DOI: 10.15276/hait.03.2020.3
UDC 004.93.1

Classification of skin lesions using multi-task deep neural networks

Borys I. Tymchenko
Odessa National Polytechnic University, Odessa, Ukraine
ORCID: http://orcid.org/0000-0002-2678-7556
Philip O. Marchenko
Odessa I.I. Mechnikov National University, Odessa, Ukraine
ORCID: http://orcid.org/0000-0001-9995-9454
Eugene M. Khvedchenya
Independent researcher, Odessa, Ukraine
ORCID: http://orcid.org/0000-0002-2363-3850
Dmitry V. Spodarets
VITech Lab, Odessa, Ukraine
ORCID: http://orcid.org/0000-0001-6499-4575

ABSTRACT

Skin cancer is the most prevalent type of cancer disease. The most of skin cancer deaths are caused by melanoma, despite being the least common skin cancer. Early and accurate detection and treatment is the best healing, however detection of this type of malignancy in the early stages is not obvious. Data-driven solutions for malignant melanomas detection can make treatment more effective. Convolutional neural networks have been successfully applied in different areas of computer vision, also in the classification of cancer types and stages. But in most cases, images are not enough to reach robust and accurate classification. Such metadata as sex, age, nationality, etc. could also be applied inside the models. In this paper, we propose an end-to-end method for the classification of melanoma stage using convolutional neural networks from an RGB photo and persons' metadata. Also, we provide a method of semi-supervised segmentation of the region of melanoma appearance. From the experimental results, the proposed method demonstrates stable results and learns good general features. The main advantage of this method is that it increases generalization and reduces variance by using an ensemble of the networks, pretrained on a large dataset, and fine-tuned on the target dataset. This method reaches ROC-AUC of 0.93 on 10982 unique unseen images.

Keywords: computer vision; convolutional neural networks; multi-task learning; skin cancer; image classification; image segmentation

For citation: **Tymchenko B. I., Marchenko P. O., Khvedchenya E. M. Classification of skin lesions using multi-task deep neural networks. *Herald of Advanced Information Technology*. 2020; Vol.3 No.3: 136–148. DOI: 10.15276/hait.03.2019.3**

INTRODUCTION

Skin cancer is the most widespread type of human malignancy, and melanoma, specifically, is responsible for most of the deaths. The worldwide problem of melanoma incidence has risen rapidly for the last 50 years and became a problem that a lot of scientists from different countries trying to deal with. This year an estimated 100350 adults (60190 men and 40160 women) in the United States are expected to be diagnosed with invasive melanoma of the skin, and around 70000 of them could be fatal. Melanoma is the fifth most common cancer among men and the sixth most common cancer among women [1].

Similar to other cancer types, early and mild stages are hardly distinguishable visually. Currently, dermatologists evaluate every one of a patient's moles to identify outlier lesions or that are most likely to be melanoma. If melanoma is caught early, most of them can be cured with minor surgery.

Existing AI approaches have not adequately considered this clinical frame of reference. Dermatologists could enhance their diagnostic accuracy if detection algorithms take into account "contextual" images within the same patient to determine which images represent a melanoma. If successful, classifiers would be more accurate and could better support dermatological clinic work.

Convolutional neural networks have been successfully applied in different areas of computer vision, also in the classification of cancer types and stages. But in most cases, images are not enough to reach robust and accurate classification. Such meta-data as sex, age, nationality, etc. is also applied inside the model [2]. Also, the way of preprocessing is significant, e.g. sometimes it's important to understand the contour of the lesion, and sometimes only the texture matters [3]. The Skin Cancer Foundation gives simple guidelines for self-check, which can be used in a computerized solution [3]:

– Most melanomas are asymmetrical.

– Melanoma borders tend to be uneven and may have scalloped or notched edges.

– Melanoma may have different shades of brown, tan, or black. The colors red, white, or blue may also appear.

– A lesion is the size of about 6 mm is a warning sign

– Any change in size, shape, color, or elevation may be a warning sign of melanoma.

Also, there are several works on multi-task models of classification and segmentation [4–6]. However, all this models learned segmentation from annotated datasets. Thus, we research a method for learning segmentation only from binary annotated dataset and without any additional info about segments.

THE AIM

The aim of this work is to propose an end-to-end method for the classification of melanoma stage using convolutional neural networks from an RGB photo and persons' meta-data, as well as to provide a method of semi-supervised segmentation of the region of melanoma appearance.

TASKS OF THIS WORK

Main tasks of this work are:

– to summarize the most influential works in this field;

– to analyze available datasets for melanoma classification;

– to implement a method for melanoma classification and unsupervised segmentation;

– to test the developed method in the experiment

ANALYSIS OF THE LATEST RESEARCH AND PUBLICATIONS

Recent research in the field of automatic malignancy detection is connected with state-of-the-art deep learning approaches in image recognition, there are much fewer works with classical machine learning and handcrafted features. Here, we state the most influential works in this field. For example, Mustafa et al. [7] created an approach with manually extracted features (GrabCut for lesion segmentation) and trained SVM with a radial basis kernel to discriminate cancerous lesions. Also, Nasiri et al. [8] tried to augment images with different algorithms and trained k-nearest neighbor models to solve the task.

CNNs have emerged to be one of the major techniques for image classification in the last few years since a large number of improvements have been made. Also, many techniques train networks to solve classification problems appeared and prove their work on a lot of outstanding results. One of the most popular techniques is transfer learning.

Brinker et al. [9] experimented with ImageNet pretrained networks, such as ResNet-50, to classify early stages of melanomas. 4204 biopsy-proven images of melanoma and nevi (1:1) were used for the training of a convolutional neural network (CNN). Also, new techniques of deep learning were integrated: differential learning rates, rather than the same learning rate for all layers, reduction of the learning rate based on a cosine function, stochastic gradient descent with restart, to avoid local minima.

Codella et al. [10] proposed a system for the segmentation and classification of melanoma from dermoscopic images of skin. For disease classification, they employed an ensemble of recent machine learning methods, including deep residual networks, convolutional neural networks, etc. They proved that ensembles are capable to perform better results, than models separately.

Nasiri et al. [4] researched skin lesions classification using deep learning for early detection of melanoma in a case-based reasoning (CBR) system. This approach has been employed for retrieving new input images from the case base of the proposed system DePicT Melanoma Deep-CLASS to support users with more accurate recommendations relevant to their requested problem (e.g., an image of the affected area). Their methodology derived from a deep CNN generates case representations for case base to use in the retrieval process. Integration of this approach to DePicT Melanoma CLASS, significantly improving the efficiency of its image classification and the quality of the recommendation part of the system.

Research in the field of multi-task learning was also performed by Song et al. [5]. They proposed framework which can perform skin lesion detection, classification, and segmentation tasks simultaneously without requiring additional pre-processing or post-processing steps. Similar work was done by Chen et al. [6], which used multitask U-Net network for detection and segmentation.

