Journal of Digital Science

ISSN 2686-8296

Volume 2 Issue 1 May 2020

CONTENTS

Neural Network Model for Assessing the Physical and Mechanical Properties of a Metal Material Based on Deep Learning

Andrei Kliuev[1[0000-0002-5168-5252]], Roman Klestov[1[0000-0003-2063-3596]]
Valerii Stolbov[1[0000-0003-3259-8660]]

[1] Perm National Research Polytechnic University, Perm, Russia, 614990

https://doi.org/10.33847/2686-8296.2.1_2

Received 01.03.2020/Revised 05.04.2020/Accepted 11.05.2020/Published 28.05.2020

Abstract. The paper investigates the algorithmic stability of learning a deep neural network in problems of recognition of the materials microstructure. It is shown that at 8% of quantitative deviation in the basic test set the algorithm trained network loses stability. This means that with such a quantitative or qualitative deviation in the training or test sets, the results obtained with such trained network can hardly be trusted.
Although the results of this study are applicable to the particular case, i.e. problems of recognition of the microstructure using ResNet-152, the authors propose a cheaper method for studying stability based on the analysis of the test, rather than the training set.

Keywords: Deep neural networks, material microstructure, image recognition, deep learning, algorithmic stability.

1. Introduction

A well-known problem is the creation of the training and testing sets in machine learning problems, which also include deep learning tasks in order to predict the physic mechanical properties of functional materials [1, 2]. Many researchers in the field of applied calculations use the rule of dividing the entire available set into a learning and testing ratio of 80/20 or 70/30. Usually the reasons of such a division is not given. As for the recommendations on the size of the training set, they are completely absent, except for the statements that the power of the set must be 3-10 times the number of parameters of the neural network. Such a wide range of opinions says only that this task, for all its relevance, is poorly investigated.

In addition, there is the problem of assessing the quality of the training set. There is no clear answer to the question: how erroneously annotated images can affect the learning outcome? For example, quite often the work of the classifying neural network is compared with the human work process. Is such an assessment fair? After all, the training set contains annotated data that was prepared by people and which, in turn, contain the same errors. How can these errors distort the result?

It is known that the fundamental theory of uniform convergence of V.N. Vapnik - A.Y. Chervonenkis [3] treats the fatal errors of the learning algorithm as a fail of algorithmic stability. Stable learning algorithms are understood as those that form only those hypotheses, the result of which changes only slightly with a small change in the training set [4, 5]. Within this theory, sets of such algorithms have been found, for example, regression, support vector machine, etc.

Learning deep neural networks can hardly be analyzed within the theory of uniform convergence, but numerical research on the stability of specific networks can be quite accessible. For example, in [6], the influence of the size of the training and testing sets on the accuracy and generalizing ability of the three-layer MLP neural network in the binary classification problem was studied. In [7], the accuracy and scatter of the GoogLeNet network in the problem of the classification of body parts by computed tomography images is investigated. The size of the training set is selected on the basis of the analysis of the accuracy of the training results. The disadvantages of this work include the fact that the dependence of the accuracy of training on the size of the training sample was calculated on the basis of 6 computational experiments (for each of the classes), and the testing set itself was obtained as a result of the separation of the training set in 75/25 proportion. However, if there is no sustainability of training, then these results can hardly be trusted.

2. Literature Review

There are successes in the recognition and spatial localization of objects by machine learning methods based on convolutional networks, first proposed by Jan LeCun [8]. With millions of pre-classified images collected in bulky databases of annotated images, such as ImageNet or COCO, and the back propagation method the results were really impressive [9,10,11].

Modern methods are also used in metallurgy. The problems of classification of microstructures are solved mainly. For example, in [12] the problem of classification of high-carbon steels by their microstructure using the VGG-16 network is studied. For training, testing and validation the base of sections' images of the CMU-UHCS (Carnegie Mellon University Ultrahigh Carbon Steel) taken by an electron microscope (the base contains 961 marked images divided into 7 categories) is used. In [13] an attempt to classify low-carbon steels is made. The training set of Material Engineering Center Saarland (MECS) was used. In [14] recognition of dendritic microstructures on the digital images of the microsections is performed. The training set was formed on the basis of micro images of the DoITPoMS project (Dissemination of it for the promotion of Materials Science) of the University of Cambridge.

