Melanoma Detection

Using Transfer Learning Techniques

Viktor Horváth
Távközlési és Médiainformatikai Tanszék,
Budapesti Műszaki és Gazdaságtudományi Egyetem,
Budapest, Hungary

Adrienn Horváth
Távközlési és Médiainformatikai Tanszék,
Budapesti Műszaki és Gazdaságtudományi Egyetem,
Budapest, Hungary

Beatrix Koltai
Távközlési és Médiainformatikai Tanszék,
Budapesti Műszaki és Gazdaságtudományi Egyetem,
Budapest, Hungary

***Abstract*—Among skin cancer diseases, melanoma is one of the most common and most dangerous ones, being difficult to detect and causing serious problems once spread deeper. Melanomas have many different shapes, sizes, and colours, thus making it harder to provide comprehensive warning signs. Although melanoma is usually curable in an early stage, it becomes serious if not treated and can spread deeper into the skin or other parts of the body making treatment even harder and eventually becomes deadly. For this reason, detecting melanoma early is of utmost importance. Neural Networks provide significant progress in this field, making it possible to detect melanomas with many previously learned patterns, dealing with larger datasets than a human specialist can handle. In our work, we analysed different models for detecting melanomas to select the most accurate method for this purpose.**

Keywords—melanoma, detection, transfer, learning, convolutional, neural, networks, resnet, inception, efficientnet

I. Introduction

Detecting signs of skin cancer is a severe problem worldwide and is an ongoing task demanding state of the art technologies to fulfil and to make relevant progress in clinical diagnostics. Detecting melanomas accurately on human skin among the many types of pigmented lesions would be one of the most significant steps towards better healthcare. For this very reason, in our work, we set a goal to design and implement a method for analyse and compare existing solutions for image recognition that could be used for melanoma detection and evaluate the results to be able to select the most accurate algorithm. To achieve this, we gathered relevant and state of the art studies and papers, and processed these to rely on them when designing our system.

II. Previous Solutions

A. Review of Literature

Melanoma is a type of cancer that originates from melanocyte cells, in most cases from the epidermis. [1][2] In case of detection at an early stage the 5-year related survival rate is 90%, which may be decreased to 9-15% by a delayed treatment started at tumor stage IV. [3][4] Therefore it is of high importance for early detection so that the prospects of a cure can be significantly favourable [5][6][7].

Dermoscopy in clinical practice is widely used for melanoma detection. The diagnosis is based on some specific morphological features of the mole, such as symmetry, border irregularity, colour variegation, [8] irregular dots of pigment, irregular peripheral extensions, a blue-white veil, and other specific irregular patterns [9].

This makes it possible to use pattern recognition techniques.

In 1994, an artificial neural network was proposed for epiluminescence microscopy pattern analysis of pigmented skin lesions, which was amongst the pioneers of utilizing Artificial Intelligence for skin cancer detection solutions. In this study, 88% of the test set of pigmented skin lesions were diagnosed by human experts and 86% by ANN. [10]

In another study, where a mobile application was developed for a similar purpose, 129,450 pictures were used with 2,032 different types of diseases, and a model was built with end-to-end training, to make it possible for users to recognize their lesions using this application. [11]

Today AI performs better than dermatologists in melanoma detection. [17]

In many machine learning algorithms, it means a great disadvantage that the algorithm expects future data to be the same feature space and have the same distribution. This is not the case in most of the real-world applications, and to address this issue, transfer learning has emerged as a new learning framework. [16]

III. System Design

In the course of our work, we did not have professional help relating the scientific background of the topic, therefore, we could only rely on a relatively small database. Our main objective was to identify the most effective model, with which we can achieve the most accurate result on even a limited database. To achieve this, we selected different models, which could work well in detecting melanomas, performed the same training on the same dataset to be able to compare these models, and evaluate the most accurate solution.

We chose the following 3 models: EfficientNetB0, ResNet50, InceptionV3.

A. EfficientNetB0

EfficientNet is a convolutional neural network that can learn rich feature representations for a wide range of images. There are many types of EfficientNet models, starting with the smallest by size, the EfficientNetB0. The network has an image input size of 224x224. Compared to other models operating with comparable accuracy,

EfficientNet is significantly smaller than those. For instance, the ResNet50 model has roughly 5 times more parameters in keras, nonetheless, it still underperforms the smallest EfficientNet, the EfficientNetB0.