Yang et al. [11] proposed even harder multitask model, which solves different tasks (e.g., lesion segmentation and two independent binary lesion classifications) at the same time by exploiting commonalities and differences across tasks.

PROBLEM STATEMENT

In the recent research, multiple ways of the classification and segmentation were presented.

However, semi-supervised multi-task learning is not researched together. Additionally, the usage of person-level meta-data is insufficiently studied. Besides, these researches do not investigate the influence of data augmentations.

In our research, we address the problem of semi supervised segmentation along with multi-task learning. Additionally, we add patient's level meta-data to improve image representations and an additional augmentation process, which is used on the source images, as an efficient way to prevent model from overfitting to the training data from different distributions.

The dataset

The image data used in this research was taken from several datasets with identical structures. We use SIIM & ISIC datasets from 2017, 2018, 2019, and 2020 years. These datasets were generated by the International Skin Imaging Collaboration (ISIC) and images were from the following sources: Hospital Clínic de Barcelona, Medical University of Vienna, Memorial Sloan Kettering Cancer Center, Melanoma Institute Australia, The University of Queensland, and the University of Athens Medical School.

All these datasets consist of around 50000 RGB images in total, from which around 3000 were malignant. The dataset contains 434 duplicate images. Besides the image data, meta-data about patients were given. Images and meta-data were provided in DICOM format, which is a commonly used medical imaging data format. Also, the dataset was available in JPEG format with images resized to a uniform 1024x1024. Meta-data was also provided outside of the DICOM format, in CSV files [12].

The tabular info was provided as follows:

1. *image_name* – unique identifier, points to filename of related DICOM image;
2. *patient_id* – unique patient identifier (string);
3. *sex* – the sex of the patient (is blank when unknown);
4. *age_approx* – approximate patient age at time of imaging (integer);
5. *anatom_site_general_challenge* – location of imaged site (string);
6. *diagnosis* – detailed diagnosis information (string);
7. *benign_malignant* – indicator of malignancy of imaged lesion (string, one of "benign" and "malignant");
8. *target* – binarized version of the target variable (boolean).

Values for *anatom_site_general_challenge* are taken from predefined finite set, so we encode it as one-hot vectors.

Meta-data is available per patient, so different images can have the same set of patient-level features. We use all available meta-data, except of *patient_id* and *diagnosis*, as **they are available only in training datasets**.

The dataset has a high class imbalance. The distribution of diagnoses is shown in *Fig. 1*. For the diagnosis, *unknown* researchers guarantee that it is not malignant [13].

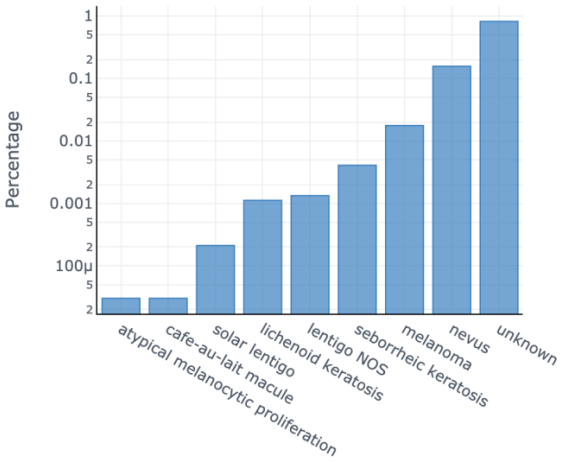

Fig. 1. **Distribution of diagnoses in the dataset. Notice the log scale**

Additionally, there are differences between train and test distributions of variables in meta-data. Melanoma is found more frequently in older men. Some images of the same patient are spread in time, and others are not.

Depending on the cancer stage, the outlook of malignant and benign lesions can be similar and different. Early-stage melanoma tends to be almost indistinguishable from benign lesions.

Image samples for benign and malignant classes are shown in *Fig. 2* and *Fig. 3* respectively.

Due to different sources of images and different imaging standards, they have structured noise in form of linear bars, regions marked with a pen, centering lines, etc. Depending on a site of neoplasm and gender, hairline could be also be observed. All of these additions could make a significant influence on the training process and may lead to overfitting, so it became an additional challenge to make models robust to it.

Train-validation data splitting

In this dataset, there are multiple factors, which can lead to the leakage of the diagnosis from the training subset to the validation subset while training. It can lead to overoptimistic results, along with the poor ability to generalize to the unseen data.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

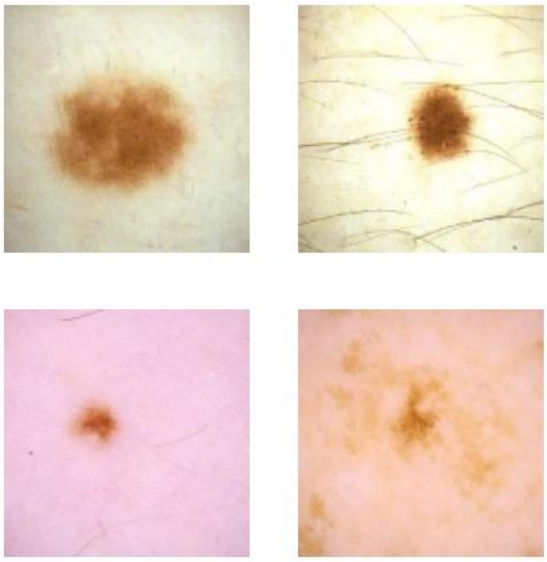

Fig. 2. **Samples of benign lesions**

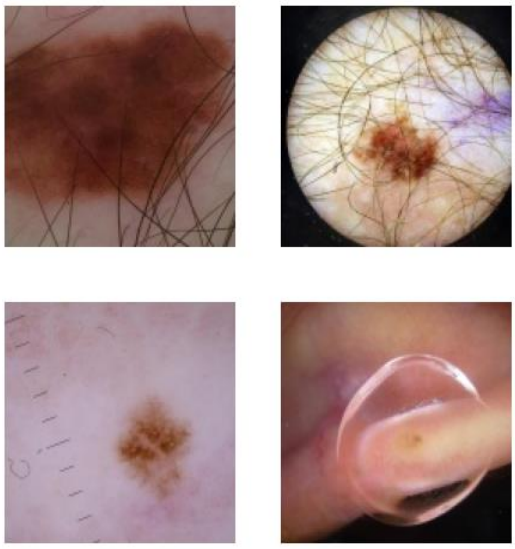

Fig. 3. **Samples of malignant lesions**

Here, under leakage we understand factors, which are spuriously correlated with diagnosis but have no causal relationship to it. E.g. images from the same patient can have similar visual features (hair type, skin color, etc.), or the number of images per patient can reveal information about the diagnosis process, that will not be available at the testing time.

