The European Proceedings of
Social & Behavioural Sciences
EpSBS

Future Academy ISSN: 2357-1330

https://dx.doi.org/10.15405/epsbs.2018.11.02.56

ICPE 2018
International Conference on Psychology and Education

DETERMINATION OF THE ARCHITECTURE OF THE NEURAL NETWORK FOR RECOGNITION ALGORITHM UHRSI

Anna A. Ostrovskaya (a)*, Nikita E. Semenov (b), Anton O. Rubtsov (c)
*Corresponding author

(a) Peoples' Friendship University of Russia, 117198, ul. Miklouho-Maclay, 6, Moscow, Russia
ostrovskaya_aa@rudn.university
(b) Peoples' Friendship University of Russia, 117198, ul. Miklouho-Maclay, 6, Moscow, Russia
(c) Russian Space Systems Corporation, 111250, ul. Aviamotornaya, 53, Moscow, Russia

Abstract

The Earth remote sensing is becoming a new and quickly developing interdisciplinary area of a practical importance and various commercial applications. The IT industry now in accordance with the academic sector is now working on developing new competences in the area which includes the development of new algorithms and calibration of existing program complexes and libraries. Investigation about using neural networks for detection geo-objects on the satellite images is the field of primary importance. In this paper we present the way of determination of the initial data, the architecture of the neural network, the accuracy characteristics of the recognized objects for the algorithm of recognition of buildings and structures based on the data of ultra-high resolution space imagery (UHRSI) which was made according to our research. The results could be used in the different application of remote data analysis including satellite application in different branches of military and civic needs.

Keywords: Image recognition, neural network, satellite images, object detection, machine learning. .

https://dx.doi.org/10.15405/epsbs.2018.11.02.56
Corresponding Author: Anna A. Ostrovskaya
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

1. Introduction

The Earth remote sensing is becoming a new and quickly developing interdisciplinary area of a practical importance. The IT industry now in accordance with the academic sector is now working on developing new competences (Chursin & Kashirin et al., 2018; Kashirin & Semenov et al., 2017) in remote data analysis and other connected fields. In machine learning applications for remote sensing, aerial image interpretation is usually formulated as a pixel labeling task. Given an aerial image the goal is to produce either a complete semantic segmentation of the image into classes such as building, road, tree, grass, and water or a binary classification of the image for a single object class. So the new algorithms should be devised and studied. Below we present an way to set the preliminary conditions for the algorithm developed in the project of RUDN University and Russian Space Systems Corporation as an approach for a new competence development.

2. Problem Statement

To develop an algorithm and approach for the problem of recognition of buildings and structures based on the data of ultra-high resolution space imagery.

3. Research Questions

To analyse the initial data form, the architecture of the neural network, the accuracy characteristics of the recognized objects for the algorithm of recognition of buildings and structures based on the data of ultra-high resolution space imagery (UHRSI).

4. Purpose of the Study

Determination of the requirements for UHRSI.

5. Research Methods

The research methods are neural network theory for ultra-high resolution space imagery.

6. Findings

After the investigation and experiments we formulate the following requirements to the initial data for the algorithm of detecting buildings and structures.

1.Initial data

- Images from the sensor of Geoton Resource-P Satellite;
- Presence of 4 channels (red, blue, green and near infrared);
- Processing level 4A is the integrated panchromatic image (processing level 2A) and multi-spectral (processing level 2A1) images of the same territory (Pansharpening);
- Cloudiness less than 70% of the image coverage;
- The angle of deviation from the nadir of the original image should not exceed 30 degrees;
- The point sizes of each satellite image channel must be pairwise identical;
- The point sizes of the channels must have at least 1024 points in each dimension;

- The information component of each satellite channel point must be 10 bits;

2. Determination of the requirements to the neural network architecture

- The neural network must have at least 10 layers;
- The neural network must have multiple inputs and one output;
- The point sizes of the result should correspond to the point sizes of the input image;
- The architecture of a neural network should provide the possibility of using methods of regularization and retraining prevention, namely, the methods of "normalizing batches" (Ioffe & Szegedy, 2015, p. 448-456) and "exceptions" (Srivastava, 2014, p. 1929-1958).
- The architecture should include links of the "passthrough characteristics" type (Chaurasia & Culurciello, 2017, p. 1-4)
- The architecture of the neural network should be adapted to the learning transfer strategy for a part of the layers;

