DOI: 10.36163/aon-2024-0010

AUTOMATIC AIRCRAFT RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK

Dominik Adamiak
Polish Air Force University, Institute of Navigation
Dywizjonu 303 no 35 Street, 08-521 Dęblin, Poland
d.adamiak6888@wsosp.edu.pl

Anna Ślesicka
Polish Air Force University, Institute of Navigation
Dywizjonu 303 no 35 Street, 08-521 Dęblin, Poland
a.slesicka@law.mil.pl

ABSTRACT

The paper presents the possibility of utilisation convolutional neural network for aircraft classification by their type. Main purpose of the study was to use a common deep learning network and modify it to correctly classify three types of general aviation aircraft. Differentiation is made based on their low quality picture with black outline on white background. Database utilized in this study is small compared to common CNN databases and results ought to be taken only as a trend. Research consisted of choosing right parameters of network to make the recognition as accurate as possible. 20 samples have been produced to evaluate accuracy of the software and eliminate deviations. Conclusions and issues have also been described.

Keywords — deep learning, aircraft, classification

1. INDRUCTION

Deep learning is consistently gaining a lot of attention nowadays due to increased employment of artificial intelligence in everyday electronics. Object differentiation is one of possibilities created by CNN, however it is crucial to distinguish level of importance with utilization AI verdicts [1]. With, for example, smartphone additional software, little errors will rather not pose a risk to software users. Artificial Intelligence needs to be under specific control when it is operational, since this kind of software is not infallible [2]. Its strongly depends on the poor information selection from the allowed database. This topic is crucial when it comes to aviation [3] – risk of mistakenly assigning low speed aircraft to high speed one might result in giving wrong recommendations, possibly with an incidents. To avoid such risks, operator of AI enhanced software need to be sure of its high accuracy [4]. That is why Artificial Intelligence ought not make decisions for people, rather just provide a form of additional information.

Research was focused on using deep learning network, which was originally used to recognize military vehicles by their radar reflection. MATLAB code was modified by me to the point it was able to differentiate any monochromatic shapes. Main purpose of this network is to distinguish three common general aviation aircraft: Cessna 152, Diamond DA20 and Diamond DA42. Every picture used in the database for analysis was made for this purpose.

This innovative feature can be used as additional Air Traffic Control support and help differentiate general aviation aircraft, which often lacks advanced equipment transmitting information about their type. Simultaneously, limiting operation of this software to the least problematic air traffic with similar specification aircraft would not pose a major risk of giving incorrect instructions to the pilot due to the software failure. In case of uncertainty, when aircraft type information is crucial for airspace management, controller may ask pilot for their type and, when AI incorrectly classify aircraft, make corrections in the system. Such incidents would be further analysed by both the software engineers and program themself, including new pictures in the constantly growing database. Monochromatic aircraft visualisations may be gathered by collecting their radar reflections.

2. RELATED WORKS

Topic of utilization deep learning network both for increasing aviation safety and image recognition are commonly found in scientific papers. A significant part of the image classifications are conducted with SAR images [5]. Chinese scientists proposed utilization of CNN network to distinguish aircraft from optical images [6] and even estimate aircraft landing gear angles [7]. Major works with deep learning used for ship recognition are also ongoing [8] [9]. Military industry also draws from intensive research with deep learning image recognition, both for identifying different objects or vehicles [10] [11] and strategically important environmental changes [12] [13].

3. NETWORK STRUCTURE

3.1. DATABASE CREATION

To create fully valuable database, a lot of different positions of the same aircraft were chosen and outlined to get black aircraft shape on a white background. Outlines have some shades of grey from the production process – creating high quality raster images from vectors.

Fig. 1. Diamond DA-20 picture included in the database.

If one type of aircraft exists with variations in equipment (e.g. different landing gear) or can be presented with different flaps or landing gear configuration, it was implemented in the database.

Database consists of 30 pictures of each type of aircraft, 90 pictures in total. It is significantly too small number of pictures for a deep learning network but it can show trends, if program is working correctly. From this base, pictures are divided for three purposes: learning, validation and classification [11]. To make it work correctly and avoid false trends, following proportions were chosen: 80% for learning, 10% for validation and 10% for classification.