We use a triple-stratified leak-free K-Fold cross-validation scheme [14]:

1. remove duplicate images;
2. isolate images from the same patient within a single fold;
3. balance folds to have the same distribution of malignant to benign images (1.8 %);
4. balance folds by the number of images per patient.

This method provides a more reliable cross-validation scheme, which is especially important when using models ensemble.

Evaluation metric

In this research, we used the area under the ROC curve [15] as our main metric. This metric is widely used by many researchers for binary classifiers. Also, in medicine, ROC analysis has been extensively used in the evaluation of diagnostic tests. The output of a binary classifier is interpreted as a probability distribution over the classes. Objects with an output value greater than 0.5 are assigned to the positive class in a binary classifier and objects with an output value less than 0.5 are assigned to the negative class. But according the ROC-AUC approach, the threshold used for classification systematically varies between 0 and 1, and the sensitivity and specificity are determined for each selected threshold. The ROC curve is calculated by plotting the sensitivity against 1-specificity and can be used to evaluate the classifier. The further the ROC curve deviates from the diagonal, the better the classifier. As a single value of classifiers quality, area under curve is calculated.

MULTI-TASK LEARNING

Multi-task learning is an approach to inductive transfer that improves generalization by using the domain information contained in the training signals of related tasks as an inductive bias. It does this by learning tasks in parallel while using a shared representation; what is learned for each task can help other tasks be learned better [16].

Here, we research multi-task learning for two tasks – classification and segmentation of the lesion.

Network architecture

In this research, we focus on the lesion classification. Presented neural networks are based on the conventional deep CNN architecture. As training deep CNNs from scratch is computationally expensive, we utilize inductive transfer from Imagenet-trained convolutional neural networks [17].

To concentrate the attention of CNN on the lesion itself, we use two branches: for classification and segmentation. We use same two-branch classification CNN structure as [18] (*Fig*. 4), replacing attention gating with the separate segmentation branch.

During training and inference time, we use the segmentation branch both as an attention mask for classification [19], and as a separate segmentation output. We utilize ImageNet-trained encoder as is, connecting segmentation decoder and the classifier to its last convolutional layer.

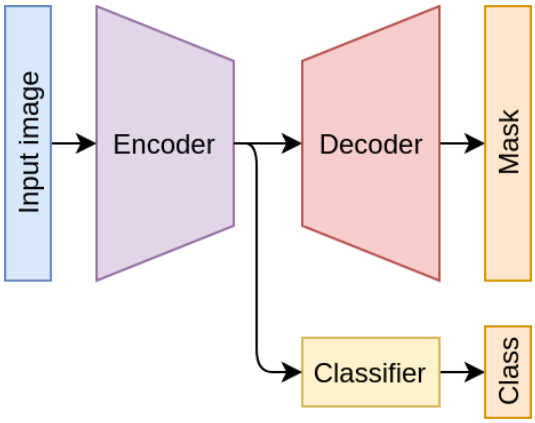

Fig. 4. **General two-branch CNN structure**

Meta-data usage

To incorporate meta-data into the model, we add separate encoder for meta-data along with the image encoder and concatenate representations from both encoders. Binary and string parameters are encoded as one-hot vectors, while numerical parameters are left as is. Then, all representations are concatenated into a single vector M_{in}.

For the meta-data encoder, we use a single linear layer with ReLU activation.

Let M_{in} be normalized onehot-encoded meta-data, F – features from image encoder, W and b are the weight and bias of the linear layer, respectively, and $\bigoplus$ is the vector concatenation. Then,

$$M_{out} = ReLU(WM_{in} + b), \quad (1)$$

$$V = M_{out} \bigoplus F. \quad (2)$$

Here, V is a result vector which is passed to the decoder and classifier. The structure of the multimodal model is shown in *Fig. 5*.

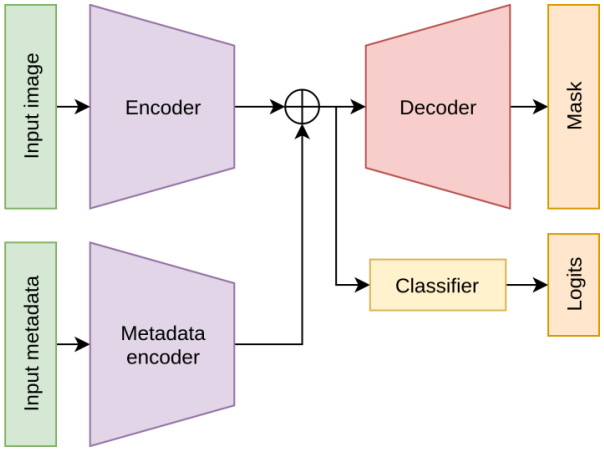

Fig. 5. **CNN structure with meta-data encoder**

Incorporation of meta-data into the model, allows it to learn discriminative image features better, than without meta-data. As meta-data is correlated with diagnosis (e.g. sex and age), we observe, that learning is improved, because the image encoder does not use its capacity to infer patient-level features from images. Thus, meta-data creates a prior, which is then refined using image features in decoders.

Unsupervised lesion region segmentation

During training, we use the segmentation branch to refine classification.

We define classifier function as $f_{classifier}$, and mask decoder function as $f_{segmentation}$. Let C and M be the classification and segmentation result tensors, respectively, and V is the encoder result vector:

$$C = f_{classifier}(V), \quad M = f_{segmentation}(V). \quad (3)$$

We define refined segmentation mask as:

$$M_{refined} = \sigma(M) \circ C. \quad (4)$$

Where $\circ$ is the element-wise matrix multiplication, σ is the element-wise sigmoid activation function:

$$\sigma = \frac{1}{1 + exp(-M)}. \quad (5)$$

Where $exp(-M)$ is the element-wise matrix exponential function.

For a single image, classifier output C is a single number, and the segmentation mask M and refined segmentation mask $M_{refined}$ are matrices of the shape *[H, W]*, where H and W are height and width of the decoder output.

To get refined classifier logits, we sum refined mask matrix $M_{refined}$ and divide by the sum of sigmoid of the segmentation mask M elements:

$$C^{refined} = \frac{\sum_h^H \sum_w^W M_{refined(hw)}}{\sum_h^H \sum_w^W \sigma\,(M_{hw})}. \quad (6)$$

Then, $C_{refined}$ is a single number.

For a batch of images, the same calculations are performed for every image in the batch. Graphical representation is show in the *Fig. 6*.