EfficientNet models [14] have a main building block of mobile inverted bottleneck MBConv, first used in MobileNetV2. The model uses direct shortcuts between bottlenecks and depthwise separable convolution, and with this, the model effectively reduces computation, in comparison to traditional layers. EfficientNet uses squeeze-and-excitation (SE) optimization with which even better performance can be achieved.

There is another great advantage of using an EfficientNet model, it can scale more efficiently, which also allows for better performance.

EfficientNets starting from B0 to B7 are steadily increasing accuracies and also remaining relatively small.

B. *ResNet50*

ResNet50 (Residual Networks [15]) is a classic neural network used for many different computer vision tasks. ResNet's main advantage was that it could train extremely deep neural networks successfully, which was not feasible prior to that, because of the vanishing gradients problem. Skip connection was first introduced in ResNets. Skip connection is when convolution layers are stacked one after another, but the original input is also added to the output of the convolution block.

The network has an image input size of 224x224.

C. *InceptionV3*

Inception-v3 is also a convolutional neural network. The architecture of an Inception v3 network consists of factorized convolutions, which reduces computation by reducing the parameters involved in the network; smaller convolutions, which leads to a much faster training; asymmetric convolutions; auxiliary classifier, which is a small CNN (convolutional neural network) inserted between layers during the training that in a case in an Inception-v3 network, acts as a regularizer; and grid size reduction usually done by pooling operations. The network has an image input size of 299x299.

The originally proposed Inception-v3 model was 42 layers deep, and the computation cost was only approximately 2.5 higher than with a GoogLeNet and it was furthermore considerably more efficient than a VGGNet. [13]

IV. Implementation

A. *Database*

The only open-source melanoma database available for us was the HAM10000 (“Human Against Machine with 10000 training images") dataset, which was created specifically for academic machine learning purposes. [12]

	lesion_id	image_id	dx	dx_type	age	sex	localization
1211	HAM_0000871	ISIC_0025964	mel	histo	40.0	female	chest
1212	HAM_0000871	ISIC_0030623	mel	histo	40.0	female	chest
1213	HAM_0000040	ISIC_0027190	mel	histo	80.0	male	upper extremity

Fig. 1. Dataset Visualization

This dataset consists of 10,015 dermatoscopic images about moles, both benign and malignant, thus being a representative collection of important diagnostic categories. Images and the corresponding meta-data were collected over a period of 20 years, after which they processed the stored data, unified the diagnoses, and formed seven generic classes of these dermatoscopic images. These generic classes were chosen particularly for supporting the use of these images as a benchmark dataset for the diagnosis of pigmented lesions by both human specialists and machines. These seven classes cover more than 95% of all pigmented lesions occurring in daily clinical practice.

Lesions in these dermatoscopic images are 600x450px at 96DPI and were enhanced with manual histogram corrections to allow a better contrast and color reproduction.

From this dataset, more than 50% of images have been confirmed by pathology and for the rest of the cases, ground truths were provided either by expert consensus, or confirmation by in-vivo confocal microscopy, thus providing an assured diagnosis in all cases.

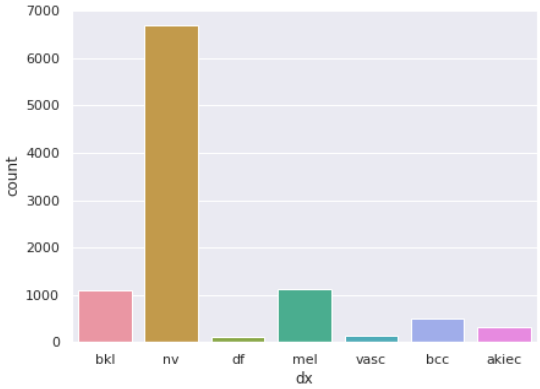

Fig. 2. Melanoma data in dataset

In our case, we have to detect melanomas among these dermatoscopic images. There are only 1113 images showing melanomas, which are only 10% of the entire dataset. Although we worked with the other lesion types as well, to be able to separate melanomas from those images, this is a significantly smaller amount, considering the fact that we also needed a separate dataset for the training and testing as well.

