METHODS FOR INCREASING THE CLASSIFICATION ACCURACY BASED ON MODIFICATIONS OF THE BASIC ARCHITECTURE OF CONVOLUTIONAL NEURAL NETWORKS

Svitlana Shapovalova[1], Yurii Moskalenko[2]

[1]Department of Automation of Design of Energy Processes and Systems, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine
lanashape@gmail.com
ORCID: http://orcid.org/0000-0002-3431-5639

[2]Department of Automation of Design of Energy Processes and Systems, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine
yuramuv@gmail.com
ORCID: http://orcid.org/0000-0002-0824-9201

Shapovalova, S., Moskalenko, Y. (2020). Methods for increasing the classification accuracy based on modifications of the basic architecture of convolutional neural networks. ScienceRise, 6 (71), 10–16. doi: http://doi.org/10.21303/2313-8416.2020.001550

ARTICLE INFO

Article history:
Received date 26.11.2020
Accepted date 24.12.2020
Published date 30.12.2020

Section:
Information Technology

DOI

10.21303/2313-8416.2020.001550

KEYWORDS

classification
convolutional neural network
ResNet
EfficientNet
deep neural network
deep learning

ABSTRACT

Object of research: basic architectures of deep learning neural networks.

Investigated problem: insufficient accuracy of solving the classification problem based on the basic architectures of deep learning neural networks. An increase in accuracy requires a significant complication of the architecture, which, in turn, leads to an increase in the required computing resources, as well as the consumption of video memory and the cost of learning/output time. Therefore, the problem arises of determining such methods for modifying basic architectures that improve the classification accuracy and require insignificant additional computing resources.

Main scientific results: based on the analysis of existing methods for improving the classification accuracy on the convolutional networks of basic architectures, it is determined what is most effective: scaling the ScanNet architecture, learning the ensemble of TreeNet models, integrating several CBNet backbone networks. For computational experiments, these modifications of the basic architectures are implemented, as well as their combinations: ScanNet+TreeNet, ScanNet+CBNet.

The effectiveness of these methods in comparison with basic architectures has been proven when solving the problem of recognizing malignant tumors with diagnostic images – SIIM-ISIC Melanoma Classification, the train/test set of which is presented on the Kaggle platform. The accuracy value for the area under the ROC curve metric has increased from 0.94489 (basic architecture network) to 0.96317 (network with ScanNet+CBNet modifications). At the same time, the output compared to the basic architecture (EfficientNet-b5) increased from 440 to 490 seconds, and the consumption of video memory increased from 8 to 9.2 gigabytes, which is acceptable.

Innovative technological product: methods for achieving high recognition accuracy from a diagnostic signal based on deep learning neural networks of basic architectures.

Scope of application of the innovative technological product: automatic diagnostics systems in the following areas: medicine, seismology, astronomy (classification by images) onboard control systems and systems for monitoring transport and vehicle flows or visitors (recognition of scenes with camera frames).

1. Introduction

1. 1. The object of research

The object of research is the basic architectures of deep learning neural networks.

1. 2. Problem statement

The solution of classification problems is a demanded task in control, monitoring and diagnostics systems. Such a task often arises in the analysis of diagnostic images, for example, in the field of medicine, seismology, cosmology and other fields, for example, diagnostics using MRI images, determining mineral deposits using seismic images, determining the types of objects using

telescope images, and others. Obviously, solving the problem requires high accuracy. However, achieving the required accuracy, as a rule, requires large computing resources: namely, the consumption of video memory and the cost of learning/output time, which are not always available to the user. Therefore, the development of methods for improving the classification accuracy on deep learning networks in conditions of limited computer resources is an urgent task and has practical significance.

The main modern approach to image analysis is the use of convolutional neural networks (CNNs). For such tasks, the basic architectures of neural networks have been developed, which have actually become standard. Modern research is aimed at modifying such architectures to improve accuracy and/or reduce the computational resources required to train and output a network.

