ISSN 2079-3316 PROGRAM SYSTEMS: THEORY AND APPLICATIONS vol. 13, No 3(54), pp. 81–98 Research Article artificial intelligence, intelligent systems, neural networks

UDC 519.68+004.89 10.25209/2079-3316-2022-13-3-81-98

# Nodules detection on computer tomograms using neural networks

Dinara Halilovna **Giniyatova**<sup>1</sup>, Vasilii Aleksandrovich **Lapinskii**<sup>2</sup> Kazan Federal University, Kazan, Russia <sup>12</sup>normaliti@qmail.com (learn more about the authors on p. 98)

Abstract. Results of neural networks (NN) application to the problem of detecting neoplasms on computer tomograms of the lungs with limited amount of data are presented. Much attention is paid to the analysis and preprocessing of images as a factor improving the NN quality. The problem of NN overfitting and ways to solve it are considered. Results of the presented experiments allow drawing a conclusion about the efficiency of applying individual NN architectures in combination with data preprocessing methods to detection problems even in cases of a limited training set and a small size of detected objects.

Key words and phrases: object detection, image processing, neural networks, YOLO

2020 Mathematics Subject Classification: 68T07; 68T45, 92C55

Acknowledgments: This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program ("PRIORITY-2030")

For citation: Dinara H. Giniyatova, Vasilii A. Lapinskii. Nodules detection on computer tomograms using neural networks // Program Systems: Theory and Applications, 2022, 13:3(54), pp. 81–98. http://psta.psiras.ru/read/ psta2022\_3\_81-98.pdf

#### Introduction

Timely detection of small nodular neoplasms in the lung tissue gives patients a chance to detect a tumor at the earliest stages. According to the World Health Organization, lung cancer topped the ranking of oncological diseases in  $2020^{\textcircled{m}}$ , becoming the most common cause of death. About 1.8

© Giniyatova D. H., Lapinskii V. A. 2022 @ Эта статья по-русски: http://psta.psiras.ru/read/psta2022 3 61-79.pdf



million people died from the disease. Late diagnosis largely reduces the effectiveness of treatment, so it is important to develop automatic methods for high-quality search for pathological areas, which will allow specialists to quickly and accurately detect lung neoplasms.

In recent years, using constantly evolving technologies and machine learning algorithms, high-tech solutions for medical problems have been developed. In particular, neural networks (NNs) that can detect melanoma with high accuracy on dermatoscopic images [1] or predict survival in patients with lung cancer based on x-ray examination [2] have been implemented. Numerous studies have shown that systems developed on the basis of artificial intelligence technologies in some cases give more accurate predictions than radiologists do. For example, the neural network for diagnosing breast cancer presented in [3] showed results better than those of 61.4% of radiologists. All this indicates the prospects of using deep learning for automatic screening, determining the stage of the disease, predicting the effect of treatment and the outcome of the disease.

The problems of analyzing medical images (ultrasound images, radiographs, magnetic resonance and computer tomography) can be divided into several types: classification, detection and segmentation. As a result of classification, one determines to which of the proposed classes an image belongs; for example, it is possible to determine the presence or absence of a disease (binary classification). The detection problem, as a rule, assumes the presence of several objects in the image, in contrast to localization, and consists in finding a bounding box for each of the objects and assigning a class label to it. This class of problems includes the detection of various anomalies in medical images. Segmentation is the division of an image into non-overlapping areas, where the class label is assigned pixel by pixel. A popular direction is, for example, the segmentation of blood vessels in images of the retina.

The purpose of the present work is to study methods for solving the detection problem with the application of neural networks using the example of detecting nodular neoplasms on computer tomograms of the chest organs.

### 1. Data Analysis and Preprocessing

One of the main problems in solving the problems of analyzing medical images using neural networks is the dataset. As a rule, it is not always possible to find labeled data with the necessary description of diseases, or there is very little data, and for their labeling it is necessary to gather a council of specialists, which requires significant amount of time and material resources. There exist several methods to overcome the problem. The first one involves the use of poorly labeled data, in which an information about the size and location of the anomaly is partially



FIGURE 1. Examples of digital tomograms from the NODE21 dataset



FIGURE 2. An example of the isolation of neoplasms

or completely absent, and only the patient's status (sick/healthy) is known. The second one involves the use of various data preprocessing and augmentation tools, which together can improve the quality and increase the original set. In the present study, we use the second approach.

