture, including layers, training parameters such as sparsity constraint and others).

One can define a tangential vector V_f in the parameter space F if an incremental change of the model is possible in the subset of parameters f associated with V. For example, V can describe an addition of neurons in a deep or latent layer of the model, addition or modification of a training parameter, and so on. The maximal set of orthogonal tangential vectors at a certain point f in F defining architecture A_f then defines a tangent space $T_V(A_f)$ in which incremental change of A_f is possible.

We will assume that learning success of models described by architecture A_f with data D can be described by a continuous function $L(D, A_f)$, possibly multi-dimensional. Methods described in Section 2.3 as well as others can be used to define the learning success functional $L(D, A_f)$. With the learning success function and a given architecture a region in the data space $D_S(m) = D(A_f, m))$ can be defined by the

condition $L(D_S, A_f) \geq m$, the minimal learning success. D_S thus defines the success region of the architecture A_f in the data space.

Suppose in the tangent space $T(A_f)$ at the point F defining an architecture A_f, an incremental vector e_f exists such that:

$$\delta L\left(D_S, A_f\right)|_{F, e_f} \geq 0$$

5

with the condition of inclusivity on the success region of the evolved model $A_{f+\delta f}$, $D_S(f + \delta f)$: $D_S(f) \subset D_S(f + \delta f)$. If additionally, the condition of complexity increase is satisfied, that is, there is a subset $H_{\delta f} \subset D_S(f + \delta f)$ of a higher conceptual complexity, $C(H_{\delta f}) > C(D_S(f))$, (5) defines a direction of incremental evolution of the architecture that improves learning success with the data of similar or higher conceptual complexity.

The set of orthogonal vectors e_f in the tangential space T_V satisfying the conditions in (5) defines a tangential subspace, or a slice $E_f \subset T_V$ in that produces an improvement of learning success as a result of incremental evolution of the architecture. The positive evolvability condition at a given point F in the architecture parameter space can then be described by:

$$Dim\left(E_f\right)|_{A_f} > 0$$

6

that is, there exists at least one direction of incremental evolution with an improvement in learning success. As long as the condition (6) is satisfied, a path or trajectory P_L can be defined in the model parameter space originating from some initial point A_0 such that learning success defined by the metric $L(D, A)$ increases monotonously along P_L:

$$\frac{\partial L(D_S, A)}{\partial v}|_{P_L} \geq 0$$

7

where v, parameters defining the trajectory P_L in the model parameter space. P_L can be called a learning trajectory or learning path.

A progression of generative learning architecture in (4) provides an example of a learning path from the data and architecture of minimal complexity to the realistic data in the MNIST dataset. The complexity of the models increased in this process from (H+ 3, 3H) to (H+ 24, 24H) for (the number of neurons, count of passive synapses) in the modeling components of the model, excluding physical rendering, with only a slight increase in the operational resources (from 3 active neurons per stimulus to 3 or 4).

Based on this example and the definitions in this section, the general problem of evolutionary learning can then be stated as:

For given data D, presuming it has non-trivial conceptual content, find a learning path $P_L(A_0, A_D)$ *from the minimal architecture* A_0 *to the architecture* A_D *that achieves successful learning with D, i.e. such that* $D \subset D_S(A_D)$.

Due to limitations of this study, evolutionary learning will be addressed in more detail in a future work.

5 Discussion

In this work we demonstrated that structured representations can be produced by biologically feasible generative neural network models of limited complexity. The size and complexity of the studied models, including physical adaptation (rendering) block, was well within the range of primitive biologic organisms such as jellyfish, snails and leeches [33,34]. Structured character of latent representations produced by these models closely correlated with characteristic patterns in the sensory inputs, as well as the ability to learn concept regions in an iterative unsupervised process [15,20] can offer an essential evolutionary advantage in massive reduction of required memory and complexity of processing of sensory inputs.

As supported by the results of experiments in Sections 3.1–3.3, the bi-directional generative architecture offers an essential ability to evolve incrementally producing continuous improvement in learning success with data of increasing complexity. These results may shed light on the question, how complex generative models capable of analyzing and modeling sensory environments of very high complexity could have emerged and evolved in the natural environments, with further results expected in the direction of evolutionary learning as defined and discussed in Section 3.5.