3. Determination of requirements to the accuracy characteristics of recognized objects

- Buildings and structures with a total linear dimension of less than 10 meters, should achieve the acquisition accuracy of 0.5 and segmentation accuracy (measure of Sørensen) of each object of no less than 0.6;
- Buildings and structures with a total linear dimension exceeding 10 meters, should achieve the acquisition accuracy of 0.88 and segmentation accuracy (measure of Sørensen) of each object of no less than 0.6;
- Buildings and structures with a total linear dimension exceeding 75 meters, should achieve the acquisition accuracy of 0.98 and segmentation accuracy (measure of Sørensen) of each object of no less than 0.6;
- Buildings and structures with a total linear dimension exceeding 200 meters, should achieve the acquisition accuracy of 0.99 and segmentation accuracy (measure of Sørensen) of each object of no less than 0.6 (accuracy = $TP + TN$/Total), where TP is the number of true object definitions, TN is the number of false object definitions, and Total is the total number of unique objects.

Measure of Sørensen = $2TP/(2TP+FP+FN)$, where TP is the number of true definitions of points in the object, FP is the number of false positive definitions of points in the object, FN is the number of false negative definitions of points in the object.

2. Development of an algorithm for recognition of buildings and structures based on ultra-high resolution space imagery

The general developed scheme of the algorithm for recognition of buildings and structures according to the data of ultra-high resolution space imagery is shown in Figures 1 and 2.

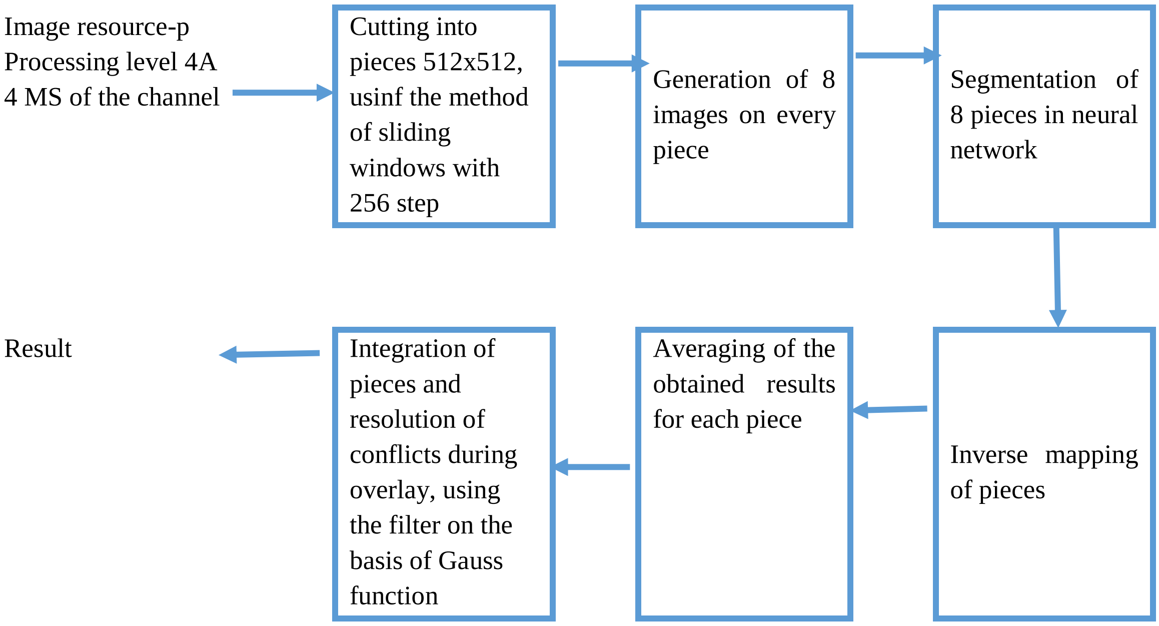


Image resource-p
Processing level 4A
4 MS of the channel
Cutting into pieces 512x512, usinf the method of sliding windows with 256 step
Generation of 8 images on every piece
Segmentation of 8 pieces in neural network
Inverse mapping of pieces
Averaging of the obtained results for each piece
Integration of pieces and resolution of conflicts during overlay, using the filter on the basis of Gauss function
Result

Figure 01. Diagram of the algorithm for recognition of buildings and structures.

