

Latent Representations of Terrain in Aerial Image Classification

Pylyp Prystavka¹, Serge Dolgikh², Olga Cholyshkina³ and Oleksandr Kozachuk¹

¹ National Aviation University, 1 Lubomyra Huzara Ave, Kyiv, 03058, Ukraine

² Solana Networks, 301 Moodie Drive, Ottawa, K2H 9C4, Canada

³ Interregional Academy of Personnel Management, 2/16 Frometivska St., Kyiv, 03039, Ukraine

Abstract

Investigation of informative representations of complex data is a rapidly developing field of research in machine learning. In this work we present a process of production and analysis of informative low-dimensional latent representations of real-world image data with neural network models of unsupervised generative learning. A model of convolutional autoencoder based on VGG-16 architecture was used to produce low-dimensional latent representations of aerial image data and the characteristics of distributions of several higher-level classes of terrain types were studied. The analysis of distributions demonstrated a landscape of compact concept clusters for most studied types of terrain with good separation between concept regions. The results of this work can be used in developing methods of effective learning with minimal labeled data based on the emergent concept-sensitive structure in the latent representations.

Keywords

Artificial Intelligence, unsupervised machine learning, clustering, image recognition, image classification

1. Introduction

Classification of complex data such as images represents a significant challenge in the areas with severe deficit of labels. In these problems and applications, unsupervised machine learning methods such as processing data with models of unsupervised generative self-learning has shown to be effective in identification and selection of in-formative latent representations that can simplify subsequent classification and significantly reduce the label requirement. The motivation of this work was to apply these methods to the practical task of aerial image recognition where a specific set of classes combined with a strong deficit of labels make application of standard methods of supervised classification challenging.

1.1. Related Research

Informative representations obtained with models of unsupervised generative self-learning were used in a number of applications to identify the concepts or classes of interest in the observable data. Artificial neural networks have strong potential in such problems and applications due to their capability of universal approximation [1,2], making them suitable for processing data of virtually any type and complexity including live image data recorded in aerial surveillance.

Recognition and classification of terrain images to identify objects and structures such as roads, pathways, tracks of transport are challenging tasks due to variety of backgrounds and environments that can be encountered in obtaining the data in the real-world environment under different conditions, that

is further complicated by poor or inconsistent between different studies formalization of classes and severe deficit of labelled samples [3-5].

Hierarchical representations of observable data were obtained in a completely unsupervised training process with Restricted Boltzmann Machines (RBM) and Deep Belief Networks (DBN) [6,7] offering a noticeable improvement in the quality of subsequent supervised learning. Different types, architectures and flavors of generative models were investigated since including autoencoder neural networks, Generative Adversarial Networks (GAN) [8,9] to name only a few in a rapidly expanding field, resulting in improved accuracy and versatility of the models with virtually unlimited range of applications. The relations between learning and statistical thermodynamics was studied in [10] leading to understanding of a deep connection between learning processes in artificial neural models and principles of information theory and statistical thermodynamics.

Previous results in unsupervised representations with generative self-learning neural network models include applications of deep autoencoder models of different architectures such as sparse, variational, convolutional and others to create informative representations of image [11,12] and other types of data [13,14]. These results have demonstrated that categorization of data by common higher-level concepts in the latent representations under certain constraints imposed in training can be considered a general effect of information processing in such models [15]. An unsupervised structure of this type that does not require massive amounts of labeled data to identify can be harnessed for more effective learning in the environments with strong deficit of labels and/or new and unknown environments where labeled data is scarce. Given the constraints of the problem and the results discussed earlier, it was hypothesized that applying the methods and models of unsupervised generative learning to this problem may allowed to obtain informative representations of image data and reduce the requirement for labeled data to achieve successful learning.

1.2. Motivation

The motivation of this work suggested by earlier results [5-8] was to investigate the structure that emerges in unsupervised representations of generative models with real-world image data and introduce methods of production, evaluation and analysis of informative latent representations that can be used in developing of effective learning models with reduced requirements of labeled training data.

To solve the problem of strong label deficit methods of creating informative representations with deep neural network models of unsupervised generative learning were applied, that are capable of learning essential patterns in the observed data in the unsupervised mode without any labeled data. The novelty of the proposed approach is a successful application of generative models to real-world data such as aerial images, allowing eventually to perform processing on board of an autonomous vehicle; secondly, designing and demonstrating the methods of evaluation and measurement of latent representations, including entirely unsupervised ones.