The detailed architecture of decoders is shown in *Fig. 7*. Here, encoders are image and meta-data encoder. For image encoder, any pretrained CNN can be used. In our experiments, we found, that EfficientNet [20] models achieve the best performance.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

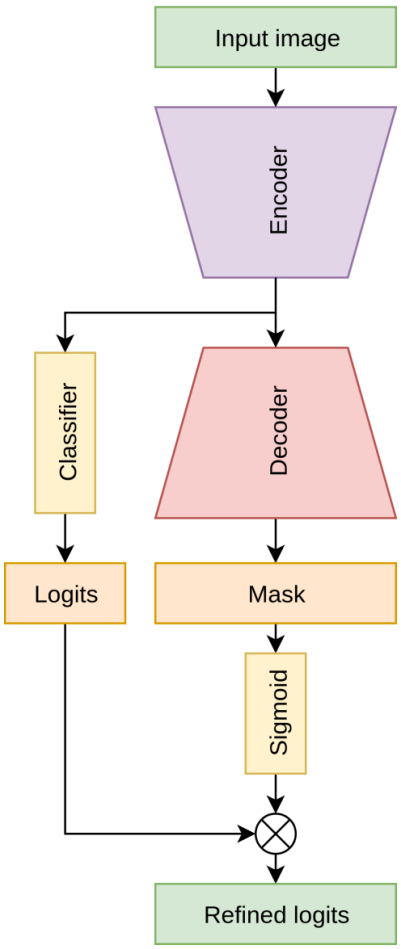

Fig. 6. **Network structure during training**

Naturally, CNN learns to assign high values (close to 1 after sigmoid activation) to regions with lesions in order to propagate gradients for the classification branch, while down-weighting regions without lesions. This way, we can use this attention map at the inference stage as a segmentation mask.

Additionally, samples with inconsistent attention maps tend to be classified wrong, so it can be a warning sign for doctors checking predictions manually.

Preprocessing

Both model training and inference are done with preprocessed versions of original images.

Because skin lesions are mostly in the center of the image, we crop the central square of the original image, and then resize it to the desired resolution, depending on the receptive field of the encoder part of the neural network.

To increase the contrast of lesions, we utilize CLAHE [21] processing on cropped and resized images.

Because skin lesions can be found in every person, with different lighting, or on different skin conditions, we apply data augmentation to increase the variability of the dataset.

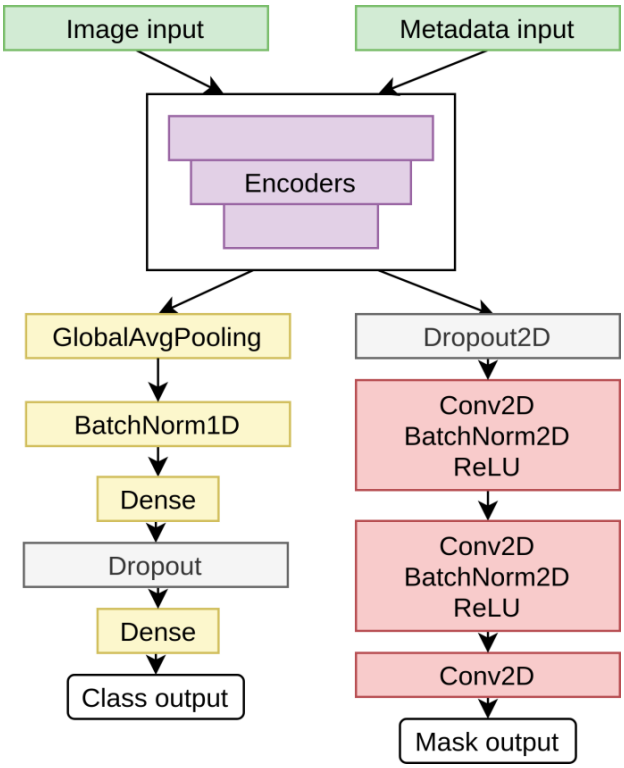

Fig. 7. **The detailed architecture of decoders**

Data augmentation

We used online augmentations, at least one augmentation was applied to the training image before inputting to the CNN. We used augmentations from Albumentations [22] library: horizontal and vertical flips; shift, scale, rotation; shift of RGB channels; random changes of brightness, contrast, and gamma.

Due to the way the dataset was collected, there is a spurious correlation of diagnosis and zoom level on the dermatoscope. To alleviate this correlation, we use the Microscope [23] augmentation, which adds a random black circle outline to the image of the lesion. The example of the true and the augmented images are shown in *Fig. 8.*

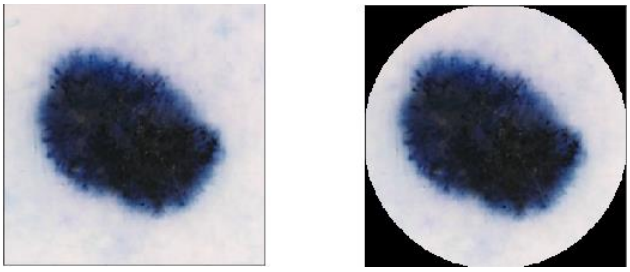

Fig. 8. **Microscope augmentation [23]: original and augmented images**

Additionally, we use Cutout regularization [24] to improve the robustness of CNN to partially

covered or obscured with hair lesions. In order not to accidentally obscure big parts of the lesion in the image, we mask several small patches at random instead of the single big one.

Training process

We utilize a single-stage training process, which includes transfer learning and multi-task learning.

Training is performed in a 3-fold cross-validation scheme. The feature extractor is initialized with noisy-student [25] trained weights. In our experiments, we observed that this initialization leads to consistently better results, than Imagenet initialization.

We randomly resample the dataset at each epoch to better capture the minority class and to reduce training time. We downsample the majority class (benign lesions) to match the number of minority class (malignant). To match the source distribution, we initialize the bias term of the last layer of the model according to class imbalance ratio.

During our experiments, CNNs were trained up to 50 epochs with early stopping [26]. Training stopped automatically in a range from 20 to 40 epochs. In this task, we used a Radam optimizer, consistently better, than Adam [27] and SGD [28] baselines.

We use cosine annealing learning rate schedule to achieve a better any-time performance of our CNNs [29].

During training, we monitor the distribution of the classifier predictions for both classes separately.

To stabilize training, we use an exponential moving average (EMA) for weights of the CNN [30]. The validation curve for the ROC-AUC [15] metric on a holdout test set for a single model is shown in *Fig. 9*. We notice a similar improvement in all experiments.

Loss functions

To train our models we used different loss functions and their combinations. As we resample the dataset to mitigate class imbalance, binary cross-entropy loss is enough to achieve good classification for the majority of samples.

However, we observed the situation when benign and malignant melanomas looked quite similar (e.g. amelanotic melanoma, where the malignant cells have very little or no pigment at all), so we opted to down-weight the influence of well-classified examples and concentrate the optimization process on hard examples.