Convolutional networks are successfully used in similar problems in the sense of determining surface defects. In the work [15] with the help of MPCNN (CNN with max-polling of subsampling layers) the problem of classification of surface defects of steel pipes is set and solved. Later, the authors [16] solved a similar problem of classification of surface defects of steel sheets. In [17,18] the method of detection of defects on the surface of rails is investigated.

The purpose of the work consists in the construction and training of a convolutional neural network able to predict the physicomechanical properties of the material at the images of the microstructure. Within the framework of the developed approach the problem of constructing a generative-adversarial network (GAN) [19,20] able to represent the structure of the material with predetermined physicomechanical characteristics is posed. In [21] is shown that with the help of trained on textures convolutional network it is possible to calculate quite accurate characteristic of texture recorded in the form of Gram's matrix. Components of the matrix are obtained from the cores of convolutions belonging to different levels of the network. In the work a well-known pre-trained VGG-19 is used. The author argues that his approach is much more accurate and flexible than the description of the texture by a two-point correlation function. By applying noise and other transformations, it is able to generate textures similar to the original using a convolutional network. The author [22] used this technique to analyze and generate the microstructure of metals and alloys (Fig.1). Noise, scaling and rotation were used during generation. There is a paper more focused on sustainability research only [23]. This paper should provide more details on experimental part of research.

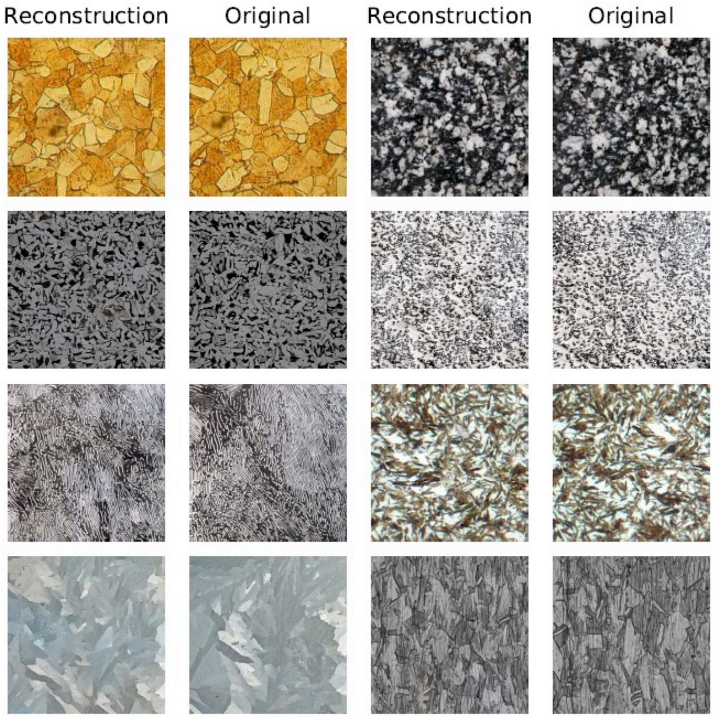


Fig. 1. The original structures and their reconstruction analogues [15].

The network constructed and trained in the work will be the basis for the construction of generative and discriminant parts of GAN.

3. Data and Methodology

To solve the problem the VGG-16 architecture was chosen as the most proven convolutional neural network in this area. A pre-trained version of the network with an input image of size 256x256 was selected. The process of preparation the training set and the learning process of the network are described below.

Training dataset

The primary data for the training set was prepared by the staff of the Institute of Nanosteels of Magnitogorsk State Technical University in the frame of international project Horizon 2020. The data was represented by 758 digital annotated images of metal alloy sections based on iron of different chemical composition and grain-phase structure. The images were made using an electron microscope with different powers from x1000 to x20000. Examples of images are shown in Fig.2. The images were annotated with physicomechanical characteristics measured during the tests.

The samples were tested for hardness and strength to determine the physicomechanical properties and the following characteristics were established: microhardness, conditional yield strength, time resistance, elongation, relative contraction. Tests were carried out at room temperature. The manufacture of the sections was performed before the test for the unloaded sample.