When developing basic architectures, the following are used:

1. Skip-connection to solve the problem of vanishing gradient;

2. Global pooling to significantly reduce the size of the global layer and the invariance of the network to the dimension of the input image;

3. Mechanisms of attention to improve the accuracy of recognition.

With these tools, learning problems are solved in the underlying architectures.

The following basic architectures are most often used for image analysis:

1) ResNet-like, in particular ResNet [1], ResNext [2], SeResNet [3].

A feature of these architectures is the use of skip connections, which after the development [1] became the standard for all networks.

2) DenseNet [4].

A feature of the network is a new scheme for using skip connections.

3) EfficientNet [5].

The first two types of architecture made it possible to scale the architecture in depth, that is, by increasing the number of layers. EfficientNet allows to scale the architecture not only in depth, but also in width, that is, by increasing the number of convolution channels.

However, when solving current applied problems, neural networks of the basic architecture can't always provide the required accuracy. The following methods are used to improve accuracy:

1) Cross-validation of the training dataset into N parts, followed by learning N neural networks of the same architecture and averaging the prediction result.

2) The use of an ensemble of neural network models consists in the simultaneous use of basic architectures of various types, such as ResNet, EfficientNet, as well as their modifications.

Using cross-validation and ensemble of models is computationally intensive as it is necessary to train and make predictions on multiple networks.

Therefore, the problem arises of determining such methods for modifying basic architectures that improve the classification accuracy and require insignificant additional computing resources.

1. 3. Approach to problem solution

The main approach to solving this problem can be methods for modifying basic architectures:

– based on the method of scaling architecture SCAN (A Scalable Neural Networks Framework), proposed in [6];

– based on the ensemble of TreeNet models proposed in [7];

– a method for integrating several CBNet backbone networks, proposed in [8].

Each of these methods has been developed to improve the accuracy of the associated classification or segmentation application. Evidence of the effectiveness of these modifications can be carried out experimentally and tested on test problems. However, no publications have yet been presented on the study of a comparative analysis of these methods from the point of view of the efficiency of increasing accuracy on the same problems. Therefore, it is necessary to prioritize the use of both each of the methods independently, and their combinations.

The aim of research is to determine such methods of modifications of the basic architectures of the convolutional neural network, as well as their combinations, which provide the maximum improvement in the classification accuracy.

2. Methods for modifying architectures

2. 1. Method of scaling architecture

An element of the basic architecture of a neural network is a block – a group of layers that process a signal of the same dimension in width and height. Usually a block contains the following layers: convolution, activation, addition layer, multiplication layer. In what follows, such blocks will be denoted by Bi, where i is the depth of the block placement relative to the input layer. The final layer is the FC (fully connected) layer.

The work [6] presents a method for modifying SCAN (A Scalable Neural Networks Framework), which proposes:

1) Using additional outputs (classifier layers) in the neural network after each of the blocks (**Fig. 1**).

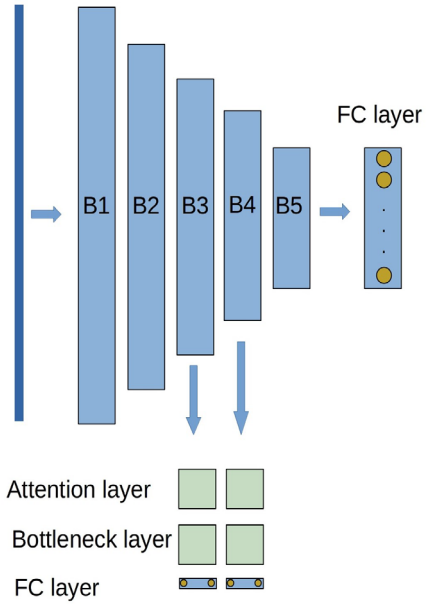

Fig. 1. SCAN network diagram

To ensure that the fully connected layer is not directly connected to the output of the block, the output signal passes through an additional bottleneck layer and an attention layer. Thus, the network will have several outputs.

Based on the computational experiments carried out in this study, it was found that the best results are obtained by the ScalaNet network, which has 3 outputs: standard FC and 2 additional.