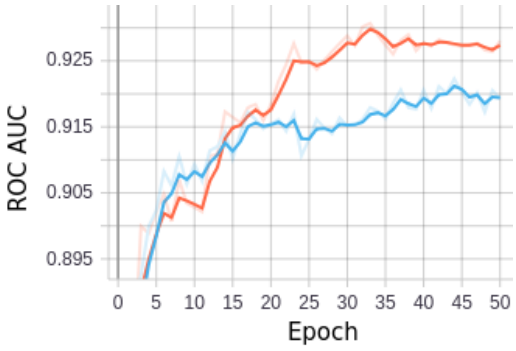

Fig. 9. **Validation curve with (orange) and without (blue) EMA. Best viewed in color**

Focal loss [31]

This loss focuses on training on hard examples and prevents the vast number of easy negatives from overwhelming the classifier during training. Focal loss reduces the weight (or impact) the values CNN predicted correctly, which often happens with the majority class.

The focal loss could be calculated as follows:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma} log(p_t). \qquad (7)$$

Where

$$p_t = p \Leftrightarrow y = 1, \text{otherwise } p_t = 1 - p.$$

In our experiments we found that $\alpha = 0.25, \gamma = 2$ worked the best.

Online Hard Example Mining (OHEM) [32]

This loss back propagates only on hard examples, which are drawn from the current state of the network. Simply, we set the threshold for the per-sample loss value and calculate the reduction function on samples, which have greater loss than a defined threshold. We also set a minimum number of examples to be selected (in our experiments, the minimal number is half the size of the batch). As a loss function here we used binary cross-entropy with mean reduction.

Experimentally, we found that with small batches (64 images), OHEM outperforms Focal loss; however, with batches larger than 100 elements, Focal loss outperforms OHEM.

Flood loss [33]

As deep neural networks overfit fast to the large number of negative samples the training dataset, we have to add regularization. Along with weight decay and dropout, we use a direct solution called flooding that intentionally prevents further reduction of the training loss when it reaches a reasonably small value, which we call the flooding level. This approach makes the loss float around the flooding

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

level by doing mini-batched gradient descent as usual but gradient ascent if the training loss is below the flooding level.

If the original learning objective is J, the proposed modified learning objective J^{flood} with flooding:

$$J^{flood} = |J(\theta) - b| + b, \quad (8)$$

where $b > 0$ is the flooding level specified by the user, and θ is the model parameter.

With flooding, the model continues to random walk with the same non-zero training loss, and we expect it to drift into an area with a flat loss landscape that leads to better generalization, which is crucial in the melanoma classification task.

Inference

During inference, we resize testing images to the size, on which models were trained. We trained models with following image sizes: 256x256, 384x384 and 512x512.

Predicted masks are resized from their native resolution, which can be from 16x16 to 64x64 depending on the encoder, to the resolution of the input image using bilinear interpolation.

We utilize mask post-processing, test-time augmentations and ensembling to achieve more stable results.

Mask post-processing

Mask output from CNN is continuous. As we train the mask segmentation branch in an unsupervised fashion, we cannot directly predict the range of the segmentation output. To alleviate the calibration of the model predictions, we binarize the mask using the Otsu threshold [34].

After the mask has been binarized, we apply a morphological opening to reduce the number of small false-positive regions.

Examples of the unsupervised segmentation for malignant and benign lesions are shown in Fig. 10 and *Fig. 11* respectively.

However, sometimes CNN fails to capture all pixels to the consistent mask, especially with big or uneven lesions. Such examples are shown in the *Fig. 12*; *Fig. 13* and *Fig. 14.*

We are going to address this issue in future research.

Test-time augmentations

To reduce the variance of predictions, we utilize test-time augmentations (TTA) [35]: we make predictions on different changed versions of the original images, and then average prediction results.

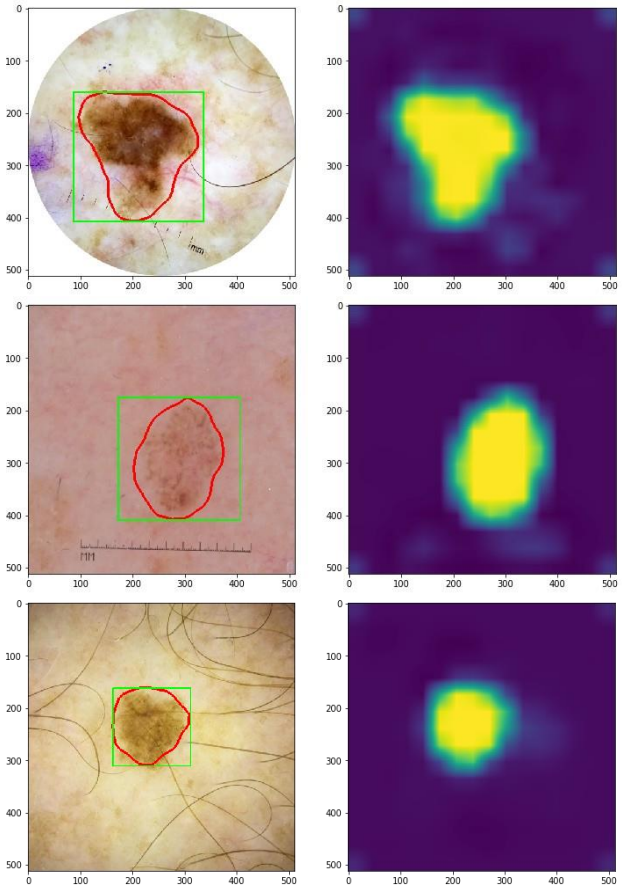

Fig. 10. **Examples of segmentation for malign lesions**
Red – outline of the binarized mask, green – bounding box around the mask. Best viewed in color

As pictures of lesions can be viewed from any angle and with different illumination, we utilize the following changes to each original image:

1. Original image;
2. Horizontal flip;
3. Vertical flip;
4. Transpose.

Which in total gives us an average of 16 predictions per single image. We use *ttach* – an efficient implementation of TTA [36].

Ensemble

For the Kaggle competition [11], we used predictions from models and setups from each fold of cross-validation in the ensemble.

Additionally, we experimented with different random seeds to get a more robust ensemble of models [37]. For the final ensemble, we selected raw predictions of low pairwise correlation from best models according to ROC-AUC and applied sigmoid activation to raw outputs. We merged all predictions into one by taking the mean of sigmoid outputs.

