

Scoring System Based on Neural Networks for Identification of Factors in Image Perception

Inna Khortiuk^[0000-0001-7525-8362], Galyna Kondratenko^[0000-0002-8446-5096],
Ievgen Sidenko^[0000-0001-6496-2469], Yuriy Kondratenko^[0000-0001-7736-883X]

Intelligent Information Systems Department, Petro Mohyla Black Sea National University,
68th Desantnykiv Str., 10, Mykolaiv, 54003, Ukraine,
innamysnyk@gmail.com, halyna.kondratenko@chmnu.edu.ua,
ievgen.sidenko@chmnu.edu.ua, yuriy.kondratenko@chmnu.edu.ua

Abstract. In the paper, the authors analyze the image memorability prediction using memorability experiments (tests) and neural networks. With the development of modern technologies for processing and presenting data, images make it easy to demonstrate the visualization of results, the visibility of materials, design features, etc. Image memorability prediction (image perception) has high scientific and practical importance and can be used in various fields, in particular, business, education, design, etc. Recently, deep convolutional neural networks (CNNs) have been attracting considerable attention in image restoration. The factors that influence the image memorability and the technologies and approaches for data processing are analyzed. This paper shows the effectiveness of applying the considered technologies and approaches on the experimental data of the public LaMem set, while calculating the quantitative assessment of image memorability. Photographers, designers, and advertisers will benefit from the results of this study directly. The peculiarities of the “Deep Learning” technology implementation are discussed in detail. Results confirm the efficiency of the considered approach for image memorability recognition. In our case studies, we have compared different neural network architectures and develop a scoring system for image memorability. To achieve this target, the authors use PHP as a programming language.

Keywords: image memorability, image perception dataset, CNN, deep learning, neural networks.

1 Introduction

People remember thousands of images and the amazing number of their details [1]. While a few pictures are remembered, others are overlooked or rapidly forgotten. Designers, promoters, clip makers, and photographers are usually stood up with the actual question "what makes an image memorable?". But how to create an image that is memorable to the watcher? While experts investigated the ability of a person to memorize visual stimuli [2], in [3], 2222 photographs were memorized, and the concept of image memorization was introduced as the speed at which viewers (watchers)

Copyright © 2020 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

determine a repeated image presentation sometime after its first demonstration on the screen. The storage of these images has been found to be the same for different subjects in different contexts, making some of the images more memorable than others. Thus, while the memory of an image can be considered as an indicator that is difficult to calculate, recent publications by scientists prove that this is not the case and there is a clear pattern.

The image storage property has the potential to change it by changing specific properties or components presented in the image. First of all you need to understand the parameters (factors), which influence on image memorability. The paper [4] shows that image memory cannot be determined by the usual properties for image descriptions such as "weirdness", "aesthetic enjoyment" or "fun", and, moreover, people are not good to predict the memory of the image. Herewith, computer vision technologies have been proven to be quite effectiveness for solving the current problem. The paper [5] describes appropriate technologies for determining the investment of certain areas of the image to the memory of the image. It is possible to change the level of image memorability by changing the areas with low or high memorability. This can have far-reaching results in a variety of areas, from advertising and games (for example, determine which commercials or company logos are more recognizable and memorable) to education (for example, change the image without losing its primary meaning) and social networks (for example, determine the best scenes for perception and memorization in order to integrate interesting objects into them in the future). There is another solution for visual perception: to check how well people remember images, thereby showing them an image that they have never seen before. The task is how much it is necessary to change the image so that a person believes that he has never seen this image before.

The paper [6] provides an overview of recent image storage work that shows that memory is a measurable and inherent image property. Next, it is necessary to define in more detail the elements that influence the image memorability. Development and globalization of the market contributes to increased competition in it. That is why many companies have to expand their sphere of influence, expanding the sales area of their goods or services. Along with this, the delivery process is becoming more expensive.

2 Related Works and Problem Statement

Recent studies on the storage of images [7] reported a high degree of consistency of factors due to which participants in the experiment can remember and forget images. This may indicate that memory is very closely related to images, despite the personal differences between the participants who observe the images. Modeling the results of a high degree of consistency of image factors was carried out for images of more than 100 different groups [8], and after that, it was also demonstrated on more specialized groups of images, in particular on people's faces [9] and visual scenes [10]. In these works, it was demonstrated that the corresponding consistency is not considered a peculiar property of factors and that it cannot be considered only as a difference be-

tween images (for example, images in rooms can be more memorable, and images outside the room - vice versa). It has been shown that high consistency is maintained in 21 different categories of scenes both inside and outside of the room, each of which consists of hundreds of images.

