Neuro-symbolic Modeling of Visual Data with Sparse Generative Embeddings*

Serge Dolgikh[1,†] and Oleksandr Sliusarenko[1,*,†]

[1]National Aviation University, Lubomyra Huzara 1, Kyiv, Ukraine

Abstract

The problem of learning in novel environments, where training data annotated with known categories, classes or concepts, can be scarce or even absent is a well-researched field in computer science and the theory of learning systems. When prior data for the conventional methods of supervised learning is not available, a promising approach can be to use methods of self-supervised learning which are based on the ability of the learning models to create informative representations of the observable data under the incentive to produce accurate generations of the observable distribution. In this work we applied methods of unsupervised learning and dimensionality reduction in combination with a neurosymbolic approach to analyze the conceptual structure of simple visual data represented by images of basic geometric shapes and develop a process for constructing conceptual models or "maps" of the domain data that is not dependent on the availability of prior knowledge about its distribution, characteristics, content and so on. Conceptual models of general data can be useful in a number of ways, such as gaining initial insights into essential content and composition of the distribution, verification of balance of the dataset, bias and other essential characteristics.

Keywords

generative learning, unsupervised concept learning, neuro-symbolic learning, clustering

1. Introduction

Conventional methods of supervised machine learning have been demonstrated to be highly effective in multiple problems, applications and domains. In this branch of the learning theory, the conceptual structure or model of the data that is being analyzed is known from the outset in the form of the structure of known categories (classes) that annotate certain ("training") subsets of samples of the observable distribution. Such information is not always available and can be counted on when facing novel problems and domains which have not accumulated considerable amounts of prior knowledge, by any of the essential characteristics of volume, representativity, density and precision of sampling and so on.

When prior data for the conventional methods of supervised learning is not avail-able or may not be counted on, a promising approach can be found in employing methods and models of self-supervised learning, which bypass or circumvent such a dependency by instead incentivizing learning models to create informative embeddings of the observable data of the problem/domain which can pro-duce generations of the original distribution with high fidelity and accuracy. While methods of unsupervised learning and dimensionality reduction, including generative, were researched extensively in prior studies, in this work we applied them to analyze the conceptual

*IVUS2025: Information Society and University Studies 2024, May 15, Kaunas, Lithuania

[1*] Corresponding author.

[†] These authors contributed equally.

sdolgikh@kai.edu.ua (S. Dolgikh); 3615640@stud.kai.edu.ua (O. Sliusarenko)

0000-0001-5929-8954 (S. Dolgikh); 0009-0003-8532-9285 (O. Sliusarenko)

structure of simple visual data and develop a process to pro-duce conceptual models or "maps" of the data in the problem that is not dependent on the availability of large quantities of the prior knowledge about the problem/domain. Such methods can be useful in working with novel domains and environments where prior data can be scarce or non-existent.

2. Methods

2.1. Related work

The idea that essential information in certain samplings of unknown distributions i.e., "data" can be extracted with learning systems incentivized to produce com-pressed representations or "images" of it, with the ability to reproduce or generate the original distribution has long history in Machine Learning. The first such models were of the type of Restricted Boltzmann Machines (RBM) and Deep Belief Net-works (DBN) [1,2]. A variety of approaches, models and architectures of generative unsupervised and/or self-supervised learning systems were introduced and examined in the period that follows including autoencoder neural networks [3,4], Generative Adversarial Networks (GAN) [5,6] and others to name only several promising directions of recent research in a rapidly expanding field.

A number of experimental studies over the years produced results that hinted at an association of the success of generative learning and the construction of a conceptual models of the data as an aggregation of samples with certain "proximity" within a group or class. Among those, of note is the "cat experiment" [7] that demonstrated the emergence of sensitivity to visual concepts in general images on the level of a single neuron in models of deep self-supervised generative learning trained with massive sets of real images. Latent distributions of several sets of images representations produced with deep variational autoencoder [4] were examined in [8] including the perspectives of geometric differentiation of different types (disentanglement). Geometric and topological structure of distributions in generative embed-dings of images of basic geometric shapes produced with deep convolutional neural models was studied in [9]. An in-depth overview of the current approaches in representation learning was given in [10].