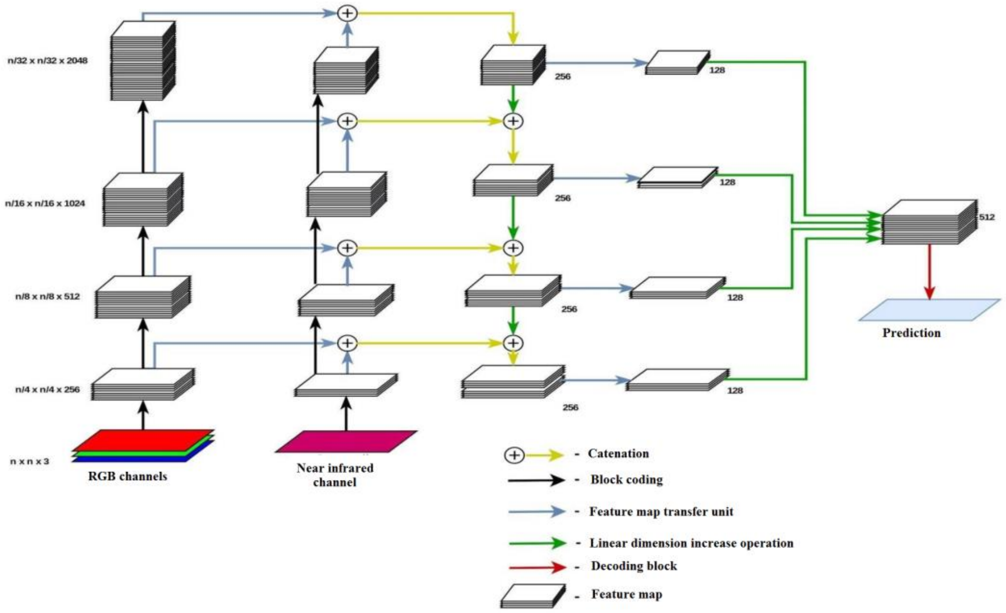


Figure 02. Neural network architecture.

The input algorithm is expected to receive a 4-channel satellite image from Geoton sensor, having the processing level 4A - a composite image of a panchromatic (processing level 2A) and a multi-spectral (level of processing 2A1) images of the same territory (Pansharpening) (Figure 3).

https://dx.doi.org/10.15405/epsbs.2018.11.02.56
Corresponding Author: Anna A. Ostrovskaya
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

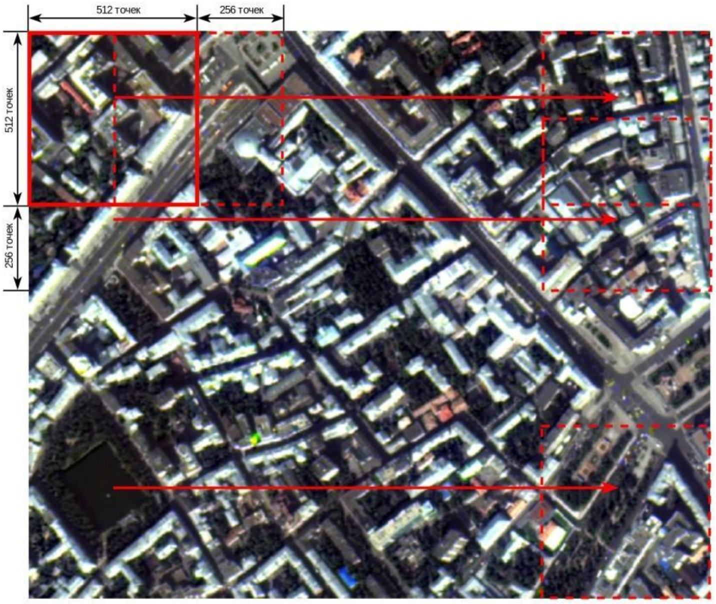


Figure 03. Splitting the original image into pieces of fixed size, using the sliding window method.

