DEVELOPMENT OF MOBILE APPLICATION BASED ON MACHINE LEARNING METHODS FOR SKIN CANCER SCREENING

A. Karataev

Saint-Petersburg State University of Aerospace Instrumentation,
Saint-Petersburg, Russia
E-mail: hol0d@me.com

Abstract

Skin cancer takes the third place among Russian men and the second place among Russian women in comparison with all cases of diagnosed cancer. Melanoma is one of the most aggressive types of skin cancer. Only half of the patients diagnosed with melanoma survive the threshold of five years. Therefore, early diagnosis of melanoma is the only way to achieve a favorable prognosis for a treatment. The solution to this problem can be a mobile application that will help make an early diagnosis of melanoma and get the necessary treatment on time. This article discusses the mobile training method for solving this problem. It also describes the practical experience obtained on the basis of a trained model as well.

***Keywords*:** skin cancer screening, melanoma, mobile technology, machine learning, convolutional neural network, vgg16, transfer learning, fine tuning.

INTRODUCTION

An oncological disease is characterized by the appearance of a special cells that have the ability of uncontrolled division; due to this property, they invade the underlying tissues and metastasize to distant organs. The disease has been connected with damage of a cellular proliferation and a cellular differentiation owing to genetic disorders. Cutaneous carcinoma is the general name for a widespread malignant diseases. These diseases has been divided in two big groups: non-melanoma skin cancer and melanoma itself. The most widespread non-melanoma types of a skin cancer: squamous cell carcinomas and basal cell carcinoma also called basalioma.

Non-melanoma skin lesions tend to grow slowly and rarely metastasize. The melanoma is the most aggressive types of skin cancer, which usually develops from skin pigment cells. Long-term solar irradiation, chemical and thermal burns, as well as repeated damage of moles can cause the progression of a tumor growth.

According to the World Health Organization (2000) 200.000 cases of melanoma were diagnosed worldwide, with 65,000 deaths associated with it. Between 1998 and 2008, the increase in the incidence of melanoma in Russia was 38.17%; the standardized incidence rate increased from 4.04 to 5.46 per 100 thousand. Melanoma mortality rate in Russia was 3159 in 2008, and the standardized mortality indicator was 2.23 per one hundred thousand [1].

Early diagnosis of melanoma is carried out either by using a simple magnifying glass or by the dermatoscope, which makes the stratum corneum of the epidermis look transparent. At the same time it is possible to determine with high probability whether the mole is dangerous or not based on the ABCDE system (Asymmetry, Border irregularity, Color, Diameter, Evolving). With the help of computer diagnostic systems one can also get a good results. Boldrick et al. compared results of diagnostics made by the expert and by the model of an artificial neural network during their research.

The results of the experiment were close to the results obtained by a specialist, where the indicators of sensitivity and specificity of the model were within 95% and 88%, respectively. At the same time expert's indicators of dermatological specificity and sensitivity were within 95% and 90% respectively [2].

There are ample opportunities for the implementation of a portable diagnostic system. The second part of the article deals with the development of a diagnostic model and a mobile application, by which is possible to make an early diagnostics and to address a necessary treatment in a right time.

METHODOLOGY

FORMULATION OF THE PROBLEM

Diagnosis of a melanoma is a problem of binary classification, therefore, function of a posteriori probability in this case takes a form of Kulbak—Leibler distance. This measure, describes the amount of information which is lost at approach of distribution of P by means of distribution of Q. Posterior probability is the conditional probability of random events, provided that there is a posteriori data, i.e. measurements after the experience.

$$KL(P||Q)=\sum_i p\left(x_i\right)log\frac{p\left(x_i\right)}{q\left(x_i\right)}.$$

Optimizing the distance in its original form is not very convenient, so it must be reduced to a form called cross-entropy. For the binary classification problem, this objective function has the form of average cross entropy over all data points:

$$L\left(\theta\right)=H\left(p_{data},q\left(\theta\right)\right)=$$
$$=-\frac{1}{N}\sum_{i=1}^{N}(y_i log\widehat{y_i}\left(\theta\right)+\left(1-\widehat{y_i}\right)*$$
$$*log\left(1-\widehat{y_i}\left(\theta\right)\right)).$$

Thus, mathematically, the problem is reduced to finding the vector of the maximum a posteriori hypothesis for the cross-entropy function:

$$\theta_{MAP}=\arg max_\theta\, p\left(\theta|D\right)=.$$
$$=\arg max_\theta\, p(D\,|\,\theta)p\left(\theta\right)$$

DATA SET AND PROCESSING

For model training the data set from International Skin Imaging Collaboration has been used: Melanoma Project ISIC. This partnership between academia and industry is designed to facilitate the use of digital skin imaging to reduce melanoma death rates [3]. Examples of cellular' growths from the Melanoma Project ISIC set are presented in Fig. 1.