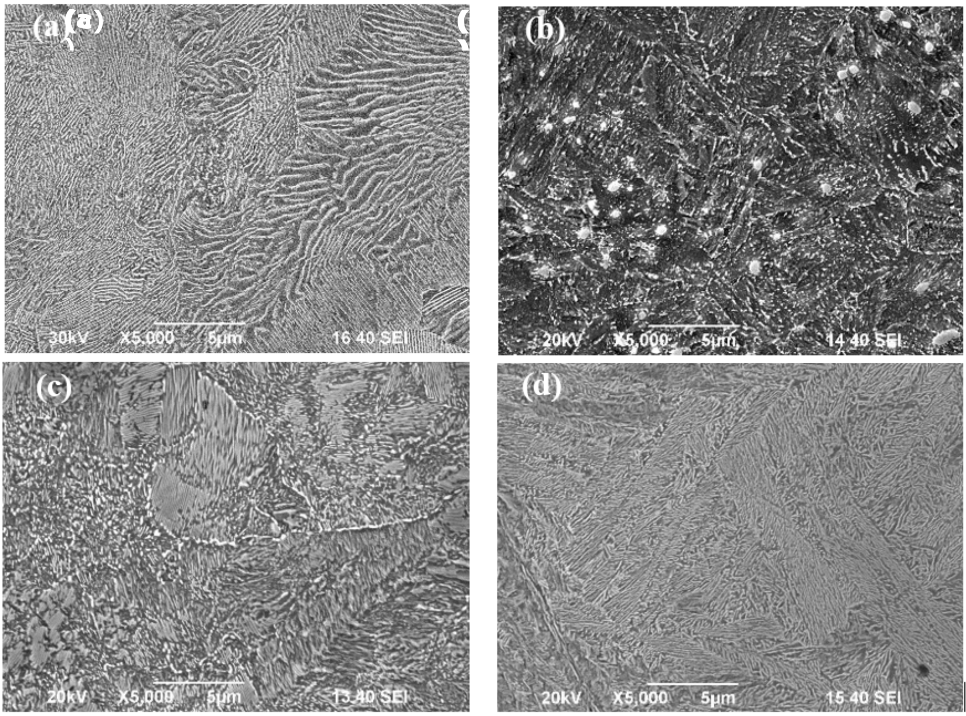


Fig. 2. Images of microsections from the training sample: (a) ferritic-carbide mixture, (b) secondary troostite-sorbite+chromium carbides, (c) lamellar pearlite +granular pearlite, (d) lower bainite.

During the process of preparing the training set of available images with the help of various transformations, learn and verification sets of 8200 and 2000 images were formed, respectively. This was done in order to increase the learning set and the accuracy of recognition. The learning set was divided into classes on microhardness. Classes of microhardness and distribution of the learning set on them are presented in Table.1.

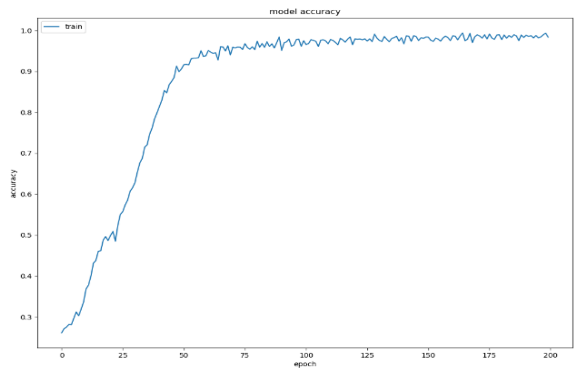
Table 1. Distribution of the training set by microhardness classes

Microhardness classes	Number of images in the learning set	Number of images in the verification set
I class, (0-2600 MPa)	940	240
II class, (2600-2900 MPa)	1420	260
III class, (2900-3200 MPa)	2620	640
IV class, (3200-3500 MPa)	1480	360

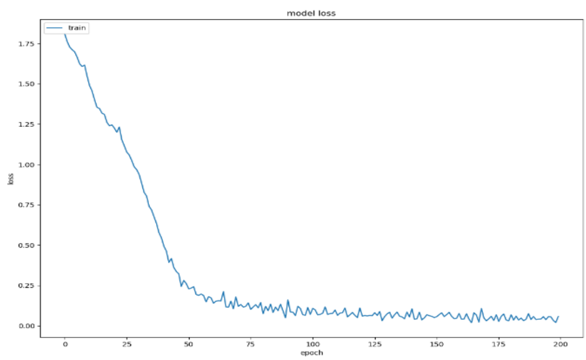
Learning process

The Keras framework in conjunction with the TensorFlow machine learning library was chosen as a software platform for the implementation of the network.