Sliding window method:

1. Make the size of the sliding window - $H_{SW} \times W_{SW}$ pixels;
2. Make the increments of the sliding window - S_H pixels (vertically) and S_W pixels (horizontally);
3. Complete the edges of the initial image I with the height H and width W to the size of the sliding window;
4. Form the pieces of the image (crop) of the size $H_{SW} \times W_{SW}$ from the complemented image in increments of $S_H \times S_W$

The following values were chosen for this algorithm:

- $\mathrm{H_{SW}} = \mathrm{W_{SW}} = 512$ pixels, to ensure the hit of rather big objects in one crop, which will allow to achieve the specified accuracy, and the practical possibility to create the software implementation on one hand, on the other hand, since the size of the window directly determines the size of the layer of the neural network, which increase leads to the increase of the requirements for the size of memory and the speed of the hardware and exponential increase of the learning time.
- $\mathrm{S_H} = \frac{\mathrm{H_{SW}}}{2} = 256$ pixels, $\mathrm{S_W} = \frac{\mathrm{W_{SW}}}{2} = 256$ pixels, to ensure the overlay of the crops so that the edge of one crop coincides with the center of the adjacent. It will enable to avoid artifacts and conflicts on the edges of the crop when restoring the segmentation of the entire image from the segmentations of certain crop areas.

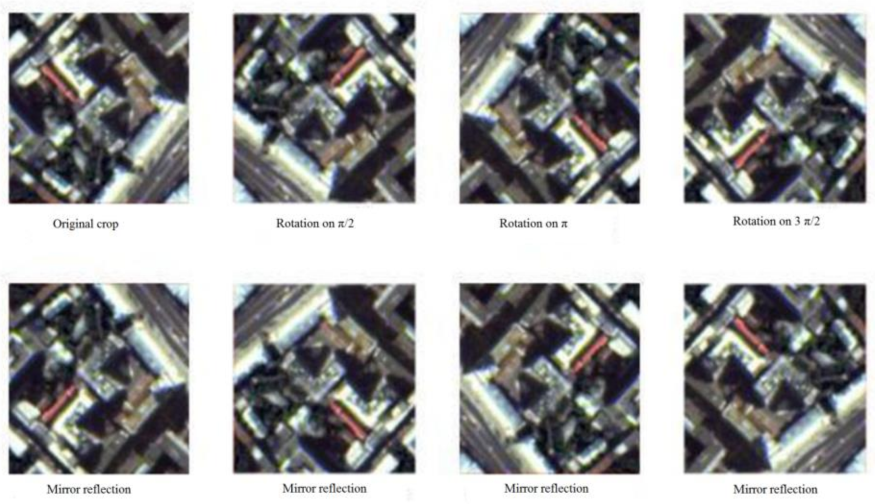


Figure 04. Turns and reflections of the crop.

Generating of 8 reflections for each crop (Figure 4) - all possible variations of the reflection of the initial piece, using the rotation operations on π2 and the mirror reflection.

Let C be the crooked image,

$\varphi(x)$ is a rotation by π_2, $\psi(x)$ is a mirror reflection, then the set 8 of reflections O can be represented as:

$$O = \{C, \varphi(C), \varphi(\varphi(C)), \varphi\left(\varphi(\varphi(C))\right), \psi(C), \psi(\varphi(C)), \psi\left(\varphi(\varphi(C))\right), \psi\left(\varphi\left(\varphi(\varphi(C))\right)\right)\}.$$

Obtain a segment map for each of the 8 reflections. At this stage, the original crop is passed through a trained deep neural web. Denote the function of obtaining the prediction P on the crop image C, as h. Then $P = h(C)$.

The network architecture is a sequence of encoding blocks that reduce the spatial resolution of the original crop, and decoding blocks that increase the spatial resolution, combining input data with feature maps, obtained by the transmission method from the encoding blocks of the corresponding resolution, which ensures the ensemble of the results of all layers and resolutions (Figures 5 and 6).

The encoding block is a set of 3 operations on feature cards.

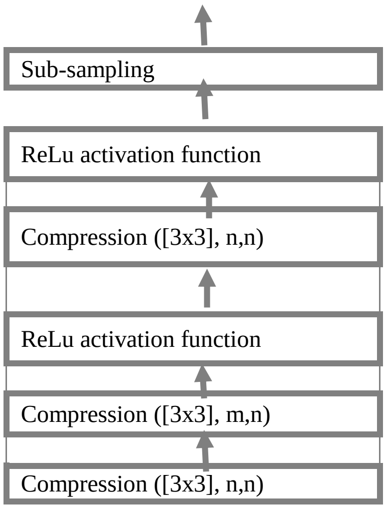


Figure 05. Block diagram of the encoding block.