Every picture included in the database is a png format, resolution 1500 x 1500 px. The CNN network however does not need such a high quality to recognize aircraft correctly [6]. Every picture is automatically downscaled by the script to 64 x 64 px. It is also possible to manipulate the size of the picture by changing one parameter in the script, not necessarily creating new pictures with different resolutions. Higher resolution results in significantly longer learning process and not necessarily in increased accuracy. It also forces user to reduce amount of filters used in layers due to amount of RAM memory utilization.

3.2. SOFTTWARE OPERATIONS

After randomly choosing pictures based on the parameters mentioned in previous paragraph, the program will display exemplary pictures chosen for the learning process. Non – proportional amount of types of aircraft on this scheme should not be taken into account as these are only samples from learning pictures separated for training, which are chosen proportionally by the software – there is no possibility to assign more e.g. C152 for training and only DA42 for classification. This operation is further showcased with classification explanation.

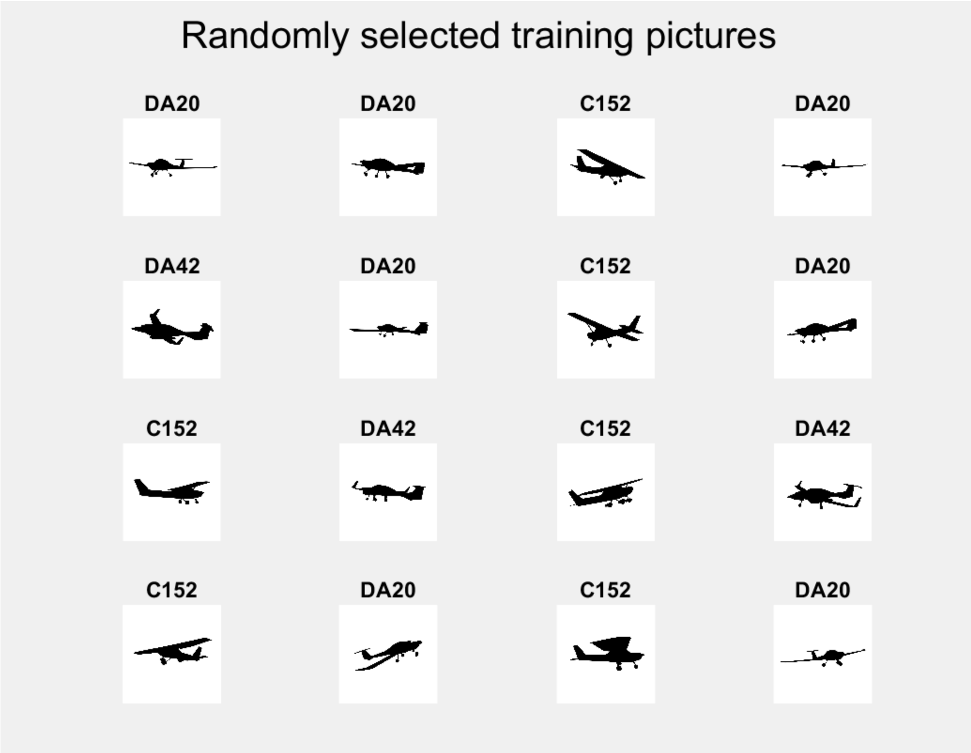

Fig. 2. Randomly selected training pictures shown for the program user.

In the figure above, many different pictures from the utilized database can be seen. In this example, software showcased more DA20 aircraft compared to other types for the user. As mentioned before, it is only illustrative image.

Neural network consist of three types of layers: convolutional, pooling and fully connected. In this network, convolutional and pooling layers are used alternately while fully connected layer is placed at the end of operation. Pooling layers objective is to reduce the spatial dimensions of the picture, while convolutional layers extract complex features [14]. Every one of three subsequent convolutional layers has also increased filter size, starting at 3x32, multiplying each next value by two.