General and theoretical principles of concept learning are being applied to a growing number of practical problems where conceptual structures were described. Without limitation these results include general and medical imaging [11,12], language and linguistics [13], network security [14] and many other areas, domains and problems.

The results quoted above and others demonstrated that generative learning with the incentive to effectively "encode" complex real-world data in informative com-pressed representations or embeddings can produce structured informative representations correlated with characteristic types, or concepts in the original data, in process that does not require or depend on extensive prior knowledge about the distribution but rather, on the capacity of the generative models to generate effective ap-proximation of the original (observable) distributions from informative representations/embeddings of significantly reduced dimensionality.

It can be noted that concurrently with these results, recent studies in experimental neuroscience demonstrated the ubiquitous character of modeling sensory data of different types, including visual, olfactory and others with small populations of active biological neurons, effectively, low-dimensional latent representations, by animals and humans [15,16]. These results may point at interesting parallels in the learning processes of artificial and natural (i.e., biological) learning systems.

In this work, we set out from the outset to examine generative learning and the structure of informative representations produced as a result of it, from the perspective of the construction of conceptual models of the data, that is, geometric structures of characteristic general types and their distribution in the informative embed-ding spaces created by the models in the process of training. Such models can be useful in a number of ways and perspectives of investigation, such as gaining initial insights into data that does not yet have a confident association to known types or classes, verification of balance, bias and others.

2.2. Generative Neural Models and Sparse Embeddings

The rationale for selection of the architecture of neural models used in this work was their success with complex realistic data of different types as discussed in the literature section. Models based on the architecture of convolutional autoencoder neural network [3] with strong dimensionality reduction to a low-dimensional embedding were used to produce representations of a dataset of images of basic geometric shapes as described in this section. Given significant progress in the area of image recognition, in this work we chose image data as representative of realistic, real-world sensory environment of a learning system.

The models in the study were trained in the self-supervised mode, with minimization of the error of generation of the input images in learning iterations as explained in the section on training. Prior information such as annotations (labels) that de-scribed the type of shapes was not used in the training phase. Thus, the embeddings of images in the low-dimensional latent spaces obtained with these models can be considered fully unsupervised (zero-shot learning), as they did not require nor de-pended on the annotations or any other prior knowledge about the data.

The selection of the method in this study was driven by the demonstrated success of neuro-generative approaches in modeling simpler visual data [8,9], as well as the feasibility of constructing lower-dimensional symbolic structure of the data within the chosen architecture.

2.3. Generative Neural Model

In this work, we used feedforward artificial neural models of the architecture of a convolutional autoencoder. The architecture of the models can be described as a feedforward sequence of the stages of convolutional acquisition of scale-invariant features (convolution-pooling layers) followed by deep layers of dimensionality reduction with a single encoding (latent) layer of size M = 3–5 (flat architectures) or M = 6–10 (sparse architectures). The decoding / generative stage was fully symmetrical to the encoder.

An architecture diagram and the hyperparameters of the models are described in **Figure 1** and Table 1.

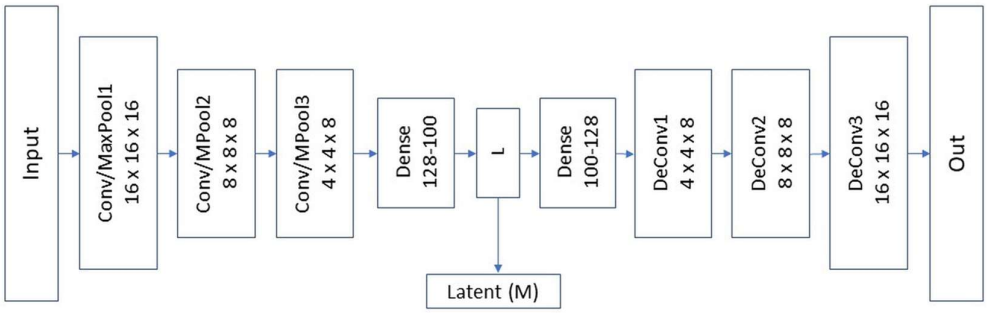

Figure 1: Convolutional autoencoder architecture with a single encoding layer.