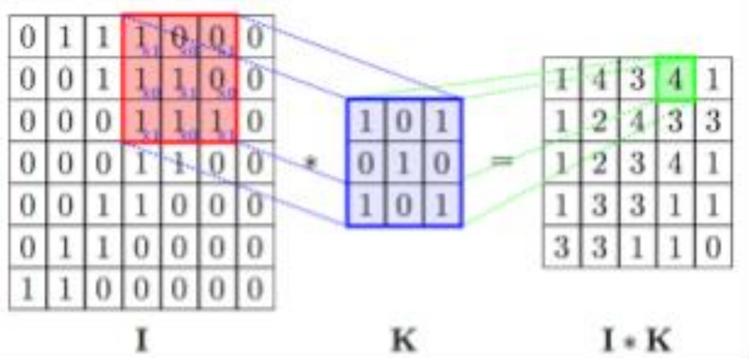


Figure 06. Example of a convolutional layer operation.

These operations are:

- The convolution layer is the main block of the convolutional neural network. Every channel of the convolution layer includes filter, its convolution core processes the previous layer fragment by fragment (summing the results of the matrix product for each fragment). It is denoted as Convolution ([*k* × *k*], *m*, *n*), where *k* × *k* is the size of the convolution core, m and n are the number of input and output filters for the layer, respectively;
- Relu activation function. The scalar result of each convolution falls on the activation function, which is a function $\text{ReLU} = x^{+} = \max(0,x)$, this function helps to evade the problems of a damped and exploding gradient, and is computationally simple;
- The sub-sampling layer (otherwise downsampling, undersubsampling, Figure 7) is a non-linear compaction of the feature map, with a group of points (usually 2 × 2 size) compressed to one point, passing a non-linear transformation. The most commonly used function is the maximum one. Transformations involve disjoint rectangles or squares, each of which is compressed to one point, and the point having the maximum value is selected. Pulling operation enables to reduce the spatial volume of the image essentially. Pulling is interpreted as follows. If some features have already been detected in the previous folding operation, then further processing does not demand a detailed image and it is condensed to a less detailed image. And it serves to generate new feature maps of greater dimension.

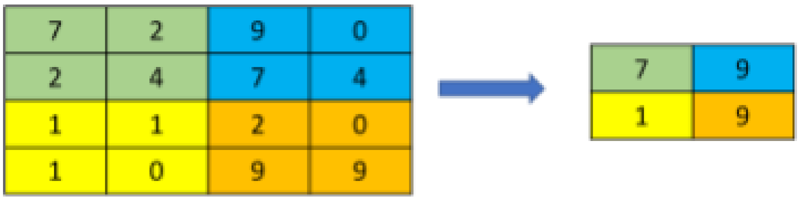


Figure 07. Example of the sub-sampling operation.

The decoding block is a set of 4 consecutive operations (Figure 8):

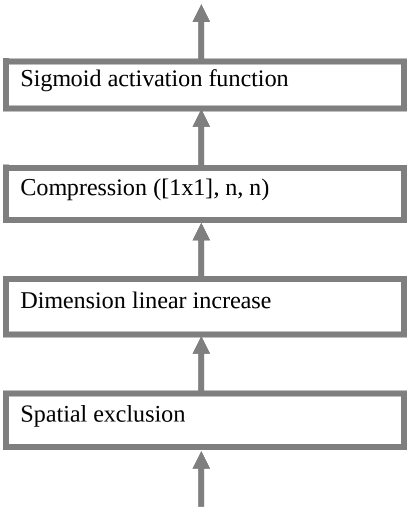
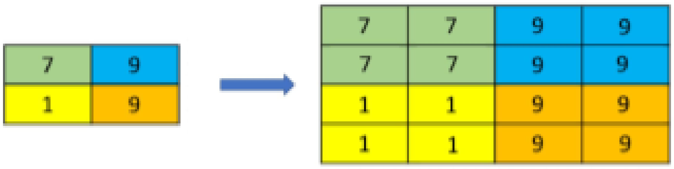

Sigmoid activation function
Compression ([1x1], n, n)
Dimension linear increase
Spatial exclusion

Figure 08. Block diagram of the decoding unit.

