

ABSTRACT

Tymchenko B. I. Neural Network Methods for Planar Image Analysis in Automated Screening Systems. — Qualification scientific work in the form of manuscript.

Thesis for doctor of philosophy degree in speciality 122 — Computer science. — ICS ONPU, Odesa, 2021.

In **introduction**, the relevance of improving automated screening systems (screening is defined as detection and classification of atypical objects or processes) by creating models of datasets and methods for their generation, as well as models of neural networks and multi-tasking learning methods to increase the accuracy of classification and segmentation of planar images without increasing time. The object, subject, tasks and methods of research are defined; the connection with research works is shown; the scientific novelty and practical significance of the obtained results are given; the personal contribution of the applicant is covered.

In the **first section** of the dissertation the analysis of problems of analysis of planar images (which do not have depth measurement, and in which it is possible to discard scale of objects) in automated screening systems on the basis of neural network technologies is carried out.

The existing systems of automated screening in the subject areas of medicine, meteorology and remote sensing and features of their functioning by analyzing planar images are analyzed. It is shown that the effectiveness of such systems directly depends on the accuracy of image classification and segmentation, as false-positive results can lead to premature response, and false-negative to the omission of existing problems and late response. In addition, no less impor-

tant factor influencing the effectiveness of screening systems is the time spent on the implementation of the educational process, which, in turn, is associated with the quality of pre-created samples of training data. However, in the tasks of verifying the receipt of sufficient data and conducting high-quality annotations by professionals is usually expensive, and sometimes impossible due to the small number and variety of training examples, subjectivity in the formation of annotations in data sets. These factors negatively affect the accuracy of classification and segmentation in the training of deep neural networks by standard methods. To eliminate such influence, it is proposed to solve the problem taking into account partially erroneous annotations when constructing models of classification and segmentation.

It is shown that in such conditions it is expedient to use deep neural networks, due to the possibility of automating the learning process and the use of multi-tasking learning that increase the accuracy of classification and segmentation.

The analysis of the existing models of datasets and methods of their generation showed the limitations of these models, which would allow to form partially erroneous annotations, which have characteristic of real automated screening tasks.

Thus, solving an important scientific and practical problem of increasing the accuracy of classification and segmentation in the analysis of planar images for automated screening systems by improving models of neural networks and methods of their multitasking, especially in terms of partially erroneously annotated data samples.

In the **second section** method for generating datasets to enable testing of neural networks in learning using partially erroneous annotations and proposed a parametric formalization of the data set model with partially erroneous annotations, which are characteristic of real automated screening tasks were developed.

Because obtaining reliable classification and segmentation annotations for test samples is often impossible or very costly, it is proposed to use model datasets to assess the reliability of models and methods of learning deep neural networks. Such datasets make it possible to conduct training and testing in controlled conditions by artificially adding errors to the annotations of the training data sample (test samples remains with precise annotations).

Proposed model of datasets \mathcal{M} has the following definition:

$$\mathcal{M} \in \{\mathcal{X}_b, \mathcal{X}_{tex}, N, S_{img}, S_{obj}, \Delta_{max}, N_{obj}, P_e, P_d, S_e, S_d\} \quad (1)$$

where \mathcal{X}_b - background images set, \mathcal{X}_f - objects set, \mathcal{X}_{tex} - object texturing set, N - number of images in the generated dataset, S_{img} - pixel size of generated images, S_{obj} - average size of the objects (pixels), Δ_{max} - maximum deviation of objects size, N_{obj} - maximum number of objects in the image, P_e and P_d - probabilities of mask erosion and dilation for every object, S_e and S_d - kernel sizes for erosion and dilation.

The last four parameters are introduced for controlled annotation errors.

Formulated **first point of scientific novelty**: for the first time a parametric formalization of the dataset model with partially erroneous annotations, which are characteristic of real problems of automated screening, was proposed, which allowed the development of a method of generating training, test and validation data sets.

Based on the proposed parametric model (??), a method for generating data sets with partly-erroneous annotations has been developed, which contains the following steps:

1. Choose random background: $x_{bg} \sim \mathcal{X}_b$
2. Choose number of objects in the image: $N_{obj} \sim \mathcal{U}(1, N_{obj})$
3. Initialize the segmentation mask: $M = 0$ so $M \in \mathcal{R}^{C \times S_{img} \times S_{img}}$
4. For n_{obj} times:
 - 4.1 Sample object size: $s \sim \mathcal{U}(S_{obj} - \Delta_{max}, S_{obj} + \Delta_{max})$