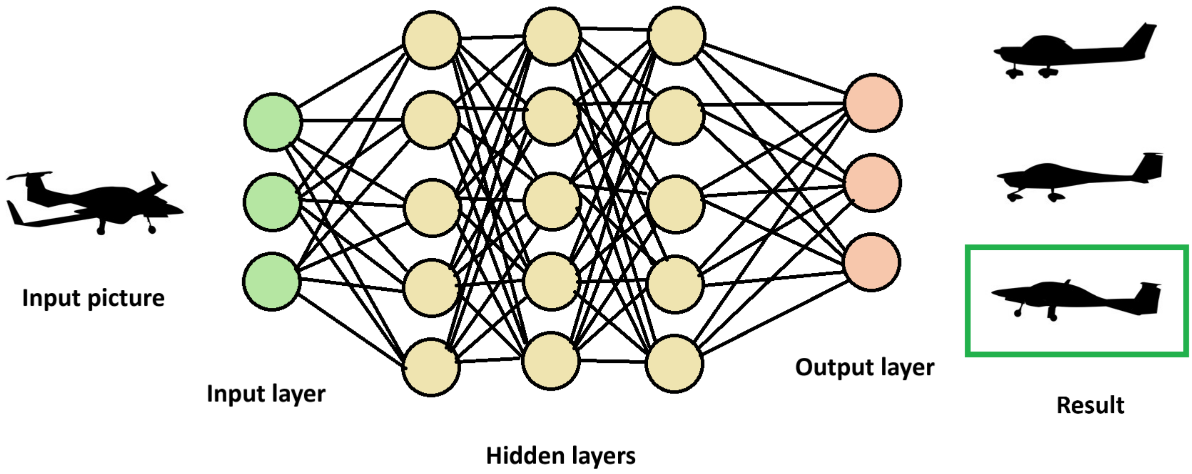

Fig. 3. Exemplary CNN layer scheme.

Every pooling layer is built with “max rate” characteristic. Deep learning network focused on recognizing objects discriminates most characteristic parts of each picture with samples that it takes [15]. With max rate, it chooses most unique local features of each sample. It is a common scheme. When utilized with aviation, max rate is crucial. Many different parts are indistinguishable in different aircraft types.

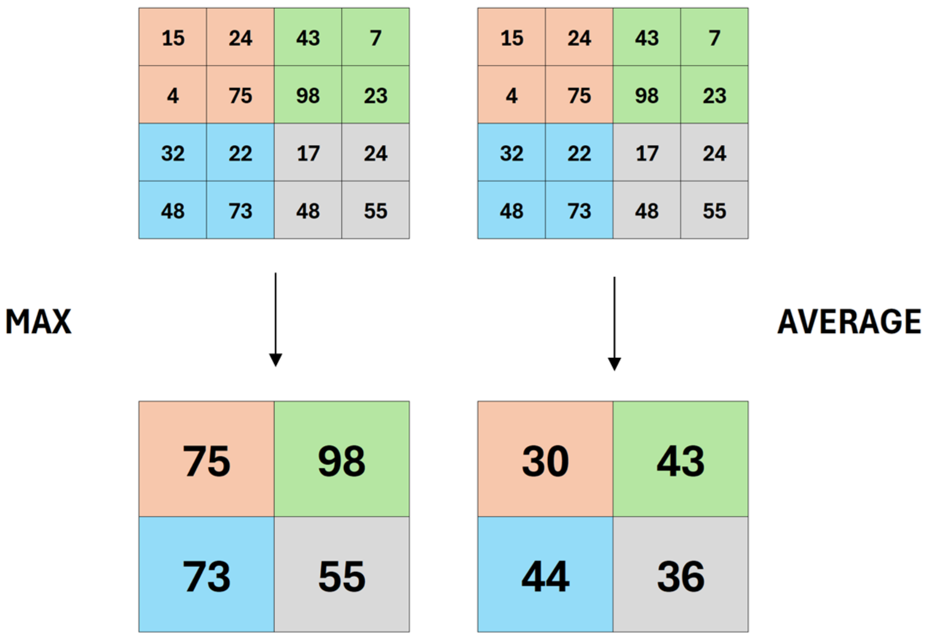

Fig. 4. “Max Rate” and “Average Rate” outputs comparison.
2x2 Pooling, Stride of 2.

Learning process is showcased as a dynamic graph, which is created simultaneously with learning, Every iteration is marked on it. When working with relatively small databases, amount of iterations is equal to amount of epochs. That is how it works in this example.

Model training progress should increase its accuracy with each epoch, ultimately finishing at around 100% after a certain amount of epochs. With example presented here, training process is basically finished at about 15 epoch. After that, no major increase or decrease in accuracy ought happen. However, forcing the network to go through more epochs is time consuming and decreasing efficiency of the program. Every sample showcased in further part of this report was made with epoch amount varying between 15 to 30. No major difference in accuracy was noted.