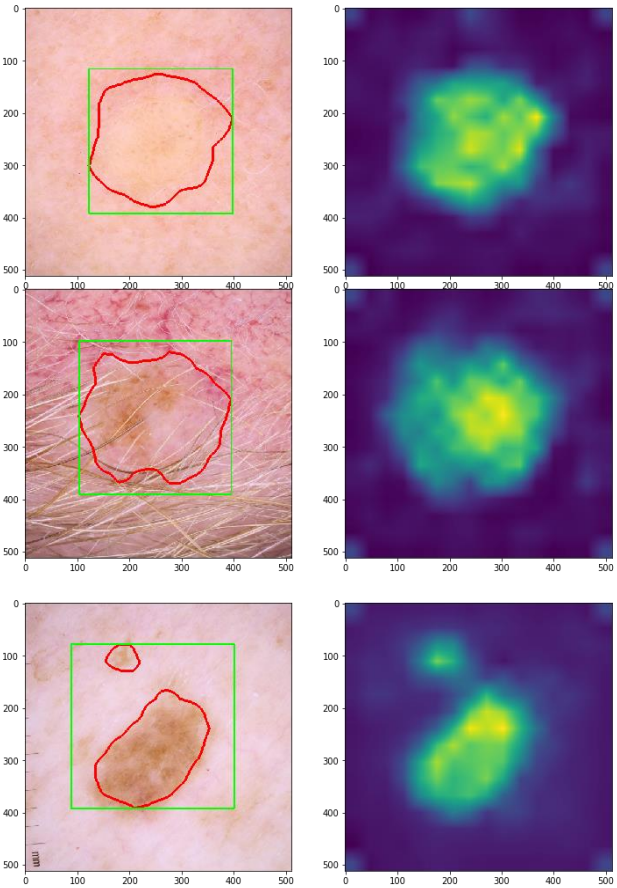

Fig. 11. **Examples of segmentation for benign lesions**

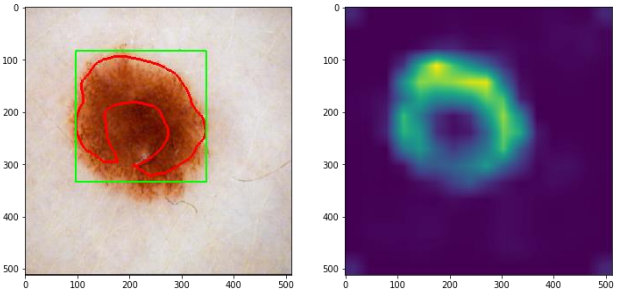

Fig. 12. **Incomplete segmentation of the malignant lesion**

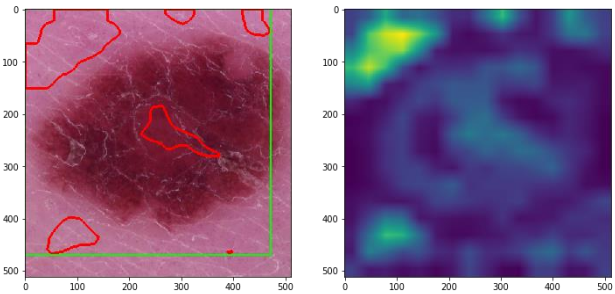

Fig. 13. **Complete failure to capture the region**

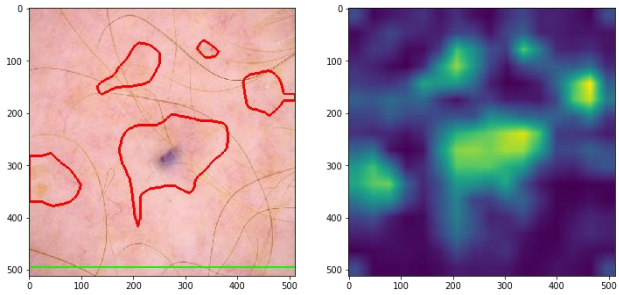

Fig. 14. **Failure case with inconsistent regions in the mask**

Our best ensemble consisted of the following encoder architectures:

1. EfficientNet-B4 [20].
2. EfficientNet-B5 [20].
3. EfficientNet-B6 [20].
4. SE-ResNext 101 [38].

Additionally, selected models have different training setups and head classifiers for output features, which increases the accuracy and robustness of the ensemble.

We used the Catalyst framework [39] based on PyTorch [40] with GPU support. Evaluation of the whole ensemble was performed on Nvidia V100 GPU in about 100 minutes, processing 4.5 seconds per image with test time augmentation.

RESULTS

Our test stage was split into two parts: holdout testing and final testing, which contained 10982 unseen images. For final test reliability, it was split on 30 % of the public test, and rest 70 % was blind. Such an approach helped us to check whether the model works stable on unseen data.

Our best single model (EfficientNet-B5, OHEM loss) scored 0.9304/0.9402 of ROC-AUC points. The same model, trained without segmentation branch reaches ROC-AUC scores of 0.9286/0.9372. The same model, trained without meta-data input reaches ROC-AUC scores of 0.9104/0.9253.

For comparison, ensembling with test time augmentation performed better on the public and blind test sets, as it has a better ability to generalize on unseen images.

We tried several types of ensembling:

1) mean ensemble of raw outputs of the model, which scored 0.9347/0.9497 of ROC-AUC points blind/public test sets;

2) mean the ensemble of sigmoid outputs of the model, which scored 0.9353/0.9511 of ROC-AUC points blind/public test sets;

3) log-mean ensemble, which scored 0.9392/0.9492 of ROC-AUC points blind/public test sets.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

In the Kaggle competition [11], our method was ranked 420/1148 on blind/public test sets. The method *with rank 1/1148* in this competition scored 0.9490/0.9586 ROC-AUC.

Future work can extend our method with modifications of segmentation head and more accurate augmentations to decrease the influence of noisy additional info (such as hair). Additionally, the method can be extended with explicit filtering of image features with meta-data features.

CONCLUSIONS

In this paper, we proposed an end-to-end method for the classification of melanoma malignancy using convolutional neural networks from an RGB photo and persons' meta-data. We provided a method of semi-supervised segmentation of the region of melanoma appearance, which also improves results of classification by concentrating attention of the lesion instead of its surroundings.

Segmented regions with melanomas could be used as a preprocessing step, as cleaning augmentation, or as an additional informative part for classification.

The main advantage of this method is that it provides solutions for the task of segmentation, even if segmented training data is not provided. This method can provide guidance to doctors and to inform them, when the diagnosis should be clarified manually. Besides, this method benefits from using an ensemble of the networks, pretrained on a large dataset, and finetuned on the target dataset increasing generalization and reducing variance.

REFERENCES

1. "Melanoma: Statistics". Available from: https://www.cancer.net/cancer-types/melanoma/statistics. [Accessed 29th September 2019].

2. Castilla, R., Rangel-Cortes, J., García-Lamon, F., Adrian, T. "CNN and Metadata for Classification of Benign and Malignant Melanomas". *Intelligent Computing Theories and Application.* 2019. p. 569–579. DOI: 10.1007/978-3-030-26969-2_54.

3. "Melanoma Warning Signs". Available from: https://www.skincancer.org/skin-cancer-information/melanoma/melanoma-warning-signs-and-images. [Accessed 15th August 2020].