2) Modification of the loss function for auxiliary classifiers is made in such a way that, in addition to the cross-entropy function, the sum of several functions is used:

$$loss = \sum_{1}^{n} \left((1-\alpha) * \text{CrossEntropy} y + \alpha * KL + \lambda * DIFF \right), \quad (1)$$

where C – the number of classifiers, α, λ – coefficients that determine the influence of each component of the sum, CrossEntropy – cross-entropy function between a real label and a predictable one, KL – Kullback–Leibler distance between the current C-th classifier and the final one, $Diff$ – difference between the output of the current classifier and the final one.

In this study, computational experiments were carried out to determine the optimal loss function. On their basis, the cross-entropy function was chosen.

3) Establishment of thresholds for additional classifiers for calculating the final result.

In this study, computational experiments were carried out to determine the best method for integrating the results obtained at all outputs. The arithmetic mean was chosen as the final result.

The basic SCAN architecture requires additional computational resources through the use of the bottleneck layer. Computational experiments carried out in this study have shown

that when this layer is abandoned, the computational costs compared to the core network are almost unchanged. In general, using the bottleneck layer results in a slight increase in accuracy.

2. 2. A way of learning an ensemble of TreeNet models

In [7], it is proposed to use an ensemble of TreeNet models, in which some of the layers will be common (**Fig. 2**). The complexity of such an architecture relative to the basic one depends on the number of joint blocks and is intermediate between the autonomous network and the ensemble of models.

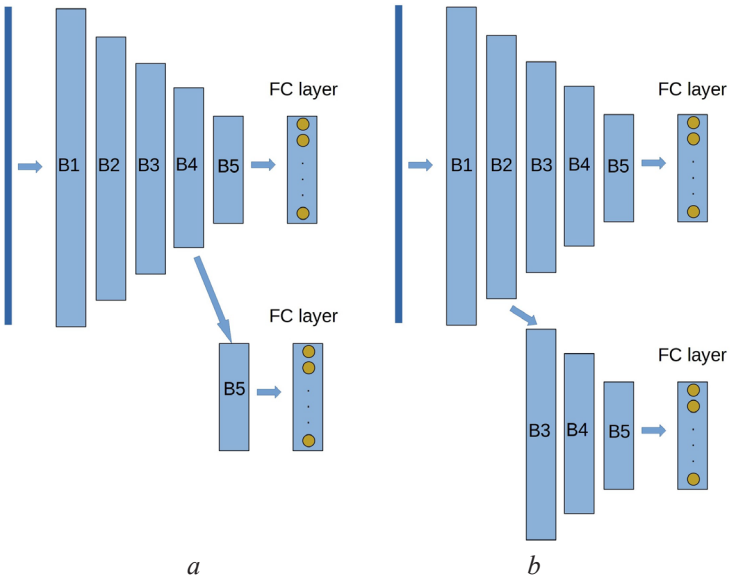

Fig. 2. Versions of the TreeNet network: *a* – with one independent block; *b* – with three independent blocks

By choosing a place for connecting additional outputs closer to the network input, it is possible, on the one hand, to increase the classification accuracy, but, on the other hand, to significantly increase the required computing resources.

2. 3. Method of integrating multiple backbone networks

The work [8] proposed a type of network designed for segmentation, which has several backbone-networks (standard networks) CBNet. Each such *j*-th network consists of *i*-th blocks. The output of each block b_i^j is directed not only to the next block b_{i+1}^j, but also to the input of the *i*-th block of the (*j*+1)-th network b_i^{j+1} after the corresponding reduction of the dimension (**Fig. 3**).

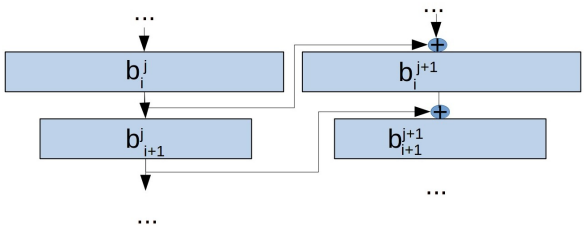

Fig. 3. Fragment of the CBNet network

Since such a modification is very demanding on computational resources, a simplified version was proposed in the study [5], in which not all blocks of the *i*-th backbone are used, but only a few last blocks. In this work, only the parallel connection of the last unit is tested.