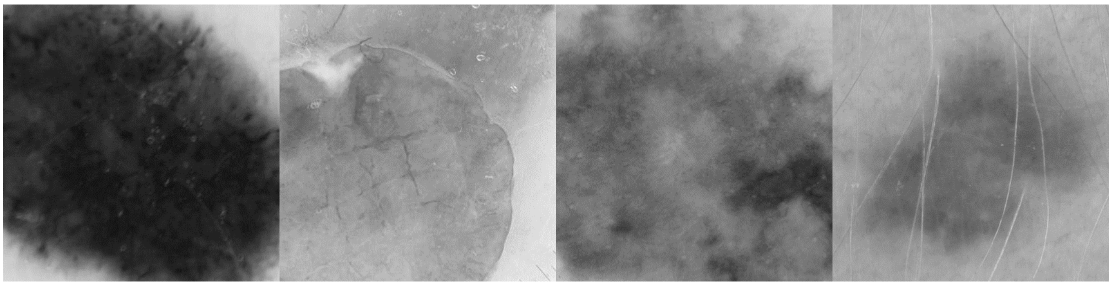

Fig. 1. Examples of cellular' growths from the Melanoma Project ISIC

Simple set of steps has been applied to pre-processing:

— Visual inspection.
— Cutting and turn of images.
— Removal of non-representative images.
— Augmentation for a data set balancing.

To combat retraining, it was decided to use an early stop strategy, so the main data set was divided into a training and validation subset. The strategy of an early stop means a training stop when the mistake on a validation set begins to grow. This method shows similar results with L2 regularization [4].

The training data set consisted of 3 thousand images; 1.5 thousand for each class, respectively. In turn, the validation collection consisted of a 1000 images, 500 for each class.

Data for test set, were taken from PH2Dataset. This set includes about 200 qualitative images with fields of damages and detail's masks [5].

.

DESIGN AND METHODS OF TRAINING

A pre-trained VGG16 model was selected for training in the generated data set. This model is a convolutional neural network that contains 13 convolutional and 3 fully connected layers [6]. The architecture of the convolutional layers of the model is presented in Fig. 2.

The first part of the network highlights the characteristic features in the image and consists of alternating stages of convolution and subsample. The size of the convolution kernel in all layers is 3 by 3. In a subsample, the maximum value is selected from a 2 by 2 square.

The second part is responsible for the classification of the object in the image according to the features that were highlighted in the previous step. This part contains 3 fully connected layers. In the first two layers are 4096 neurons, the number of neurons in the last layer depends on the number of classes of the problem being solved.