We used two flavors of the architecture in this study: flat and sparse. Flat models had a fully interconnected embedding layer of dimension M = 3–5, that is, three to five latent neurons. Sparse models had a larger central layer, M = 6–10, with L1 sparsity activation penalty imposed in training that effectively reduced the number of the activated neurons (effective neurons) to approximately three for the images in the dataset.

Table 1
Architectural parameters of generative neural models

Type	Layers	Trainable parameters	Cost function	Comment
Flat	11	4×10^4	MSE, cross-entropy	
Sparse	15	9×10^5	Cross-entropy	Sparsity penalty L1

In the input layer of the model ("Input" layer, **Figure 1**) were the batches of color im-ages of the shape (64 × 64 × 3), which were reproduced in the output layer by the learning model. The distance between the inputs and outputs generated by the model represented the learning gradient that was backpropagated to reduce the cost function with standard learning methods of artificial neural models [17].

The feedforward type of the neural models ensured that all information necessary for generation of the images in the original (input) distribution was contained in the central (encoding) layer of the model, after the completion of the training phase. Thus, the informative embedding of the images was produced by the activations of the neurons in the central (embedding) layer of the models. The inverse, generative function of the trained models is realized by propagating the latent activation vectors: (a_1, .. a_M), through the generative (right of the central encoding layer) part of the model (**Figure 1**), with the weights adjusted during the training phase. It follows then that a trained feedforward generative model, after completion of the training phase could realize two types of mappings or transformations: the encoding, E, from the original images to their latent vector embeddings; and the generative one, *G*, in the inverse direction. The effect of self-supervised learning is that the embeddings of sets of images Ik produced by the encoding transformation, $e_k = E(I_k)$ can be effectively (that is, with high accuracy and precision) reproduced by the generative part of the model

$$I_k \cong G(e_k)=G(E(I_k)) \quad (1)$$

where $E(x)$: the embedding transformation.

2.4. Data

In this work, we used a datasets of color images of basic geometric shapes [18]. It contained color images of size 64 × 64 of seven distinct types of basic geometric shapes as follows:

- Circle: red and blue, with size variation of 0.3 – 0.9 of the image size, and a varia-tion of the intensity of the gray background.
- Triangle: red, blue, 0.3 – 0.9 of the image size, with a variation of the intensity of the gray background.
- Horizontal stripe: blue, centered horizontally, wide (0.5 – 0.9 of the image size); red, centered horizontally, narrow (0.2 – 0.5 of the image size).
- Vertical stripe: blue, centered vertically, narrow (0.2 – 0.5 of the image size).

As well, the training set contained greyscale backgrounds of size 64 × 64, of varying intensity. Examples of the images in the dataset are shown in **Figure 2**: Accuracy of generation in self-supervised generative learning.

The composition and variation of the dataset was aimed at modeling of simple yet realistic natural visual sensory environments.

2.5. Training the Generative Model

Following the phase of self-supervised training with the color images dataset, most of the models in the trained group achieved success in learning and generalization. The success of training was verified by the quality of reproduction of a representative subset of samples in the dataset that were not used in the training phase; thus, both the accuracy and generalization of the models was verified. The characteristics and results of the training process are described in Table 2.

Table 2
Training metrics

Type	Training epochs	Final accuracy	Visual accuracy	Success rate	Latent activations, max/mean
Flat	20-50	0.90-0.95	high	70-80%	0.
Sparse	15-30	0.95-0.98	high, very high	~ 90%	0.85

Notably, the models in the sparse architecture group achieved high incidence of learning success: approximately 90% and an impressive accuracy and precision in interpretation and generation of the images in the dataset as shown in **Figure 2**.

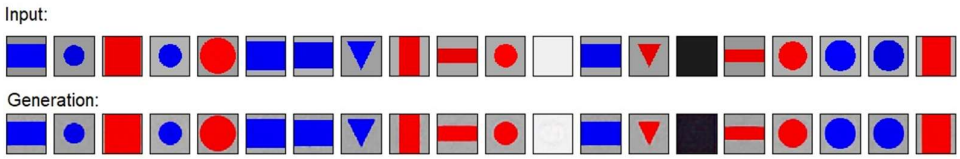

Figure 2: Accuracy of generation in self-supervised generative learning.