The learning was conducted on a hardware platform with the following characteristics: VGG16 pretrained network with imagenet weights on 200 epochs. One computational experiment took 9 hours. Network learning lasted no more than 525 iterations. The calculation of accuracy was performed every epoch and is equal to the percentage of correctly recognized images from the testing set formed from the learning set using random selection. Fig.3 shows a graph of the accuracy changes in the learning process. At the end of learning, the network recognition accuracy reached 62.5%. Figure 3, b shows the dependence of the change of the error function on the epoch number.



(a)



(b)

Fig. 3. Learning process: a) accuracy changes, 6) changes of the error function.

From Fig. 3 it can be seen that the learning process is typical for this kind of problems and does not have any features.

Improvement

Second part of experimental calculations is done with training dataset divided to 14 classes in terms of microhardness and tensile strength. The new class ranges and counts are shown in Table 2.

Table 2. Training dataset class distribution

Class #	Microhardness range, MPa	Tensile strength range, MPa	Image count
0	2100-2444	660-830	66
1	2100-2444	831-1001	36
2	2445-2789	831-1001	160
3	2445-2789	1002-1171	47
4	2445-2789	1172-1342	32
5	2790-3134	831-1001	158
6	2790-3134	1002-1171	166

7	2790-3134	1172-1342	30
8	3135-3479	1002-1171	55
9	3135-3479	1172-1342	188
10	3480-3824	1002-1171	28
11	3480-3824	1172-1342	89
12	3825-4169	1343-1513	19
13	4170-4514	1343-1513	22

This data set was used for training VGG-16[1], VGG-16[2] and ResNet-152. Learning process shown on Figure 4. It can be noticed that non pretrained VGG-16[2] acting unstable while training. Several times calculation experiment was restarted due to significant accuracy drops. Sometimes training process lead to memory crashes.

The process of studying stability was to find such a neighborhood of a stable solution, in which there is a violation of the uniform nature of convergence to the mean. A stable solution was understood as the trained deep network ResNet-152. The network was trained to solve the problem of classifying microstructure images by the hardness of a metal iron-based alloy. Annotated images of the microstructure of the alloys were used as a training set, examples of which are shown in Fig.1. In this case, the microstructure shown in Fig. 4 on the left corresponds to an alloy with a microhardness of 1900 MPa, and on the right – 7000 MPa.

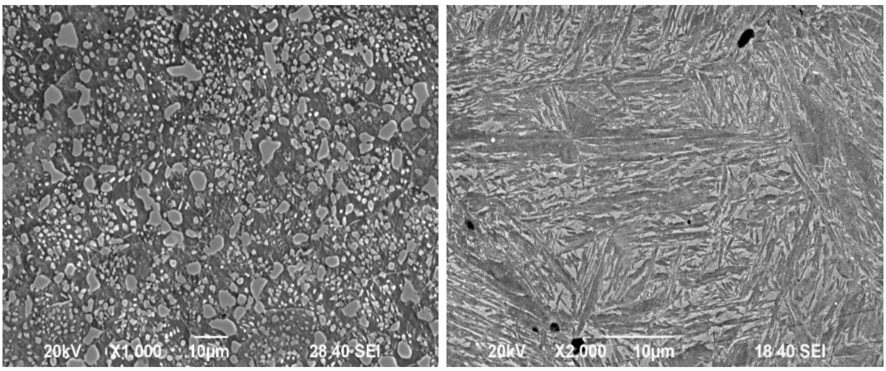
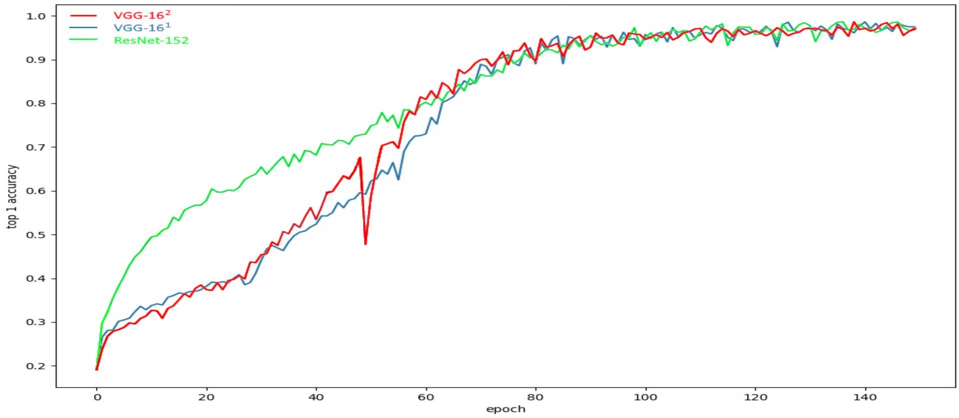