4.2 Sample object coordinates:

$$i_f \sim \mathcal{U}(0, S_{img} - s)$$

$$j_f \sim \mathcal{U}(0, S_{img} - s)$$

4.3 Sample object image $x_{fg} \sim \mathcal{X}_f$ and the corresponding object class $c_{fg} \sim \mathcal{Y}_f$

4.4 Resample the image with bilinear interpolation:

$$\hat{x}_{fg} = R_{bilinear}(x_{fg})$$

4.5 Sample the texture image $x_{tex} \sim \mathcal{X}_{tex}$

4.6 Modify object image with the object texture:

$$\hat{x}_{fg} = x_{fg} \circ x_{tex}[i_f : i_f + s, j_f : j_f + s]$$

4.7 Place object image into the background image:

$$x_{bg}[i_f : i_f + s, j_f : j_f + s] = (1 - x_{fg}) \circ x_{bg} + \hat{x}_{fg}$$

4.8 Form the segmentation mask:

$$M_{seg} = x_{fg} > \theta_{seg}$$

where θ_{seg} - is the source object binarization threshold. For the MNIST dataset $\theta_{seg} = 0.2$, and for FashionMNIST $\theta_{seg} = 0.1$.

4.9 Modify the segmentation mask according to the desired error level:

$$M_{seg} = \begin{cases} M_{seg} \oplus K^{S_d \times S_d} & \text{iff } p_d \sim \mathcal{U}(0, 1) < P_d \\ M_{seg} \ominus K^{S_e \times S_e} & \text{iff } p_e \sim \mathcal{U}(0, 1) < P_e \end{cases}$$

where $K^{S_d \times S_d}$ - dilation kernel matrix and $K^{S_e \times S_e}$ - erosion kernel matrix.

4.10 Place modified object mask into the common mask:

$$M[c_{fg}, i_f : i_f + s, j_f : j_f + s] = \max \{M[c_{fg}, i_f : i_f + s, j_f : j_f + s], M_{seg}\}$$

4.11 Save image x_{bg} and mask M

5. Finish generation

Thus, the method of generating datasets with partially erroneous annotations based on the parametric model was improved, which due to the generation of annotations: partially erroneous for the training sample and accurate - for the test, made it possible to test the impact of annotation errors on neural networks.

The method allows to obtain many data sets with similar characteristics and to use non-parametric statistical methods (such as bootstrapping) to evaluate models in the absence of real training data. The method is **the second point of scientific novelty**.

In the **third section** of the dissertation neural network models and methods of multi-tasking learning were developed to simultaneous increase in accuracy of classification and segmentation without decrease in efficiency.

To implement multi-task learning methods, it is proposed to improve the models of deep neural networks using the encoder-decoder architectures (UNet, LinkNet) by introducing an additional decoder with the normalization layer. Thus, the advanced model consists of an encoder and two decoders (for segmentation and classification tasks, respectively).

The model of the deep neural network is represented by the following expression:

$$v_1, v_2 \dots v_n = F_{encoder}(x, \theta_{enc}) \quad (2)$$

$$M_{seg} = F_{seg}((v_1, v_2 \dots v_n), \theta_{seg}) \quad (3)$$

$$C_{cls} = F_{cls}((v_n), \theta_{cls}) \quad (4)$$

where θ_{enc} - are encoder parameters, θ_{seg} and θ_{cls} - parameters of encoder and decoder, respectively, F_{seg} and F_{cls} - decoder neural networks.

The proposed model is **the third point of scientific novelty**.

With the introduction of an additional classification decoder with a normalization layer, it is possible to implement methods of multi-tasking learning and inference for deep neural networks.

During the practical use of neural network methods in screening tasks, it is established that partially erroneous annotations of training data prevent obtaining high accuracy of classification and segmentation.

The method of multi-task learning in the conditions of partially-erroneous annotations of training data, which is based on use of tasks that are connected with original task is offered. It is shown that for the segmentation problem there is a similar classification problem, for which the annotations of training data are of better quality than for the original segmentation problem.

The proposed method of multi-task learning consists of two stages.

Stage 1. Generating the derivative task. In the context of multi-task learning, the task of classification is defined as a multiple instance learning task, instead of marking each of the objects for all classes. In this case, the annotation of the original planar image is a set with one or more objects.

The loss function is calculated separately for each of the tasks (classification and segmentation). The top-limited loss function L_{seg}^{\top} is used to train the segmentation decoder, while the classification decoder uses the normal L_{cls} loss function, and the total value of the loss function is defined as the arithmetic mean between the individual values:

$$L_{total} = \frac{L_{seg}^{\top} + L_{cls}}{2} \quad (5)$$

Accordingly, the total gradient of the loss function will also be the arithmetic mean of the gradients of the components ∇L_{seg}^{\top} and ∇L_{cls} :

$$\nabla L_{total} = \frac{\nabla L_{seg}^{\top} + \nabla L_{cls}}{2} \quad (6)$$

Thus, the presence of non-zero gradients is provided to update the parameters from at least one loss function for each input example.