Another observation based on the results of the experiments in this work can be made on the extraordinary effectiveness of neural networks as functions of processing and modeling sensory environments in natural biological systems. In the authors view, it is not only due to their universal approximation power that many other methods possess; but also, their capacity of incremental modification, that increases the space of positive incremental variation, as discussed in Section 3.5, resulting in higher effectiveness in producing learning paths to architectures capable of learning sensory data of increasing complexity.

Structured representations with strong correlation with characteristic patterns in the sensory data and the capacity to learn concept regions as a conceptual model of the sensory environment can provide a natural basis for a number of intelligent functions and behaviors, including encoding, communicating and sharing semantic information about sensory observations in a collective of learners with similar generative architecture [35].

The results presented in this work show that structured conceptual representations can emerge naturally in the process of unsupervised generative learning with sensory inputs from the environment under the

constraints of generative accuracy, that is, effective interpretation and reproduction of sensory inputs, and strong redundancy reduction. They can provide a natural platform for development of intelligent behaviors including abstraction, conceptual modeling of the environment and communicating symbolic information about observations in a collective of learners. For these reasons, in the authors view, further investigation of conceptual representations in generative learning and their role in formation of intelligent functions and behaviors in both artificial and biologic systems merits attention of the research community.

References

1. Bruner J.S., Goodnow J.J. and Austin G. A.: A study of thinking. New York, USA: John Wiley and Sons, 1956.
2. Rosch, E.H.: Natural categories, Cognitive Psychology, 4, 328–350, 1973.
3. Y. Bengio, A. Courville, P. Vincent, P.: Representation Learning: a review and new perspectives. arXiv:1206.5538 [cs.LG], 2014.
4. Hinton, G., Osindero, S., Teh Y.W.: A fast learning algorithm for deep belief nets. Neural Computation 18(7), 1527–1554 (2006).
5. Fischer, A., Igel, C.: Training restricted Boltzmann machines: an introduction. Pattern Recognition 47, 25–39 (2014).
6. Coates, A., Lee, H., Ng, A.Y.: An analysis of single-layer networks in unsupervised feature learning. In: Proceedings of 14th International Conference on Artificial Intelligence and Statistics 15, 215–223 (2011).
7. Bengio, Y.: Learning deep architectures for AI. Foundations and Trends in Machine Learning 2(1), 1–127 (2009).
8. Welling M. and Kingma D.P.: An introduction to variational autoencoders. Foundations and Trends in Machine Learning, 12(4), 307–392, 2019.
9. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A.A. Bharath: Generative adversarial networks: an overview". IEEE Signal Processing Magazine, 35(1) 53–65, 2018.
10. Partaourides, H., Chatzis, S.P.: Asymmetric deep generative models, Neurocomputing, 241, 90–96, 2017.
11. Hinton G. E. and Zemel R.S.; Autoencoders, minimum description length and Helmholtz free energy. Advances in neural information processing systems, 6 3–10, 1994.
12. Ranzato, M.A., Y.-L. Boureau, S. Chopra and Y. LeCun: A unified energy-based framework for unsupervised learning. in: 11th International Conference on Artificial Intelligence and Statistics (AISTATS), San Huan, Puerto Rico, 2007, 2 371–379.
13. Le, Q.V., Ransato, M. A., Monga, R. et al. Building high level features using large scale unsupervised learning. arXiv 1112.6209 (2012).