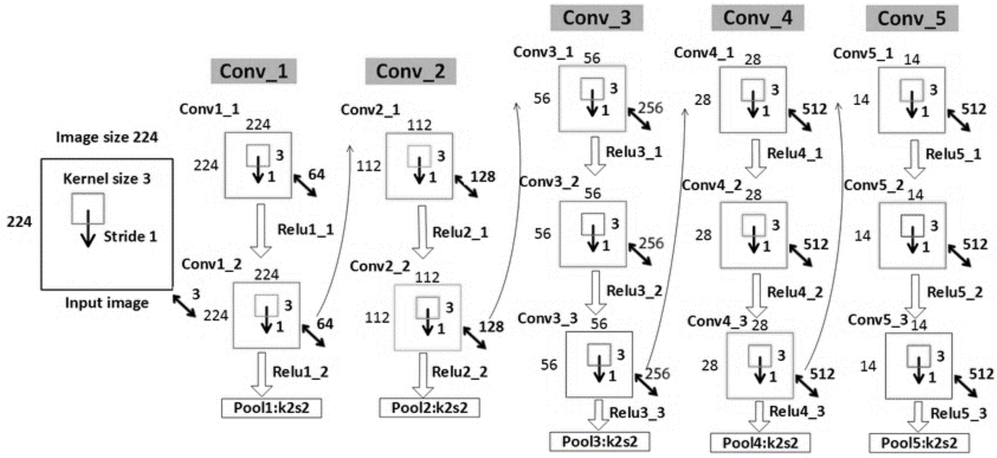

Figure 2. Convolution layers of VGG16 model

At the entrance, the VGG16 network receives an image of 224x224 pixels in size, 3 channels of color (red, green and blue). At the output, the network calculates the classification probabilities of an object in the one hot encoding format [6].

Since the size of the collected training sample does not allow to fully train a network of this size from scratch, it was decided to use such learning approaches as transfer learning and fine tuning.

Transfer learning allows you to use the weights of a pre-trained deep convolutional neural network, either as an initialization or as a distinction of distinctive features for the task. Thus, this ultimately allows the use of a much smaller data set for training, for a network of the required depth [7].

Fine tuning allows to adapt parameters of the pre-trained deep convolution neural network to objects of a solvable task, thereby reducing a final error of the posteriori probability's optimized function [8].

DEVELOPMENT OF APPLICATION

Treatment of melanoma in the early stages does not pose a serious problem for modern medicine. However, timely treatment is possible only with an early diagnosis. Thus, the development of a mobile application for the early diagnosis of melanoma should reduce mortality rates from this type of disease. The application usage diagram is presented in Fig.3.

The mechanism of application's operation for the end user, is very simple:

— The user scans problem section of his skin by the device camera.

— The neural network algorithm analyzes the image.

— The user receives a diagnosis (classification result).

If the algorithm detects a high probability for the diagnosis of cancer, the application will select the nearest dermatological clinic and arrange an appointment for a consultation.

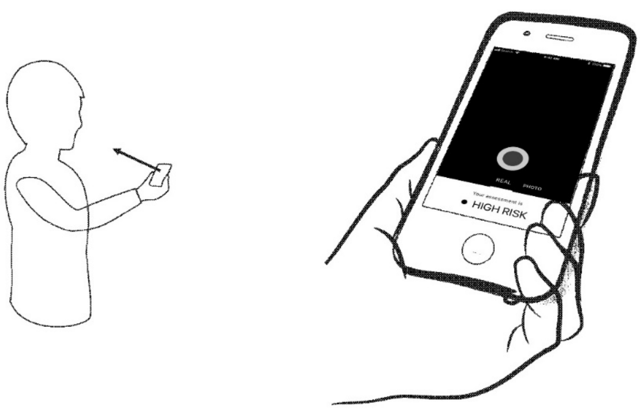

Figure 3. Screening by the mobile devices

RESULTS OF MODEL TRAINING

The chosen model studied 40 lessons which included 30 lessons of carrying out transfer learning and 10 lessons of fine tuning respectively. By results of training the value of mistake function on a validation subset was about 0.2.; an accuracy indicator on the same subset was 0.9. Indicators of accuracy and error of model are presented in Fig. 4.

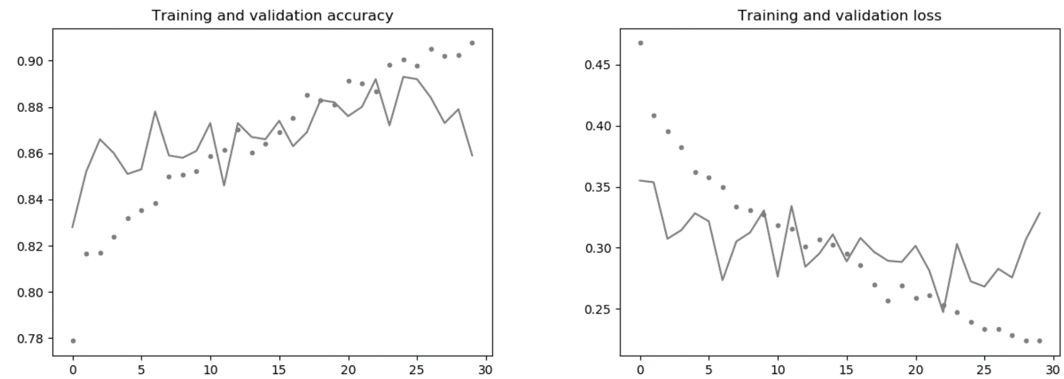

Figure 4. Indicators of accuracy and mistake for a training and validation subset