4. Nasiri, S., Helsper, J., Jung, M. & Fathi, M. "DePicT Melanoma Deep-CLASS: a deep convolutional neural networks approach to classify skin lesion images". *BMC Bioinformatics.* 2020; Vol. 21(2):84. DOI: 10.1186/s12859-020-3351-y.

5. Song, L., Lin J., Wang Z. & Wang H. "An End-to-end Multi-task Deep Learning Framework for Skin Lesion Analysis". *IEEE J Biomed Health Inform.* Epub ahead of print. PMID: 32071016. (13 Feb. 2020). DOI: 10.1109/JBHI.2020.2973614.

6. Chen, E., Dong, X., Li, X., Jiang, H., Rong, R. & Wu, J., "Lesion Attributes Segmentation for Melanoma Detection with Multi-Task U-Net". *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019),* Venice, Italy: 2019. p. 485-488. DOI: 10.1109/ISBI.2019.8759483.

7. Mustafa, S. & Kimura, A., "A SVM-based diagnosis of melanoma using only useful image features", *2018 International Workshop on Advanced Image Technology (IWAIT),* Chiang Mai, 2018. p. 1-4. DOI: 10.1109/IWAIT.2018.8369646.

8. Nasiri, S., Jung, M., Helsper, J. & Fathi M. "Detect and Predict Melanoma Utilizing TCBR and Classification of Skin Lesions in a Learning Assistant System". *Bioinformatics and Biomedical Engineering, Lecture Notes in Computer Scienc*e. *Springer International Publishing.* 2018; Vol. 10813: 531–542.

9. Brinker, T., Hekler, A. & Alexander, H. "Deep neural networks are superior to dermatologists in melanoma image classification". *European Journal of Cancer.* 2019; Vol.119. DOI: 10.1016/j.ejca.2019.05.023.

10. Codella, N., Nguyen, Q., Pankanti, S. & Gutman, D. "Learning Deep Ensembles for Melanoma Recognition in Dermoscopy Images", eprint arXiv:1610.04662. USA. 2016.

11. Yang, X., Zeng, Z., Yeo, S., Tan, C., Tey, H. & Su, Y. "A Novel Multi-task Deep Learning Model for Skin Lesion Segmentation and Classification", eprint arXiv:1703.01025. USA. 2017.

12. "SIIM-ISIC Melanoma Classification | Data". Available from: https://www.kaggle.com/c/siim-isic-melanoma-classification/data. [Accessed 19th August 2020].

13. "What is "diagnosis=unknown" in the CSV train file?". Available from: https://www.kaggle.com/c/siim-isic-melanoma-classification/discussion/155296#875346. [Accessed 15th August 2020].

14. "Triple Stratified Leak-Free KFold CV". Available from: https://www.kaggle.com/c/siim-isic-melanoma-classification/discussion/165526. [Accessed 15th August 2020].

15. Bragley, A. “The use of the area under the ROC curve in the evaluation of machine learning algorithms”. *Pattern Recognition.* 1997; Vol. 30 Issue 7. DOI: 10.1016/S0031-3203(96)00142-2.

16. Caruana, R. “Multitask Learning”. *Machine Learning* 28. 1997. p. 41–75. DOI: 10.1023/A:1007379606734.

17. Iglovikov, V. & Shvets, A. “TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation”, eprint arXiv: 1801.05746. USA. 2018.

18. Schlempe,r J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B. & Rueckert D. “Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images”, eprint arXiv:1808.08114. USA. 2018.

19. Li, K., Wu, Z., Peng, K., Ernst, J. & Fu, Y. “Tell Me Where to Look: Guided Attention Inference Network”, eprint arXiv:1802.10171. USA. 2018.

20. Tan, M. & Le, Q. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, eprints arXiv: 1905.11946. USA. 2019.

21. Zuiderveld, K. “Contrast Limited Adaptive Histogram Equalization”. *Graphics Gems IV*. *Academic Press.* 1994.

22. “Albumentations: fast image augmentation library and easy to use wrapper around other libraries”. Available from: https://github.com/albumentations-team/albumentations/. [Accessed 15th August 2020].

23. “Microscope augmentation”. Available from: https://www.kaggle.com/c/siim-isic-melanoma-classification/discussion/159476. [Accessed 15th August 2020].

24. DeVries, T. & Taylor, G. “Improved Regularization of Convolutional Neural Networks with Cutout”, eprint arXiv:1708.04552. USA. 2017.

25. Xie, Q., Luong, M., Hovy, E. & Le, Q. “Self-training with Noisy Student improves ImageNet classification”, eprint arXiv:1911.04252. USA. 2019.

26. Rich, C., Lawrence, S. & Giles, C. “Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping”. *Advances in Neural Information Processing Systems*. USA. 2001. p. 402–408.

27. Kingma, D. & Ba, J. “Adam: A Method for Stochastic Optimization”, eprint arXiv:1412.6980. USA. 2014.

28. Robbins, H. “A Stochastic Approximation Method”. *Annals of Mathematical Statistics* Volume 22, Number 3 (1951): 400-407. DOI: 10.1214/aoms/1177729586.

29. Loshchilov, I. & Hutter, F. “SGDR: Stochastic Gradient Descent with Warm Restarts”, eprint arXiv: 1608.03983. USA. 2016.

30. Tarvainen, A. & Valpol, H. “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results”, eprint arXiv: 1703.01780. USA. 2017.

31. Lin, T., Goyal, P., Girshick, R., He, K. & Dollár, P. “Focal Loss for Dense Object Detection”, eprint arXiv: 1708.02002. USA. 2017.

32. Shrivastava, A., Gupta, A. & Girshick, R. “Training Region-based Object Detectors with Online Hard Example Mining”, eprint arXiv: 1604.03540, USA. 2016.

33. Ishida, T., Yamane, I., Sakai, T., Niu, G. & Sugiyama M. “Do We Need Zero Training Loss After Achieving Zero Training Error?”, eprint arXiv: 2002.08709. USA. 2020.

34. Otsu, N. “A Threshold Selection Method from Gray-Level Histograms”. *IEEE Transactions on Systems, Man, and Cybernetics.* Vol. 9 , Issue 1, Jan. 1979. p. 62-66. DOI: 10.1109/TSMC.1979.4310076.

35. Moshkov, N., Mathe, B., Kertesz-Farkas, A., Hollandi, R. & Horvath, P. “Test-time augmentation for deep learning-based cell segmentation on microscopy images”. Eprint bioRxiv 814962. USA. 2020. DOI: 10.1101/814962.

36. “Image Test Time Augmentation with PyTorch”. Available from: https://github.com/qubvel/ttach/. [Accessed 15th September 2020].