14. Higgins, I., Matthey, L., Glorot, X., Pal, A. et al.: Early visual concept learning with unsupervised deep learning. arXiv 1606.05579 (2016).
15. Dolgikh, S.: Topology of conceptual representations in unsupervised generative models. In: 26th International Conference Information Society and University Studies, Kaunas, Lithuania (2021).
16. Shi, J., Xu, J., Yao, Y., and Xu, B.: Concept learning through deep reinforcement learning with memory augmented neural networks. Neural Networks 110, 47–54 (2019).
17. Gondara, L.: Medical image denoising using convolutional denoising autoencoders, in: 16th IEEE International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, 2016, 241–246.
18. A P S. C., Lauly S., H. Larochelle H., Khapra M. M., Ravindran B. et al.: An autoencoder approach to learning bilingual word representations. In: 27th International Conference on Neural Information Processing Systems (NIPS'14), Montreal, Canada, 2 1853–1861 (2014).
19. Zhou C. and Paffenroth R.C.: Anomaly detection with robust deep autoencoders. In: 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada, 665–674 (2017).
20. Dolgikh S.: Categorized representations and general learning. In: 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions (ICSCCW-2019), 1095 93–100 (2019).
21. Yoshida, T., Ohki, K.: Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. Nature Communications 11, 872 (2020).
22. Bao, X., Gjorgiea, E., Shanahan, L.K. et al.: Grid-like neural representations support olfactory navigation of a two-dimensional odor space. Neuron 102 (5), 1066–1075 (2019).
23. Le, Q.V.: A tutorial on deep learning: autoencoders, convolutional neural networks and recurrent neural networks. Stanford University, 2015.
24. Rodriguez, R.C., Alaniz, S., and Akata, Z.: Modeling conceptual understanding in image reference games. In: Advances in Neural Information Processing Systems, Vancouver, 13155–13165 (2019).
25. Liu W., Wang Z., Liu X. et al.: A survey of deep neural network architectures and their applications. Neurocomputing 234 11–26 (2017).
26. He K., X. Zhang S., Ren S., and J. Sun J: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778 (2016).
27. Dolgikh S.: Can jellyfish dream? Conceptual representations in unsupervised generative learning. Cambridge Open Engage doi:10.33774/coe-2021-6mh95 (2021).
28. LeCun Y.: The MNIST database of handwritten digits. Courant Institute, NYU Corinna Cortes, Google Labs, New York Christopher J.C. Burges, Microsoft Research, Redmond.
29. Keras: Python deep learning library. https://keras.io/, last accessed: 2020/11/21.
30. Dolgikh S.: Sparsity constraint in unsupervised generative learning. In: 21st Conference ITAT (Information technologies - Applications and Theory), Slovakia, 188–194 (2021).

31. Fukunaga, K., Hostetler, L.D.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory 21 (1), 32–40 (1975).
32. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), 175–185 (1992).
33. Garm, A., Poussart, Y., Parkefelt, L., Ekström, P., Nilsson, D-E.: The ring nerve of the box jellyfish Tripedalia cystophora. Cell and Tissue Research 329 (1), 147–157 (2007).
34. Roth G, Dicke U.: Evolution of the brain and intelligence. Trends in Cognitive Science 9 (5), 250 (2005).
35. Dolgikh S: Synchronized conceptual representations in unsupervised generative learning. In: 13th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2021), accepted.

Declarations

Competing interests statement The authors declare no competing interests.