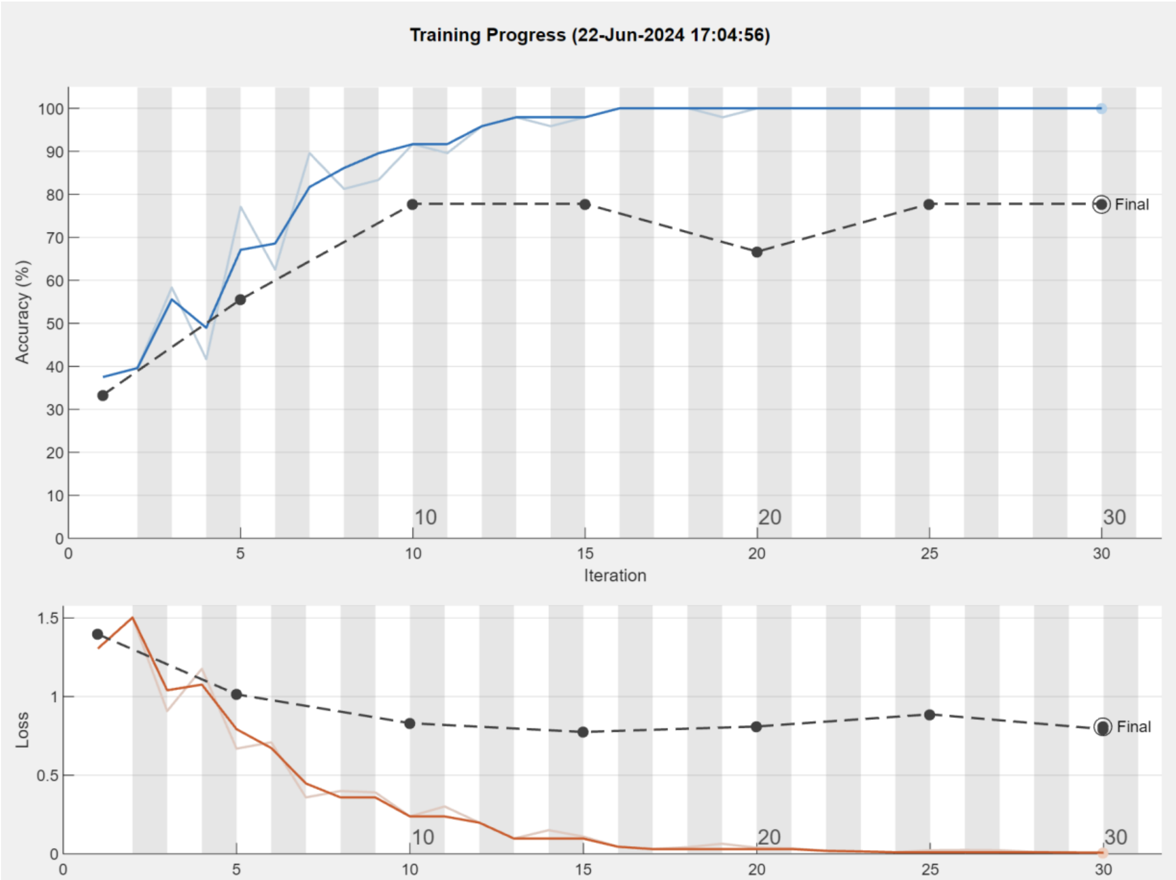

Fig. 5. Dynamic training progress visualized by MATLAB.

15 epoch training takes around one minute on an average laptop that was used for this study. During this process, around 12 GB of RAM memory was used at peek moments.

The final part of the software operation is to classify remaining pictures to one of three available types of aircraft. This attempt is showcased on a graph - "Confusion Matrix". In this example, 10% of database is used for classification, which means 9 pictures, 3 of each type of aircraft. While this amount may be considered as insufficient, increasing classification value will only reduce network efficiency due to the need to decrease learning value. With such a small database, this proportions are necessary for overall representative result. Percentage values shown on the right side of the graph indicate accuracy of the recognition based on one dimensional compatibility. Worth noting is that it showcases full accuracy when less than three pictures are assigned to the category, but also indices less accuracy when three correct pictures are correct with few confusions.