The dataset NODE21<sup>®</sup> is used as the initial data, which is assembled from several other datasets, ChestX-ray8 [4], PadChest [5], JSRT<sup>®</sup>, Open-I<sup>®</sup>. The base dataset named above was proposed by the organizers of the NODE21–Grand Challenge competition. The NODE21 set consists of 4882 computer tomography (CT) scans in MHA format, of which only 1134 have neoplasms. A total of 1476 neoplasms were found on tomography.

Figure 1 shows examples of CTs from the NODE21 set, and Figure 2 shows an example with marked neoplasms.

Table 1 contains information about how many neoplasms are present in what number of images. The table shows that most of the images have only one neoplasm, and the maximum number of neoplasms present on the tomogram is 3.

Figure 3 shows a heatmap of the distribution of neoplasms based on CT in the dataset under consideration. According to the distribution obtained,



TABLE 1. The number of images to the number of neoplasms

FIGURE 3. Distribution of neoplasms by lung area in the NODE21 dataset

it can be said that nodular neoplasms are distributed evenly over the entire plane of the lungs; most often neoplasms are located in the central part.

As noted above, preprocessing is an important step in preparing data for training neural networks. Next, the main approaches to image preprocessing for the NODE21 dataset will be described.

Images without the desired objects are not of great value, since quality metrics evaluate the accuracy of the location of the bounding box, which prevents accidental detections. Therefore, as a preprocessing, it was decided to discard pure tomograms that do not contain data on the desired objects. In addition, images were used for training, for which the contrast was increased and parts of the image that did not belong to the lungs were cropped. Such areas do not carry useful information for solving the problem under consideration, and a smaller image size will speed up the learning process.

The next important stage of data preprocessing is augmentation in order to increase the amount of data and increase the stability of the model. After analyzing the dataset, it was decided to apply five augmentation algorithms: Gaussian noise, motion blur, horizontal and vertical rotations, and perspective.

Gaussian noise is random noise that has a Gaussian distribution. An example of an image with Gaussian noise is shown in 4a. This algorithm is used to make the neural network resistant to random noise that occurs

#### NODULES DETECTION





(a) Gaussian noise (b) with a motion blur effect FIGURE 4. An example of an image with noise



(a) original image (b) rotated horizontally (c) rotated vertically FIGURE 5. Mirror transformations caused by 3D-rotation

in images.

Applying motion blur is necessary to simulate blurry shots. In our case, noise with a small force was applied in order to make the algorithm resistant to small shifts. An example of an image with motion blur is shown in Figure 4b.

In object detection problems, the transformation specified by vertical and horizontal rotations (see Figure 5) allows the network to be trained on the same object, taking into account the rotation, thereby making it resistant to such changes.

A random projective transformation (perspective) is used for the same purpose as rotations. An example of applying a perspective is shown in Figure 6.



(a) original image (b) the result of transformation FIGURE 6. example of projective transformation

# 2. Quality Metrics in Problems of Object Detection in Medical Images

To assess the quality of detection, indicators of true positive and false positive results were used. With their help, the accuracy (*Precision*) and recall (*Recall*) were calculated using the following formulas:

(1) 
$$Precision = \frac{TP}{TP + FP},$$

(2) 
$$Recall = \frac{TP}{TP + FN},$$

where TP is the number of true-positive results, FP is the number of false-positive results, and FN is the number of false-negative results.

The *Precision* metric reflects the reliability of the model in classifying positive results, while *Recall* measures the ability of the model to detect samples that belong to the positive class. In addition, to analyze the quality of the algorithm, the *Precision* – *Recall* (*PR*) curve and the *AveragePrecision* (*AP*) metric, which is calculated as the area under the *Precision* – *Recall* curve, are used.

Additionally, the metrics used by the organizers of the NODE21 -Grand Challenge competition to evaluate the participant's decisions were calculated, namely, *TruePositiveRate*:

(3) 
$$TPR = TruePositiveRate = Recall = \frac{TP}{TP + FN}$$

and FalsePositiveRate:

(4) 
$$FPR = FalsePositiveRate = \frac{FP}{FP + TN},$$

where TN is the number of true negatives, Receiver Operating Characteristic (ROC) curve, which is the ratio of TPR to FPR, and Free Response Operating Characteristic (FROC) curve, which is the ratio of TPR to the average number of false positives.

The final target metric is calculated via the formula

(5) TargetMetric = 0.75AUC + 0.25FROC@0.25,

where AUC is the area under the ROC curve;  $FROC@\alpha$  is the value of FROC computed with a confidence threshold equal to  $\alpha$ .