To analyze the errors of the first and second types, a ROC/AUC-graph was built, allowing to assess the quality of the binary classification. This graph displays the ratio between the share of objects in the total number of carriers of a trait, correctly classified as bearing a trait, and the share of objects in the total number of objects that do not carry a trait, mistakenly classified as bearing a trait. The ROC/AUC graph is presented in Fig. 5.

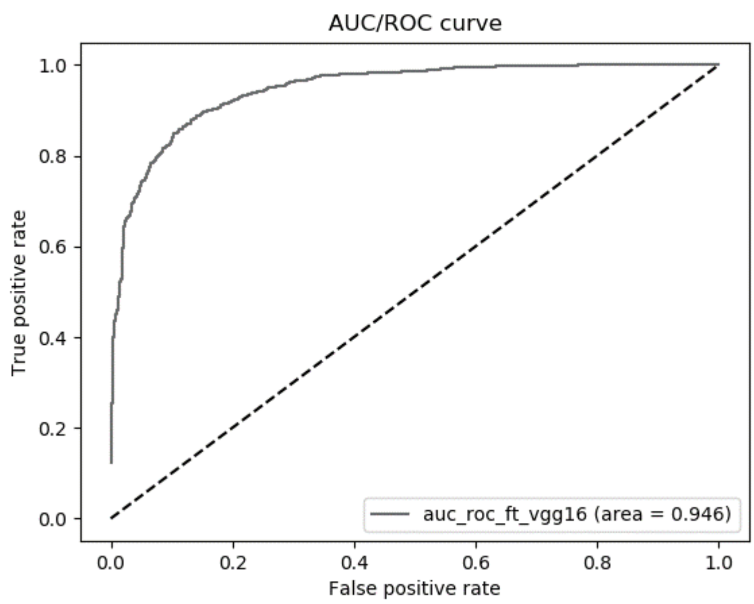

Figure 5. The ROC/AUC graph for the trained model

The quantitative interpretation of this graph is given by the AUC indicator – this is the area bounded by the ROC curve and the axis of the fraction of false positive classifications. The higher the AUC, the better the classifier. The AUC index for the trained model of the classifier is presented in Fig. 5 and is

0.946. This indicator value indicates the possibility of using the classifier for diagnostic purposes.

RESULTS OF A RESEARCH

The study successfully developed a mobile application based on a trained model. You can learn more about the application source code or take part in an open source project by going to the github repository (https://github.com/akarataev/gleam-ios).

REFERENCES

[1] World Health Organization: Cancer Prevention [electronic resource] URL: https://www.who.int/cancer/prevention/ru/ (15.10/2018).

[2] J. C. Boldrick, C. J. Layton, J. Nguyen, and S. M. Swetter, "Evaluation of digital dermoscopy in a pigmented lesion clinic: Clinician versus computer assessment of malignancy risk,"Journal of the American Academy of Dermatology, vol. 56, pp. 417-421, 2007/03/01/ 2007.

[3] International Skin Imaging Collaboration: Melanoma Project ISIC [electronic resource] URL: https://www.isic-archive.com/ (15.10/2018).

[4] Collobert R., Bengio S. Links Between Perceptrons, MLPs and SVMs // Proc. 21st ICML, New York, NY, USA: ACM, 2004. — P. 23–.

[5] PH² Database – Universidade do Porto [electronic resource] URL: http://www.fc.up.pt/addi/ph2%20database.html (15.10/2018).

[6] Keras Application [electronic resource] URL: https://keras.io/applications/ (15.10/2018).

[7] Liu, J., Wang, Y., Qiao, Y.: Sparse Deep Transfer Learning for Convolutional Neural Network. In: Association for the Advancement of Artificial Intelligence (2017).

[8] Ge, W.,

[9] Yu, Y.: Borrowing Treasures from the Wealthy: Deep Transfer Learning through Selective Joint Fine-Tuning (2017).