Previous studies have shown that objects are better remembered when they stand out from the main scene or background [10]. For example, in the works [5, 11], the authors demonstrated the results of long-term memory when working with images depicting strange things.

Konkle et al. have shown that categories of objects with conceptually prominent samples showed less memory interference when increasing the number of samples [8]. In addition, for specific groups of images, in particular, people's faces, the results showed that another or unusual (non-trivial) face will be remembered more than usual [12, 13]. Borkin et al. analyzed that unique categories of visualizations have much higher memory performance than regular graphics, and that new and unique visualizations are better remembered [3, 14, 15].

3 Methods for Solving the Problem of Image Recognition

Image recognition tasks are characterized by a number of difficulties. These are: external factors due to the fact that the images of the same objects can vary greatly in different conditions and depend on the point of observation, lighting, scale, overlap, background complexity, and internal - variability in the middle of the class and deformation (various forms of object) [16, 17].

Depending on the above factors, different groups of methods are used to solve the problems of image recognition in an image. The color / brightness filter method can be used in cases where the subject is substantially different from the background and the illumination is uniform and does not change. By separating such objects from the background, you can use geometric analysis to make classification decisions [18].

If the object on the background does not stand out significantly or has a complex color, then applying the method of color characteristics of the brightness of the filters will not give good results. In this case, a more efficient method of contour selection and analysis. To do this, we highlight the border in the image, that is, the place of a sharp change in the brightness gradient, for example, using the Canny method. Then we check the highlighted border for matching the geometric contours of the object, this can be done using the Hough Transform method, to search within circles, ellipses, lines or to apply contour analysis methods [2].

If the image has many small details, then the contour analysis can be complicated. In this case, you can apply the pattern mapping method. This method is used to find the areas of images that are most similar to a given template - a specific object in one class. The pattern mapping method finds the exact points of the pattern with the image points [4].

However, if the image is rotated or resized with respect to template settings, this method does not work well. Key points based methods are used to overcome these limitations. A key point is a small area that is significantly highlighted in an image.

There are several methods for determining such points, these can be Harris corner detector or blob, that is, small areas of equal brightness, with a fairly clear border stand out against the general background. For the key point, the descriptor is calculated - the characteristic of the key point. The descriptor is calculated over a given point area as the direction of the brightness gradients of different parts of this area. There are several methods for calculating descriptors for special points: SIFT, SURF, ORB, etc. Key points can be used to find an object in an image. To do this, we need to take an image of the desired object and do the following [5, 7, 15]:

- find specific points of the object in the image with the object and calculate their descriptors;
- find special points in the image already mentioned and calculate their descriptors;
- compare the descriptors of the special points of the object and the descriptors of the special points found in the image;
- if a sufficient number of matches is found, then mark the area with the corresponding points.

The scope of the above methods is limited due to the fact that they all have poor generalizability. This means that the model describes one particular object or a very narrow class of objects. In order to successfully recognize not one object, but large and diverse classes of objects, you can apply machine learning methods [12].

Deep learning is a set of feature/representation learning methods of machine learning. In the context of image recognition, the main algorithm of deep learning is a convolutional neural network.

4 Image Memorability Experiments

A new dataset was created by sampling figures from 21 various categories of inside and outside scenes. The choice of category was dependent on having at least 300 images of the required size. Duplicate images were manually removed. Images were truncated and cropped to 700x700px, 25% of each scene category was randomly selected as targets, and other images became fillers). The targets are images for which they have received memory ratings [19, 20].

The Amazon Mechanical Turk (MTurk) was used to collect the experimental data. MTurk is a site that allows individuals and entrepreneurs to coordinate the use of the human mind to solve problems that computers cannot currently solve. This is one of the famous Amazon Web Services sites. On this site, employers can upload Human Intelligence Tasks, for example, choosing the best photo available, adding product descriptions, or identifying artists on music CDs. Staff members can later review current actual tasks and solve them at the employer-determined cash payment.