Two notable characteristics where significant differences were observed between the flat and sparse architecture groups were the success rate, that was higher for the sparse group (Table 1) and the volume (the absolute value) of the latent activations that were notably higher (by up to two orders of magnitude) for the models with the flat architecture of the encoding layer. This difference can be essential for realistic natural learning systems where activations of neural signals are represented by electrical microcurrents and therefore, higher activations require higher consumption of energy and other essential physical resources

2.6. Sparse Generative Embeddings

Generative models with sparse embedding layers achieve high level of learning success by modeling, i.e., encoding samples in the input distribution with sparse activations of neurons in the embedding layer. Sparsity generally means a low ratio of significant (such as, non-zero) elements relative to the size of the data vector. This is characteristic of generative neural systems regardless of origin, natural or artificial [15,16].

In artificial generative neural systems sparsity can be induced by addition of the sparsity penalty in training (for example, L1 regularization) that incentivizes models to reduce the sum of the absolute values of significant elements of the activation vector. The advantages of sparse embeddings in a self-supervised learning process will be discussed further in this work.

As a result of a sparsity penalty imposed in training, most images in the dataset produced activations of 2 to 4 latent neurons, with the average activation "dimensionality" i.e., the average number of active latent neurons of 3.2.

A sparse embedding space can be described geometrically as a set or a "stack" of effective latent projections or "slices". Indeed, considering an *M*-dimensional latent space, i.e. a model with the size

of the latent layer M, and the effective sparsity S, latent projections correspond to distinct groups of latent neurons of the size S that are activated by inputs and can be identified by the unique indexes of latent neurons.

The projections with higher populations of activations can be identified with a sufficiently representative sample of input images encoded in the embedding space; inputs are placed in a given projection if their most significant activations match the projection's unique index i.e. the tuple of latent neuron indices, such as (i, j, k), 1 .. M. Additionally, filtering for significant activations should be applied, as $\sum_L |A(l_k)| \geq a_{min}$, where $A(l_k)$: activations of the latent neurons; a_{min}: certain threshold of minimal activation in the projection determined from the distributions of latent activations in the general sample.

The result of this process is a structure of representative projections and their populations, $L = \{ (I_k, P_k) \}$ that describe the distribution of the original data (images) in the sparse embedding space. To exclude random noise in the distribution the structure of projections can be truncated at some minimal threshold of the population of the projections.

The essential projections of the sparse embedding space, identified by this method reflect characteristic structure in the distribution of the encoded dataset in the sparse embedding space. As was commented earlier, due to the ability to restore the original samples with high accuracy and precision, it has to contain or "encode" significant information about the distribution of samples in the original data.

2.7. Conceptual Structure of Generative Embeddings

An essential assumption in generative learning is that observable samples that are similar in the space of the observable parameters will be associated by proximity, in some measure, in an informative representation of reduced dimension. It is substantiated by the results of [7,8,19] and other studies. Based on it, it can be conjectured that the latent positions corresponding to the samples in the original dataset could form structures of higher concentration or density, in the latent regions associated with certain characteristics of observable parameters.

Distributions of density of the encoded samples in latent embedding spaces can be studies with methods of density clustering, that do not require prior knowledge about the data, and thus can be considered entirely unsupervised. Then a structure of density concentrations, combined with the analysis of the distribution of activations as outlined in the preceding section, can produce an accurate description or "conceptual model" of the original data. As noted, the model or factorization of data can be produced in this process without prior knowledge of its content such as annotation with known types, categories or classes or other forms of prior knowledge.

3. Results

In this study the main focus was on development of reliable methods of construction of conceptual models of data that do not depend nor require massive prior knowledge about the domain or the distribution. The data was chosen as an example of a visual sensory environment that is both simple and yet can be realistic for some intelligent systems. Construction of conceptual models from samplings of the environment can be useful in interpretation of inputs and construction of differentiated intelligent responses to sensory stimuli.

For experiments described in this section we chose a small subset of generative models that achieved success as described in 2.4. Data. Due to limitation of for-mat, the focus of this study was on verification of the viability of the approach in construction of conceptual models of sensory data, whereas other relevant directions such as consistency of conceptual models between individual instances of learning systems, statistical confidence and others will be addressed in another work.