The dataset of images recorded in real surveillance of terrain from an aerial vehicle was chosen for two main reasons: first to demonstrate that the methods developed in this work can be applied to realistic complex data types; and not in the least, because the tasks aerial image classification and interpretation are becoming increasingly common in many practical applications.

2. Methods and Data

This section contains the description of the model, data and methods used to produce and analyze the distributions of the characteristic classes of data in the input image dataset.

2.1. Methodology

The models were of the type of deep convolutional autoencoder neural networks [16] based on VGG architecture that achieved good results in supervised learning of image data [17]. The models were implemented in Tensorflow and Keras [18] with a number of common machine learning packages and libraries.

Generative neural network models were used as described further in this section to produce compressed latent representations of aerial terrain observation images represented in the dataset to evaluate distributions of classes identified by a type of terrain the latent space. Methods of geometrical analysis were applied to samples transformed to latent representation to identify characteristic parameters of distributions, such as compactness and separation between distribution regions of different classes.

2.2. Generative Model Architecture

The architecture diagram of the models used in this work is given in Figure 1. It can be described as a deep convolutional autoencoder neural network that contained the encoder model with several convolutional blocks, activation and normalization layers producing a flattened numerical representation with dimensionality $8 \times 8 \times 256$; and the generator model with up-sampling blocks with the resulting output layer of the same dimensionality as the input layer.

The models were trained in the unsupervised mode, with unlabeled raw image data, to reduce the generative error, i.e. the mean deviation of input images in the training dataset from their regeneration by the model as:

$$E = \text{mean}(|G(S) - S|) \rightarrow \min \quad (1)$$

where S is the training sample, $G(S)$, the output of the model on the training sample.

The architecture of the model is shown in **Figure 1**.

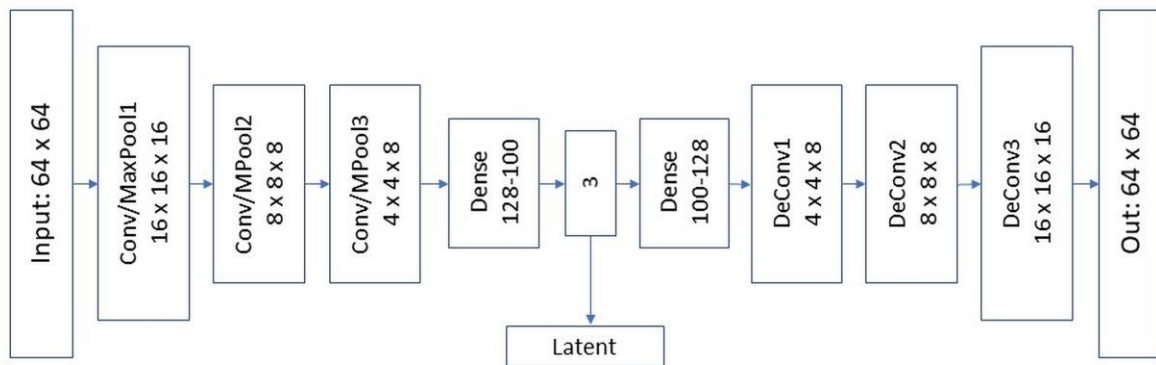


Figure 1 Deep convolutional autoencoder with latent representation of image data

A trained model can produce encoded representation of the input sample X in the observable space by transforming it to the latent representation space defined by activations of the neurons in the latent layer of the model as:

$$r = E(X) \quad (2)$$

where $E(X)$ is the generating stage of the model, from observable input to the latent representation layer (**Figure 1**).

It needs to be noted that in a trained generative model the encoding transformation (1) is defined in completely unsupervised process and does not require for training any samples labeled with known categories of the observable data.

2.3. Data

The dataset of images was obtained in live aerial surveillance of the terrain with preprocessing of scaling to the standard size (64×64) and augmentation by rotation. Images in the dataset were classified in semi-automatic process into classes representing characteristic types of terrain with significant representation in the dataset. In the rest of the study, classes or categories of images were denoted with a symbol, such as “T” for transport tracks, “W” for wooded areas and so on.

The detailed composition of the dataset is described in Table 1.