B. *Image preprocessing*

Since our aim was to detect only those pictures that contained information about melanomas we made a binary categorization on the HAM10000 dataset, using the two labels ‘melanoma’ for malignant melanoma and for all the other lesions ‘others’, which contained actinic keratoses, basal cell carcinoma, benign keratosis, dermatofibroma, melanocytic nevi, and vascular skin lesions as well.

The HAM10000 dataset contained only 1113 images about melanomas, thus we decided to use data augmentation for a better operation. We used the Image Data Generator for this purpose. With data augmentation, the chance of overfitting was also reduced.

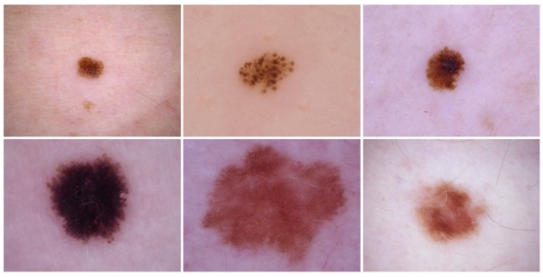

Fig. 3. Melanoma images (top row) and non-melanoma images (bottom row)

We considered it especially important to use a balanced dataset and for this reason, we manually multiplied the melanoma image dataset separated for the model training, to produce even datasets for melanoma and for the other kinds of pigmented lesions.

We used two ways of data augmentation. Firstly, we manually multiplied only the train dataset with respect to the morphology of the melanomas and after that, we used an Image Data Generator for on-the-fly augmentation. Manual alterations consisted of horizontal- and vertical flipping of the images.

A few of the more relevant features of a mole are the following: shape, symmetry, border irregularity, color spectrum, irregular dots of pigment.

Due to this, we did not use those parameters in the Image Data Generator, which would have caused significant deviations in these features.

At first, loading this produced dataset was unfeasible for us due to the limit in memory and so we decided to load and train our models in batches. For this purpose, we used the flow_from_directory() function with a batch size of first 32, then 16 images.

C. *Training*

After preprocessing of images was carried out, the next step was to decide which state-of-the-art models could perform better in recognizing our still relatively small dataset of images.

At first, we experimented with freezing the layers differently, and tried to achieve better results.

One idea was to freeze the entire base models. Thus, in both the building of the EfficientNetB0 and the InceptionV3 network, we froze the base model, making it untrainable, and trained only the top of the models, which consisted of a GlobalAveragePooling2D layer, a Dense layer with Rectified Linear Unit (ReLU) activation, and then another Dense layer with sigmoid activation. This way, we got two models, which had different, untrainable base models (EfficientNetB0 and InceptionV3), and had the same, trainable layers on top of this base model.

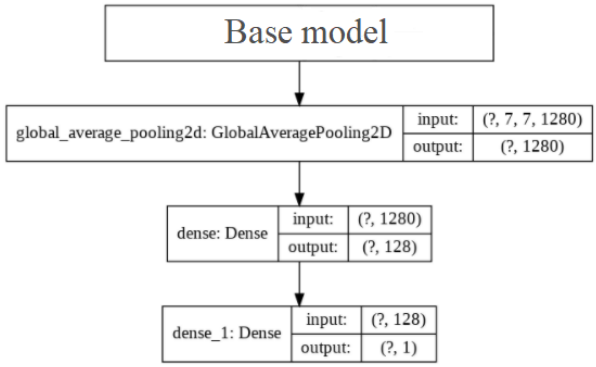

Fig. 4. Visualization of our models

Another idea was to freeze some layers of the base models and leave roughly half of layers trainable.

These did not provide the results we expected, this is why we eventually decided to leave every layer trainable. This configuration seemed to perform best.

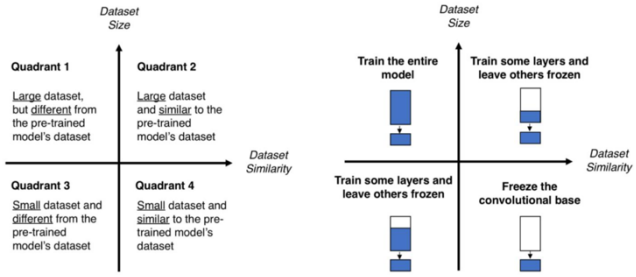

Fig. 5. Explaining the effect of freezing different layers [1]

We used an Adam optimizer. We also decided to use early stopping and saved the most accurate model's weights.