### 3. Selecting a Neural Network Architecture

At present time, object detection problems are solved using a variety of neural network architectures, each of which has its own characteristics. Convolutional neural networks [6] successfully cope with a number of problems, while others require more complex architectures or even ensembles of several neural networks. To solve the problem, two actual NN architectures were chosen and their comparative analysis was carried out.

### 3.1. DETR

For the first NN, we consider the machine learning algorithm called End-to-End Object Detection with Transformers [7] or «DETR» for short. The DETR algorithm was introduced in 2020. Its release coincided with the beginning of the wide development of the NN [8] in speech processing problems. Although transformers were originally used in sequence processing, due to their excellent results, researchers have been looking for ways to apply transformers to other areas of machine learning. Thus, the DETR neural network has become a reference model for the use of transformers in the field of computer vision and, in particular, for object detection. To solve the problem of detecting neoplasms, the DETR-R50 network was used, which was previously trained on the COCO dataset [9].

In this network, the ResNet-50 architecture [10] acts as convolutional layers. The total number of trainable parameters is approximately 41 million.

Ten training epochs were conducted, which took about 10 hours in total in the Google Colaboratory cloud service with GPU connection. As a result,





FIGURE 7. Precision—Recall curve on the validation set



(a) without filtering by confidence



FIGURE 8. An example of network operation

the network showed the following results on the validation set: *Precision* = 0.09, *Recall* = 0.33, AP = 0.055; *PR* curve is shown in Figure 7.

To improve the predictive power of the network, low-confidence predictions can be discarded. The 50% confidence was taken as the threshold. Figure 8 shows examples of neoplasm detection in the lungs without and with filtering by confidence.



FIGURE 9. Precision—Recall curve on the validation set with 50% confidence threshold for predictions

After filtering by confidence, the network showed the following results on the validation set: Precision = 0.09, Recall = 0.36, AP = 0.059; PR curve is shown in Figure 9.

According to the results obtained on the validation set, this network shows low accuracy. The first possible reason for the unsatisfactory results is still insufficient data set. The second possible reason is the specificity of the problem itself. The presented architecture does not cope well with the detection of small objects, such as nodular neoplasms in the lungs.

# 3.2. YOLOv5

The  $YOLOv5^{\textcircled{m}}$  neural network is an improved fifth version of the  $YOLO^{\textcircled{m}}$  algorithm implemented in the PyTorch framework. The YOLO («You Only Look Once») algorithm has become one of the most popular architectures due to its good combination of detection accuracy and low resource consumption. Studies have shown that it is also effective for a number of medical tasks, in particular, when examining mammograms for suspicious lesions [11].

To solve the problem, we used the YOLOv5s6 network, previously trained on the COCO dataset. This neural network takes 1280x1280 pixel images as input and contains about 12.6 million trainable parameters.

For comparison with the DETR architecture, ten training epochs were also conducted, which took about 4 hours. The following results were obtained on the validation set: Precision = 0.65, Recall = 0.67, AP = 0.25; PR curve is shown in Figure 10.

Low-confidence predictions were also discarded to improve detection quality. With filtering by confidence with a threshold of 50%, the





FIGURE 10. Precision—Recall curve after 10 training epochs



FIGURE 11. Precision—Recall curve after 10 training epochs and discarding Low-confidence predictions

following results were achieved: Precision = 0.67, Recall = 0.81, AP = 0.31 and the PR curve is shown in Figure 11.

Additionally, an experiment was carried out on the same model, but already on 20 epochs. It showed the following results filtered by 50% confidence: Precision = 0.71, Recall = 0.81, AP = 0.31 and the PR curve is shown in Figure 12.

The YOLOv5 version allows to collect training data and metrics at each epoch. Let us take a closer look at the training process for 20 epochs, shown in Figure 13.

An analysis of the graphs of loss functions on the test and validation sets shows that the neural network starts to retrain after the 10th epoch. This



FIGURE 12. Precision—Recall curve after 20 training epochs and discarding Low-confidence predictions



FIGURE 13. Graphs of the loss function during training on 20 epochs

indicates a quantitative deficiency and still weak augmentation of the available data, which is why it is not yet possible to take into account the results obtained in this experiment.

# 3.3. Comparative analysis

Above, two architectures of neural networks were considered in relation to the problem of detecting nodular neoplasms on computer tomography. Under the same conditions, two algorithms DETR and YOLOv5 were used in order to compare their effectiveness and the possibility of further use for solving the problems under consideration.