Table 1
Terrain image dataset

Class	Symbol	Size	Description
Transport tracks	T	9327	Tracks on soil, field or dirt road
Narrow dirt road	N	11700	Dirt roads
Built areas	B	1600	Roads in build areas
Wooded areas	W	9766	Roads in wooded areas
Paved roads	H	955	Paved roads, highways
Footpaths	F	0.2	Footpaths, trails
Wide dirt road	D	1411	
Other	O	22466	Not roads; general background

2.4. Training

The models were trained in unsupervised process with minimization of the deviation between the training set of images and their regenerations by the model. Cost functions used for unsupervised training were Mean Squared Error (MSE) and binary cross entropy (BCE), both showing strong improvement in the process of training (**Figure 2**).

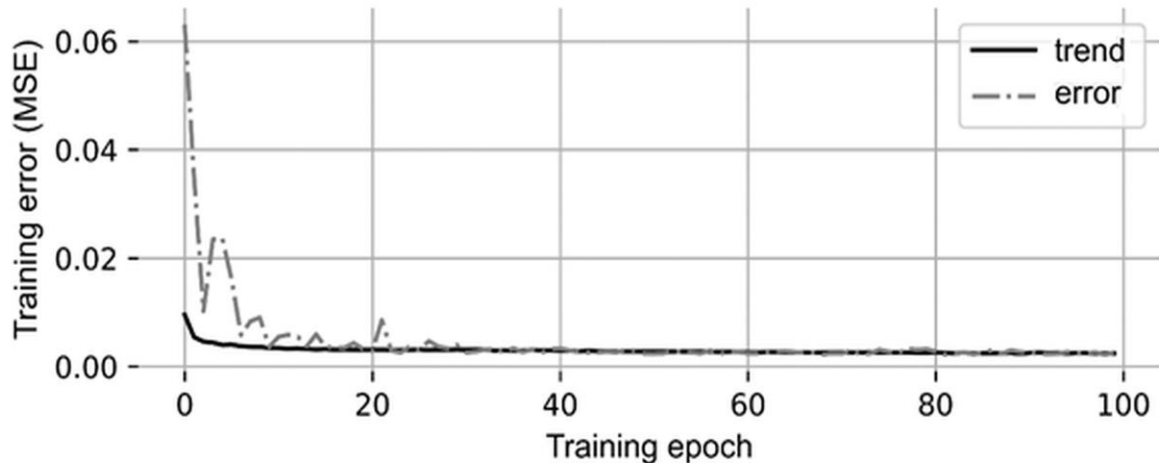


Figure 2. Training error and trend, unsupervised generative self-learning.

Strong reduction of generative error in the process of unsupervised training indicated that the latent representations created by the models contained sufficient information to regenerate observable data, because in a feedforward neural network of the type used in the study the output has to be generated entirely from the information contained in the latent representation.

3. Results

In this section we present the results of measurement and analysis of distributions of higher-level concepts in the original (observable) dataset in the latent representations produced with generative training of unsupervised autoencoder models as described in the previous sections.

3.1. Overall Characteristics

The characteristics of the general representative set of samples in the latent representation, without breakdown by higher-level concepts were as follows:

The analysis of principal components [19] produced the following results:

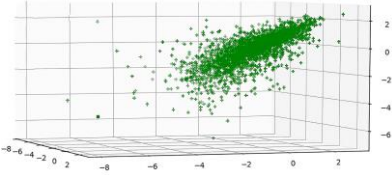
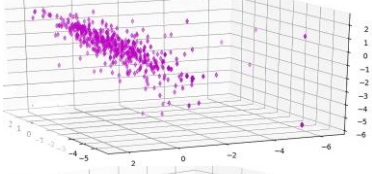
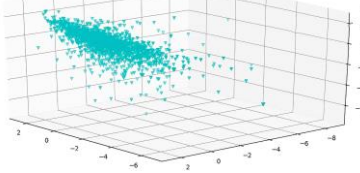
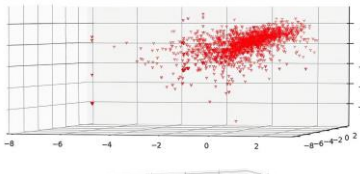
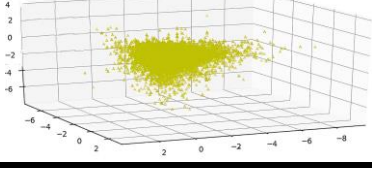
First three components: 54.4% of overall variation
 First 10 components: 72.7 %
 First 100 components: 98.2% of overall variation.