Figures

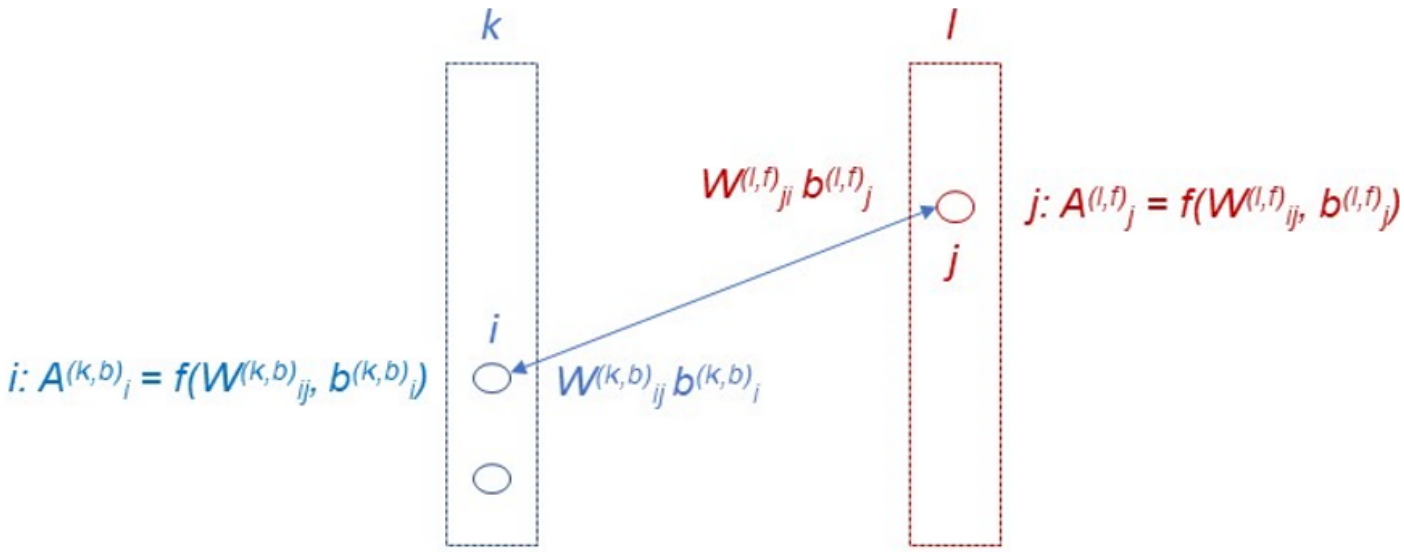

Figure 1

Bi-directional synapse.

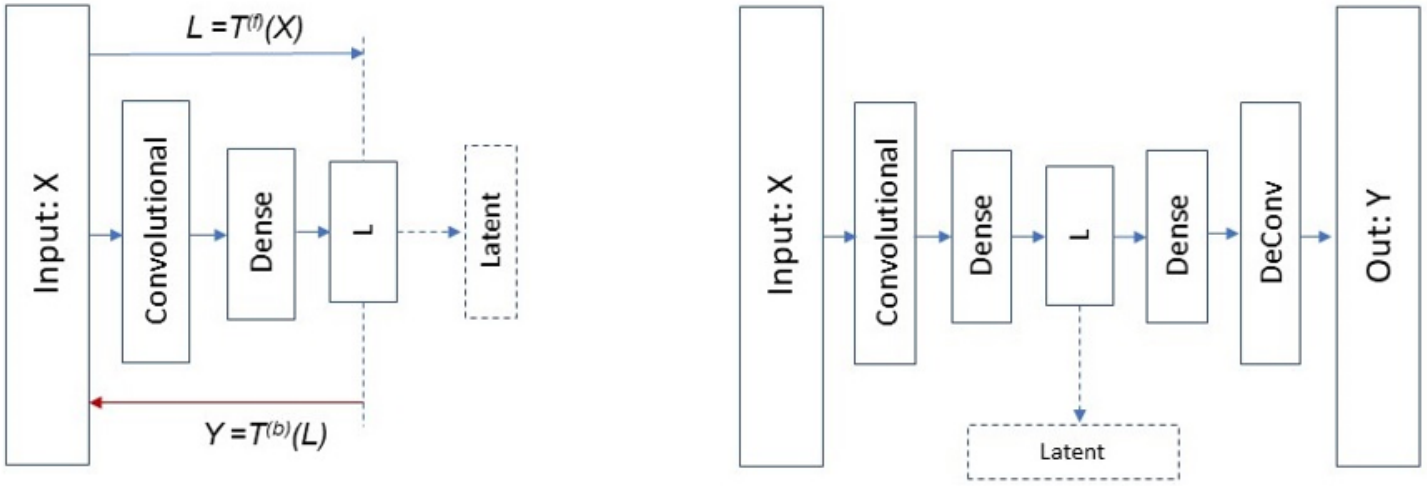

Figure 2

Training of bi-directional generative model (left) and equivalent autoencoder model (right).