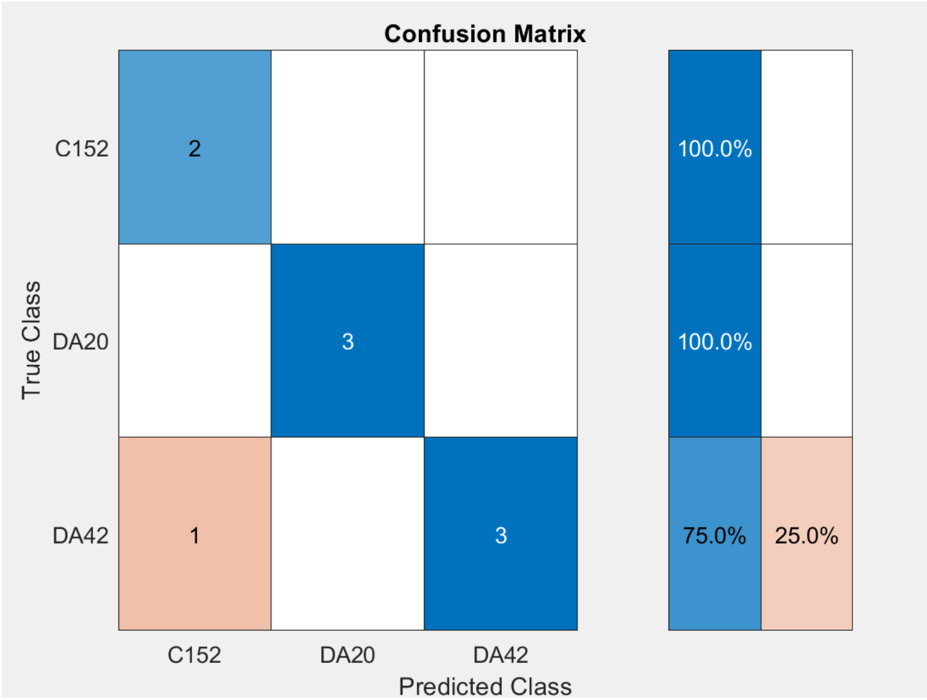

Fig. 6. Confusion Matrix of 19th trial – the highest overall accuracy during study.

Confusion Matrix presented in fig. 6. has the highest overall accuracy of every 20 samples taken into account during this study. However, during process of adaptation network to aircraft recognition, fully 100% results were achieved a few times. Such a case however did not occurred during this study.

Percentage values on the right side of the matrix may serve as an example of accuracy calculation. One Cessna 152 was mistakenly classified as Diamond DA42. In C152 category, accuracy shown is 100% even with one missing object. Reduction of accuracy is however observed in DA42 category due to excessive amount of pictures assigned to this category.

2.3 CLASSIFICATION ERRORS

The major flaw of this program is its inconsistency. 20% of samples are totally wrong and cannot be used for recognition. To make sure that our results fit into permissible range of error, program need to perform this same procedure a few times and choose most optimal result via algorithm, rejecting such deviations. The only possibility of happening wrong result from software site is due to randomly choosing unfavourable pictures for learning, which substantially differ from pictures chosen for classification. This is the only random factor in the entire script. This error ought to disappear with the increase of database size.

To prevent software from choosing deviations into results by the software, simple algorithm may be used, which would check regularity of placement classification images on confusion matrix and, in case of major deviation, reject such sample from verdict. It is also necessary, as mentioned multiple times, to not only base verdict on one sample - the program should be ran three to four times to maximally increase accuracy and, at the same time, reject deviations with second verification method comparing it to correct results.