3. Computational experiments

3. 1. Applied neural networks

EfficientNet-b5 is chosen as the basic architecture for carrying out computational experiments. This choice is due to the fact that the efficientNet-b5 network does not require a large amount of resources and provides high accuracy results. The following modifications of this architecture were used:

1. ScanNet – modification of the network according to method 1 with two additional outputs of classifiers according to method 1 (**Fig. 1**).

2. TreeNet – modification of the network according to method 2, in which the last block is duplicated twice (**Fig. 2**).

3. ScanNet+TreeNet – modification according to methods 1 and 2.

4. CBNet – modification of the network, in which the last block is duplicated twice according to methods 3 (**Fig. 3**).

5. ScanNet+CBNet– modification according to methods 1 and 4.

The sixth network that took part in the experiments is the basic efficientNet-b5 as a reference.

3. 2. Test problem

Computational experiments are carried out on the SIIM-ISIC Melanoma Classification dataset provided on the Kaggle platform [9]. The images show pictures of birthmarks. The original label is {0, 1}, where 1– the birthmark is melanoma.

The dataset is provided by ISIC (The International Skin Imaging Collaboration). The training part of the data contains 33126 images, testing – 10982. In addition, a collection of images from competitions of previous years is provided on the platform [10]. Thus, the size of the training dataset is 60487 images.

In addition to the images, the patient's metadata are also available: gender, approximate age, location of the birthmark, and others. Metadata is not used in this study.

In the dataset examples, there is a significant data imbalance: among the 60487 training data, there are 55008 labeled 0 (melanoma) and 5479 labeled 1 (melanoma).

The area under the ROC-curve is specified as the metric of the accuracy of the competition.

The size of the images in the examples is 6000×4000 pixels. Before being fed to the neural network, the image size was reduced to 384*384. During learning, balancing was used: the same number of examples of both classes was fed to the network. Also, during learning, augmentation was used, which consisted in random distortion of the input image before feeding it to the neural networks. Among the methods of augmentation were used distortions, rotations, reflections, random changes in brightness/contrast, and others. In the final prediction, TTA (test time augmentation) was used, which consisted in anticipating the label not only of the input image, but also of its augmentations from the D4 symmetry group. That is, the neural network provided 8 marks for one image, and the result was averaged as an arithmetic mean.

3. 3. Experiment parameters

To determine the optimal loss function, a series of computational experiments was carried out, based on the results of which, the following loss function was determined:

$$loss = \text{BinaryCrossEntropy} + 0.15 * \text{FocalLoss}(\text{gamma}=2), \quad (2)$$

where BinaryCrossEntropy is the result of binary cross-enropy between the provided labels and the real ones; FocalLoss is a loss function [11] for examples with a larger error (the coefficient of the gamma degree is chosen 2).

To increase the accuracy, the training sample was divided into 5 parts using cross-validation.

The learning was carried out on an Nvidia GTX 1080ti video card, Batch size 8.

4. Results

The proposed network modifications took part in the Kaggle competition. The value of the metric for evaluating the results of the network is shown in **Table 1**. Kaggle public score – the public score displayed on the Kaggle platform immediately after loading the pre-

diction result, Kaggle private score – the final result, which is announced after the end of the competition.

Table 1

The accuracy of the SIIM-ISIC Melanoma Classification problem solution

#	EfficientNet-b5 network modification type	Area under the ROC curve		
		Local Cross-validation Score	Kaggle public Score	Kaggle private score
1	–	0.94489	0.9109	0.9218
2	ScanNet	0.95635	0.9141	0.9286
3	TreeNet	0.94717	0.9123	0.9214
4	ScanNet+TreeNet	0.95975	0.9140	0.9303
5	CBNet	0.94673	0.9115	0.9216
6	ScanNet+CBNet	0.96317	0.9133	0.9337

For the competition, the result was presented, which consisted of the predictions of each of the networks.