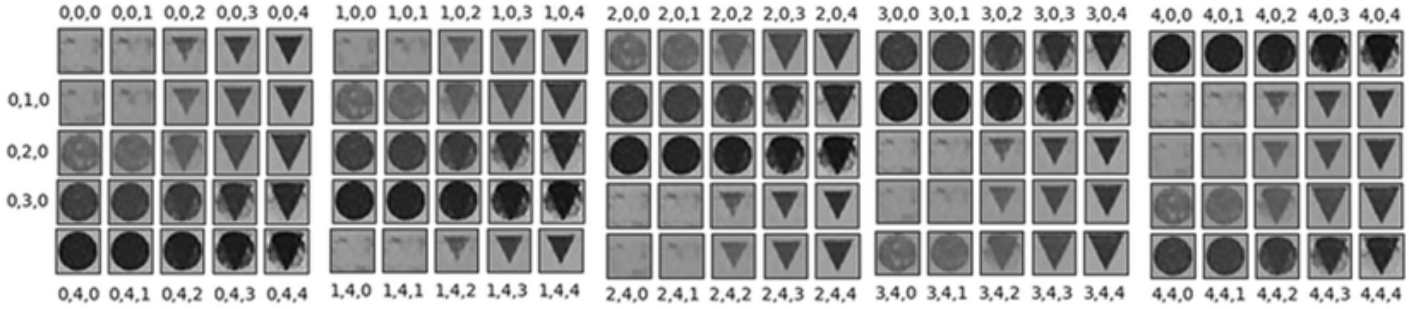

Figure 3

Generative structure of conceptual representations, L_3 model, Shapes-1,2 datasets, [14]

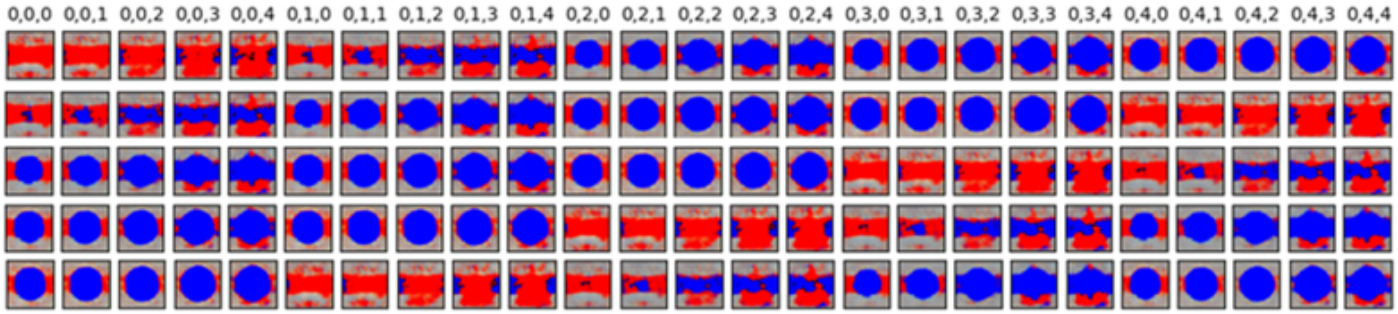

Figure 4

Generative scan, Shapes-C dataset, L_{10S} model [29]

Figure 5

Learning success, MNIST dataset (top row: inputs; bottom: generation), L_{24S} model

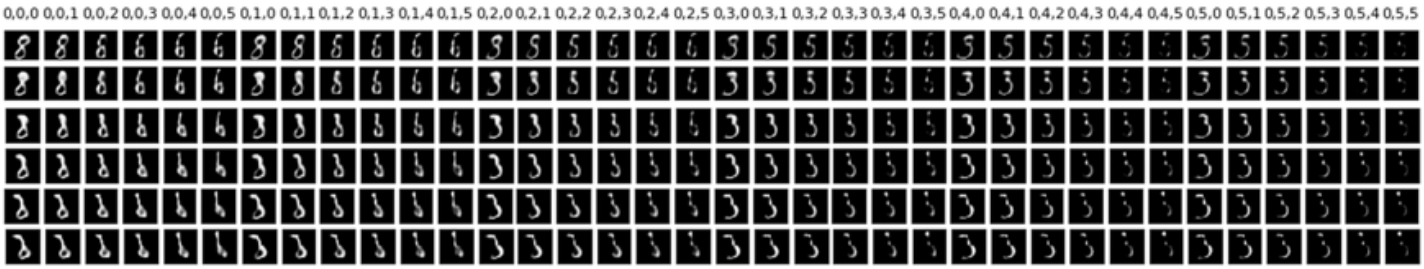

Figure 6

Generative scan, L_{24S} model, MNIST dataset