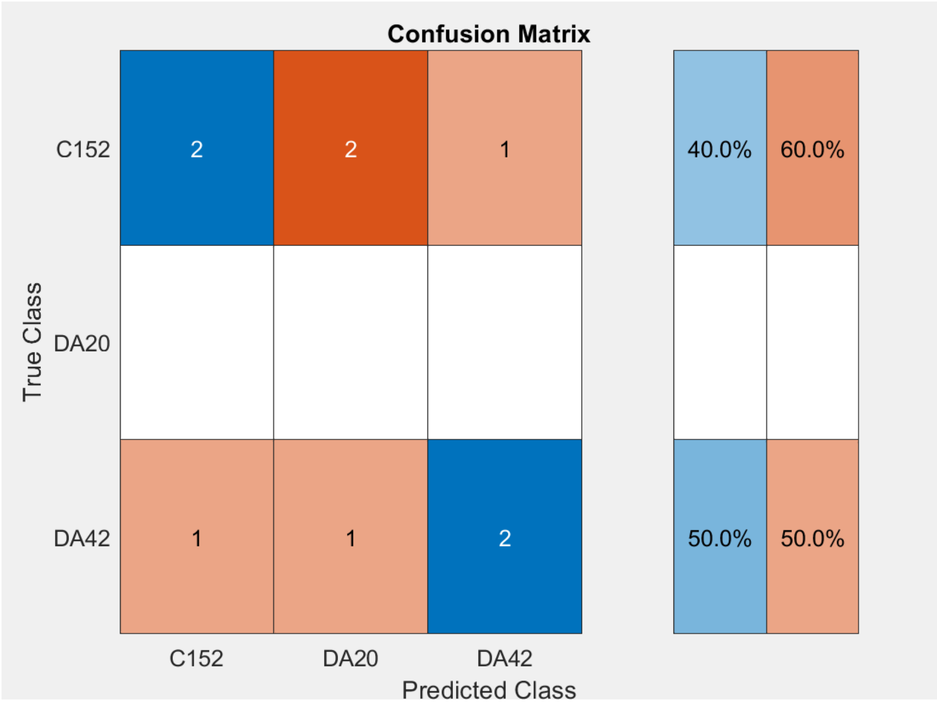

Fig. 7. Confusion Matrix of 9th trial – example of misleading result.

4. SAMPLE ANALYSIS

To showcase the real possibilities of the program, 20 following classifications were produced. During study, no networks parameters have been changed besides lowering epoch amount from 30 to 15 for 11 – 20 attempts. Sample results are showcased below. Values correspond confusion matrixes' accuracy returned by MATLAB.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
C 152	100%	75%	50%	100%	33,30%	100%	75%	66,70%	40%	100%
DA 20	66,70%	66,70%	0%	60%	50%	66,70%	50%	75%	0%	100%
DA 42	66,70%	50%	50%	100%	50%	66,70%	66,70%	100%	50%	50%

	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
C 152	75%	60%	25%	75%	100%	66,70%	100%	100%	100%	100%
DA 20	66,70%	100%	66,70%	50%	75%	100%	60%	60%	100%	100%
DA 42	50%	66,70%	50%	66,70%	75%	75%	100%	100%	75%	60%

Tab. 1. Classification accuracy of 20 exemplary samples.

3rd, 5th, 9th and 13th samples are examples of incorrect classification, it is the reason why they are highlighted. There is no rule when such mistake can happen. Early samples had significantly more errors when compared to final ones. This happened due to the only random factor this software has – random image assignment from the database for different purposes.

	C 152	DA 20	DA 42
AVERAGE	77%	65,68%	68,43%
MEDIAN	75%	66,70%	66,70%

Tab. 2. Combined statistics of 20 samples with deviations included.

First combined statistic shows accuracy including false results. In case of multiple attempts, deviations ought to be excluded from final verdict and accuracy in recognition of certain aircraft will be higher.

	C 152	DA 20	DA 42
AVERAGE	87%	75%	73%
MEDIAN	100%	66,70%	66,70%

Tab. 3. Combined statistics of 16 samples, deviations rejected.

Rejecting deviations noticeably increased accuracy, especially in case of Cessna 152. However, medians of DA20 and DA42 remain unchanged. Simple conclusion arising is the program is very consistent with the classifications and even deviations appearing occasionally in the results does not ruin the operation.

5. REMARKS AND CONCLUSIONS

From sample analysis, it is simple to conclude that convolutional neural network shown in this paper tends to correctly recognise basic general aviation aircraft. Besides generally positive results, not a single sample from 20 attempts made during study was characterized by 100 % accuracy. It occurred a few times during early stages of research abut most efficient network parameters, but have not been repeated in any sample. The highest probability of such an error is database size – not every aircraft position is covered and network struggles to discriminate the type with a new, previously unknown placement on a picture. With deep learning network being used for classification of complicated objects, such as aircraft, every part of it could be taken by network as a defining factor. It becomes even more problematic when one aircraft exists in multiple configurations. Database, besides having as many possible placements as possible, should also cover multiple configurations of every type.