Fig. 4. Examples of images of microsection of material included in the training set

In the process of training the network, at the 200th epoch, an accuracy of 83.9% was achieved according to Top-3, after which the weights of the network were frozen. Accuracy assessment was made on a set consisting of 1097 elements not presented to the network during its training. The process of convergence in accuracy calculated on the training set is depicted in Figure 2.

ResNet-152 network shows a high rate of convergence. For almost a few epochs, an accuracy of 80% on Top-3 was achieved (Figure 5). The learning process is stable.



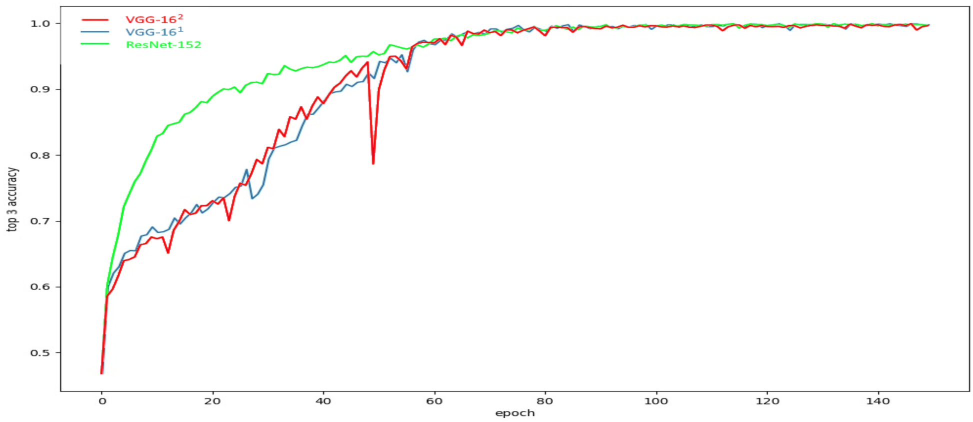


Fig. 5. Accuracy on the training set on the evaluation of Top-1 and Top-3

Table 3 shows the learning outcomes performed by different estimates on the training and validation sets.

Table 3. Network Learning Outcomes

Network	Training accuracy		Validation accuracy	
	Top-1	Top-3	Top-1	Top-3
VGG-16[1]	0.9751	0.9975	0.4704	0.8117
VGG-16[2]	0.9712	0.9970	0.5203	0.8142
ResNet-152	0.9593	0.9978	0.6243	**0.8933**

A definite surprise was that, despite the instability of learning, the unexperienced VGG-162 network showed a much better Top-1 result compared to the pre-trained version of the same network. This can be explained by the fact that the first layers of the convolutional untreated network were able to better adapt to specific images of microstructure. The pre-trained network has previously been trained on a very large and diverse set of photos. This improved the universal properties of the VGG-161 network (in particular, its first layers), but worsened the degree of recognition of structures of a special type.

As expected, the ResNet-152 network showed the best result of 83.9% in the Top-3 rating. The shown accuracy allows the use of a trained network as the core of an intelligent system for the integrated assessment of the strength properties of functional and structural materials. For the above computational experiments, the training set was divided into 14 classes of metals and alloys microstructures. Each of the 14 classes corresponds to the range of values of microhardness and tensile strength. Thus, the classification problem within the framework of these parameters was solved [21, 22], but the accuracy indicated in Table 1 can be improved by changing the training set.

