Transfer Learning of ImageNet Neural Network for Pigmented Skin Lesion Detection

Pedro M. M. Pereira[13]
pedrommpereira@co.it.pt
Rui Fonseca-Pinto[12]
rui.pinto@ipleiria.pt
Rui Pedro Paiva[3]
ruipedro@dei.uc.pt
Luis M. N. Tavora[2]
luis.tavora@ipleiria.pt
Pedro A. A. Assuncao[12]
amado@co.it.pt
Sergio M. M. de Faria[12]
sergio.faria@co.it.pt

[1] Instituto de Telecomunicações
Portugal
[2] ESTG
Polytechnic Institute of Leiria
Portugal
[3] DEI - FCTUC
University of Coimbra
Portugal

Abstract

Traditional Artificial Neural Networks (ANN) have been investigated in the past for skin lesion classification and nowadays their performance is already quite useful to assist in medical diagnosis and decision processes. In the field of visual object recognition, recent developments of such networks (Deep Convolutional Neural Networks) are currently the winners of the ImageNet competition. This work extends the use of CNN for classification of pigmented skin lesions, by investigating a training methodology based on transfer learning on pre-trained networks.

1 Introduction

The importance of skin lesion classification arises from the fact that one of the most dangerous skin cancers, the melanoma, is developed from pigmented melanocytes [5] and its incidence in the world population is increasing very fast. Skin cancer can be either benign and malignant. Since the melanoma is malignant, it is very likely to cause death after some time. However, if diagnosed at early stages, high cure rates are achievable. Thus, early detection and full characterisation of suspicious skin lesions is the key to reduce mortality rates associated to this type of skin cancer. The development of computer vision techniques to automatically identify melanoma has been under study for decades [3] and automatic techniques for detection and classification is becoming increasingly useful to assist dermatologists and to support expert systems [6, 9].

Recent advances in visual recognition led to the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [2, 10], which uses a dataset comprising more than 14 million images (of which 1 million have bounding box annotations with around 100 hundred words) that can be divided into 1000 different labels – manually validated by crowd-sourcing. The ImageNet Challenge is currently considered to be one of the most important initiatives and the dataset has therefore become a benchmark standard for large-scale object recognition, i.e., image classification, single-object location and object detection. Due to its competition-based approach, many authors are constantly improving their image classification/recognition algorithms every year. This has led to an exponential growth of related research and significant advances in state-of-the-art techniques [10].

This work focuses on studying the performance of skin cancer detection using highly-accurate networks, developed in recent years for ImageNet. Relevant comparisons are made with the performance obtained for the 1000 categories in ImageNet. To this end, the ISIC dataset [1] is selected, as the collection of skin lesion images. This dataset contains a total of 3438 images that can be divided into: 2380 benign and 1058 malignant lesions. These malignant lesions are classified as melanoma, basal cell carcinomas and squamous cell carcinoma, while the remaining ones are benign. Such classification was obtained from an unspecified number of skin cancer experts.

The transfer learning approach used in this research study uses some selected pre-trained networks from ImageNet to first extract a number of abstract features, which are fed forward to several different classifiers. Then the classification performance is evaluated and discussed.

2 Proposed Approach

The proposed approach follows a processing pipeline from the input image data to the output classification results. Firstly, before entering in the network, a pre-processing stage is responsible for performing data augmentation and then image resizing to match the network intake. Secondly, these data enters in the pre-trained network whose output is fed to the final classifier. Several classifiers are studied in this work. Different alternatives are separately trained resorting to both original data and augmented data with 20% random information holdout for later evaluation of the trained network.

2.1 Architectures

In ILSVRC history there are several pre-trained networks, already capable of image classification over 1000 different categories. This work elects 5 of the must frequently used networks, which have shown to be able to adapt to other identification and classification problems. These networks are: Alexnet [7], pioneering networking comprising 25 layers and it was the winner of the 2012 ILSVRC; VGG16 and VGG19 Net [11], reinforced the notion that convolutional neural networks must have layers in depth such that visual data present a hierarchical representation; GoogLeNet [12], has the Inception module that deviates from the standard sequential layer-stacking approach and it was the winner in 2014; and ResNet50 [4], presents an innovative way of solving the vanishing gradient problem, it comprises 177 layers and it was the winner in 2015.

2.2 Pre-processing: data augmentation and image resizing

To increase accuracy, data augmentation is performed by using a limit set of random transformations [8]. In this work the following transformations were selected: *Intensity Values Adjustment*: increases the contrast of the image; *Contrast-Limited Adaptive Histogram Equalization*: enhances the contrast of a given grayscale image by transforming the values so that its distribution matches a uniform/flat histogram (256 bins); *Random Brightness*: induces brightness variation to the image; *Random Edge-Aware Local Contrast*: enhances or flattens the image local contrasts; *Random Sharpness*: sharpens the image using the unsharp-masking method; *PCA Colour Jitter*: modifies the intensities of the RGB channels in the image according to the PCA transformation; *Random Affine Transformations*: operation between affine spaces that preserves points, straight lines and planes. As a final note, the augmentation strategies are not all used at the same time. The *PCA Colour Jitter* and *Random Affine Transformations* are always used at the end of the augmentation step, but the remaining operators are only randomly applied with a 10% change (each). After this stage, each image is augmented 200 times, thus effectively making the dataset 200 times larger.

After a possible augmentation step, and before entering the network, all input data (images) is resized to fit the network intake. Apart from Alex-Net, which receives a 277x277 (pixel) RGB image, all other networks accept a 224x244 (pixel) RGB image. Therefore, as a final step before entering the network, the images are resized to their smallest dimension (maintaining aspect ratio) and then centre-cropped to remove the outer border in excess (if any).