These results indicated the possibility of strong redundancy reduction in the latent representation without significant loss of information. Based on these results in the analysis of class distribution three principal dimensions with the highest variation were used, which allowed to produce direct visualizations of concept distributions in the latent representation.

3.2. Latent Concept Distributions

In this section the results of the measurement and analysis of concept distributions in the latent representation of generative models are presented. The parameters of distributions such as the characteristic size, standard deviation and density of the concept distribution regions in the latent representation coordinates are given relative to the maximum dimension of the overall latent dataset, and the uniform density.

Table 2
 Concept distributions in the latent space

Class	Parameters (Size, STD, Density)	Visualizations
Narrow dirt roads	0.22, 0.41, 0.28 0.53 – 0.91 39.6	
Wide dirt roads	0.28, 0.26, 0.35 0.69 – 0.9 39.2	
Tracks	0.39, 0.37, 0.16 0.73 – 1.18 43.3	
Paved roads highways	0.25, 0.26, 0.05 0.15 – 0.73 307.7	
Footpaths trails	0.36, 0.36, 0.23 1.0 – 1.15 33.5	

For most concepts with significant representation in the dataset, a compact and well-defined character of latent concept distributions was observed with the density of the concept region (i.e. the region of distribution of samples associated with the studied concept in the latent representation space) significantly higher than uniform.

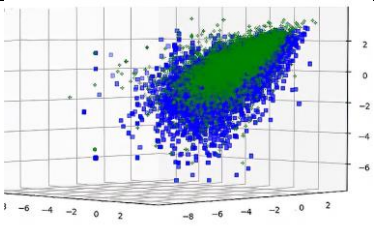
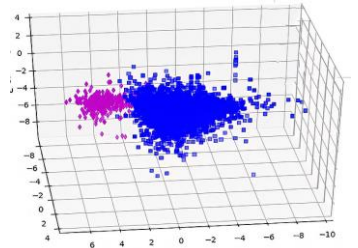
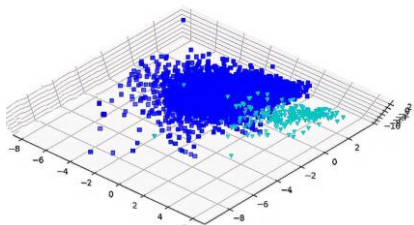
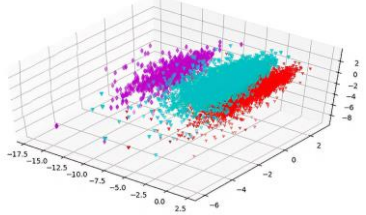
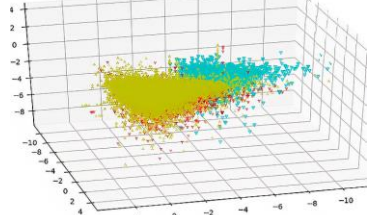
These results confirm the earlier observed effect of correlation between unsupervised latent distributions and higher-level concepts with strong representation in the observable dataset [14]. An in-depth analysis of distributions will be attempted in a future study.

3.3. Unsupervised Categorization

In this section, measurements related to categorization capacity of the models are presented and discussed. Shown in Table 3 are visualizations of intersections of concept regions, with the highest relative volume. A cross-concept intersection matrix can be defined to indicate the degree of disentanglement of concept regions in the representation, as the ratio of the latent volume of the overlapping region between concepts A and B , $O_{a,b}$ to the volume of the concept region A , H_a :

$$Y_{a,b} = \frac{V(O_{a,b})}{V(H_a)} \quad (2)$$

Table 3
Cross-concept overlap matrix Y , latent representation

Class	Concept intersection Y , maximum	Visualization
N	0.22	
B	0.03	
W	0.08	
H	0.11	
F	0.18	

As can be seen from the results in Table 3, a good separation of concept regions was observed for most categories of images with significant representation in the dataset, indicating strong categorization achieved by the models in the process of unsupervised generative learning.