37. Fort, S., Hu, H. & Lakshminarayanan, B. “Deep Ensembles: A Loss Landscape Perspective”, eprint arXiv: 1912.02757. USA. 2019.

38. Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. “Squeeze-and-Excitation Networks”, eprints arXiv: 1709.01507. USA. 2017.

39. “Accelerated DL R&D”. Available from: https://github.com/catalyst-team/catalyst/. [Accessed 15th September 2020].

40. “PyTorch”. Available from: https://pytorch.org/. [Accessed 15th August 2020].

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

DOI: 10.15276/hait.03.2020.3
UDC 004.93.1

Класифікація уражень шкіри з використаннями багатозадачних глибкоких нейронних мереж

Б. І. Тимченко
Одеський Національний Політехнічний Університет, Одеса, Україна
ORCID: http://orcid.org/0000-0002-2678-7556
Ф. О. Марченко
Одеський Національний Університет ім. І.І. Мечникова, Одеса, Україна
ORCID: http://orcid.org/0000-0001-9995-9454
Є. М. Хведченя
Незалежний дослідник, м. Одеса, Україна
ORCID: http://orcid.org/0000-0002-2363-3850
Д. В. Сподарець
VITech, Одеса, Україна
ORCID: http://orcid.org/0000-0001-6499-4575

АНОТАЦІЯ

Рак шкіри є найбільш поширеним видом онкологічних захворювань. Більшість випадків смерті від раку шкіри спричинені меланомою, хоча це найменш поширений рак шкіри. Раннє та точне виявлення та лікування є найкращим зціленням, однак виявлення цього виду злоякісної пухлини на ранніх стадіях не є легким. Рішення, засновані на даних для виявлення злоякісних меланом можуть зробити лікування більш ефективним. Згорткові нейронні мережі успішно застосовуються в різних областях комп'ютерного зору, а також у класифікації типів та стадій раку. Але в більшості випадків зображень недостатньо для досягнення надійної та точної класифікації. Такі метадані, як стать, вік, національність тощо, також можуть бути застосовані всередині моделей. У цій роботі ми пропонуємо end-to-end метод класифікації стадії меланоми за допомогою згорткових нейронних мереж із фотографії RGB та метаданих пацієнтів. Також ми пропонуємо метод напівавтоматичного навчання сегментації області новоутворення. На основі експериментальних результатів запропонований метод демонструє стабільні результати та вивчає ознаки, що добре описують новоутворення. Головною перевагою цього методу є те, що він збільшує узагальнення та зменшує дисперсію, використовуючи ансамбль мереж, попередньо навчений на великому наборі даних та донавчений на цільовому наборі даних. Цей метод досягає ROC-AUC 0.93 на 10982 унікальних нових зображеннях.

Ключові слова: комп'ютерний зір; згорткові нейронні мережі; багатозадачне навчання; рак шкіри; класифікація зображень; сегментація зображень

DOI: 10.15276/hait.03.2020.3
UDC 004.93.1

Классификация поражений кожи с помощью многозадачных глубоких нейронных сетей

Б. И. Тимченко
Одесский Национальный Политехнический Университет, Одесса, Украина
ORCID: http://orcid.org/0000-0002-2678-7556
Ф. А. Марченко
Одесский Национальный Университет им. И.И. Мечникова, Одесса, Украина
ORCID: http://orcid.org/0000-0001-9995-9454
Е. Н. Хведченя
Независимый исследователь, Одесса, Украина
ORCID: http://orcid.org/0000-0002-2363-3850
Д. В. Сподарец
VITech Lab, Одесса, Украина
ORCID: http://orcid.org/0000-0001-6499-4575

АННОТАЦИЯ

Рак кожи – наиболее распространенный вид онкологических заболеваний. Большинство смертей от рака кожи вызваны меланомой, несмотря на то, что это наименее распространенный вид рака кожи. Раннее и точное обнаружение и лечение - лучшее лечение, однако обнаружение этого типа злокачественного новообразования на ранних стадиях затруднено.

Решения на основе данных для обнаружения злокачественной меланомы могут сделать лечение более эффективным. Сверточные нейронные сети успешно применяются в различных областях компьютерного зрения, а также при классификации типов и стадий рака. Но в большинстве случаев изображений недостаточно для надежной и точной классификации. Такие метаданные, как пол, возраст, национальность и т.д., также могут применяться внутри моделей. В этой статье мы предлагаем end-to-end метод классификации стадии меланомы с использованием сверточных нейронных сетей на основе RGB-фотографии и метаданных пациентов. Также мы предлагаем метод полуконтролируемой сегментации области появления меланомы. По результатам экспериментов предлагаемый метод демонстрирует стабильные результаты и изучает хорошие признаки на изображениях. Основное преимущество этого метода заключается в том, что он уменьшает дисперсию за счет использования ансамбля сетей, предварительно обученных на большом наборе данных и дообученных на целевом наборе данных. Этот метод достигает ROC-AUC 0.93 на 10982 уникальных изображениях.

Ключевые слова: компьютерное зрение, сверточные нейронные сети, многозадачное обучение, рак кожи, классификация изображений, сегментация изображений

ABOUT AUTHORS

Borys I. Tymchenko – Ph. D. student of the Institute of Computer Systems
Odessa National Polytechnic University, Odessa, Ukraine
tymchenko.b.i@opu.ua
Борис І. Тимченко – аспірант інституту Комп'ютерних систем, Одеський національний політехнічний Університет, Одеса, Україна

Борис И. Тимченко – аспирант Института Компьютерных систем, Одесский национальный Политехнический Университет, Одесса, Украина

Philip O. Marchenko – Ph.D student of the Faculty of Mathematics, Physics and Information Technology, Odessa I.I. Mechnikov National University, Odessa, Ukraine
p.marchenko@stud.onu.edu.ua

Філіп О. Марченко – аспірант Факультету Математики, фізики та інформаційних технологій, Одеський національний університет ім. І.І. Мечникова, Одеса, Україна

Филип А. Марченко – аспирант факультета Математики, Физики и Информационных технологий, Одеский национальный университет им. И.И. Мечникова, Одесса, Украина

Eugene M. Khvedchenya – independent researcher ekhvedchenya@gmail.com,Odessa, Ukraine
ekhvedchenya@gmail.com

Євген М. Хведченя – незалежний дослідник, Одеса, Україна

Евгений Н. Хведченя – независимый исследователь, Одесса, Украина

Dmytro V. Spodarets – Head of R&D VITech, VITech Lab, Odessa, Ukraine
dmitry.spodarets@vitechlab.com

Дмитро В. Сподарець – керівник відділу досліджень та розробок, VITech Lab, Одеса, Україна

Дмитрий В. Сподарец – руководитель отдела исследований и разработок, VITech Lab, Одесса, Украина

Received 07.08.2020
Received after revision 17.09.2020
Accepted 21.09.2020

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)