3.1. Construction of Conceptual Models of Visual Data

We constructed conceptual models of generative embeddings of data in the study according to the process described earlier with several instances of sparse generative models with the sparse latent layer $M = 6$ that successfully completed the phase of self-supervised training. Density clustering method MeanShift [20] was used to identify concentration clusters in the dominant projections of sparse generative embeddings.

Table 3 provides a summary of the observed distribution of the concentration clusters in the sparse projections with the population above 50% of the overall sample (the dominant projections).

Table 3
Distributions in the dominant projections of sparse generative embeddings.

Projection, by rank, neuronal index	Relative population	Number of clusters	Description
(1,4,6)	0.76	23	blue types, minus BHS
(4,5,6)	0.68	18	all red types
(1,4,5)	0.66	16	blue types
(1,5,6)	0.55	16	BHS, blue and red circles
(1,2,4)	0.51	11	similar to (1,4,5)

Observable images that correspond to the centers of the concentration (density) clusters identified in the dominant projections are shown in **Figure 3**: Concentration clusters in the dominant projections: generations of cluster centers. Outlier clusters with the population below 3% of the projection sample were "pruned" and excluded from the subsequent analysis.

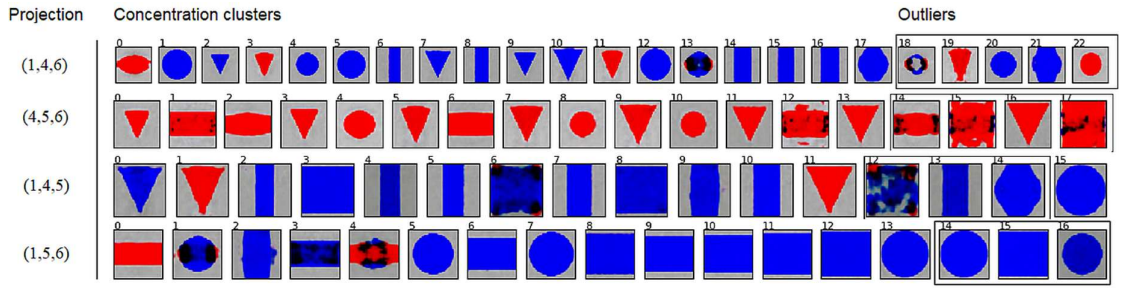

Figure 3: Concentration clusters in the dominant projections: generations of cluster centers.

The consistency of the composition of the concentration clusters identified with this method was verified as well: over 90% of the samples that were placed in the same concentration cluster were of the same observable type (i.e., the type and color of the geometric shape).

The character of density clusters identified with this method was similar for all models used in the study, though positions, number and specific type of clusters were model-specific.

3.2. From Concentration Clusters to Concept Prototypes

The structure of concentration clusters C_L obtained in the preceding section represent an initial course approximation of the conceptual model of the data. One can recall that a) the compressed or "encoded" information about the original distribution is contained, entirely in the latent layer of the models and can be restored or generated with the generative transformation (1); and b) the concentration clusters in the dominant projections represents characteristic regions in the sparse

latent space where significant fractions of the samples in the encoded distribution were concentrated. Thus, the structure of concentration clusters represents a model of the original data D that can be translated to an approximating distribution $\tilde{D}$ by inflating it into the embedding dimension (that can be achieved by padding the remaining latent positions with zeros) and propagating to the observable space via generative transformation:

$$\tilde{D}=G(C_L \uparrow M) \tag{2}$$

In the next step, we attempted to further refine the conceptual model of the data by removing the remaining redundancy of multiple clusters representing the same type of the images. An effective approach was found by applying a "neuro-symbolic aggregation" of the concentration clusters leading to the resulting structure of the concept prototypes of the original distribution D.

In general terms, the process can be described as follows:

We begin traversing the structure of concentration clusters C_L described earlier in the descending order of the dominant projections by the population and in the projection, in the descending order of the clusters by population (with the population cutoff applied). This traversal sequence ensures that every cluster in the structure will be visited in the process.