For clarity, the values of the target metrics for the considered neural

Architecture	Precision	Recall	AP
DETR YOLOv5	$\begin{array}{c} 0.09 \\ 0.67 \end{array}$	$\begin{array}{c} 0.36\\ 0.81 \end{array}$	$\begin{array}{c} 0.059 \\ 0.31 \end{array}$

TABLE 2. Results of neural networks on the validation set after 10 training epochs

networks are given in table 2. From the obtained results, it follows that the YOLOv5 neural network demonstrates higher accuracy in comparison with the DETR architecture. First of all, this is due to the difference in approaches to the detection problem and the specificity of the data on which neural networks are trained. The DETR neural network uses a «transformer» type architecture, which, as already mentioned, is very demanding on the amount of data. Because of this, the YOLO architecture solves this problem much better.

Based on the results obtained, the YOLOv5 architecture was chosen for further study. However, the current solution using this neural network and dataset is not satisfactory as the neural network tends to get overfitted quickly. Therefore, it is necessary to solve the problem of overfitting in order to increase the generalizing ability and the quality of detection of the chosen model on the available data.

### 3.4. Solving the Problem of Overfitting

One way to solve the problem of overfitting is to increase the training sample. This can be achieved either by an additional data set or by searching for similar images; however, due to the specificity of the problem under consideration, such an option to increase the current dataset is not suitable.

Another approach to solving this problem is to use more advanced algorithms for augmentation and additional image filtering. To enhance the contrast and normalize the data, the CLAHE filter [12] was used, which is a variant of adaptive histogram equalization in which contrast enhancement is limited to avoid introducing or amplifying noise in the image. Further, additional augmentation algorithms were used. The YOLOv5 model allows researchers to augment images while loading data. This method allows to generate new images every time, increasing the generalizing ability of the model, while there is no need to store a lot of data on disk, and all operations occur without much loss of performance. In addition to the augmentation algorithms already considered (vertical and horizontal reflections—flipud, fliplr, translate, scale), the following have been added:

- $\bullet$  hsv\_h-HSV-Hue augmentation,
- hsv s-HSV-Saturation augmentation,
- hsv v-HSV-Value augmentation,

#### NODULES DETECTION



(a) the original image



(c) HSV-Saturation



(b) HSV-Hue





FIGURE 14. The shifts in the corresponding space of the HSV color model

- mosaic mosaic augmentation,
- mixup mixup augmentation,
- $\bullet$  copy\_paste overlay random segments between images.

To illustrate the operation of these algorithms, let us take the original image Figure 14a, for which the CLAHE filter has already been applied. The HSV-Hue, HSV-Saturation, HSV-Value algorithms perform shifts in the corresponding space of the HSV color model; examples are shown in Figure 14.



(a) mosaic



FIGURE 15. An example of operation of the mixup algorithm

The mosaic algorithm performs random movement of image elements (see Figure 15a). The mixup and copy\_paste algorithms overlay elements between images, but mixup also performs random transformations of the mixed elements; an example of how this algorithm works is shown in Figure 15.

As a result of the experiments, the following augmentation parameter values were chosen, which are specified in the configuration file:

• hsv_h: 0.015,	• scale: 0.9,	• mosaic: 1.0,
• hsv_s: 0.7,	• flipud: 0.5,	• mixup: 0.1,
• hsv_v: 0.4,	• fliplr: 0.5,	• copy_paste: 0.1.

• translate: 0.1,

## 4. Results

In the conducted studies, the YOLOv5s6 neural network described above was taken as a training model. The algorithms and augmentation parameters were chosen in such a way that the model was not prone to rapid retraining.

The training was carried out on 300 epochs. The training process is shown in Figure 16. As can be seen from the data obtained, the overtraining problem was solved with the help of augmentation.

#### NODULES DETECTION



FIGURE 16. Graphs of loss functions and accuracy metrics of the final model



FIGURE 17. Examples of operation of the trained neural network YOLOv5

The results of the network operation are shown in Figure 17. When filtering by confidence with a threshold of 0.5 on the validation set, the model demonstrates the following indicators: Precision = 0.70, Recall = 0.90, AP = 0.38;

Figure 18 shows the corresponding PR curve.

In addition, let us calculate the metrics used to evaluate the quality of decisions made by NODE21-Grand Challenge participants. Metric AUC

DINARA H. GINIYATOVA, VASILII A. LAPINSKII



FIGURE 18. Precision—Recall curve on the validation set after 300 training epochs



FIGURE 19. ROC curve on the validation set

= 0.898, FROC@0.25 = 0.741, TargetMetric = 0.859; the ROC curve is shown in Figure 19.