According to the results of the competition, the ensemble of models took 65th place among 3314, which allowed it to enter the top 2 % of participants and receive a virtual silver medal.

5. Discussion of research results

The final result of the prediction for each of the networks is a generalization of the solutions of the additional and final outputs of the classifiers.

According to the research results, the total end result for all networks was more accurate than each of the independent outputs. For example, for the ScalaNet network the final result was 0.95635, and each of the i-th outputs was respectively: 0 (the last final output) – 0.95602; 1 – 0.95106; 2 – 0.90868.

The best accuracy results were obtained using a combination of ScanNet and CBNet methods.

The disadvantages of using the combination of ScanNet+CBNet methods include a significant increase in video memory consumption when analyzing large-scale images.

Thus, the limitation of the use of this combination of methods is associated with the limitations of the user's computing resources, primarily the amount of video memory.

The prospect for further research is the analysis and improvement of neural networks, which contain two branches. The first branch has a classic architecture and receives a thumbnail image as input. The second one has a simplified architecture, but receives a larger source image as input. The end result is a union of the results from both branches.

6. Conclusions

1. Based on the analysis of the existing methods for improving the classification accuracy on the starter networks of basic architectures, it has been determined that the following methods are the most effective: architecture scaling (ScanNet), learning an ensemble of TreeNet models, integration of several backbone networks (CBNet).

2. Computational experiments implemented the following modifications of the basic architectures of the SGort networks: ScanNet, TreeNet, CBNet and their combinations: ScanNet+TreeNet, ScanNet+CBNet.

3. On the basis of the computational experiments, the effectiveness of these methods has been proved in comparison with the basic architectures: according to the area under the ROC curve metric, the accuracy value has increased from 0.94489 (network without modifications) to 0.96317 (network with ScanNet+CBNet modifications).

References

[1] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778. doi: http://doi.org/10.1109/cvpr.2016.90

[2] Xie, S., Girshick, R., Dollar, P., Tu, Z., He, K. (2017) Aggregated residual transformations for deep neural networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1492–1500. doi: http://doi.org/10.1109/cvpr.2017.634

[3] Hu, J., Shen, L., Sun, G. (2018) Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7132–7141. doi: http://doi.org/10.1109/cvpr.2018.00745
[4] Huang, G., Liu, Z., Weinberger, K., Maaten, L. (2017). Densely connected convolutional networks. CVPR, 2261–2269. doi: http://doi.org/10.1109/cvpr.2017.243
[5] Tan, M., Le, Q. (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning, 6105–6114.
[6] Zhang, L., Tan, Z., Song, J., Chen, J., Bao, C., Ma, K. (2019). SCAN: A Scalable Neural Networks Framework Towards Compact and Efficient Models. Advances in Neural Information Processing Systems (NeurIPS), 4029–4038.
[7] Lee, S., Purushwalkam, S., Cogswell, M., Crandall, D., Batra, D. (2015). Why M heads are better than one: Training a diverse ensemble of deep networks. Available at: https://arxiv.org/pdf/1511.06314
[8] Liu, Y., Wang, Y., Wang, S., Liang, T., Zhao, Q., Tang, Z., Ling, H. (2020). CBNet: A Novel Composite Backbone Network Architecture for Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 34 (7), 11653–11660. doi: http://doi.org/10.1609/aaai.v34i07.6834
[9] SIIM-ISIC Melanoma Classification. Kaggle. Available at: https://www.kaggle.com/c/siim-isic-melanoma-classification
[10] Merge External Data. Kaggle. Available at: https://www.kaggle.com/shonenkov/merge-external-data
[11] Lin, T., Goyal, P., Girshick, R., He, K., Dollar, P. (2017) Focal Loss for Dense Object Detection. IEEE International Conference on Computer Vision, 2980–2988. doi: http://doi.org/10.1109/iccv.2017.324