At first, we used the same learning rate in both the EfficientNetB0 and the IneptionV3 based models, setting its value to 0.0001. After some experimenting with the parameters, we found out that this value was too large in the case of the InceptionV3 model, and to solve this, we set its value to the tenth of the original size, to 0.00001.

We used the same batch size in both cases, which was 32 at first and 16 in the final parametrization.

We experimented with the value of the 'step per epoch' parameter a little more. This parameter sets the number of batch iterations before the model considers a training epoch finished. In general, the proper setting of this parameter could be important in larger datasets, or as in our case, when using random data augmentations during the training progress.

Therefore, setting the step per epoch was important for us as well. At first, we used the value of the length of the training dataset divided by the size of the batches, which was 7128/16=446, making epochs a bit too large. This way the model reached its best accuracy through a relatively few epochs, which is not a disadvantage, but for drawing conclusions from the training process, it is better to see more details, so setting the step per epoch parameter to a smaller value had a practical advantage.

[1] https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751

The final value resulted in 100 batch iterations in one epoch, and we also shuffled them making it randomized, through a total of 50 epochs of training.

D. Evaluation and Testing

We tested our models on the same set of test data since evaluating them on different images could result in the incomparability of the two built models. As we mentioned in the image preprocessing section, we found it important to use a balanced dataset, so for testing, we also separated a corresponding dataset, a total of 222 images consisting of 111 melanoma and 111 non-melanoma images called in our case 'others'.

During the evaluation of the InceptionV3 model, we got the following results: the value of the loss was 0.4101 with an accuracy value of **82.43%.** These were the best results.

During the evaluation of the EfficientNet-B0 model, the loss value was even smaller, 0.3233, with a much better **87.84%** accuracy.

It can be stated that during the training, the EfficientNet-B0 model was more precise and converged faster, reaching its best accuracy much faster than the other model with an InceptionV3 base.

We optimized hyperparameters step by step, considering and altering only one at a time, which made it possible for us to draw conclusions from each impact caused by the individual parameter alterations.

At first, we became curious about the correlation between the resolution of images and the accuracy, because we accidentally produced better results training the model on images in worse resolution and testing it with images in better resolution, than when testing with the same resolution as used during training. This resulted in a loss value of 0.3056 with 88.74% accuracy. After this, we tried other resolutions as well, but eventually, we did not find any obvious correlation with the accuracy.

Testing with same resolution, the loss value was 0.3233, and the accuracy was 87.84%.

Fig. 6. Loss function

The loss values as a function of epoch number during the evaluation of the EfficientNetB0 model could be seen in Fig. 6. During training, we used the earlystop callback function with a patience value of 5. It can be seen, that after the minimum of the validation loss function at 11. epoch, there is 5 worse, and then the training stops. At the same time, the loss of the train was improving constantly. This can be explained as overfitting.

Similarly to the change of the accuracy as a function of epochs, the best accuracy is at 11. Epoch (which is shown as 10. epoch in the figure due to the different indexing of epochs).

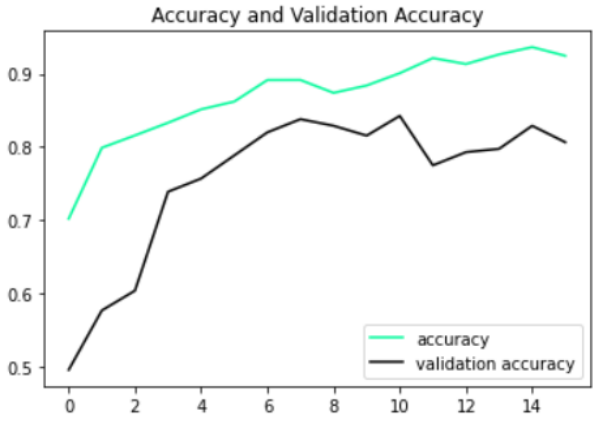

Fig. 7. Accuracy function

V. Future plans, conclusion

To sum up, by modeling two convolutional networks with different base layers, we could successfully compare them for a more accurate image recognition to detect melanoma malignum (a type of severe skin cancer).