Let us clarify that the value of the target metric for the leaders of the competition lies in the range from 0.801 to 0.839  $NODE21^{(R)}$ .

# Conclusion

Approaches to solving the problem of detecting nodular neoplasms in the human chest area on computer tomograms using neural networks were considered. The NODE21 dataset was used as a data sample. A comparative analysis of two current architectures, DETR and YOLO, was carried out, which showed that an insufficient dataset and its specificity affect the quality of detection and the tendency of the model to get overfitted.

The YOLO algorithm showed the best results and was chosen as the

main one for solving the problem. The overfitting problem was solved with stronger augmentation algorithms and additional image filtering.

The obtained values of the target metrics indicate the possibility of using neural networks to solve the problem of neoplasm detection and other similar problems of medical image analysis.

# References

- L. Wei, K. Ding, H. Hu. "Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network", *IEEE Access*, 8 (2020), pp. 99633–99647.
- [2] K. L. Kehl, H. Elmarakeby, M. Nishino, Van Allen E. M., E. M. Lepisto, M. J. Hassett, B. J. Johnson, D. Schrag. "Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports", JAMA Oncol., 5:10 (2019), pp. 1421–1429. dcl↑s2
- [3] A. Rodriguez-Ruiz, K. Lang, A. Gubern-Merida, M. Broeders, G. Gennaro, P. Clauser, Th. H. Helbich, M. Chevalier, T. Tan, Th. Mertelmeier, M. G. Wallis, I. Andersson, S. Zackrisson, R. M. Mann, I. Sechopoulos. "Stand-Alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists", *Journal* of the National Cancer Institute, **111**:9, pp. 916–922. €2↑82
- [4] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, Summers R. M.. "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2097–2106. 60%<sup>↑</sup>83
- [6] D. I. Kaliyev, O. Ya. Shvets. "Convolutional neural networks for solving fire detection problems based on aerial photography", *Program Systems: Theory and Applications*, 13:1 (2022), pp. 195–213 (in Russian). Computer Statement (In Russian).
- [7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko.

DE: TR: End-to-end object detection with transformers, GitHub. URL 187

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin. "Attention is all you need", Advances in Neural Information Processing Systems 30, 31st Conference on Neural Information Processing Systems (NIPS 2017) (Long Beach, CA, USA), eds. I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, 2017, ISBN 9781510860964, pp. 6000–6010.

- [9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick. "Microsoft COCO: common objects in context", Computer Vision – ECCV 2014, Lecture Notes in Computer Science, vol. 8693, eds. Fleet D., Pajdla T., Schiele B., Tuytelaars T., Springer, Cham, 2014, ISBN 978-3-319-10601-4, pp. 740–755. Construction of the statement of
- [10] K. He, X. Zhang, Sh. Ren, J. Sun. Deep residual learning for image recognition, 2015, 12 pp. arXiv: 1512.03385 [cs.CV] 187
- [11] A. Kolchev, D. Pasynkov, I. Egoshin, I. Kliouchkin, O. Pasynkova, D. Tumakov. "YOLOv4-based CNN model versus nested contours algorithm in the suspicious lesion detection on the mammography image: A direct comparison in the real clinical settings", *Journal of Imaging*, 8:4 (2022), 88. (2)<sup>489</sup>
- [12] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, ter Haar Romeny B., J. B. Zimmerman, K. Zuiderveld. "Adaptive histogram equalization and its variations", *Computer Vision*, *Graphics and Image Processing*, **39**:3 (1987), pp. 355–368. (1)<sup>92</sup>

Received	17.06.2022;
approved after reviewing	12.08.2022;
accepted for publication	10.09.2022.

Recommended by

prof. A. M. Elizarov

# Information about the authors:



### Dinara Halilovna Giniyatova

Senior Lecturer of the Department of Applied Mathematics and Artificial Intelligence of the ICM&IT, KFU. Research interests include high-performance computing, machine learning, electrodynamics, mathematical modeling, programming

0000-0002-0853-0984
e-mail: normaliti@gmail.com

### Vasilii Aleksandrovich Lapinskii

4th year B.S. Student of ICM&IT, KFU. Research interests include machine learning, image processing, high performance computing

(D) 0000-0003-2880-9218 e-mail: vasilylapinsky@gmail.com

The authors contributed equally to this article. The authors declare no conflicts of interests.