Stage 2. Loss function bounding. To reduce the impact of the erroneous part of the annotations, it is proposed to introduce a constraint to the loss function for the segmentation problem with lower quality data annotations. That allowed to reduce the influence of gradients of the loss function on examples with erroneous annotations when learning on several tasks:

$$\mathcal{L}] = \min(L, \theta)$$

where θ - loss function threshold.

Thus, the introduced boundaries are represented by the following expression:

$$\nabla \min(L, \theta) = \begin{cases} 1, & \text{iff } L \in (-\infty, \theta] \\ 0, & \text{iff } L \in (\theta, \infty) \end{cases}$$

In the absence of markup for examples in segmentation problems, iterative refinement of this markup is possible. To do this, a refined map of classification features is calculated by the Hadamard product between the output of the segmentation decoder normalized by the sigmoid function and the classification logs:

$$M_{unsup} = \hat{M}_{seg} \circ C_{cls} \quad (7)$$

Next, to obtain the classification result, the summation of the elements M_{unsup} is performed with normalization by the sum of the elements of the original non-normalized localization map:

$$C_{unsup} = \frac{\sum_{h=0}^H \sum_{w=0}^W M_{unsup}(h,w)}{\sum_{h=0}^H \sum_{w=0}^W M_{seg}(h,w) + c} \quad (8)$$

Based on the proposed model of the neural network and the method of its training, a method of multi-taski prediction has been developed. This

allowed to increase the accuracy of classification and segmentation of planar images without increasing time needed.

Method consists of the following stages:

Stage 1. Normalization. Let $C_{cls} \in \mathcal{R}^C$ and $M_{seg} \in \mathcal{R}^{C \times H \times W}$ - logits of classification and segmentation decoders, respectively and lie in the interval $(-\infty, +\infty)$.

To obtain normalized to the interval $[0, 1]$, sigmoid activation function is used:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Stage 2. Renormalization. Renormalization is the weighing of a segmentation map using normalized classifier logits. The first step is the transformation of segmentation and classification logs into uncalibrated estimates between $[0, 1]$:

$$\begin{aligned}\hat{M}_{seg} &= \sigma(M_{seg}) \\ \hat{C}_{cls} &= \sigma(C_{cls})\end{aligned}$$

These estimates have the same dimensions as original segmentation mask and classification vector, additional dimensions are added for simplicity: $\hat{M}_{seg} \in \mathcal{R}^{C \times H \times W}$ and $\hat{C}_{cls} \in \mathcal{R}^{C \times 1 \times 1}$

Segmentation map is weighted using the Hadamard product: \hat{M}_{seg} and \hat{C}_{cls}

$$M_{refined} = \hat{M}_{seg} \circ \hat{C}_{cls}$$

Using the model, the proposed methods were tested in different conditions, ablation studies were performed and the analysis of the resistance of the proposed method to different levels of errors in the annotations was performed. On average, the increase in the Dice coefficient relative to the base method was 13 %.

Thus, **fourth point of scientific novelty** is formulated, namely, improved methods of multi-task learning and inference based on an improved model of convolutional neural networks by combining classification and segmentation and introducing a constraint when calculating the segmentation loss function, which allowed to increase the accuracy of segmentation and classification in automated screening tasks.

In **fourth section** developed tools that implement the proposed solutions. The developed method was tested as part of experiments on synthetic data generated using the proposed model, as well as in experiments in real problems: screening of diabetic retinopathy, melanoma screening, and screening of cloud formations.

The tools are developed in the Python programming language using the PyTorch automatic differentiation framework. Based on the developed tools, effective software modules have been created, which are integrated with cloud services for solving resource-intensive learning tasks of neural networks, which provides high computing power and speed of prediction in automated screening tasks.

For the task of automated screening in the cloud organization patterns segmentation in satellite images (within the project "Understanding Clouds from Satellite Images" on the platform for competitions in data science Kaggle) the proposed models of neural networks were used, as well as methods for training and inference. The increase in Dice measure relative to the base model was 3.9%.

For the task of automated screening in the diabetic retinopathy classification (within the project "APTOS 2019 Blindness Detection" on the platform for competitions in data science Kaggle) used the proposed methods of multi-tasking and inference. The increase in F1-measure relative to the base model was 2.1 %.

For the task of recognizing skin lesions in melanoma screening (within

the SIIM-ISIC Melanoma Classification project on the Kaggle Data Science Competition Platform), the proposed localization of image-relevant image features was used. The use of the proposed method simplified the process of monitoring the learning of neural networks, which helped prevent retraining and increase the reliability of the classification by 3.5%.

The methods and tools developed in the work were introduced into the educational process of ONPU and the SafetyRadar software product of VITech Lab, the main purpose of which is to screen the presence of elements of personal protective equipment on people in construction sites, or hospitals and laboratories.

Key words: image analysis, dataset models, deep neural networks, multi-tasking machine learning, segmentation, classification, loss functions.