- Spatial exclusion – shuts down the layer of neurons with probability p;
- A convolution layer with a 1x1 core is necessary to reduce the dimension of the feature map;
- Activation function. The scalar result of each convolution falls on the activation function, which is a nonlinear function $\text{sigmoid} = \frac{1}{1+e^{-x}}$, this function allows both to amplify weak signals and to avoid saturation from strong signals.
- Linear increase (Figure 9) of dimension - reverse sub-sampling operation is a linear repetition of a feature map, with each point being transformed into a group of 2 × 2 points, passing a linear transformation. Transformations affect all points, each of them turns into a group of points, while they have the same value. This operation enables to increase the image dimension;



7	9
1	9

7	7	9	9
7	7	9	9
1	1	9	9
1	1	9	9

Figure 09. Example of the linear dimension increase.

As a result, the output of the neural network is a set of 8 probability maps that each point of the original crop belongs to the class "Buildings and Constructions".

Reverse mapping operations (rotation to $-\pi/2$ and mirror image) are applied to the resulting set of probability maps in order to obtain the preimages of the images used with respect to output probability maps. Since $x = \psi(\psi(x))$ and $x = \varphi^{-1}(x) = \varphi\left(\varphi(\varphi(x))\right)$, the desired multitude O_P will have the form:

$$O_P = \{P_1, \varphi^{-1}\left(\varphi^{-1}(\varphi^{-1}(P_2))\right), \varphi^{-1}(\varphi^{-1}(P_3)), \varphi^{-1}(P_4), \psi(P_5),$$

$$\psi\left(\varphi^{-1}\left(\varphi^{-1}(\varphi^{-1}(P_6))\right)\right), \psi\left(\varphi^{-1}(\varphi^{-1}(P_7))\right), \psi(\varphi^{-1}(P_8))\}$$

Refine the boundaries of the obtained segments by averaging the predictions. Then the final prediction for each point of the crop image will be calculated using the following formula:

$$P_{result}(i,j) = \frac{\sum_{k=1}^{8} P_k(i,j)}{8}$$

This approach improves the segmentation results obtained in the previous stage.

Combine the obtained intersecting prediction maps with the help of a weighted sum using a two-dimensional Gaussian distribution with zero in the center of the crop and a root-mean-square deviation $\sigma = \frac{H_{SW}}{2*3} \approx 85$, calculated at points corresponding to the centers of the crop pixels to get the segmentation of the original image. It will eliminate conflicts and artifacts on the crop borders, since the crop, which is closest to a pixel will have the greatest contribution to its value, and the contribution of the extreme points of the crop is ~ 0.2.

7. Conclusion

In this paper we presented the way of determination of the initial data, the architecture of the neural network, the accuracy characteristics of the recognized objects for the algorithm of recognition of buildings and structures based on the data of ultra-high resolution space imagery (UHRSI) which was made according our research. The results could be used in the different application of remote data analysis

Acknowledgments

The paper was prepared with the financial support of the Ministry of Education of the Russian Federation in the framework of the scientific project No. 14.575.21.0167 connected with the implementation of applied scientific research on the following topic: «Development of applied solutions for processing and integration of large volumes of diverse operational, retrospective and the thematic data of Earth's remote sensing in the unified geospace using smart digital technologies and artificial intelligence» (identifier RFMEFI57517X0167)

References

Chaurasia, A. & Culurciello, E. (2017). Linknet Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE Visual Communications and Image Processing (VCIP)* (p. 1-4).

Chursin, A, Kashirin, A. et al. (2018). The approach to detection and application of the company's technological competences to form a business-model. *IOP Conference Series Materials Science and Engineering*, *313*, 012003. doi:10.1088/1757-899X/312/1/012003

Ioffe, S. & Szegedy, C. (2015). Batch Normalization Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *International Conference on Machine Learning* (p. 448-456).

Kashirin, A., Semenov, A., Ostrovskaya, A. & Kokuytseva, T. (2016). The Modern Approach to Competencies Management Based on IT Solutions. *JIBC-AD - Journal of Internet Banking and Commerce, 01,* 813075.

Srivastava N. et al. (2014). Dropout: a simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, *15*(1), 1929-1958.