In our work, we implemented an EfficientNetB0 based model and, an InceptionV3 based model, on top of which we added the same top layers. This way, we could evaluate which base model performs more accurately at detecting melanomas. With our hyperparameter configuration, the EfficientNetB0 based model proved to be more accurate and produced better results.

In the future, implementing a ResNet50 based model with the same top layers could give further information about which convolutional network is optimal for this purpose.

We could further improve the performance of our models with hyperparameter optimization, moreover, a richer and more divergent dataset would also provide new opportunities.

References

[1] Somlai Beáta. Melanoma malignum. In Kárpáti Sarolta, Kemény Lajos, Remenyik Éva (szerk.), Bőrgyógyászat és Venerológia, Medicina, Budapest, 2013, 753-765

[2] Klaus Wolff, Richard.A.Johnson, Arturo P. Saavedra. Melanoma precursors and Primary Cutaneous Melanoma. In Klaus Wolff, Richard.A.Johnson, Arturo P. Saavedra (szerk.), Fitzpatrick's color atlas and synopsis of clinical dermatology, Mc Graw Hill Education, United States of America., 2013, 252-283.

[3] Garbe, C., K. Peris, A. Hauschild, P. Saiag, M. Middleton, L. Bastholt, J.J. Grob, J. Malvehy, J. Newton-Bishop, A.J. Stratigos, H. Pehamberger, A.M. Eggermont, F. European Dermatology, European Association of Dermato-Oncology (EADO), European Organisation for Research and Treatment of Cancer (EORTC).(2016) Diagnosis and treatment of melanoma. European consensus-based interdisciplinary guideline - Update. Eur J Cancer, 63: p. 201-17.

[4] Balch, C.M., J.E. Gershenwald, S.J. Soong, J.F. Thompson, M.B. Atkins, D.R. Byrd, A.C. Buzaid, A.J. Cochran, D.G. Coit, S. Ding, A.M. Eggermont, K.T. Flaherty, P.A. Gimotty, J.M. Kirkwood, K.M. McMasters, M.C. Mihm, Jr., D.L. Morton, M.I. Ross, A.J. Sober, and V.K. Sondak. (2009) Final version of 2009 AJCC melanoma staging and classification. J Clin Oncol, 27(36): p. 6199-206.

[5] Lawler PE, Schreiber S. Cutaneous malignant melanoma: nursing's role in prevention and early detection. Oncol Nurs Forum. 1989 May-Jun;16(3):345-52. PMID: 2660118.

[6] Edman RL, Wolfe JT. Prevention and early detection of malignant melanoma. Am Fam Physician. 2000 Nov 15;62(10):2277-85. PMID: 11126854.

[7] de Vries E, Coebergh JW. Cutaneous malignant melanoma in Europe. Eur J Cancer. 2004 Nov;40(16):2355-66. doi: 10.1016/j.ejca.2004.06.003. PMID: 15519506.

[8] Friedman RJ, Rigel DS, Kopf AW. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA Cancer J Clin. 1985 May-Jun;35(3):130-51. doi: 10.3322/canjclin.35.3.130. PMID: 3921200.

[9] Russo, T., V. Piccolo, G. Ferrara, M. Agozzino, R. Alfano, C. Longo, and G. Argenziano. (2017) Dermoscopy pathology correlation in melanoma. J Dermatol, 44(5): p. 507-514.

[10] Binder, M. et al. Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. *Br J Dermatol* 130, 460–465 (1994).

[11] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017 Feb 2;542(7639):115-118. doi: 10.1038/nature21056. Epub 2017 Jan 25. Erratum in: Nature. 2017 Jun 28;546(7660):686. PMID: 28117445.

[12] Tschandl, P. *et al*. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* 5:180161 doi: 10.1038/sdata.2018.161 (2018).

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

[14] Tan, M. and Quoc V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.” ICML (2019).

[15] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[16] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[17] European Society for Medical Oncology. (2018, May 28). Man against machine: AI is better than dermatologists at diagnosing skin cancer. *ScienceDaily*. Retrieved December 11, 2020 from www.sciencedaily.com/releases/2018/05/180528190839.htm