AMT 1 experiment.

For this:

- AMT memory games were set up, in which successions of 120 images (a combination of goal images and filler images selected from one group) were displayed for 1

- second each, the original image and its repetition were at a distance of 91-109 images from each other, and sequential images, which were separated by a special sign (cross) with a time interval of 1.4 sec;
- some placeholder images were repeated with shorter intervals of 1-7 images and were used as attentiveness tests to find out when a person did not take part in the game (did not pay the necessary attention);
 - experiment participants pressed a key when they saw a repeated image, and then received feedback (cross of different colors). In this case, the image was repeated only once;
 - experiment participants could pass this memory test several times, but each time another group of images was displayed;
 - false alert is triggered when the experiment participant shows at the first presentation that the image has been repeated. Not a single keystroke during the first presentation is a correct deviation. The answer is considered correct when the repeated image is correctly remembered, otherwise, the answer is considered incorrect (mistake).

New set of data for AMT 1 includes 1754 targets / 7674 filler images. The partial set of corresponding data is presented in Table 1.

Table 1. The partial set of corresponding data for AMT 1.

| Category | Targets | Fillers | HR (%) | FAR (%) | HR cons. (ρ) | FAR cons. (ρ) | Datapoints /Target |
|----------------|---------|---------|-----------------|-----------------|---------------------|----------------------|--------------------|
| Amusement park | 68 | 296 | 64.2 (SD: 15.5) | 10.2 (SD: 9.7) | 0.85 (SD: 0.3) | 0.80 (SD: 0.3) | 84.8 (SD: 3.2) |
| Playground | 74 | 330 | 63.3 (SD: 14.4) | 14.7 (SD: 12.7) | 0.78 (SD: 0.4) | 0.84 (SD: 0.3) | 86.4 (SD: 2.7) |
| | | | ... | ... | ... | | |
| Cockpit | 68 | 320 | 49.5 (SD: 17.2) | 18.2 (SD: 14.7) | 0.70 (SD: 0.5) | 0.88 (SD: 0.2) | 80.6 (SD: 3.5) |

The HR and FAR values are calculated over the goals (targets), for which identified an average of 85 data points per image.

AMT 2 experiment.

AMT 2 experiment was conducted on the grouped set of target and placeholder images in all categories, herewith a new set of memory factors was collected following the same protocols as before.

New dataset statistics for AMT 2 (Table 2).

Table 2. New data for AMT 2 experiment

| Category | Targets | Fillers | HR (%) | FAR (%) | HR cons. (ρ) | FAR cons. (ρ) | Datapoints /Target |
|-----------|---------|---------|-----------------|----------------|---------------------|----------------------|--------------------|
| 21 scenes | 1754 | 7296 | 66.0 (SD: 13.9) | 11.1 (SD: 9.5) | 0.74 (SD: 0.2) | 0.72 (SD: 0.1) | 74.3 (SD: 7.5) |

Comparing memory performance across data sets shows consistency of results and stability of memory performance. For images in 21 scenes context (AMT 2), the image memorization value is higher than in 7 scenes context (in the laboratory) and even

higher than in 1 scene context (AMT 1). At the same time, the overall trend remains the same.

5 Practical Implementation

Since the main purpose of our practical implementation was to simplify and accelerate the results it was suggested to divide the system into three main parts: the user interface, the client service and the REST API of the scoring system itself [21, 22].

The REST API service was deployed on the IBM AI network and was the only call to accept image links. Given that the neural network model can only analyze one image per call, the entire call interface is only one image parameter. The image parameter took a direct link to the image for analysis. Given this restriction, the processing of multiple images was simultaneously translated into client service (Fig. 1).