Next, from the input structure of the concentration clusters C_L, the structure of concept prototypes C_P is calculated by the following sequence of operations:

1. For a position i in C_L, let $V(i)$ be the observable representative of the cluster;
2. to $V(i)$: apply uniform scaling to the standard size (e.g. 0.75 of the image size) if the non-background area is smaller than the frame of the image (that is of constant size, 64 × 64)
3. else if the non-background area equals the frame of the image in one dimension, apply standard scaling in the opposite direction to the standard size (e.g. 0.75).
4. compare the resulting "standardized" representative image of the cluster $V_S(i)$ to the prototypes in P_j in C_P for example, by the standard metric such as L2 (squared distance) norm. If there is a match, i.e., the $L2(V_S(i), P_j) \leq \varepsilon$, where ε, predetermined threshold, assume that cluster i is associated with a known type image, P_j.
5. else, append $V_S(i)$ to C_P.
6. continue until every cluster in the structure C_L has been visited.

As a result of this process, one obtains the new structure of concept prototypes, or common general types in the domain data associated with the regions of distribution of the type in the sparse embedding space represented by clusters, possibly multiple, in $\tilde{D}$.

As a demonstration of the outlined method with the model used to construct the cluster structure the following conceptual model of the dataset of images was obtained. The description of the distribution region of a certain distinct type of shape is the index of the projection ordered by population, followed by a list of indices of clusters identified as being in the relation of similarity by the neurosymbolic algorithm.

The resulting conceptual model of seven types and 58 concentration clusters in significant projections/slices of the sparse embedding space contains all types of shapes that were present in the dataset. It can be added that the geometric positions of the clusters (geometric centers) can be obtained with the proposed method as well, how-ever while conceptual frameworks were consistent across several tested models, it was observed that latent coordinates have significance only for each individual model and cannot be interpreted directly, without some process of symbolic synchronization by other models even of the same architecture. Further examination of this question will be attempted in a future study.

It can be noted in the conclusion that at no point in the analysis of the conceptual structure annotated data was used. Thus, conceptual modeling of complex realistic data can be performed with this method in a completely unsupervised, zero-shot learning process that does not depend on nor require significant prior information about the data or the domain.

Table 4
Concept model, geometric shapes dataset

Concept prototype	Concentration clusters (projection rank, list of clusters)
Red circle (RC)	(2, (4, 9, 11))
Red triangle (RT)	(1, (4, 12)), (2, (2, 3, 7 ,13)), (3, (2, 12))
Blue circle (BC)	(1, (3, 8, 10, 11)), (2, (1, 4, 8, 10, 12, 14)), (4, (2, 6, 8, 14))
Blue triangle (BT)	(1, (4, 12)), (3, (1))
Red horizontal stripe (RHS)	(1, (2, 5, 6, 13, 18)), (2, (2, 3, 7, 13)), (4, (1))
Blue horizontal stripe (BHS)	(3, (4, 9)), (4, (4, 7, 9–13))
Blue vertical stripe (BVS)	(1, (7, 9, 15–17)), (3, (5, 6, 8, 10, 11)), (4, (3))

4. Conclusions

The method of conceptual analysis and modeling of sensory data described and demonstrated in this work has proven effective in determining the conceptual structure of a sampling of images of basic geometric shapes that can model some simple yet realistic visual environments. Conceptual models obtained with the method contained all types of shapes in the dataset and allowed to associate them to characteristic regions in sparse generative embeddings. The constructed conceptual structures were consistent between different learning modes.

The results of this study with sparse generative embeddings of relatively simple generative neural models underline the extraordinary ability of sparse embeddings to model and store information about the conceptual structure of the data. Indeed, a single sparse latent layer of dimension M contains ~ M^3 or M^4 of distinct three- or four-dimensional projections, each having the capacity for modeling several distinct visual types of images (visual types or concepts). Then, the entire English vocabulary of concrete nouns that correspond to distinct types of concrete observable objects can be modeled by a generative system of this type with only about 25 latent neurons [21]. Such efficiency in conceptual modeling can be a key factor that made neural cognitive architectures the solution of choice for intelligent biological systems. Coincidentally or not, recent results in experimental neuroscience pointed at the ubiquity of sparse low-dimensional representations of sensory data in animals and humans as cited earlier.

The results of our work connect with other results in the analysis of visual data with neurosymbolic models, such as those by Khan et al. [22] and Abdessaied et al. [23], demonstrating the promise of combining deep learning with symbolic reasoning for visual understanding and dialog.