Fig. 8. Comparison of DA42 type in two landing gear configurations.
On the left: DA42 with landing gear deployed. On the right: DA42 with landing gear retracted.

Fig. 9. Comparison of DA20 with two types of landing gear, without and with landing gear fairing.

Increasing database by at least few times or possibly rebuilding network to accept 3D learning samples should fix the issue of single misleading results during generally successful attempts.

In case size of the database is also the reason for fully incorrect classifications, increasing amount of pictures ought to also fix second major issue with program. All the circumstantial evidence points to this being the main issue, but in case expansion of the database will not fix the issue, other actions could be taken. As mentioned in previous paragraphs, software may utilize algorithm to detect and reject obvious deviations from final results. It would also be crucial to test every aircraft multiple times to create as unambiguous statistics for algorithm as possible and make sure that final result is correct.

It should be noted that manual implementation of every single high quality aircraft picture valid for deep learning utilization is very time consuming. Many different aircraft positions need to be chosen with different configurations, then manually outlined and converted to format supported by MATLAB. Both selection and graphical processes are time consuming and that is the reason why only limited size database was used for this study.

REFERENCES

1. Y. Li, Y. Xiao, Y. Gong, R. Zhang, Y. Huo, Y. Wu. Explainable AI: A Way to Achieve Trustworthy AI. IEEE International Conference on Big Data Security on Cloud (BigDataSecurity), High Performance and Smart Computing (HPSC) and Intelligent Data and Security (IDS), 2024.
2. Q. Lu, L. Zhu, X. Xu, J. Whittle, D. Douglas, C. Sanderson. Software Engineering for Responsible AI: An Empirical Study and Operationalised Patterns. IEEE/ACM International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP). 2022.
3. K. Tejasen, P. Raju, W. Ryan, A. Jordan, R. Yang. Defining an Initial Classification Scheme for Non-Deterministic AI Technologies. IEEE Integrated Communications, Navigation and Surveillance Conference, ICNS, 2022.
4. Y. Camgözlü, Y. Kutlu. Analysis of Filter Size Effect in Deep Learning. Journal of Artificial Intelligence with Applications, 1(1), 20-29, 2020.
5. A. Passah, S. Sur, B. Paul, D. Kandar. SAR Image Classification: A comprehensive Study and Analysis. IEEE Access, 2022.
6. Y. Wang, H. Li, P. Jia, G. Zhang, T. Wang, X. Hao. Multi-Scale DenseNets-Based Aircraft Detection from Remote Sensing Images. Sensors 2019.
7. F. Li, Z. Wu, J. Li, Z. Lai, B. Zhao, C. Min. A Multi-Step CNN-Based Estimation of Aircraft Landing Gear Angles. Sensors, 2021.

8. J. Li, Z. Yu, J. Chen, H. Jiang. A SAR Ship Detection Method Based on Adversarial Training. Naval Submarine Academy, Qingdao 264001, China, 2024.
9. J. Iu, C. Xu, H. Su, L. Gao, T. Wang. Deep Learning for SAR Ship Detection: Past, Present and Future. Remote Sens 2022.
10. D. Chae, H. Lim, J. Park, Y. Choi, J. hee Yoo. Application of Satellite SAR Image to Ground Vehicle Target Dataset and Detection and Object Segmentation". Journal of the KIMST, 25(1):30–44, 2022.
11. T. Hiippala. Recognizing Military Vehicles in Social Media Images Using Deep Learning. IEEE, 2017.
12. F. Chen, A. Zhang, H. Balzter, P. Ren, H. Zhou. Oil Spill SAR Image Segmentation via Probability Distribution Modelling. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15:533–554, 2022.
13. B. Choe. Polarimetric Synthetic Aperture Radar (SAR) Application for Geological Mapping and Resource Exploration in The Canadian Arctic. 2017.
14. Y. Liu, H. Pu, D. Sun. Efficient Extraction of Deep Image Features Using Convolutional Neural Network (CNN) for Applications in Detecting and Analyzing Complex Food Matrices, Trends in Food Science & Technology, 2021.
15. P. Lang, X. Fu, M. Martorella, J. Dong, R. Qin, X. Meng, M. Xie. A Comprehensive Survey of Machine Learning Applied to Radar Signal Processing. 2020.