A significant improvement in ResNet-152 results was achieved while improving the quality of the training set. Samples of images with magnification of x40 were excluded from it. Firstly, their number was small, and secondly, images with such multiplicity were not present in all categories. It was suggested that these images impair learning outcomes. Table 4 presents the recognition accuracy by material class.

Table 4. Recognition Accuracy by Material Class

Class number	**Validation accuracy**	
	Top-1	Top-3
0	0.718	1.000
1	0.812	1.000
2	0.593	0.843
3	0.562	0.812
4	0.375	0.937
5	0.937	1.000
6	0.687	0.937
7	0.500	0.933
8	0.562	0.875
9	0.906	0.906
10	**0.937**	**1.000**
11	**0.375**	**0.843**
12	0.894	1.000
13	0.590	1.000

The best accuracy according to both estimates was achieved in class 10. The worst accuracy was estimated at Top-1 in classes 11 and 4. At the same time, according to Top-3, acceptable accuracy was reached at the same classes, which was 84.3%.

Sustainability research

Determining the accuracy of the network on the test set showed that the density of the error distribution can be approximated by the lognormal distribution (Fig.6), which once again confirms the correctness of the network. The optimization criterion in the learning process was cross-entropy, which is known to be proportional to the logarithm of the network error probability distribution.

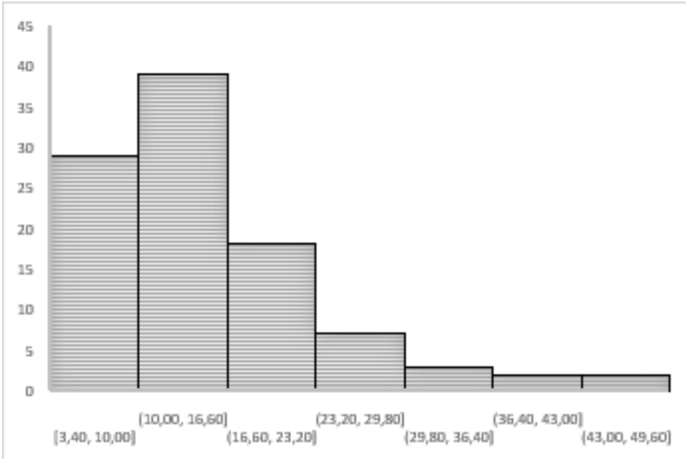


Fig. 6. The histogram of the distribution of accuracy in the test set

To study the stability of a trained network, the test set was subjected to variations. The neighborhood of a stable solution was formed by excluding a certain number of elements from the test set. Elements were randomly selected on 100 implementations. Such a neighborhood is called Leave-one-out (LOO) [4]. Usually its depth is 1 element. In this work, several neighborhoods were built: with a 2, 4, 6, 8, 10 and 12 percent deviation from the base number of elements in the test set, which was used to assess the accuracy of a stable solution. After running and determining the accuracy for all implementations in the neighborhood, the resulting accuracy was obtained by averaging. The results of the computational experiment are shown in Figure 7.

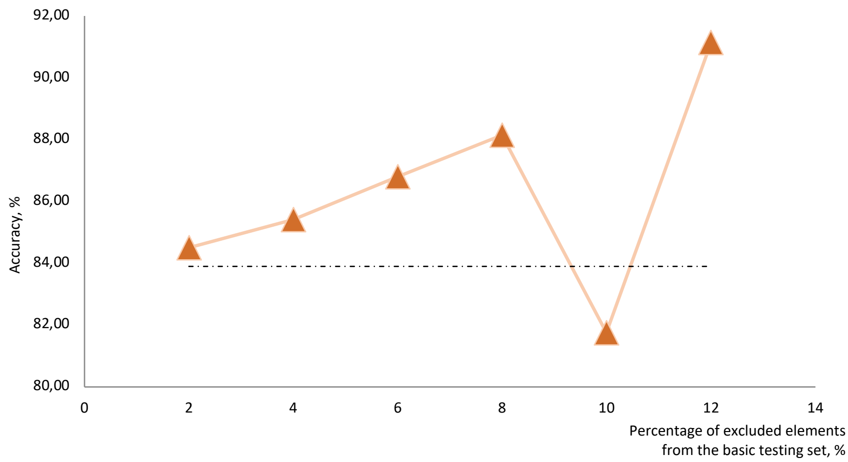


Fig. 7. Dependence of the accuracy of the trained neural network in the LOO neighborhood of the basic testing set