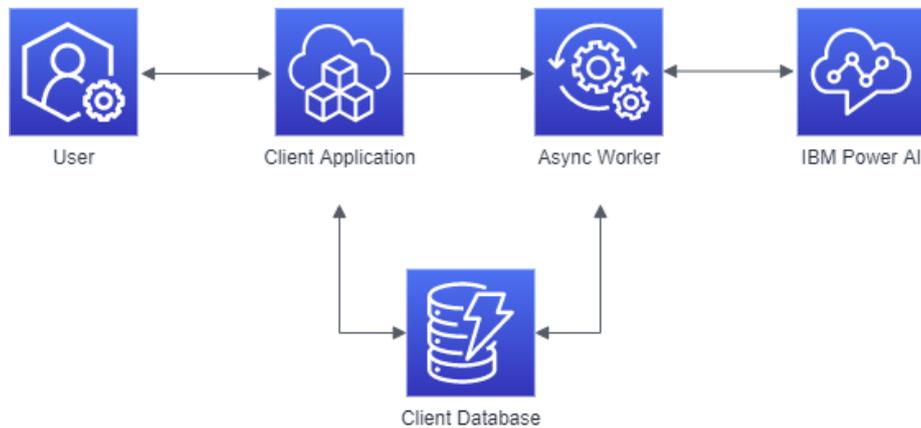


Fig. 1. Schematic representation of the organizational structure of the scoring system [21]

Client service is implemented as an application in PHP programming language, using Laravel web Framework [10] and following the MVC web development paradigm [23-26]. Downloaded image data and analysis results are stored in a small SQLite database to simplify deployment and save computing resources. The client service interface is four separate REST API calls, namely:

- analyze is accepts n-number images and sends an asynchronous Redis queue for background processing;
- index is returns a list of downloaded images and data about them. Data such as upload date, file name, image processing step, and memory rating are displayed;
- view is returns a description of a particular image;
- compare is generates a comparative characteristic between two different, previously loaded and analyzed images in the system.

The end-user visible system is implemented in the form of an interactive Web Interface with a cross-cutting of VueJs, HTML and CSS libraries. The entire user interface is designed using the SPA approach to maximize comfort and speed when analyzing image memory evaluation [27].

The service suite was deployed using Docker Compose [28-30]. The Nginx open source server was used as the web server. For the first time when downloading the implemented system, the user will see a window asking to download images for further analysis (Fig. 2).

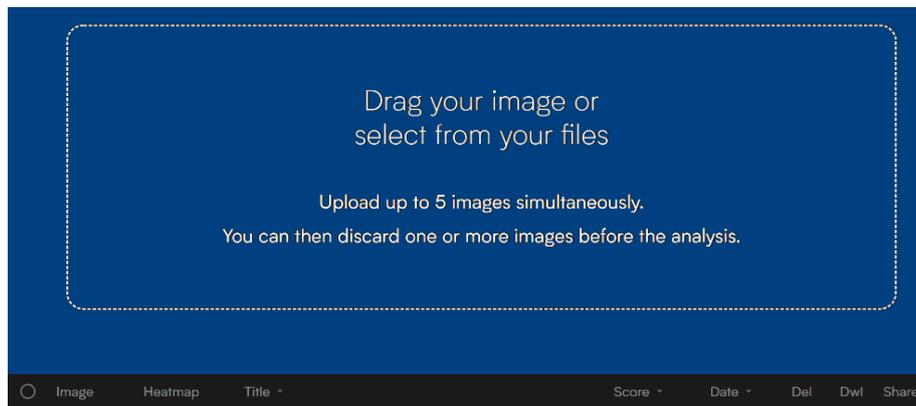


Fig. 2. Image upload form

The user can drag and drop the desired images through the standard operating system window. After the images have been selected, the system will prompt you to delete the unnecessary images or add additional ones. When the user is ready to send the image for analysis, just click the Analyze button (Fig. 3).

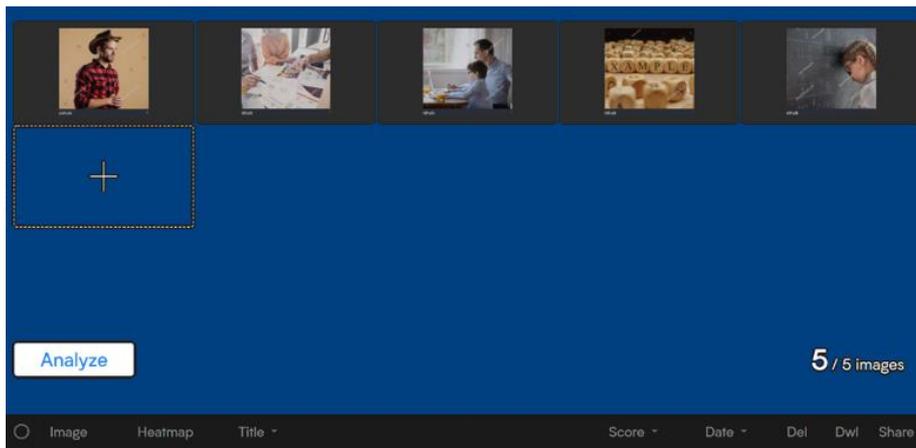


Fig. 3. Image analysis interface

After successfully downloading all the images, the system will queue each image for analysis. The user is not required wait for completion as the analysis runs asynchronously and the system will send a successful completion message.