	AlexNet			VGG16			VGG19			GoogleNet			ResNet50		
	Acc	SE	SP	Acc	SE	SP	Acc	SE	SP	Acc	SE	SP	Acc	SE	SP
SVM	30.9	99.5	0.4	34.6	97.6	6.7	46.7	69.2	36.8	30.7	100	0.0	30.7	100	0.0
KNN	74.5	61.1	80.5	67.7	56.9	72.5	71.3	60.2	76.3	73.8	57.3	81.1	72.8	60.2	78.4
Tree	68.7	46.9	78.4	64.2	40.8	74.6	64.0	40.3	74.6	67.7	51.2	75.0	61.3	49.3	66.6
Linear	70.3	4.3	99.6	69.4	55.9	75.4	55.5	86.3	41.8	73.9	52.1	83.6	75.8	47.9	88.2
NaiveBayes	64.9	73.5	61.1	62.6	66.4	60.9	64.8	64.5	64.9	64.9	73.9	60.9	72.5	55.9	79.8

Table 1: Test Results without using Augmented Data in training

	AlexNet			VGG16			VGG19			GoogleNet			ResNet50		
	Acc	SE	SP	Acc	SE	SP	Acc	SE	SP	Acc	SE	SP	Acc	SE	SP
SVM	72.6	54.5	80.7	71.9	58.8	77.7	71.9	58.8	77.7	71.6	60.7	76.5	72.8	53.1	81.5
KNN	75.1	62.6	80.7	70.3	52.1	78.4	70.3	52.1	78.4	71.8	55.5	79.0	75.4	64.9	80.0
Tree	65.5	48.8	72.9	68.0	46.9	77.3	68.0	46.9	77.3	67.5	46.0	77.1	69.3	53.1	76.5
Linear	78.0	52.1	89.5	60.8	80.1	52.3	60.8	80.1	52.3	69.4	69.2	69.5	78.5	43.1	94.1
NaiveBayes	67.1	66.8	67.2	69.3	0.0	100	69.3	0.0	100	62.6	70.6	59.0	67.7	72.5	65.5

Table 2: Test Results using Augmented Data in the training

2.3 Learning Strategy

As mentioned before, the overall architecture includes ImageNet networks and a transfer learning scheme for feature extraction using alternative classifiers. Since the selected pre-trained architectures already provide highly accurate predictions in the ImageNet challenge, it is assumed that they are also able to extract a great variety of abstract knowledge/features from the given images containing skin lesions. In this transfer learning strategy, the output of the last convolutional layer in the pre-trained ImageNet network is connected to several alternative classifiers. The classifiers used in this work are: the SVM classifier, the K-Nearest Neighbours, the Tree classifier, a Linear classifier and a NaiveBayes classifier.

3 Results and Discussion

Using the ImageNet networks as feature extractor on the original 3438 images, while holding out 20% of this data for later testing, the network knowledge provides an average accuracy of 61% on the testing data, while the accuracy obtained in training data is 87% on average. The overall results are shown in Table 1, where it can be observed that the best performing classifier is the KNN with an average accuracy of 72% on unseen test data across the different networks and 100% on the training data. Still regarding the training data performance, the SVM and the Tree classifiers achieve accuracies of 99% and 98%, respectively. However, only 62% and 61% accuracy is obtained on unseen test data.

When data augmentation is used, the performances increase by 9% on the test-set and lose 12% accuracy on the training-set. Table 2 is presented for comparison with the previous results. In this case the training-set only comprises augmented images, while the test-set is the same as before. It is observed that image augmentation provides some improvement to the classification results. Despite the small improvement of the KNN classifier, which only gains 0.6% accuracy on test data, the SVM classifier more than double's its performance. Taking into account the training results (not shown here), this increase in performance is justified by the reduction of overfitting resulting from data augmentation.

4 Conclusion

ImageNet winning networks already achieve an accuracy greater than 95%, but when adapted to classify skin lesions their performance drops to quite modest results, even using data augmentation. This work performed transfer learning to classify skin lesions as malignant or benign using 5 cornerstone neural network architectures that have been proven to produce high results on other domains. The results demonstrate that there is significant room for further research, using highly accurate networks and transfer learning for specific classification in the field of medical imaging. In particular, it is necessary to investigate how to improve transfer learning performance using networks trained on completely different domains.

Acknowledgments

This work was supported by the Fundação para a Ciência e Tecnologia, Portugal, under PhD Grant SFRH/BD/128669/2017 and PlenoISLA project in the scope of R&D Unit 50008, through national funds and where applicable co-funded by FEDER – PT2020 partnership agreement.

References

[1] International skin imaging collaboration: Melanoma project website. `https://isic-archive.com/`. Accessed: 01.12.2017.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[3] Robert J. Friedman, Darrell S. Rigel, and Alfred W. Kopf. Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. *CA: A Cancer Journal for Clinicians*, 35(3):130–151, 1985.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conf. on comp. vision and pattern recognition*, pages 770–778, 2016.

[5] Howard L Kaufman. *The melanoma book: a complete guide to prevention and treatment*. Gotham, 2005.

[6] K. Korotkov and R. Garcia. Computerized analysis of pigmented skin lesions: a review. *Artif. Intel. in Medicine*, 56(2):69–90, 2012.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[8] Joseph Lemley, Shabab Bazrafkan, and Peter Corcoran. Smart augmentation learning an optimal data augmentation strategy. *IEEE Access*, 5:5858–5869, 2017.

[9] Sameena Pathan, K Gopalakrishna Prabhu, and PC Siddalingaswamy. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions – a review. *Biomedical Signal Processing and Control*, 39:237–262, 2018.

[10] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[11] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[12] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.