4. Results

From Figure 7 it follows that the process of validation the trained network evenly converges to the accuracy obtained on the base sample (indicated by the dotted line). When the number of elements in the test set decreases by more than 8%, there is a sharp oscillation of the accuracy estimate, which, in our opinion, is the loss of the algorithmic stability of the neural network in the classification problem under consideration, which it solves.

This statement may be applied to the learning process, since the test set is an independent random homogeneous sample. If the algorithm lost stability on the test sample during the control test of the accuracy of the network, then the network learning process will lose stability when the number of elements in the training sample decreases by more than 8%. Deviations of 8% cannot be considered random, since the accuracy obtained by averaging over 100 computational experiments for each percent distribution where on each iteration random elements were removed from full elements set in according to that percent count.

It should be noted that the search for the loss of stability of the learning process in deep networks by the direct method is a difficult task. If the network learning process on a training set with a size of decades of thousands of elements can take several days, then the study of learning sustainability is difficult from a practical point of view. The authors of this paper propose a simplified approach — investigate uniform convergence on a test set, and then extend the findings to the entire training set.

The result obtained in this paper may impose not only quantitative but also qualitative limitations on the training set. The resulting limit may indicate that in the training set there cannot be more than 8% of erroneously annotated images. Otherwise, the learning process will lose its stability and the learning outcomes cannot be trusted.

5. Conclusion

The paper studies the problem of uniform convergence of the learning process and the accuracy assessment of the deep ResNet-152 network in the problem of analysis of microsections of iron-based alloys. It is shown that with 8% quantitative deviation in the base set, the algorithm of the trained network loses stability. This means that with so many elements in the test set an adequate assessment of the accuracy of the network is impossible.

The methodology for assessing the stability of the deep network, applied in this work, can be extended to other networks and tasks. It does not require volumetric computations, since it allows one to estimate the sufficiency of the number of elements in the training set without performing network training on training sets having various capacities.

Acknowledgments

The reported study was funded by the Ministry of Science and Higher Education of the Russian Federation (the unique identifier RFMEFI58617X0055) and by the EC Horizon 2020 is MSCA-RISE-2016 FRAMED Fracture across Scales and Materials, Processes and Disciplines. The authors are grateful to the staff of the Institute of Nanosteels of MSTU named after G.I. Nosov, in particular to M. P. Baryshnikov, for the experimental data provided, which made it possible to train the constructed neural network with a given accuracy.

References

1. Gitman M.B., Klyuev A.V., Stolbov V.Y., Gitman I.M. Complex Estimation of Strength Properties of Functional Materials on the basis of the Analysis of Grain-Phase Structure Parameters // Strength of Materials. 2017. Vol. 49. No.5. – PP. 710-717.
2. Kliuev A., Klestov R., Bartolomey M., Rogozhnikov A. (2019) Recommendation System for Material Scientists Based on Deep Learn Neural Network. In: DSIC 2018. Advances in Intelligent Systems and Computing. vol 850, pp. 216-223.Springer, Cham.
3. Vapnik V. N., Chervonenkis A. YA. Teoriya raspoznavaniya obrazov / M.: Nauka (1974).
4. Bousquet O., Elisseeff A. Algorithmic Stability and Generalization Performance. Advances in Neural Information Processing Systems. 13, 196-202 (2001).