After the results are obtained, the user can view all the estimates (for example, the score of the first image is 8.81) and sort them by the order (Fig. 4).

| Image | Heatmap | Title | Score | Date | Del | Dwl | Share |
|-------|-----------------------|-------|-------|------|-----|-----|-------|
| | Processing heatmap... | | 8.81 | | | | |
| | Uploaded... | | | | | | |

Fig. 4. Scoring table

By clicking on a specific image, you can see a heat map that reflects the impact of the image elements on the image's memory rating (Fig. 5).

← Compare □|□ Add

you can compare up to 5 images

shutterstock

shutterstock

8.98 / 10

stock-photo-examples-word-written-on-wood-block-415964203.jpg

shutterstock

shutterstock

8.81 / 10

stock-photo-american-cowboy-western-life-man-unshaven-cowboy-beige-background-unshaven-guy-in-cowboy-hat-1503282395.jpg

Fig. 5. Comparison table with heat maps

To compare the image with the other analyzed results, the user can click the compare button and select the images that interest him. As a result, the system will generate a comparative table with visualized heat maps for each image (Fig. 5). A heat map is a tool that uses a color palette to visualize data in an image. The heat map uses the color spectrum from warm to cool tones to show areas of attention that attract users (warm tones attract the most attention, cold tones attract the least attention).

6 Conclusions

In the paper, authors analyze the image memorability prediction using memorability experiments (tests) and neural networks. The developed scoring system not only allows to evaluate and compare the results of storing the images of the automatically trained system with real experimental data, but also has practical application, as it is a full-featured application with a user-friendly interface, which widens the circle of users. The system can be used by designers, SMM specialists, teachers and students, advertising industry workers, and regular users, who aim to address the issue of choosing the image that the audience will pay attention. The studies presented in the paper offer great potential for future use as they provide tools to determine the conditions in which images can be best remembered (and how the image memorability may vary in different contexts).

As a result, the existing technologies and approaches to image recognition were analyzed; existing studies on the detection of factors of image storage were considered; scoring system using PHP programming language and convolutional neural networks for identification of factors in image perception was developed; experiments and results analysis using Amazon Mechanical Turk were performed.

References

1. Bartlett, J., Hury, S., Thorley, W.: Typicality and familiarity of faces. *Memory & Cognition* 12(3), 219-228 (1984).
2. Borji, A., Itti, L.: Defending Yarbus: Eye movements reveal observers' task. *Journal of vision* 14(3), 29-29 (2014).
3. Borkin, M., Vo, A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., Pfister, H.: What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* 19(12), 2306-2315 (2013).
4. Brady, T., Konkle, T., Alvarez, G.: A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision* 11(5), 4-4 (2011).
5. Standing, L.: Learning 10000 pictures. *The Quarterly journal of experimental psychology* 25(2), 207-222 (1973).
6. Wiseman, S., Neisser, U.: Perceptual organization as a determinant of visual recognition memory. *The American journal of psychology*, 675-681 (1974).
7. Bainbridge, W., Isola, P., Oliva, A.: The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General* 142(4), 1323-1334 (2013).