Unsupervised categorization or decoupling of higher-level concepts in the unsupervised latent representations is the effect observed in a number of experiments with unsupervised self-learning models [6-8] that is evident as compact and well-separated concept distributions in the latent space.

Accordingly, categorized distributions tend to minimize the volume and consequently, maximize the relative density of latent concept distributions while minimizing the overlapping between the distributions of different concepts.

3.4. Generative Ability

Models that were successful in identifying characteristic patterns or concepts in the observable data as a result of generative self-learning can be expected to be able to regenerate input samples with close resemblance / low deviation from original input samples. To evaluate generative ability of the models, experiments were performed with samples of the concepts present in the training dataset, as well as those that were not classified into concepts (i.e. general background). Examples of generative results of the models are shown in Table 4.

Table 4
Generative ability of self-learning models

Sample	Original	Generated
Concept (Field)		
Concept (Built area)		

A successful generative ability across multiple classes of data indicates that the distributions in the latent representation were correlated with characteristic patterns in the original (i.e. observable) data. It follows from the architecture of feed-forward artificial neural networks that all the information necessary for generation of the output of the model must be contained in the latent representation layer and therefore the process of unsupervised generative training was able to produce informative representation with substantially reduced redundancy in the observable data represented in the training dataset.

Generative ability of self-learning models such as those studied in this work can be used in augmentation of training datasets for models of supervised learning, to improve the accuracy and extend classification ability to classes of data under represented in the datasets. It will be investigated in more detail in another study.

4. Discussion

The results presented in this work are in agreement with the growing number of reports with observations of the effects of concept-correlated latent representations emerging in unsupervised generative learning and provide strong additional arguments in support of the general character of this effect. Given the wide range of models and types of data, from lower complexity [14] to massive and complex architectures [11,12] where unsupervised categorization in the latent representations in generative self-learning has been observed, this conclusion appears to be well substantiated.

It was demonstrated that under certain constraints such as generative accuracy and information compression or redundancy reduction, low-dimensional latent representations of models learning to generate input distributions in a completely unsupervised learning process can produce distinct structure that is correlated with principal higher-level concepts in the observable data. The results in Sections 3.2, 3.3 on measurement of latent distributions of main concept regions appear to support this conclusion.

Identification and description of unsupervised latent structure that emerges in generative learning can be a valuable instrument in the analysis of general data, in particular, of less known origin where massive prior knowledge such as large labeled datasets used in supervised machine learning may not be available.

A number of recent results indicated that similar low-dimensional representations can play an important role in processing of sensory data by humans [20,21]. Demonstration of success of low-dimensional representations obtained with models of generative self-learning supports the conclusion about the general nature of the observed effect in the learning systems of biological and artificial origin and provides an intriguing possibility of a connection to bioinformatics [22] with learning systems able to learn intuitively, incrementally and with minimal prior knowledge of the environment.

5. Conclusions

Based on the results reported in this work, several essential observations can be made on the process of unsupervised generative learning with real-work image data and the conceptual structure in the latent representations of data produced by such models:

1. The models of generative unsupervised learning used in the study were capable of producing well-defined categorized representations correlated with the principal (i.e. strongly represented) higher-level concepts in the training dataset.
2. The observed latent representations showed good categorization and separation of principal concepts and appear to support the hypothesis [5] of a correlation between the unsupervised representation structure emergent in unsupervised generative learning and higher-level concepts with significant representation in the training data.
3. Methods of measurement of categorization capacity of unsupervised generative models in the latent representations were defined and validated.
4. The models showed good generative capacity for some principal concepts in the training data. Optimization of models for generative ability will be further investigated in a future study.
5. Methods of evaluation of categorization ability of models are general and can be applied to different types of data and model architectures.

Overall, the observed latent representations showed good categorization and separation of principal concepts and appear to support the hypothesis [5] of a correlation between the categorization performance and architecture of the model.

The methods of evaluation of latent distributions of data classes, demonstrated and verified in this work are of general nature not limited to a specific type of data and can be instrumental in evaluation of the learning capacity and performance of generative models. The unsupervised latent structure demonstrated in this and other works can be used to enhance learning ability of the models in the environments with strong deficit of labels. These findings can therefore be instrumental in development of learning models and methods that are capable of acquiring knowledge in a flexible and environment-driven process that is closer to learning of biological systems.