5. Bousquet O., Elisseeff A. Stability and Generalization. Journal of Machine Learning Research. 2, 499-526 (2002).
6. Jason Brownlee (2019) Impact of Dataset Size on Deep Learning Model Skill and Performance Estimates [Digital resource]. URL (https://machinelearningmastery.com/impact-of-dataset-size-on-deep-learning-model-skill-and-performance-estimates) (Request date 04.02.2019).
7. Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy, Synho Do (2015) How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? arXiv:1511.06348 – 2015.- preprint arXiv.
8. LeCun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W. and Jackel L. D.: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, (1989).
9. Szegedy Christian, Liu Wei, Jia Yangqing, Sermanet Pierre, Reed Scott, Anguelov Dragomir, Erhan Dumitru, Vanhoucke Vincent, Rabinovich Andrew: Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1-9 (2015).
10. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp.770-778 (2016).
11. Jie Hu, Li Shen, Gang Sun: Squeeze-and-Excitation Networks. ILSVRC 2017 image classification winner. arXiv:1709.01507 (2017).
12. Brian L. DeCost, Toby Francis, Elizabeth A. Holm: Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures. Acta Materialia. 133, 20-40 (2017).
13. Azimi S. M., Britz D., Engstler M., Fritz M., Mücklich F.: Advanced Steel Microstructure Classification by Deep Learning Methods. arXiv: 1706.06480 (2017).
14. G.V. Aristov, A.V. Klyuev: Detection and classification metal's and alloy's microstructure using deep neural network. Proceeding of 27th International conferencion of computer graphics and machine vision GraphiCon. p.180-183 (2017)
15. Masci J., Meier U., Ciresan D., Schmidhuber J., Fricout G.: Steel defect classification with Max-Pooling Convolutional Neural Networks. The 2012 International Joint Conference on Neural Networks (IJCNN), pp.1-6-131 (2012).
16. Shiyang Zhou, Youping Chen, Dailin Zhang, Jingming Xie, Yunfei Zhou: Classification of surface defects on steel sheet using convolutional neural networks. Materiali in tehnologije, vol.51, pp. 123 (2017).
17. Soukup D., Huber-Mork R.: Convolutional Neural Networks for Steel Surface Defect Detection from Photometric Stereo Images. ISVC 2014: Advances in Visual Computing, pp. 668-677 (2014).
18. Faghih-Roohi S., Hajizadeh S., Nu´nez A., Babuska R., De Schutter B.: Deep convolutional neural networks for detection of rail surface defects. Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN 2016), pp. 2584–2589 (2016).
19. Goodfellow Ian J., Pouget-Abadie Jean, Mirza Mehdi, Xu Bing, Warde-Farley David, Ozair Sherjil, Courville Aaron, Bengio Yoshua: Generative Adversarial Networks. arXiv:1406.2661 (2014).
20. Alec Radford, Luke Metz, Soumith Chintala: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv:1511.06434 (2015).
21. Gatys L. A., Ecker A. S., Bethge M.: Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks. arXiv:1505.07376 (2015).
22. Nicholas Lubbers, Turab Lookman, Kipton Barros: Inferring low-dimensional microstructure representations using convolutional neural networks. arXiv:1611.02764v1 (2016).
23. Klyuev A.V., Stolbov V.Y., Gitman M.B., Klestov R.A. (2020) Stability of the Deep Neural Networks Learning Process in the Recognition Problems of the Material Microstructure. In: ICIS 2019. Lecture Notes in Networks and Systems, vol 78. pp. 307-313. Springer, Cham. DOI: 10.1007/978-3-030-22493-6_27.
24. Klestov R., Klyuev A., Stolbov V. About some approaches to problem of metals and alloys microstructures classification based on neural network technologies .Advances in Engineering Research.157, 292-296 (2018).
25. Gitman I.M., Klyuev A.V., Gitman M.B. and Stolbov V.Yu. Multi-scale Approach for Strength Properties Estimation in Functional Materials. ZAMM Zeitschrift fur Angewandte Mathematik und Mechanik, vol.98. No.6. 945-953 (2018).

Aims and Objectives
Published online by ICS two times a year, Journal of Digital Science (JDS) is an international peer-reviewed journal which aims at the latest ideas, innovations, trends, experiences and concerns in the field of digital science covering all areas of the scholarly literature of the sciences, social sciences. The main topics currently covered include: Digital Communications and Network; Digital Economics, Education, Engineering, Finance, Health Care.

The main goal of the journal is the effective dissemination of original incites/results generated by the human brain and presented/reflected in articles using modern information/digital technology.

Contact information

Website: https://ics.events

Email: conf@ics.events