8. Konkle, T., Brady, T., Alvarez, G., Oliva, A.: Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General* 139, pp. 558-578 (2010).
9. Xiao, J., Hays, J., Ehinger, K., Oliva, A., Torralba, A.: Sun database: Large-scale scene recognition from abbey to zoo. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3485-3492, San Francisco, CA (2010).
10. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Object detectors emerge in deep scene CNNs. *arXiv preprint arXiv:1412.6856* (2014).
11. Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 806-813, Columbus, OH (2014).
12. Basturk, B., Karaboga, D.: An Artificial Bee Colony (ABC) Algorithm for Numeric function Optimization. In: *IEEE Swarm Intelligence Symposium*, Indianapolis, Indiana, USA (2006).
13. Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., Oliva, A.: Intrinsic and extrinsic effects on image memorability. *Vision research* 116, 165-178 (2015).
14. Khosla, A., Bainbridge, W., Torralba, A., Oliva, A.: Modifying the memorability of face photographs. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3200-3207 (2013).
15. Worthen, R. R. H. J. B.: *Distinctiveness and memory*. University Press, Oxford (2006).
16. Mikhov, D., Kondratenko, Y., Kondratenko, G., Sidenko, I.: Fuzzy logic approach to improving the digital images contrast. In: *IEEE 2nd Ukraine Conference on Electrical and Computer Engineering, UKRCON*, pp. 1183-1188, Lviv, Ukraine (2019).
17. Pomanysochka, Y., Kondratenko, Y., Kondratenko, G., Sidenko, I.: Soft computing techniques for noise filtration in the image recognition processes. In: *IEEE 2nd Ukraine Conference on Electrical and Computer Engineering, UKRCON*, pp. 1189-1195, Lviv, Ukraine (2019).
18. Wang, W., Sun, J., Li, J., Wu, Q., Liu, J.: Investigation on the Influence of Visual Attention on Image Memorability. In: Zhang, Y. (eds) *Image and Graphics. ICIG 2015. Lecture Notes in Computer Science*, vol. 9219, pp. 573-582. Springer, Cham (2015).
19. Basavaraju, S., Sur, A.: Multiple instance learning based deep CNN for image memorability prediction. *Multimed Tools Appl* 78, 35511-35535 (2019). DOI: 10.1007/s11042-019-08202-y.
20. Leonardi, M. et al.: Image Memorability Using Diverse Visual Features and Soft Attention. In: Ricci, E., Rota Bulò, S., Snoek, C., Lanz, O., Messelodi, S., Sebe, N. (eds) *Image Analysis and Processing. Lecture Notes in Computer Science*, vol. 11752, pp. 171-180. Springer, Cham (2019).
21. Glanzer, M., Adams, J. K.: The mirror effect in recognition memory: data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(1), 5-16 (1990).
22. Ji, J., Chen, G., Sun, L.: A novel Hough transform method for line detection by enhancing accumulator array. *Pattern Recognition Letters* 32(11), 1503-1510 (2011).
23. Pomanysochka, Y., Kondratenko, Y., Sidenko, I.: Noise filtration in the digital images using fuzzy sets and fuzzy logic. In: *15th International Conference on ICT in Education, Research, and Industrial Applications: PhD Symposium (ICTERI 2019: PhD Symposium)*, vol. 2403, pp. 63-72, Kherson, Ukraine (2019).
24. Watanabe, K., Imamura, M., Asami, K., Amanuma, T.: A Web Application Development Framework Using Code Generation from MVC-Based UI Model. In: Omatu, S. et al. (eds) *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Am-*

- bient Assisted Living. IWANN 2009. Lecture Notes in Computer Science, vol. 5518, pp. 404-411. Springer, Berlin, Heidelberg (2009). DOI: 10.1007/978-3-642-02481-8_57.
25. Kondratenko, Y., Gordienko, E.: Neural Networks for Adaptive Control System of Caterpillar Turn. In: Annals of DAAAM for 2011 & Proceeding of the 22th Int. DAAAM Symp. "Intelligent Manufacturing and Automation", 2011, pp. 0305-0306.
 26. Sidenko, I., Kondratenko, G., Kushneryk, P., Kondratenko, Y.: Peculiarities of Human Machine Interaction for Synthesis of the Intelligent Dialogue Chatbot. In: 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 1056-1061, Metz, France (2019).
 27. Amengual, X., Bosch, A., de la Rosa J.L.: Review of Methods to Predict Social Image Interestingness and Memorability. In: Azzopardi, G., Petkov, N. (eds) Computer Analysis of Images and Patterns. CAIP 2015. Lecture Notes in Computer Science, vol. 9256, pp. 64-76. Springer, Cham (2015).
 28. Cook, J.: Docker for Data Science. Apress, Berkeley, CA (2017).
 29. Bhat, S.: Practical Docker with Python. Apress, Berkeley, CA (2018).
 30. Freeman, A.: Essential Docker for ASP.NET Core MVC. Apress, Berkeley, CA (2017).