There are several directions in which the approaches and ideas developed in this work can be extended. Conceptual models developed with the proposed methods can be used in the initial classification and recognition of characteristic types of sensory inputs with minimal need for prior knowledge, that is, zero-shot learning. The effectiveness and accuracy of these methods in the recognition and classification of data into the identified general types can be studied in future works. Another interesting direction of future study is the consistency of conceptual models between individual learners and the possibilities for synchronization and exchange of information in an ensemble of learners as briefly mentioned earlier. While some interesting initial results have been obtained in this area, they certainly merit another, dedicated study of their own.

Declaration on Generative AI

The authors have not employed any Generative AI tools in preparation of this work.

References

[1] Hinton, G., Osindero, S., Teh Y.W.: A fast-learning algorithm for deep belief nets. Neural Computation 18(7), 1527–1554 (2006).

[2] Fischer, A., Igel, C.: Training restricted Boltzmann machines: an introduction. Pattern Recognition 47, 25–39 (2014).

[3] Le, Q.V.: A tutorial on deep learning: autoencoders, convolutional neural networks and recurrent neural networks. Stanford University (2015).

[4] Welling M. and Kingma D.P.: An introduction to variational autoencoders. Foundations and Trends in Machine Learning, 12(4), 307–392 (2019).

[5] Creswell A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta B. et al: Generative adversarial networks: an overview. IEEE Signal Processing Magazine, 35(1) 53–65 (2018).

[6] Partaourides, H., Chatzis, S.P.: Asymmetric deep generative models. Neurocomputing, 241, 90–96 (2017).

[7] Le, Q.V., Ranzato, M.A., Monga, R., Devin, M., Chan, K. et al. Building high level features using large scale unsupervised learning. In: 29th International Conference on International Conference on Machine Learning ICML'12, pp. 507–514 (2012).

[8] Higgins, I., Matthey, L., Glorot, X., Pal, A., Uria, B. et al.: Early visual concept learning with unsupervised deep learning. arXiv 1606.05579 (2016).

[9] Dolgikh, S.: Topology of conceptual representations in unsupervised generative models. In: 26th International Conference Information Society and University Studies (IVUS-2021), Kaunas, Lithuania CEUR-WS.org 2915, pp. 150–157 (2021).

[10] Bengio, Y., Courville, A., Vincent, P.: Representation Learning: a review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35, 1798–1828 (2012).

[11] Gondara, L.: Medical image denoising using convolutional denoising autoencoders. In: 16th IEEE International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, pp. 241–246 (2016).

[12] Shi, J., Xu, J., Yao, Y., Xu, B.: Concept learning through deep reinforcement learning with memory augmented neural networks. Neural Networks 110, 47–54 (2019).

[13] APS, 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, pp. 1853–1861 (2014).

[14] Zhou C., Paffenroth R.C.: Anomaly detection with robust deep autoencoders. In: 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada, pp. 665–674 (2017).

[15] Yoshida, T., Ohki, K.: Natural images are reliably represented by sparse and variable populations of neurons in visual cortex. Nature Communications 11, 872 (2020).

[16] 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).

[17] Spall, J.C. Introduction to stochastic search and optimization: estimation, simulation, and control. Wiley, Hoboken, New Jersey (2003).

[18] Dolgikh, S.: Geometric shapes datasets, Mendeley Data, V1 (2024). doi: 10.17632/ypfcrfxjgk.1.

[19] Dolgikh, S.: Categorization in unsupervised generative self-learning systems. International Journal of Modern Education and Computer Science, 13 (3) 68–78 (2021).

[20] 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).

[21] Brysbaert, M., Warriner, A.B., Kuperman, V.: Concreteness ratings for 40 thousand generally known English word lemmas. Behavioral Research 46, 904–911 (2014).
[22] Khan MJ, Ilievski F, Breslin JG, Curry E.: A survey of neurosymbolic visual reasoning with scene graphs and common sense knowledge. Neurosymbolic Artificial Intelligence 1, (2025).
[23] Abdessaied, Adnen & Bâce, Mihai & Bulling, Andreas. Neuro-Symbolic Visual Dialog. 10.48550/arXiv.2208.10353. (2022).