References

- [1] 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, pp. 215– 223, 2011.
- [2] Hornik, K., Stinchcombe M., White H., “Multilayer feedforward neural networks are universal approximators”, *Neural Networks*, vol. 2 (5), pp. 359–366, 1989.
- [3] Marfil R., Molina-Tanco L., Bandera A., Rodriguez J.A. and Sandoval F., “Pyramid segmentation algorithms revisited”, *Pattern Recognition*, vol. 39 (8), pp. 1430 – 1451, 2006.
- [4] Chyrkov A., Prystavka P., “Suspicious Object Search in Airborne Camera Video Stream”, in: Hu Z., Petoukhov S., Dychka I., He M. (eds) *Advances in Computer Science for Engineering and Education, ICCSEEA 2018, Advances in Intelligent Systems and Computing*, vol. 754, Springer, Cham, pp. 340 – 348, 2018.
- [5] Prystavka P., Cholyskhina O., Dolgikh S. and Karpenko D., "Automated object recognition system based on convolutional autoencoder," 10th International Conference on Advanced Computer Information Technologies (ACIT-2020), Deggendorf, Germany, pp. 830–833, 2020.
- [6] Fischer A., Igel C., “Training restricted Boltzmann machines: an introduction”, *Pattern Recognition*, vol. 47, pp. 25–39, 2014.
- [7] Hinton G. E., Osindero S., Teh Y.W., “A fast learning algorithm for deep belief nets”, *Neural Computation*, vol. 18 (7), pp. 1527–1554, 2006.
- [8] Welling M. and Kingma D.P., “An introduction to variational autoencoders”, *Foundations and Trends in Machine Learning*, vol. 12 (4), pp. 307–392, 2019.
- [9] Creswell A., White T., Dumoulin V., Arulkumaran K., Sengupta B., Bharath A.A., “Generative adversarial networks: an overview”, *IEEE Signal Processing Magazine*, vol. 35, (1), pp. 53–65, 2018.
- [10] Ranzato, M. A., Boureau Y.-L., Chopra S., LeCun Y., “A unified energy-based framework for unsupervised learning”, in: 11th International Conference on Artificial Intelligence and Statistics (AISTATS), San Juan, Puerto Rico, 2007, vol. 2, pp. 371–379.
- [11] Le, Q.V., Ransato, M. A., Monga R., et al., “Building high-level features using large scale unsupervised learning”, arXiv 1112.6209 [cs.LG] 2012.
- [12] Higgins, I., Matthey, L., Glorot, X., Pal, A., et al., “Early visual concept learning with unsupervised deep learning”, arXiv1606.05579 [cs.LG], 2016.
- [13] Shi, J., Xu, J., Yao, Y., and Xu, B., “Concept learning through deep reinforcement learning with memory-augmented neural networks”, *Neural Networks*, vol. 110, pp. 47–54, 2019.
- [14] Dolgikh S., “Categorized representations and general learning”, in: Proceedings of 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions, vol. 1095, pp. 93–100, 2019.
- [15] Tishby N., Pereira F. C., Bialek W., “The Information Bottleneck method”, arXiv:physics/0004057, 2000.
- [16] Kavukcuoglu K., Sermanet P., Boureau Y. L., Gregor K., Mathieu M., Cun Y., “Learning convolutional feature hierarchies for visual recognition”, *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, vol. 1, pp. 1090-1098 Vancouver, Canada, 2010.
- [17] Simonyan K. and Zisserman A., “Very deep convolutional networks for large-scale image recognition”, arXiv,1409.1556 [cs.LG], 2014.
- [18] Keras: The Python Deep Learning library, URL: <https://keras.io>.
- [19] Jolliffe, I.T., “Principal Component Analysis”, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28.
- [20] Yoshida, T., Ohki, K., “Natural images are reliably represented by sparse and variable populations of neurons in visual cortex”, *Nature Communications*, vol. 11, pp. 872 2020.
- [21] Bao X., Gjorgieva E., Shanahan L.K., Howard J. D., T. Kahnt T., Gottfried J. A., “Grid-like neural representations support olfactory navigation of a two-dimensional odor space”, *Neuron*, vol. 102 (5), pp. 1066–1075, 2019.

[22] Hassabis D., Kumaran D., Summerfield C. and Botvinick M., “Neuroscience inspired Artificial Intelligence”, *Neuron* vol. 95, 245-258, 2017.